UNIVERSITY OF CALIFORNIA SAN DIEGO

# Understanding pancreatic mechanisms of diabetes risk with genetic association and single cell epigenome data

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Joshua Chiou

Committee in charge:

Professor Kyle Gaulton, Chair
Professor Eran Mukamel
Professor Bing Ren
Professor Maike Sander
Professor Xin Sun

2021

The dissertation of Joshua Chiou is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

# DEDICATION

This dissertation is dedicated to my family and friends, who have supported me throughout my graduate studies. To my loving parents John and Kimberly, who have always encouraged me to pursue my interests. To my sister Jocelyn, who has inspired me to become more creative. To my girlfriend Na, who has been by my side every step along the way. And to my friends, who have shared many memories with me along this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am deeply grateful to my dissertation advisor, Dr. Kyle Gaulton, for his invaluable mentorship and support throughout my PhD program. Through his guidance, I have grown to become a fledging scientist. I am also thankful to my committee members, Dr. Eran Mukamel, Dr. Bing Ren, Dr. Maike Sander, and Dr. Xin Sun, and former committee member Dr. Wenxian Fu, for their insightful discussion about my thesis projects and career aspirations.

My thesis projects would not have been possible without the support of former and current members of the Gaulton lab (namely Mei-Lin Okino and Serina Huang), our collaborators at the Center for Epigenomics (namely David Gorkin, Zhang “Frank” Cheng, Jee Yun Han, and Sebastian Preissl), and the Sander lab (namely Ryan Geusz and Chun Zeng).

Chapter 2, in full, has been accepted for publication of the material in “Single cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk“ in *Nature Genetics*, 2021. Joshua Chiou, Chun Zeng, Zhang Cheng, Jee Yun Han, Michael Schlichting, Michael Miller, Robert Mendez, Serina Huang, Jinzhao Wang, Yinghui Sui, Allison Deogaygay, Mei-Lin Okino, Yunjiang Qiu, Ying Sun, Parul Kudtarkar, Rongxin Fang, Sebastian Preissl, Maike Sander, David U Gorkin, Kyle J Gaulton. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been accepted for publication of the material in “Cell type-specific genetic mechanisms of type 1 diabetes risk“. Joshua Chiou, Ryan J Geusz, Mei-Lin Okino, Jee Yun Han, Michael Miller, Rebecca Melton, Elisha Beebe, Paola Benaglio, Serina Huang, Katha Korgaonkar, Sandra Heller, Alexander Kleger, Sebastian Preissl, David U Gorkin, Maike Sander, Kyle J Gaulton. The dissertation author was the primary investigator and author of this paper.

## VITA

2011-2015 Bachelor of Science in Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles

2016-2021 Doctor of Philosophy in Biomedical Sciences, University of California, San Diego

## PUBLICATIONS

First author publications (*co-first authors)

**Chiou J**, Geusz RJ, Okino M, Han JY, Miller M, Melton P, Beebe E, Benaglio P, Huang S, Korgaonkar K, Heller S, Kleger A, Preissl S, Gorkin DU, Sander M, Gaulton KJ. Cell type-specific genetic mechanisms of type 1 diabetes risk. Accepted to *Nature*.

**Chiou J***, Zeng C*, Cheng Z, Han JY, Schlichting M, Mendez R, Huang S, Wang J, Sui Y, Deogaygay A, Okino M, Qiu Y, Sun Y, Kudtarkar P, Fang R, Preissl S, Sander M, Gorkin DU, Gaulton KJ. Single cell chromatin accessibility reveals pancreatic islet cell type-and state-specific regulatory programs of diabetes risk. Accepted to *Nature Genetics*.

Geusz RJ*, Wang A*, **Chiou J***, Lancman JJ*, Wetton N, Kefalopoulou S, Wang J, Qiu Y, Yan J, Aylward A, Ren B, Dong PDS, Gaulton K, Sander M. Pancreatic progenitor epigenome maps prioritize type 2 diabetes risk genes with roles in development. *eLife* 2021.

Wang A*, **Chiou J***, Poirion OB*, Buchanan J*, Valdez MJ*, Verheyden JM, Hou X, Kudtarkar P, Narendra S, Newsome JM, Guo M, Faddah DA, Zhang K, Young RE, Barr J, Sajti E, Misra R, Huyck H, Rogers L, Poole C, Whitsett JA, Pryhuber G, Xu Y, Gaulton KJ, Preissl S, Sun X, NHLBI LungMap Consortium. Single cell multiomic profiling of human lung reveals cell type-specific and age-dynamic control of SARS-CoV2 host genes. *eLife 2020;9:e62522*.

Greenwald WW*, **Chiou J***, Yan J*, Qiu Y*, Dai N*, Wang A, Nariai N, Aylward A, Han JY, Kadakia N, Regue L, Okino M, Drees F, Kramer D, Vinckier N, Minichiello L, Gorkin DU, Avruch J, Frazer KA, Sander M, Ren B, Gaulton KJ. Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nature Communications* 10, 2078. 2019.

Aylward A*, **Chiou J***, Okino M, Kadakia N, Gaulton KJ. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Human Molecular Genetics* ddy314. 2019.

Other publications

Okino M, Bengalio P, Zhu H, Yan J, Beebe E, Qiu Y, **Chiou J**, Newsome J, Kaur J, Corban S, Aylward A, Frazer KA, Ren B, Sander M, Gaulton KJ. Type 1 diabetes risk genes mediate pancreatic beta cell survival in response to inflammatory cytokine exposure. In preparation.

Zhang K*, Hocker JD*, Miller M, Hou X, **Chiou J**, Poirion OB, Qiu Y, Li YE, Gaulton KJ, Wang A, Preissl S, Ren B. A cell atlas of chromatin accessibility across 25 adult human tissues. *bioRxiv*.

Philippi A*, Heller S*, Costa IG*, Senée V*, Breunig M, Li Z, Kwon G, Russell R, Illing A, Lin Q, Hohwieler M, Degavre A, Zalloua P, Liebau S, Schuster M, Krumm J, Zhang X, Geusz R, Benthuysen JR, Wang A, **Chiou J**, Gaulton KJ, Neubauer H, Simon E, Klein T, Wagner M, Nair G, Besse C, Dandine-Roulland C, Olaso R, Deleuze J, Kuster B, Hebrok M, Seufferlein T, Sander

M, Boehm BO, Oswald F, Nicolino M, Julier C, Kleger A. *ONECUT1* mutations and variants in diabetes. In revision.

Li YE, Preissl S, Hou X, Zhang Z, Zhang K, Qiu Y, Poirion O, Li B, **Chiou J**, Liu H, Pinto-Duarte A, Kubo N, Yang X, Fang R, Wang X, Han JY, Lucero J, Yan Y, Kuan S, Gorkin DU, Gaulton KJ, Nunn M, Mukamel EA, Behrens MM, Ecker JR, Ren B. An Atlas of Gene Regulatory Elements in Adult Mouse Cerebrum. In revision.

Muus C, Luecken MD, Eraslan G, Sikkema L, Waghray A, Heimberg G, Kobayashi Y, Vaishnav ED, Subramanian A, Smillie C, Jagadeesh KA, Duong ET, Fiskin E, Triglia ET, Ansari M, Cai P, Lin B, Buchanan J, Chen S, Shu J, Haber AL, Chung H, Montoro DT, Adams T, Aliee H, Allon SJ, Andrusivova Z, Angelidis I, Ashenberg O, Bassler K, Bécavin C, Benhar I, Bergenstråhle J, Bergenstråhle L, Bolt L, Braun E, Bui LT, Callori S, Chaffin M, Chichelnitskiy E, **Chiou J**, Conlon TM, Cuoco MS, Cuomo ASE, Deprez M, Duclos G, Fine D, Fischer DS, Ghazanfar S, Gillich A, Giotti B, Gould J, Guo M, Gutierrez AJ, Habermann AC, Harvey T, He P, Hou X, Hu L, Hu Y, Jaiswal A, Ji L, Jiang P, Kapellos TS, Kuo CS, Larsson L, Leney-Greene MA, Lim K, Litviňuková M, Ludwig LS, Lukassen S, Luo W, Maatz H, Madissoon E, Mamanova L, Manakongtreecheep K, Leroy S, Mayr CH, Mbano IM, McAdams AM, Nabhan AN, Nyquist SK, Penland L, Poirion OB, Poli S, Qi C, Queen R, Reichart D, Rosas I, Schupp JC, Shea CV, Shi X, Sinha R, Sit RV, Slowikowski K, Slyper M, Smith NP, Sountoulidis A, Strunz M, Sullivan TB, Sun D, Talavera-López C, Tan P, Tantivit J, Travaglini KJ, Tucker NR, Vernon KA, Wadsworth MH, Waldman J, Wang X, Xu K, Yan W, Zhao W, Ziegler CGK, The NHLBI LungMap Consortium, The Human Cell Atlas Lung Biological Network. Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nature Medicine*. 2021.

Yan J*, Qiu Y*, dos Santos AMR*, Yin Y, Li YE, Vinckier N, Nariai N, Benaglio P, Raman A, Li X, Fan S, **Chiou J**, Chen F, Frazer KA, Gaulton KJ, Sander M, Taipale J, Ren B. Systematic analysis of binding of transcription factors to noncoding variants. *Nature*. 2021.

Gorkin DU*, Barozzi I*, Zhao Y*, Zhang Y*, Huang H*, Lee AY, Li B, **Chiou J**, Wildberg A, Ding B, Zhang B, Wang M, Strattan JS, Davidson JM, Qiu Y, Afzal V, Akiyama JA, Plajzer-Frick I, Novak CS, Kato M, Garvin TH, Pham QT, Harrington AN, Mannion BJ, Lee EA, Fukuda-Yuzawa Y, He Y, Preissl S, Chee S, Han JY, Williams BA, Trout D, Amrhein H, Yang H, Cherry JM, Wang W, Gaulton KJ, Ecker JR, Shen Y, Dickel DE, Visel A, Pennacchio LA, Ren B. An atlas of dynamic chromatin landscapes in the developing mouse fetus. *Nature* 583(7818), 744-751. 2020.

Benaglio P, Newsome J, Han JY, **Chiou J**, Aylward A, Corban S, Okino M, Kaur J, Gorkin DU, Gaulton KJ. Mapping genetic effects on cell type-specific chromatin accessibility and annotating complex trait variants using single nucleus ATAC-seq. *bioRxiv*.

Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, Yu GZ, Rüeger S, Speidel L, Kim YJ, Horikoshi M, Mercader JM, Taliun D, Moon S, Kwak S, Robertson NR, Rayner NW, Loh M, Kim B, **Chiou J**, Miguel-Escalada I, Parolo PdB, Lin K, Bragg F, Preuss MH, Takeuchi F, Nano J, Guo X, Lamri A, Nakatochi M, Scott RA, Lee J, Huerta-Chagoya A, Graff M, Chai J, Parra EJ, Yao J, Bielak LF, Tabara Y, Hai Y, Steinthorsdottir V, Cook JP, Kals M, Grarup N, Schmidt EM, Pan I, Sofer T, Wuttke M, Sarnowski C, Gieger C, Nousome D, Trompet S, Long J, Sun M, Tong L, Chen W, Ahmad M, Noordam R, Lim VJY, Tam CHT, Joo YY, Chen C, Raffield LM, Lecoeur C, Maruthur NM, Prins BP, Nicolas A, Yanek LR, Chen G, Jensen RA, Tajuddin S, Kabagambe E, An P, Xiang AH, Choi HS, Cade BE, Tan J, Abaitua F, Adair LS, Adeyemo A,

Aguilar-Salinas CA, Akiyama M, Anand SS, Bertoni A, Bian Z, Bork-Jensen J, Brandslund I, Brody JA, Brummett CM, Buchanan TA, Canouil M, Chan JCN, Chang L, CheeM, ChenJ, Chen S, Chen Y, Chen Z, Chuang L, Cushman M, Das SK, de Silva HJ, Dedoussis G, Dimitrov L, Doumatey AP, Du S, Duan Q, Eckardt K, Emery LS, Evans DS, Evans MK, Fischer K, Floyd JS, FordI, Fornage M, Franco OH, Frayling TM, Freedman BI, Fuchsberger C, Genter P, Gerstein HC, Giedraitis V, González-Villalpando C, González-Villalpando ME, Goodarzi MO, Gordon-Larsen P, Gorkin DU, Gross M, Guo Y, Hackinger S, Han S, Hattersley AT, Herder C, Howard A, Hsueh W, Huang M, Huang W, Hung Y, Hwang MY, Hwu C, Ichihara S, Ikram MA, Ingelsson M, Islam MT, Isono M, Jang H, Jasmine F, Jiang G, Jonas JB, Jørgensen ME, Jørgensen T, Kamatani Y, Kandeel FR, Kasturiratne A, Katsuya T, Kaur V, Kawaguchi T, Keaton JM, Kho AN, Khor C, Kibriya MG, Kim D, Kohara K, Kriebel J, Kronenberg F, Kuusisto J, Läll K, Lange LA, Lee MS, Lee NR, Leong A, Li L, Li Y, Li-Gao R, Ligthart S, Lindgren CM, Linneberg A, Liu C, Liu J, Locke AE, Louie T, LuanJ, Luk AO, Luo X, Lv J, Lyssenko V, Mamakou V, Mani KR, Meitinger T, Metspalu A, Morris AD, Nadkarni GN, Nadler JL, Nalls MA, Nayak U, Ntalla I, Okada Y, Orozco L, Patel SR, Pereira MA, Peters A, Pirie FJ, Porneala B, Prasad G, Preissl S, Rasmussen-Torvik LJ, Reiner AP, Roden M, Rohde R, Roll K, Sabanayagam C, Sander M, Sandow K, Sattar N, Schönherr S, Schurmann C, Shahriar M, Shi J, Shin DM, Shriner D, Smith JA, So WY, Stančáková A, Stilp AM, Strauch K, Suzuki K, Takahashi A, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tomlinson B, Torres JM, Tsai F, Tuomilehto J, Tusie-Luna T, Udler MS, Valladares-Salgado A, van Dam RM, van Klinken JB, Varma R, Vujkovic M, Wacher-Rodarte N, Wheeler E, Whitsel EA, Wickremasinghe AR, van Dijk KW, Witte DR, Yajnik CS, Yamamoto K, Yamauchi T, Yengo L, Yoon K, Yu C, Yuan J, Yusuf S, Zhang L, Zheng W, FinnGen, Raffel LJ, Igase M, Ipp E, Redline S, Cho YS, Lind L, Province MA, Hanis CL, Peyser PA, Ingelsson E, Zonderman AB, Psaty BM, Wang Y, Rotimi CN, Becker DM, Matsuda F, Liu Y, Zeggini E, Yokota M, Rich SS, Kooperberg C, Pankow JS, Engert JC, Chen YI, Froguel P, Wilson JG, Sheu WHH, Kardia SLR, Wu J, Hayes MG, Ma RCW, Wong T, Groop L, Mook-Kanamori DO, Chandak GR, Collins FS, Bharadwaj D, Paré G, Sale MM, Ahsan H, Motala AA, Shu X, Park K, Jukema JW, Cruz M, McKean-Cowdin R, Grallert H, Cheng C, Bottinger EP, Dehghan A, Tai E, Dupuis J, Kato N, Laakso M, Köttgen A, Koh W, Palmer CNA, Liu S, Abecasis G, Kooner JS, Loos RJF, North KE, Haiman CA, Florez JC, Saleheen D, Hansen T, Pedersen O, Mägi R, Langenberg C, Wareham NJ, Maeda S, Kadowaki T, Lee J, Millwood IY, Walters RG, Stefansson K, Myers SR, Ferrer J, Gaulton KJ, Meigs JB, Mohlke KL, Gloyn AL, Bowden DW, Below JE, Chambers JC, Sim X, Boehnke M, Rotter JI, McCarthy MI, Morris AP on behalf of the DIAMANTE Consortium. Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *medRxiv*.

Hocker JD, Poirion OB, Zhu F, Buchanan J, Zhang K, **Chiou J**, Wang TM, Zhang Q, Hou X, Li YE, Zhang Y, Farah EN, Wang A, McCulloch AD, Gaulton KJ, Ren B, Chi NC, Preissl S. Cardiac cell type-specific gene regulatory programs and disease-risk association. Accepted to *Sciences Advances*. 2021.

Geusz RJ, Wang A, Lam DK, Vinckier NK, Alysandratos K, Roberts DA, Wang J, Kefalopoulou S, Qiu Y, **Chiou J**, Gaulton KJ, Ren B, Kotton DN, Sander M. A dual mechanism of enhancer activation by FOXA pioneer factors induces endodermal organ fates. *bioRxiv*.

Aylward A*, Okino M*, Benaglio P, **Chiou J**, Beebe E, Padilla JA, Diep S, Gaulton KJ. Glucocorticoid signaling in pancreatic islets modulates gene regulatory programs and genetic risk of type 2 diabetes. *bioRxiv*.

Gorkin DU*, Qiu Y*, Hu M*, Fletez-Brant K, Liu T, Schmitt AD, Noor A, **Chiou J**, Gaulton KJ, Sebat J, Li Y, Hansen KD, Ren B. Common DNA sequence variation influences 3-dimensional conformation of the human genome. *Genome Biology* 20(1), 1-25. 2019.

Jin W, Mulas F, Gaertner B, Sui Y, Wang J, Matta I, Zeng C, Vinckier N, Wang A, Nguyen-Ngoc K, **Chiou J**, Kaestner KH, Frazer KA, Carrano AC, Shih H, Sander M. A Network of microRNAs Acts to Promote Cell Cycle Exit and Differentiation of Human Pancreatic Endocrine Cells. *iScience* 21: 681-694. 2019.

Wu V, Yeerna H, Nohata N, **Chiou J**, Harismendy O, Raimondi F, Inoue A, Russell RB, Tamayo P, Gutkind JS. Illuminating the Onco-GPCRome: Novel G protein–coupled receptor-driven oncocrine networks and targets for cancer immunotherapy. *Journal of Biological Chemistry* 294(29): 11062-11086. 2019.

Zhang Q, Chao T, Patil VS, Qin Y, Tiwari SK, **Chiou J**, Dobin A, Tsai C, Li Z, Dang J, Gupta S, Urdahl K, Nizet V, Gingeras TR, Gaulton KJ, Rana TM. The long noncoding RNA *ROCKI* regulates inflammatory gene expression. *The EMBO journal* 38(8): e100041. 2019.

Wortham M, Liu F, Fleischman JY, M Wallace, Mulas F, Vinckier NK, Harrington AR, Cross BR, **Chiou J**, Patel NA, Sui Y, Jhala US, Shirihai OS, Huising MO, Gaulton KJ, Metallo CM, Sander M. Nutrient regulation of the islet epigenome controls adaptive insulin secretion. *bioRxiv.*

Udler MS, Kim J, Grotthuss M, Bonàs-Guarch S, Cole JB, **Chiou J**, Anderson CD, Boehnke M, Laakso M, Atzmon G, Glaser B, Mercader JM, Gaulton KJ, Flannick J, Getz G, Florez JC. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLOS Medicine* 15(9): e1002654. 2018.

ABSTRACT OF THE DISSERTATION

Understanding pancreatic mechanisms of diabetes risk with genetic association and single cell epigenome data

by

Joshua Chiou

Doctor in Philosophy in Biomedical Sciences

University of California San Diego, 2021

Professor Kyle J. Gaulton, Chair

Although genome-wide association studies (GWAS) have demonstrated the importance of the pancreas in diabetes risk, mechanistic insight remains challenging in part due to the non-coding nature of risk variants and the lack of cell type-resolved regulatory annotations. Beta cells within pancreatic islets are central to both forms of diabetes and are destroyed by autoimmune mechanisms in T1D or dysfunctional due to insulin resistance in T2D. However, the relevance of beta cell states or other pancreatic cell types to diabetes is relatively unknown. Here I present two

studies, both of which use single cell chromatin accessibility to annotate cell type-specific genetic risk mechanisms of diabetes. In the first study, I use snATAC-seq of pancreatic islets to show that endocrine cell types are heterogenous at the gene regulation level, and beta cell states are differently enriched for T2D risk. I highlight a causal T2D variant (rs231361) at the *KCNQ1* locus that was predicted to have state-specific effects on regulatory function, was located in a beta cell enhancer co-accessible with *INS*, and affected expression levels of *INS* in embryonic stem cell-derived beta cells. In the second study, I generate the largest-to-date GWAS of T1D in 520k samples to enable comprehensive fine mapping of T1D risk signals. I show that pancreatic exocrine cells are causal contributors of T1D through integration of fine mapping with a snATAC-seq reference map of pancreatic and immune cell types. I highlight a likely-causal T1D risk variant (rs7795896) at the *CFTR* locus that mapped within a ductal-specific enhancer co-accessible with the *CFTR* promoter, had allele-specific effects on enhancer activity and transcription factor binding in ductal cell models, and was associated with decreased *CFTR* expression in ductal cells. These two studies highlight the power of integrating comprehensive GWAS fine mapping with single cell epigenomics to understand how pancreatic cell types contribute to diabetes risk.

# Chapter 1: Introduction

## 1.1 Diabetes is a complex disease

Diabetes is a disease that affects over 400 million individuals worldwide. There are two major forms of diabetes: type 1 diabetes (T1D), characterized by autoimmune destruction of the insulin-producing beta cells in pancreatic islets[1], and type 2 diabetes (T2D), characterized by insulin resistance in peripheral tissues and gradual beta cell dysfunction[2]. Both forms of diabetes are complex diseases caused by a combination of environmental and genetic risk factors.

## 1.2 Genetics of diabetes

Genome-wide association studies (GWAS) of T2D have identified >400 independent risk signals mapping to 243 genomic loci[3], most of which have small effects on T2D risk. GWAS of T1D have identified >60 risk signals including the major histocompatibility complex (MHC) locus[4], the largest component of T1D risk. Genetic variants at these risk signals predominately map to non-coding regions and are often in high linkage disequilibrium (LD) with each other, making it challenging to pinpoint the causal variant. Bayesian fine mapping approaches[5] can assign causal probabilities to risk variants; however, they require well-powered GWAS with comprehensive variant coverage. Current genetic studies for T1D[6–8] are inadequate for fine mapping because they are not genome-wide or do not have sufficient variant coverage.

## 1.3 Diabetes-relevant cell types

Previous fine mapping studies have demonstrated that T2D risk variants are enriched within active enhancers of pancreatic islets[9–12]. Like many other autoimmune diseases, T1D risk variants are enriched within active enhancers for lymphoid cells[7]. In addition, a handful of T1D risk signals

have been suggested to affect beta cells within pancreatic islets through mechanisms such as fragility[13]. To date, studies integrating genetic fine mapping with epigenomic annotations have largely relied on bulk tissue annotations. With the advent of new technologies such as single nucleus ATAC-seq (snATAC-seq)[14,15], component cell types within a tissue can now be assayed for active regulatory elements. These assays, when combined with genetic fine mapping, have the potential to reveal disease-relevant cell types and cell type-specific mechanisms of risk variants. However, they have not been widely applied to study disease-relevant human tissues such as the pancreas. In this dissertation, I apply snATAC-seq assays to the pancreas to systemically study cell type-specific mechanisms of diabetes.

## 1.4 References


1. Katsarou, A., Gudbjörnsdottir, S., Rawshani, A., Dabelea, D., Bonifacio, E., Anderson, B. J., Jacobsen, L. M., Schatz, D. A. & Lernmark, Å. Type 1 diabetes mellitus. *Nat. Rev. Dis. Primer* **3**, 17016 (2017).

2. DeFronzo, R. A., Ferrannini, E., Groop, L., Henry, R. R., Herman, W. H., Holst, J. J., Hu, F. B., Kahn, C. R., Raz, I., Shulman, G. I., Simonson, D. C., Testa, M. A. & Weiss, R. Type 2 diabetes mellitus. *Nat. Rev. Dis. Primer* **1**, 15019 (2015).

3. Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M. H., Prins, B. P., Guo, X., Bielak, L. F., Below, J. E., Bowden, D. W., Chambers, J. C., Kim, Y. J., Ng, M. C. Y., Petty, L. E., Sim, X., Zhang, W., Bennett, A. J., Bork-Jensen, J., Brummett, C. M., Canouil, M., Ec Kardt, K.-U., Fischer, K., Kardia, S. L. R., Kronenberg, F., Läll, K., Liu, C.-T., Locke, A. E., Luan, J., Ntalla, I., Nylander, V., Schönherr, S., Schurmann, C., Yengo, L., Bottinger, E. P., Brandslund, I., Christensen, C., Dedoussis, G., Florez, J. C., Ford, I., Franco, O. H., Frayling, T. M., Giedraitis, V., Hackinger, S., Hattersley, A. T., Herder, C., Ikram, M. A., Ingelsson, M., Jørgensen, M. E., Jørgensen, T., Kriebel, J., Kuusisto, J., Ligthart, S., Lindgren, C. M., Linneberg, A., Lyssenko, V., Mamakou, V., Meitinger, T., Mohlke, K. L., Morris, A. D., Nadkarni, G., Pankow, J. S., Peters, A., Sattar, N., Stančáková, A., Strauch, K., Taylor, K. D., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Witte, D. R., Dupuis, J., Peyser, P. A., Zeggini, E., Loos, R. J. F., Froguel, P., Ingelsson, E., Lind, L., Groop, L., Laakso, M., Collins, F. S., Jukema, J. W., Palmer, C. N. A., Grallert, H., Metspalu, A., Dehghan, A., Köttgen, A., Abecasis, G. R., Meigs, J. B., Rotter, J. I., Marchini, J., Pedersen, O., Hansen, T., Langenberg, C., Wareham, N. J., Stefansson, K., Gloyn, A. L., Morris, A. P., Boehnke, M. & McCarthy, M. I. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).

4. Pociot, F. Type 1 diabetes genome-wide association studies: not to be lost in translation. *Clin. Transl. Immunol.* **6**, e162 (2017).

5. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet. Epidemiol.* **33**, 79–86 (2009).

6. Barrett, J. C., Clayton, D. G., Concannon, P., Akolkar, B., Cooper, J. D., Erlich, H. A., Julier, C., Morahan, G., Nerup, J., Nierras, C., Plagnol, V., Pociot, F., Schuilenburg, H., Smyth, D. J., Stevens, H., Todd, J. A., Walker, N. M., Rich, S. S., & Type 1 Diabetes Genetics Consortium. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).

7. Onengut-Gumuscu, S., Chen, W.-M., Burren, O., Cooper, N. J., Quinlan, A. R., Mychaleckyj, J. C., Farber, E., Bonnie, J. K., Szpak, M., Schofield, E., Achuthan, P., Guo, H., Fortune, M. D., Stevens, H., Walker, N. M., Ward, L. D., Kundaje, A., Kellis, M., Daly, M. J., Barrett, J. C., Cooper, J. D., Deloukas, P., Type 1 Diabetes Genetics Consortium, Todd, J. A., Wallace, C., Concannon, P. & Rich, S. S. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).

8. Bradfield, J. P., Qu, H.-Q., Wang, K., Zhang, H., Sleiman, P. M., Kim, C. E., Mentch, F. D., Qiu, H., Glessner, J. T., Thomas, K. A., Frackelton, E. C., Chiavacci, R. M., Imielinski, M., Monos, D. S., Pandey, R., Bakay, M., Grant, S. F. A., Polychronakos, C. & Hakonarson, H. A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci. *PLOS Genet.* **7**, e1002293 (2011).

9. Gaulton, K. J., Nammo, T., Pasquali, L., Simon, J. M., Giresi, P. G., Fogarty, M. P., Panhuis, T. M., Mieczkowski, P., Secchi, A., Bosco, D., Berney, T., Montanya, E., Mohlke, K. L., Lieb, J. D. & Ferrer, J. A map of open chromatin in human pancreatic islets. *Nat. Genet.* **42**, 255–259 (2010).

10. Gaulton, K. J., Ferreira, T., Lee, Y., Raimondo, A., Mägi, R., Reschen, M. E., Mahajan, A., Locke, A., Rayner, N. W., Robertson, N., Scott, R. A., Prokopenko, I., Scott, L. J., Green, T., Sparso, T., Thuillier, D., Yengo, L., Grallert, H., Wahl, S., Frånberg, M., Strawbridge, R. J., Kestler, H., Chheda, H., Eisele, L., Gustafsson, S., Steinthorsdottir, V., Thorleifsson, G., Qi, L., Karssen, L. C., van Leeuwen, E. M., Willems, S. M., Li, M., Chen, H., Fuchsberger, C., Kwan, P., Ma, C., Linderman, M., Lu, Y., Thomsen, S. K., Rundle, J. K., Beer, N. L., van de Bunt, M., Chalisey, A., Kang, H. M., Voight, B. F., Abecasis, G. R., Almgren, P., Baldassarre, D., Balkau, B., Benediktsson, R., Blüher, M., Boeing, H., Bonnycastle, L. L., Bottinger, E. P., Burtt, N. P., Carey, J., Charpentier, G., Chines, P. S., Cornelis, M. C., Couper, D. J., Crenshaw, A. T., van Dam, R. M., Doney, A. S. F., Dorkhan, M., Edkins, S., Eriksson, J. G., Esko, T., Eury, E., Fadista, J., Flannick, J., Fontanillas, P., Fox, C., Franks, P. W., Gertow, K., Gieger, C., Gigante, B., Gottesman, O., Grant, G. B., Grarup, N., Groves, C. J., Hassinen, M., Have, C. T., Herder, C., Holmen, O. L., Hreidarsson, A. B., Humphries, S. E., Hunter, D. J., Jackson, A. U., Jonsson, A., Jørgensen, M. E., Jørgensen, T., Kao, W.-H. L., Kerrison, N. D., Kinnunen, L., Klopp, N., Kong, A., Kovacs, P., Kraft, P., Kravic, J., Langford, C., Leander, K., Liang, L., Lichtner, P., Lindgren, C. M., Lindholm, E., Linneberg, A., Liu, C.-T., Lobbens, S., Luan, J., Lyssenko, V., Männistö, S., McLeod, O., Meyer, J., Mihailov, E., Mirza, G., Mühleisen, T. W., Müller-Nurasyid, M., Navarro, C., Nöthen, M. M., Oskolkov, N. N., Owen, K. R., Palli, D., Pechlivanis, S., Peltonen, L., Perry, J. R. B., Platou, C. G. P., Roden, M., Ruderfer, D., Rybin, D., van der Schouw, Y. T., Sennblad, B., Sigurðsson, G., Stančáková, A., Steinbach, G., Storm, P., Strauch, K., Stringham, H. M., Sun, Q., Thorand, B., Tikkanen, E., Tonjes, A., Trakalo, J., Tremoli, E., Tuomi, T., Wennauer, R., Wiltshire, S., Wood, A. R., Zeggini, E., Dunham, I., Birney, E., Pasquali, L., Ferrer, J., Loos, R. J. F., Dupuis, J., Florez, J. C., Boerwinkle, E., Pankow, J. S., van Duijn, C., Sijbrands, E., Meigs, J. B., Hu, F. B., Thorsteinsdottir, U., Stefansson, K., Lakka, T. A., Rauramaa, R., Stumvoll, M., Pedersen, N. L., Lind, L., Keinanen-Kiukaanniemi, S. M., Korpi-Hyövälti, E., Saaristo, T. E., Saltevo, J., Kuusisto, J., Laakso, M., Metspalu, A., Erbel, R., Jöcke, K.-H., Moebus, S., Ripatti, S., Salomaa, V., Ingelsson, E., Boehm, B. O., Bergman, R. N., Collins, F. S., Mohlke, K. L., Koistinen, H., Tuomilehto, J., Hveem, K., Njølstad, I., Deloukas, P., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., de Faire, U., Hamsten, A., Illig, T., Peters, A., Cauchi, S., Sladek, R., Froguel, P., Hansen, T., Pedersen, O., Morris, A. D., Palmer, C. N. A., Kathiresan, S., Melander, O., Nilsson, P. M., Groop, L. C., Barroso, I., Langenberg, C., Wareham, N. J., O'Callaghan, C. A., Gloyn, A. L., Altshuler, D., Boehnke, M., Teslovich, T. M., McCarthy, M. I., Morris, A. P., & DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).

11. Thurner, M., van de Bunt, M., Torres, J. M., Mahajan, A., Nylander, V., Bennett, A. J., Gaulton, K. J., Barrett, A., Burrows, C., Bell, C. G., Lowe, R., Beck, S., Rakyan, V. K., Gloyn,

A. L. & McCarthy, M. I. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *eLife* **7**, (2018).

12. Greenwald, W. W., Chiou, J., Yan, J., Qiu, Y., Dai, N., Wang, A., Nariai, N., Aylward, A., Han, J. Y., Kadakia, N., Regue, L., Okino, M.-L., Drees, F., Kramer, D., Vinckier, N., Minichiello, L., Gorkin, D., Avruch, J., Frazer, K. A., Sander, M., Ren, B. & Gaulton, K. J. Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nat. Commun.* **10**, 1–12 (2019).

13. Dooley, J., Tian, L., Schonefeldt, S., Delghingaro-Augusto, V., Garcia-Perez, J. E., Pasciuto, E., Di Marino, D., Carr, E. J., Oskolkov, N., Lyssenko, V., Franckaert, D., Lagou, V., Overbergh, L., Vandenbussche, J., Allemeersch, J., Chabot-Roy, G., Dahlstrom, J. E., Laybutt, D. R., Petrovsky, N., Socha, L., Gevaert, K., Jetten, A. M., Lambrechts, D., Linterman, M. A., Goodnow, C. C., Nolan, C. J., Lesage, S., Schlenner, S. M. & Liston, A. Genetic predisposition for beta cell fragility underlies type 1 and type 2 diabetes. *Nat. Genet.* **48**, 519–527 (2016).

14. Cusanovich, D. A., Daza, R., Adey, A., Pliner, H., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. & Shendure, J. Multiplex Single Cell Profiling of Chromatin Accessibility by Combinatorial Cellular Indexing. *Science* **348**, 910–914 (2015).

15. Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D. U., Zhang, Y., Sos, B. C., Afzal, V., Dickel, D. E., Kuan, S., Visel, A., Pennacchio, L. A., Zhang, K. & Ren, B. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).

# Chapter 2: Single cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk

## 2.1 Abstract

Single nucleus ATAC-seq (snATAC-seq) creates new opportunities to dissect cell type-specific mechanisms of complex diseases. As pancreatic islets are central to type 2 diabetes (T2D), we profiled 15,298 islet cells using combinatorial barcoding snATAC-seq and identified 12 clusters, including multiple alpha, beta and delta cell states. We cataloged 228,873 accessible chromatin sites and identified transcription factors underlying lineage- and state-specific regulation. We observed state-specific enrichment of fasting glucose and T2D GWAS for beta cells as well as enrichment for other endocrine cell types. At T2D signals localized to islet accessible chromatin, we prioritized variants with predicted regulatory function and co-accessibility with target genes. A causal T2D variant rs231361 at the *KCNQ1* locus had predicted effects on a beta cell enhancer co-accessible with INS, and genome editing in embryonic stem cell-derived beta cells affected INS levels. Together our findings demonstrate the power of single cell epigenomics for interpreting complex disease genetics.

## 2.2 Introduction

Gene regulatory programs are largely orchestrated by *cis*-regulatory elements that direct the expression of genes in response to specific developmental and environmental cues. Genetic variants associated with complex disease are highly enriched within putative *cis*-regulatory elements[1]. The activity of regulatory elements is often restricted to specific cell types and/or cell states, limiting the ability of ATAC-seq and other "ensemble" (or "bulk") epigenomic technologies to map regulatory elements in individual cell types within disease-relevant tissues. To overcome this limitation, new approaches to obtain ATAC-seq profiles from single nuclei allow for the disaggregation of open chromatin from heterogenous samples into component cell types and subtypes[2–4]. These developments create opportunities to dissect the molecular mechanisms that underlie genetic risk of disease. However, to date snATAC-seq data from disease-relevant human tissues are limited[5–8].

Type 2 diabetes (T2D) is a multifactorial disease with a highly polygenic inheritance[9]. Pancreatic islets are central to genetic risk of T2D, as evidenced by shared association between T2D risk and quantitative measures of islet function[10–13] and enrichment of T2D risk variants in islet regulatory sites[14–18]. Islets are comprised of multiple endocrine cell types with distinct functions[19–21] and are heterogeneous[22–24] in gene expression and other molecular signatures which likely reflect different functional cell states[22,25,26]. Heterogeneity in the epigenome of islet cell types has not been described, however, which is necessary to understand islet regulation and interpret the molecular mechanisms of non-coding T2D risk variants. In this study, we map accessible chromatin profiles of individual islet cells using snATAC-seq, define the regulatory programs of islet cell types and cell states, describe their relationship to T2D risk and fasting glycemia, and predict the molecular mechanisms of T2D risk variants.

## 2.3 Results

### 2.3.1 Islet snATAC-seq reveals 12 distinct cell clusters

We performed snATAC-seq on human pancreatic islets from three donors using a combinatorial barcoding snATAC-seq approach optimized for use on tissues[2,4] (**Table 2.1**). To confirm library quality, we first analyzed the data as ensemble ATAC-seq by aggregating all high-quality mapped reads irrespective of barcode. Ensemble snATAC-seq from all three samples showed the expected insert size distribution (**Figure S2.1a**), strong enrichment of signal at transcription start sites (TSS) (**Figure S2.1b**), and high concordance of signal with published islet ATAC-seq data[14,27–29] (**Figure S2.1c**).

To obtain high-quality single cell profiles, we first filtered out barcodes with less than 500 reads (**Figure S2.1d**), resulting in a total of 18,544 cells across the three samples. We then clustered accessible chromatin profiles from these cells, making key modifications to previous approaches[4] (see **Figure S2.2**). After filtering out cells with aberrant quality metrics, we retained 15,298 cells which mapped to 12 clusters (**Figure 2.1a**). To determine the cell type represented by each cluster, we examined chromatin accessibility at the promoter region of the cognate hormone genes for endocrine cells and known marker genes for non-endocrine cell types. We identified clusters representing beta (*INS-IGF2*/insulin), alpha (*GCG*/glucagon), delta (*SST*/somatostatin), gamma (*PPY*/pancreatic polypeptide), ductal (*CFTR*), acinar (*REG1A*), immune (*NCF2*)[30], stellate (*PDGFRB*)[30], and endothelial (*CD93*)[31] cells (**Figure 2.1b-c**). We defined a broader set of marker genes for each cluster by identifying promoters with accessibility most specific to each cluster. We observed highly specific correlations between marker genes defining cell types in snATAC-seq and islet scRNA-seq[23] (**Figure 2.1d**, **Figure S2.3a-e**).

To characterize regulatory programs in each cell type, we aggregated reads for cells within each cluster and identified accessible chromatin sites for the cluster using MACS2[32]. In total, we identified 228,873 accessible chromatin sites merged across the 12 clusters. We next used

chromVAR[33] to identify TF motifs from JASPAR[34] enriched within accessible chromatin of each cell. Analysis of motif enrichments averaged across cells for each cell type revealed cell type-specific patterns of motif enrichment (**Figure 2.1e, Table 2.2**). For example, we observed enrichment of PDX1 in beta and delta cells[35], MAF in alpha and beta cells[36–38], IRF in immune cells[39] and ETS in endothelial cells[40] (**Figure 2.1e**). Hierarchical clustering of cell types based on motif enrichment revealed that regulatory programs of beta and delta cells were closely related as were alpha and gamma cells (**Figure 2.1f**), consistent with single cell expression data[30,41,42]. Motifs highly enriched in delta cell chromatin relative to beta cells included SCRT (SCRT1 -$\log_{10}$(FDR)=86.49) and ELF TFs (ELF5 -$\log_{10}$(FDR)=79.41), and motifs enriched in gamma cell chromatin relative to alpha cells included HOX (Hoxa9 -$\log_{10}$(FDR)=20.92) and IRF TFs (IRF1 -$\log_{10}$(FDR)=20.22) (**Figure 2.1h,g, Figure S2.4**).

### 2.3.2 Heterogeneity in islet endocrine cell regulatory programs

A major strength of single cell approaches is the ability to reveal heterogeneity within a cell type. Indeed, our initial clustering showed that alpha, beta and delta cells segregated into sub-clusters. We identified gene promoters with variable accessibility between sub-clusters (see Methods). Notably, *INS* had amongst the most variable promoter accessibility between beta cell sub-clusters (*INS-IGF2* beta OR=4.74, two-sided Fisher's exact $p=1.78\times10^{-40}$), and therefore for clarity we renamed the clusters $INS^{high}$ and $INS^{low}$ beta cells, respectively (**Figure 2.1b,c; Figure 2.2a**). Similarly, *GCG* promoter accessibility was variable between alpha cell sub-clusters (*GCG* alpha OR=3.67, $p=3.45\times10^{-22}$), which we renamed $GCG^{high}$ and $GCG^{low}$ alpha cells, and *SST* promoter accessibility was variable between delta cell sub-clusters (*SST* delta OR=1.86 p=0.02), which we renamed $SST^{high}$ and $SST^{low}$ delta cells (**Figure 2.1b-c; Figure 2.2a**).

We found significant overlap in the genes that distinguish hormone-high from hormone-low alpha, beta and delta cells by gene set enrichment analysis (GSEA) (**Figure 2.2b**). Genes with increased promoter accessibility in hormone-high states were enriched for hormone secretion

and glucose response; in contrast, genes with increased promoter accessibility for hormone-low states were enriched for stress-induced signaling response (**Figure 2.2a,c**). We also observed enriched TF motifs that distinguished different states (**Figure 2.2d**, **Table 2.3**). For example, RFX family motifs were enriched in hormone-high but not in hormone-low states (Rfx1 mean enrichment $INS^{high}$=0.36, $INS^{low}$=-0.95, p=0; $GCG^{high}$=0.52, $GCG^{low}$=-1.16, $p=7.3\times10^{-260}$; $SST^{high}$=0.76, $SST^{low}$=-1.24, $p=3.9\times10^{-58}$) (**Figure 2.2d**). In contrast, FOS/JUN family motifs were prominently enriched in hormone-low but not hormone-high states (FOS::JUN mean enrichment $INS^{high}$=-1.78, $INS^{low}$=3.90, p=0; $GCG^{high}$=-2.86, $GCG^{low}$=5.50, p=0; $SST^{high}$=-0.21, $SST^{low}$=7.62, $p=1.1\times10^{-121}$) (Figure 2d). These data reveal epigenomic differences between endocrine cell states among genes involved in hormone production and stress-induced signaling responses and point to an underlying commonality in regulatory networks that govern state-specific functions of endocrine cell types.

We next sought to determine whether the observed heterogeneity in the epigenome of endocrine cells correlated with heterogeneity in islet gene expression and function. We first compared our states to beta cell sub-clusters from a previous scRNA-seq study[23]. Genes with increased promoter accessibility in hormone-low cells were enriched in a beta cell sub-cluster (β-sub.4) associated with ER stress and protein folding and low *INS* expression, whereas genes with increased promoter accessibility in hormone-high cells were enriched in the other beta cell sub-clusters (β-sub.1-3) (**Figure 2.2b**). We further found enrichment of genes with differential promoter accessibility among gene sets preferentially expressed in beta cell sub-clusters from a recent scRNA-seq meta-analysis[43] (**Figure 2.2b**). Finally, we found significant overlap in genes with differential promoter accessibility between states and genes that correlate with electrophysiological measures of beta cell function from a recent Patch-seq study[44] (**Figure 2.2b**). These results thus provide a link between epigenomic heterogeneity in endocrine cells and heterogeneity in gene expression and electrophysiological function.

To explore potential gradations among endocrine cells as a continuum rather than as binary states[23,45], we used Cicero[8] to order cells from each cell type along trajectories based on chromatin accessibility. We observed cells on a gradient between hormone-high and hormone-low states (**Figure 2.2e, Figure S2.5a-c**). These trajectories allowed us to examine gene promoter accessibility and TF motif enrichment as a function of pseudo-state (**Figure 2.2e**, **Figure S2.5d**). Consistent with binary sub-clusters, lineage-specifying genes and TF family motif enrichments such as RFX and NFAT decreased along the trajectory from hormone-high to -low cells, whereas motif enrichment in TF families such as FOS/JUN increased (**Figure 2.2e**). Structurally-related TFs often have similar motifs, and thus to assign motifs to specific TFs we correlated promoter-accessibility of TFs within the structural subfamily with motif enrichments across the state trajectory. Motif enrichment for the FOS/JUN family correlated with the promoter accessibility of *FOSL1, FOSL2* and *JUN* (**Figure 2.2f**), supporting a role for these specific TFs in hormone-low cell regulation.

### 2.3.3 Islet cell type enrichment for diabetes and glycemia GWAS

Genetic variants influencing diabetes and fasting glucose level are enriched in pancreatic islet regulatory elements[14,17,18,46]. Using our islet cell type- and state-resolved accessible chromatin profiles, we determined the enrichment of variants associated with diabetes[9,47] and related quantitative phenotypes[11,12,48–51] as well as other complex traits[52–59]. We first determined the enrichment of variants in accessible chromatin sites for each cell type and state using stratified LD score regression[60–62]. We observed significant enrichment (FDR<0.1) of fasting glucose (FG) level association in $INS^{high}$ beta cells and T2D association for both $INS^{high}$ and $INS^{low}$ beta cell states (FG $INS^{high}$ Z=3.58 FDR=0.013; T2D $INS^{high}$ Z=4.41 FDR=0.001, $INS^{low}$ Z=4.19 FDR=0.002) (**Figure 2.3a**). We also observed more nominal evidence for enrichment of T2D association for $GCG^{low}$ alpha cells and both delta cell states, as well as multiple glycemic traits for endocrine cells (**Figure 2.3a**).

In these analyses, we observed differences in enrichments between $INS^{high}$ and $INS^{low}$ beta cells for fasting glucose (**Figure 2.3a**). To further resolve the heterogeneity of genetic association enrichment, we tested enrichment of T2D and fasting glucose as well as several negative control traits within single cell profiles (**Figure 2.3b, Figure S2.6a**). We observed marked heterogeneity among beta cells for fasting glucose-associated variants, where $INS^{high}$ cells had significantly stronger enrichment ($INS^{high}$ median Z=2.58, $INS^{low}$ median Z=0.68, $p=1.19\times10^{-225}$) (**Figure 2.3b**). We further calculated the average enrichment for cells binned across 'pseudostate' (see **Figure 2.2e**), which revealed decreasing enrichment from $INS^{high}$ to $INS^{low}$ beta cells (**Figure 2.3b**). In contrast, for T2D we observed consistent enrichment across $INS^{high}$ and $INS^{low}$ beta cells ($INS^{high}$ median Z=0.93, $INS^{low}$ median Z=0.91, p=0.44) (**Figure 2.3b**). As many variants affecting fasting glucose also affect T2D[12,13,63], we grouped T2D loci associated with fasting glucose ('T2D/FG' loci) and tested these loci for enrichment of $INS^{high}$ and $INS^{low}$ beta cell sites using fgwas[64]. We observed strong enrichment of T2D/FG loci for $INS^{high}$ beta cells only ($INS^{high}$ beta ln(enrichment)=4.54, $INS^{low}$ beta ln(enrichment)=-25.7), suggesting that T2D loci affecting glucose levels have state-specific effects (**Figure 2.3c**). For example, at the *ADCY5* locus associated with both fasting glucose and T2D, variant rs11708067:A>G (PPA=0.79) overlapped a site specific to $INS^{high}$ beta cells (**Figure 2.3d**).

Outside of beta cells we also observed evidence for enrichment of T2D association for variants in chromatin sites for other endocrine cell states including $GCG^{low}$ alpha cells and $SST^{high}$ and $SST^{low}$ delta cells (**Figure 2.3a**). In order to understand the potential role of these cell states in T2D risk, we performed additional enrichment analysis at known T2D risk loci using fgwas[64]. We tested alpha, delta and gamma sites not overlapping beta cell sites for enrichment at fine-mapped T2D loci from DIAMANTE. Here we again observed enrichment of T2D association for $GCG^{low}$ alpha cells (ln(enrichment)=1.75) as well as $SST^{high}$ and $SST^{low}$ delta cells (ln(enrichment)=0.86, 1.30) (**Figure 2.3e**). In further support of the likely role of these cell types

in T2D, several fine-mapped risk variants overlapped sites specific to these cell types; for example, rs1111875:C>T (PPA=0.16) mapped in a delta cell-specific site at the *HHEX* locus (**Figure 2.3f**).

Given our ability to map both complex trait and TF motif enrichments to single cells, we reasoned that joint analysis could provide insight into TFs regulating trait-relevant chromatin. We correlated single cell fasting glucose and T2D enrichment z-scores with TF motif enrichments from chromVAR[33]. Across all 15.3k cells, we observed positive correlation between fasting glucose and T2D enrichment and motifs for beta cell TFs such as PDX (**Figure 2.3g, Figure S2.6b**). Across the 7.6k beta cells only, we observed strongest positive correlation between fasting glucose and motifs in TF families enriched for $INS^{high}$ beta cells such as bHLH (NEUROD1 $\rho$=0.21, $p=9.43\times10^{-78}$) and RFX (Rfx1 $\rho$=0.21, $p=8.83\times10^{-74}$) (**Figure 2.3g**). For T2D, strongest positive correlations included motifs for TF families such as RFX (Rfx1 $\rho$=0.052, $p=3.72\times10^{-9}$), NFAT (NFATC2 $\rho$=0.047, $p=4.69\times10^{-5}$), and MEF2 (MEF2A $\rho$=0.062, $p=6.99\times10^{-8}$) (**Figure S2.6b**). We further determined whether TF motifs preferentially harbored associated variants directly. For fasting glucose, we identified strongest enrichment for $INS^{high}$ state-specific TF motifs, most notably RFX (RFX2 $p=1.3\times10^{-10}$) and NEUROD (p=0.049). For T2D, we observed enrichment for TF motifs positively correlated with T2D association including RFX (Rfx1 $p=4.0\times10^{-15}$), NFAT (NFATC2 $p=2.2\times10^{-4}$), and MEF2 (MEF2D $p=1.9\times10^{-3}$) (**Figure S2.6c**). These motifs remained significantly enriched for T2D when considering only variants predicted to disrupt the motif (Rfx1 $p=1.1\times10^{-9}$, NFATC2 $p=2.2\times10^{-4}$, MEF2D $p=7.7\times10^{-5}$).

Together these results provide state-resolved insight into the role of beta cells and their TFs in both T2D risk and fasting glucose level and implicate other endocrine cell types in T2D risk.

### 2.3.4 Predictions of variant effects on islet cell type chromatin

Predicting the effects of non-coding genetic variants on regulatory activity remains a major challenge, in large part because the sequence vocabularies that encode regulatory function differ between cell types and states. We therefore used deltaSVM[65] to predict the effects of genetic variants from the Haplotype Reference Consortium panel[66] on chromatin accessibility in each endocrine cell type and cell state. We identified 432,072 variants genome-wide with predicted allelic effects (FDR<0.1), encompassing between 115k-161k variants (7.8%-10.9% of tested variants) per cell type or state (**Figure 2.4a**).

To validate that our predictions captured true allelic effects on islet chromatin accessibility, we first compared alpha and beta cell predictions to direct measurements of allelic effects on chromatin accessibility. We found significant correlations between predicted allelic effects and allelic imbalance estimates for all alpha and beta cell states ($GCG^{high}$ Spearman $\rho=0.255$, $p=1.20\times10^{-34}$, $GCG^{low}$ $\rho=0.214$, $p=2.35\times10^{-7}$, $INS^{high}$ $\rho=0.275$, $p=1.03\times10^{-34}$, $INS^{low}$ $\rho=0.334$, $p=4.73\times10^{-38}$) (**Figure 2.4b**). We further validated five likely causal T2D variants predicted to affect beta cell chromatin which had directionally consistent effects on enhancer activity using gene reporter assays in the MIN6 beta cell line (**Figure 2.4c**). We also compared predictions to islet chromatin accessibility quantitative trait loci (caQTLs)[28], and observed significant enrichment of caQTLs among variants with predicted effects in alpha or beta cells (observed=38.4%, expected=19.7%, two-sided Fisher's exact $p=2.78\times10^{-97}$) (**Figure 2.4d**). When sub-dividing predictions based on shared, cell type-specific or state-specific effects we observed significant enrichment of caQTLs only among shared effect variants (**Figure 2.4d**).

We further characterized genetic variants predicted to have cell type- and state-dependent effects on islet chromatin. Variants with state-specific effects tended to disrupt motifs for TF families such as NEUROD and RFX for hormone-high states ($-\log_{10}(p)=59.0, 24.5$) (**Figure 2.4e**). Similarly, variants with alpha or beta cell-specific effects tended to disrupt motifs for lineage-

defining TF families including GATA for alpha cells ($-\log_{10}(p)=21.6$), and NKX6 and PDX1 for beta cells ($-\log_{10}(p)=11.2, 10.3$) (**Figure 2.4e**). To assign motifs to specific TFs, we examined promoter-accessibility of TFs within the structural TF subfamily[67]. Among GATA subfamily members only GATA6 had high promoter accessibility in alpha cells ($GCG^{high}=1.00$, $GCG^{low}=0.96$, $INS^{high}=0.22$, $INS^{low}=0.14$), suggesting that GATA6 binding is likely disrupted in alpha cells. Similarly, among RFX family members RFX6 had promoter accessibility in hormone-high state cells ($GCG^{high}=0.92$, $GCG^{low}=0.70$, $INS^{high}=0.93$, $INS^{low}=0.80$) (**Figure 2.4e**).

We evaluated whether our predictions could prioritize lower frequency (defined as minor allele frequency MAF<0.05) functional variants involved in T2D risk. We observed enrichment of genome-wide significant T2D associations among lower frequency variants with predicted effects in any endocrine cell type compared to background (**Figure 2.4f**). When considering each cell type, we observed enrichment of T2D association among variants with predicted effects in beta and delta cells, even at sub-genome-wide significant p-values (**Figure 2.4f**). For example, at the *IGF2BP3* locus, rs78840640:C>G (MAF=0.02) had allelic effects on beta cell chromatin ($INS^{high}$ beta FDR=$1.15\times10^{-4}$; $INS^{low}$ beta FDR=$6.93\times10^{-3}$), and fine-mapping supported a causal role in T2D (PPA=0.33) (**Figure 2.4g**). This variant affected enhancer activity in gene reporter assays where the alternate (and T2D risk) allele G had reduced activity (**Figure 2.4c**). These results reveal that cell type-specific chromatin can provide accurate functional predictions of lower frequency variants.

### 2.3.5 Co-accessibility links regulatory variants to target genes

Defining the genes affected by regulatory element activity remains a major challenge, as enhancers can regulate gene activity over large, non-adjacent distances[68]. A number of approaches have been developed to link regulatory elements to target genes[69,70], but are not typically cell type-resolved[27,71]. Recently, a new approach was developed to link regulatory elements at cell type resolution based on co-accessibility across single cells[8]. We thus leveraged

single cell accessible chromatin profiles to define co-accessibility between accessible chromatin sites in alpha, beta and delta cells.

To calibrate the extent to which co-accessibility reflected physical interactions between regulatory elements, we performed a distance-matched comparison between co-accessible sites stratified by co-accessibility threshold and chromatin loops identified from Hi-C and promoter capture Hi-C (pcHi-C) in primary islets[27,71]. We observed strong enrichment for sites with co-accessibility scores >0.05 for islet chromatin loops compared to non-co-accessible sites (**Figure 2.5a, Figure S2.7a-e**). We therefore used this threshold (0.05) to define co-accessible sites. Among co-accessible sites were 47,871 (alpha), 46,036 (beta) and 42,234 (delta) distal sites co-accessible with a gene promoter (**Figure S2.7f-g**), and the majority (71.9%) were cell type-specific (**Figure S2.8a-c**). For example, the *PDX1* promoter was co-accessible with 31 sites in beta and 39 sites in delta cells, including sites that directly coincided with islet pcHi-C, only 7 of which were in alpha cells (**Figure 2.5b**).

Distal sites with co-accessibility links to gene promoters harbored risk variants for T2D at many loci. At the *KCNQ1* locus, an islet chromatin site located in intron 3 of *KCNQ1* had beta cell-specific co-accessibility with the *INS* promoter over 500 kb distal and harbored a causal T2D variant rs231361:G>A (PPA=1) (**Figure 2.5c**). The site containing rs231361 was more accessible in $INS^{high}$ compared to $INS^{low}$ beta cells, and rs231361 was predicted to have state-specific effects on beta cell chromatin accessibility ($INS^{high}$ beta FDR=0.060; $INS^{low}$ beta FDR=0.40). Furthermore, rs231361 disrupted an *RFX* motif, which itself was enriched in $INS^{high}$ beta cells (**Figure 2.5c**). Published 4C data from the EndoC-βH1 human beta cell line[72] revealed physical proximity between this site and the *INS* promoter (**Figure S2.9**), but there was not similar evidence in Hi-C or other 3C-based data from EndoC-βH1 and primary islets[27,71,73].

Although we observed physical proximity between rs231361 and the *INS* promoter in beta cells, the absence of a canonical chromatin loop necessitated further validation. We therefore

deleted a 2.6 kb region flanking the site in hESCs by CRISPR/Cas9-mediated genome editing in three bi-allelic clones (*KCNQ1*$^{\Delta Enh}$) (**Figure 2.5c, Figure S2.10a,b**), and differentiated *KCNQ1*$^{\Delta Enh}$ clones and unedited control clones into beta cells using a modified version of an established protocol[74]. Analysis of cultures at the beta cell stage revealed similar numbers of INS$^{+}$/NKX6-1$^{+}$ cells in *KCNQ1*$^{\Delta Enh}$ and control clones (47.1±13.4% vs 56.5±7.6%) (**Figure S2.10c**), suggesting that the enhancer deletion had no effect on beta cell differentiation. We determined effects of the enhancer deletion on expression of all genes within 2 Mb of the enhancer and observed a significant decrease in the expression of *INS* ($p=3.02\times10^{-4}$; FDR=0.066) and *CDKN1C* ($p=1.96\times10^{-4}$; FDR=0.059) in *KCNQ1*$^{\Delta Enh}$ compared to control cells, and not for other genes (all P>.05) (**Figure 2.5d**). Analysis of INS protein levels by immunofluorescence staining, flow cytometry, and ELISA further revealed reduced INS protein abundance in *KCNQ1*$^{\Delta Enh}$ beta cells (**Figure 2.5e-g**).

To next determine whether rs231261 itself had distal effects on *INS* regulation in beta cells, we used targeted base editing to generate hESC lines that were homozygous for either the major allele G (*KCNQ1*$^{G/G}$, two clones) or the minor (and T2D risk) allele A (*KCNQ1*$^{A/A}$, three clones) (**Figure S2.10d-f**). We then differentiated the *KCNQ1*$^{A/A}$ and *KCNQ1*$^{G/G}$ clones into beta cells using the same protocol as for *KCNQ1*$^{\Delta Enh}$ with additional modifications. We determined effects of the variant alleles on expression of the two genes *INS* and *CDKN1C* with significant changes in the enhancer deletion (in **Figure 2.5d**) using qPCR. We observed a significant difference in *INS* expression between alleles (p=0.022), as well as evidence for a difference in *CDKN1C* expression although this was not significant (p>0.05) (**Figure 2.5h**). Analysis of insulin protein levels by flow cytometry and ELISA revealed significant differences in insulin abundance between variant alleles (**Figure 2.5i,j**).

## 2.4 Discussion

Over 400 known risk signals for T2D have been identified, yet only a handful have been characterized molecularly[16,17,27,75–81]. Our findings provide a roadmap demonstrating how single cell accessible chromatin derived from disease-relevant primary tissue can be utilized to define cell types, cell states, *cis* regulatory elements and genes involved in the genetic basis of T2D and other complex disease.

The *KCNQ1* locus has a complex contribution to T2D involving at least 10 independent signals[9]. Among these was causal T2D variant rs231361, where genome editing in hESC beta cell models revealed effects on insulin transcript and protein levels. Chromatin conformation from 4C in EndoC-βH1 cells[72] revealed physical proximity between the enhancer and *INS* promoter, although we did not find corresponding evidence for an interaction in other 3C-based data from EndoC-βH1 cells[73], hESC beta cells or primary islets[27,71]. Thus, while our results support a possible *cis*-regulatory effect of rs231361 on *INS* expression, we cannot currently rule out that the observed effects instead occur in *trans*, secondary to other effects. We anticipate that future studies to resolve phase between variant alleles and *INS* allelic expression in heterozygous samples will further clarify the nature of these effects.

Single cell accessible chromatin uncovered heterogeneity in the regulatory programs of endocrine cell types, pointing to TFs that likely drive cell state-specific functions. Integrating single cell heterogeneity with large-scale genetic association data revealed that genetic variants modulating fasting glucose levels likely act through the high insulin-producing beta cell state, whereas genetic risk of T2D is mediated through effects on both the high insulin-producing state and other functional beta cell state(s) likely related to stress and signaling responses. Moreover, given similar heterogeneity in the epigenomes of alpha and delta cells, our results reveal that endocrine cell regulation involves both lineage-specific programs as well as an additional layer of state-specific programs common across endocrine cell types. Our results also implicate these

other endocrine cell types in genetic risk of T2D independent of beta cells, most prominently delta cells.

Previous studies have characterized heterogeneity in beta cell physiological function, cell surface markers, and gene expression[22,44,82–84]. We found that heterogeneity in the beta cell epigenome mapped to cellular states related to insulin production and stress-related signaling response and was tightly linked to heterogeneity in beta cell gene expression[23] and electrophysiology[44]. However, there is often not perfect correspondence between sub-clusters identified by different techniques and/or studies, and we anticipate that using multi-omic methods will help to further clarify these differences. Regardless of the technology used, heterogeneity defined from single cell data is by nature dependent on computational clustering or ordering. Separating true heterogeneity from other sub-structure therefore ultimately requires experimental validation, for example by profiling cell populations sorted using sub-type markers. As the sub-clusters described in our study have not been linked directly to functional differences, experiments will be necessary to determine the relationship between epigenomic heterogeneity and cellular function. Furthermore, while we observed heterogeneity in endocrine cells from cryopreserved tissue in addition to purified islets, determining the true extent and nature of heterogeneity *in vivo* will require more extensive studies across a broader range of samples and conditions.

In summary, we present the most detailed characterization of islet cell type and state regulatory programs to date. When combined with genetic fine-mapping and whole genome sequencing, as well as additional cell type-specific data in islets[85], this resource will greatly enhance efforts to define molecular mechanisms of T2D risk. More broadly, our study provides a framework for using single cell chromatin from disease-relevant tissues to interpret the genetic mechanisms of complex disease.

## 2.5 Methods

### 2.5.1 Islet processing and nuclei isolation

We obtained islet preparations for three donors from the Integrated Islet Distribution Program (**Table 2.1**). Islet preparations were further enriched using zinc-dithizone staining followed by hand picking. For experiments performed with whole pancreas tissue, a cryopreserved tissue sample was obtained from the Network for Pancreatic Organ Donors with Diabetes (nPOD) biorepository. Studies were given exempt status by the Institutional Review Board (IRB) of the University of California San Diego.

### 2.5.2 Generation of snATAC-seq libraries

Combinatorial barcoding single nucleus ATAC-seq was performed as described previously[2,4] with several modifications. Libraries were quantified using a Qubit fluorimeter (Life technologies) and the nucleosomal pattern was verified using a Tapestation (High Sensitivity D1000, Agilent). Libraries were sequenced on a HiSeq 2500 sequencer (Illumina) using custom sequencing primers, 25% spike-in library and following read lengths: 50+43+40+50 (Read1+Index1+Index2+Read2).

Droplet-based single cell ATAC-seq libraries were generated using the Chromium Chip E Single Cell ATAC kit (10x Genomics, 1000086) and indexes (Chromium i7 Multiplex Kit N, Set A, 10x Genomics, 1000084) following manufacturer instructions. Final libraries were quantified using a Qubit fluorimeter (Life Technologies) and the nucleosomal pattern was verified using a Tapestation (High Sensitivity D1000, Agilent). Libraries were sequenced on a NextSeq 500 and HiSeq 4000 sequencer (Illumina) with following read lengths: 50+8+16+50 (Read1+Index1+Index2+Read2).

### 2.5.3 Raw data processing and quality control

For each read, we first appended the cell barcode metadata to the read name. The cell barcode consisted of four pieces (P7, I7, I5, P5) which were derived from the index read files. We

first corrected for sequencing errors by calculating the Levenshtein distance between each of the four pieces and a whitelist of possible sequences. If the piece did not perfectly match a whitelisted sequence, we took the best matching sequence if it was within 2 edits and the next matching sequence was at least 2 additional edits away. If none of these conditions were met, we discarded the read from further analyses.

### 2.5.4 Cluster analysis for snATAC-seq

We split the genome into 5kb windows and removed windows overlapping blacklisted regions (v2) from ENCODE[86,87]. For each experiment, we created a sparse *m* x *n* matrix containing read depth for *m* cells passing read depth thresholds at *n* windows. Using scanpy[88] (v.1.4.4.post1), we extracted highly variable windows using mean read depth and normalized dispersion ('min_mean=0.01, min_disp=0.25'). After normalization to uniform read depth and log-transformation, for each experiment, we regressed out the log-transformed read depth within highly variable windows for each cell. We then performed PCA and extracted the top 50 principal components. We used Harmony[24] to correct the principal components and remove batch effects across experiments, using donor-of-origin as a covariate. We used Harmony-corrected components to calculate the nearest 30 neighbors using the cosine metric, which were subsequently used for UMAP dimensionality reduction ('min_dist=0.3') and Leiden clustering[89] ('resolution=1.5').

We performed iterative clustering to identify and remove cells with abnormal features prior to the final clustering results. After removing these cells, we ended up with 15,298 cells mapping to 12 clusters. We used chromatin accessibility at windows overlapping promoters for marker hormones to assign cell types for the endocrine islet cell types and chromatin accessibility at windows around marker genes from scRNA-seq to assign cluster labels for non-endocrine islet clusters.

### 2.5.5 Comparison to bulk and sorted islet ATAC-seq

We processed sequence data of bulk islet ATAC-seq[14,27–29] and bulk pancreas from ENCODE[86]. We calculated the Spearman correlation between normalized read coverages and used hierarchical clustering to assess similarity between bulk islet samples. To check peak call overlap between aggregated snATAC-seq and bulk ATAC-seq, we split peaks based into promoter proximal (±500 bp from TSS) and distal peaks based on promoter overlap. For each cluster, we calculated the percentage of aggregate peaks that overlapped merged autosomal bulk peaks and individual sample-level autosomal bulk peaks. We processed raw sequence data of ATAC-seq from flow-sorted pancreatic cells from two prior studies[90,91]. We calculated Spearman correlations between read coverages and used hierarchical clustering to assess similarity between sorted and snATAC-seq samples.

### 2.5.6 Identifying marker peaks of chromatin accessibility

To identify peaks for each cell type, we aggregated reads for all cells within a cluster. We shifted reads aligning to the positive strand +4bp and reads aligning to the negative strand -5bp, extended reads to 200 bp, and centered the reads. We used MACS2[32] to call peaks for each cluster ('--nomodel --keep-dup all'). We removed peaks that overlapped the ENCODE blacklist v2[86,87]. We then used bedtools[92] (v.2.26.0) to merge peaks from clusters and create a set of merged peaks. We generated a sparse *m* x *n* matrix containing binary overlap between *m* peaks in the merged set of islet regulatory peaks and *n* cells. We calculated t-statistics of peak specificity for each cluster using linear regression models. For each peak and each cluster, we used binary encoding of read overlap with the peak as the outcome and whether a cell was in the cluster as the predictor, and included the log read depth of each cell as a covariate in the model.

### 2.5.7 Matching islet snATAC-seq with scRNA-seq clusters

To verify that clusters definitions and labels from single cell chromatin accessibility data matched those from single cell expression data, we obtained published scRNA-seq data from 12

non-diabetic islet donors[23]. Because cluster definitions for all cell types were not available, we re-analyzed the data and performed clustering analyses. We used the Spearman correlation between t-statistics from islet snATAC-seq and islet scRNA-seq data to verify cluster labels.

### 2.5.8 Single cell motif enrichment

We used chromVAR[33] (v.1.5.0) to estimate TF motif enrichment z-scores for each cell. First, we constructed a merged peak by cell sparse binary matrix as the input for chromVAR. We corrected for GC bias based on hg19 (BSgenome.Hsapiens.UCSC.hg19) using the 'addGCBias' function. For TF motifs within the non-redundant JASPAR 2018 CORE vertebrate set, we calculated bias-corrected deviation z-scores for each cell. Across all cell types, we selected variable TF motifs (N=141) with variability>1.2. For each cell type, we then computed the average TF motif enrichment z-score across single cells, collapsing values for cell types with more than one state. We compared motif enrichment z-scores between single cells using Benjamini-Hochberg corrected p-values (FDR<0.01) from 2 sample T-tests.

### 2.5.9 Ordering alpha, beta, and delta cells along a pseudo-state trajectory

We used Cicero[8] (v.1.3.3 with Monocle 3) to order alpha, beta, and delta cells along separate trajectories. Starting from the merged peak by cell sparse binary matrix, we extracted beta cells and filtered out peaks that were not present in beta cells. We then preprocessed the data with latent semantic indexing (LSI) and continued onto dimensionality reduction, cell clustering, and trajectory graph learning using default parameters. We then chose the root state (i.e. the start of the trajectory) based on the highest average *INS* promoter accessibility. We repeated the same procedure for beta and delta cells, instead choosing the root state based on *GCG* or *SST* promoter accessibility respectively.

### 2.5.10 Comparison of endocrine cell states

To identify TF motifs variable between cell states, we performed one-sided Student's T-test on motif z-scores between cells in each state. We adjusted p-values with the Benjamini-

Hochberg procedure and defined motifs with FDR<0.05 and absolute difference in z-score>0.5 as differential. To analyze differential promoter accessibility between cell states, we performed two-sided Fisher's exact tests between hormone-high and hormone-low states for each promoter against the null hypothesis that the promoter had similar accessibility across states. We used the Benjamini-Hochberg adjusted p-values (FDR<0.01 for alpha or beta cells and FDR<0.1 for delta cells) to identify gene promoters with differential accessibility across states. Differentially-accessible promoters were used to perform GO term enrichment on biological processes (2018 version) using Enrichr[93] (v.1.0). We filtered for terms that contained <150 genes.

We collected gene sets from Xin et al.[23], Mawla & Huising[43], and Camunas-Soler et al[44] (details in Supplementary Note). For each gene list, we performed gene set enrichment analysis[94] (GSEA) using $\log_2$ odds ratios from previous Fisher's exact tests. We also performed GSEA using significantly differential promoters (from **Figure 2.2a**) as the gene sets to assess whether cell states showed concordant differences across cell types.

### 2.5.11 GWAS enrichment genome-wide with aggregate peak annotations

We used stratified LD score regression[60–62] (v.1.0.1) to calculate enrichment for GWAS traits. We obtained GWAS summary statistics for quantitative traits[11,12,48–51], diabetes[9,47], and control traits[52–59]. To create custom LD score files, we annotated variants using peaks for each cluster as a binary annotation. In addition to the annotations included in the baseline-LD model v2.2, we included LD scores estimated from merged peaks across all clusters as the background. For each trait, we used stratified LD score regression to estimate the enrichment z-scores of each annotation relative to background. We computed one-sided p-values for enrichment based on the z-scores and corrected for multiple tests using the Benjamini-Hochberg procedure.

### 2.5.12 GWAS enrichment with single cell annotations

We determined genetic enrichment of accessible chromatin profiles in individual cells for fasting glucose level[12], T2D[10], and control traits major depressive disorder[56] and systemic lupus erythematosus[54] GWAS using polyTest[95].

To identify TFs correlated with trait enrichments, we regressed out log read depth, fraction of reads in peaks, and fraction of promoters used from the single cell enrichments. We calculated the Spearman correlation between the residuals of fasting glucose or T2D enrichment z-scores and motif enrichment z-scores using all cells or only beta cells. We used Bonferroni correction to correct p-values for multiple tests. To identify motifs directly enriched for T2D or FG association in beta cells, we identified all variants mapping in a beta cell site and using fimo[96] predicted instances of each motif in JASPAR[34] using sequence surrounding each allele. We considered variants disrupting motifs where there was a motif prediction for only one allele. We tested for enrichment of variants in a predicted motif or disrupting the motif for T2D or FG using polyTest[95].

### 2.5.13 GWAS enrichment at known T2D loci using aggregate peak annotations

We identified primary T2D risk signals[10] where the highest causal probability variant was associated with fasting glucose level[12] at genome-wide significance. We annotated variants at each signal with $INS^{high}$ and $INS^{low}$ beta cell sites and tested for enrichment using fgwas[64] (v0.3.6) [71] ('-fine'). For alpha, delta and gamma cells, we retained sites that did not overlap a beta cell site and annotated variants at all T2D risk signals[9]. To exclude the possibility that enrichments could be driven by other relevant tissues, we annotated variants in liver and adipose ATAC-seq from ENCODE[86]. We tested for enrichment using fgwas[64] ('-fine') including other tissue annotations in the model. We considered annotations with positive enrichment estimates enriched for T2D risk.

### 2.5.14 Predicting genetic variant effects on chromatin accessibility

We used deltaSVM[65] to predict the effects of non-coding variants on chromatin accessibility in each cell type and cell state. From variant z-scores we calculated p-values and FDR and considered variants significant at FDR<0.1.

### 2.5.15 Luciferase gene reporter assays

We cloned sequences containing reference alleles in the forward orientation upstream of the minimal promoter of firefly luciferase vector pGL4.23 (Promega) using KpnI and SacI restriction sites. For rs34584161:A>G, we cloned the alternative allele in the same manner as the reference alleles. For other variants, we introduced the alternative alleles via site-directed mutagenesis (SDM) using the NEB Q5 Site-Directed Mutagenesis kit (New England Biolabs). We normalized Firefly activity to Renilla activity and compared it to empty vector, and normalized results were expressed as fold change compared to control per allele. We used a two-sided Student's T-test to compare the two alleles.

### 2.5.16 Mapping allelic imbalance within clusters

We extracted genomic DNA either from spare islet nuclei (donors 1 and 2), or acinar cells (donor 3). We used the DNeasy Blood & Tissue Kits (Qiagen) according to manufacturer's protocol for purification of total DNA and performed genotyping on the Infinium Omni2.5-8 v1.4 array (Illumina). Additional details provided in Supplementary Note. Using cluster assignments for each cell, we split mapped reads for each donor into cluster-specific reads. We used WASP[97] (v.0.3.0) to correct for reference mapping bias at heterozygous variants. We then used a two-sided binomial test to assess imbalance assuming a null hypothesis where both alleles were equally likely to be observed. For each variant, we combined imbalance z-scores across donors using Stouffer's z-score method with sequencing depth as a weight for each sample.

### 2.5.17 Grouping of predicted effect variants and enrichment for islet caQTLs

We categorized variants with predicted effects on alpha or beta cells based on effects across cell type and states: “alpha” (n=8,552), “beta” (n=11,650), “hormone-high” (n=12,874), and “hormone-low” (n=10,808), and “shared” (n=27,140). We also determined the concordance in the direction of effect for variants across cell states. For the set of variants with significant effects in each state, we calculated the fraction of variants where the effect allele had predicted effect in other states. We determined significance using a two-sided binomial test assuming an expected fraction of 50%. We assessed enrichment of predicted effect variants in alpha or beta cell states for islet caQTLs[28] compared to any caQTL in alpha or beta cell sites using two-sided Fisher’s exact tests. We stratified variants with predicted effects by category and assessed enrichment of caQTLs with predicted effects within each category with two-sided Fisher’s exact tests.

### 2.5.18 TF motif enrichment within predicted effect variant categories

For each cell- or state-resolved category of variants with predicted effects, we extracted 29bp sequences (± 14bp around variant) corresponding to the higher or lower predicted effect allele. We used AME[96] in MEME (v.4.12.0) to predict motif enrichment, using position weight matrices from JASPAR 2018[34]. We used sequences from the higher or lower effect alleles as the test or background set respectively. We used TFClass (http://tfclass.bioinf.med.uni-goettingen.de/) to group motifs by family. To determine TFs most likely driving the enrichment, we checked normalized promoter accessibility within family members in alpha or beta cells.

### 2.5.19 Enrichment of predicted variants for lower frequency variants

We obtained summary statistics of T2D[9] and performed LD pruning. We identified variants that had significant effects in any endocrine cell type, as well as for each cell type for either state. We created a background set of variants without significant effects in any endocrine cell type (FDR>0.1). We created thresholds based on T2D association p-values ($5\times10^{-8}$, $1\times10^{-7}$, $1\times10^{-6}$, $1\times10^{-5}$, $1\times10^{-4}$, $1\times10^{-3}$). For each threshold, we calculated fold enrichment of predicted effect

variants as compared to the background and determined significance using two-sided binomial tests.

### 2.5.20 Single cell chromatin co-accessibility

We used Cicero[8] to calculate co-accessibility for pairs of peaks for alpha, beta, and delta cells. As in the trajectory analysis, we started from the merged peak by cell sparse binary matrix, extracted beta cells, and filtered out peaks that were not present in beta cells. We used 'make_cicero_cds' to aggregate cells based on 50 nearest neighbors. We then used Cicero to calculate co-accessibility using a window size of 1 Mb and a distance constraint of 500 kb. We repeated the same procedure for alpha and delta cells. We used co-accessibility>0.05 to define pairs of peaks as co-accessible.

### 2.5.21 Enrichment of islet Hi-C and pcHi-C loops in co-accessible peaks

We obtained islet promoter capture Hi-C (pcHi-C) and merged Hi-C loops[27,71]. For both datasets, we used coordinates for anchors directly from the loop calls. To compare co-accessibility with pcHi-C or Hi-C, we used direct overlap of peaks with anchors. For binned thresholds of co-accessibility in 0.05 increments, we calculated distance-matched odds ratios for alpha, beta or delta cell co-accessible peaks containing pcHi-C or Hi-C loops versus non-co-accessible peaks (co-accessibility<0). We used two-sided Fisher's exact tests to assess significance.

### 2.5.22 Hi-C library construction and data analysis

We performed *in situ* Hi-C as previously described using MboI on two batches of hESC-derived beta cells using the differentiation protocol described below cultured in high glucose (20mM mg/mL) or low glucose (5mM mg/mL). Hi-C libraries were sequenced to read counts of 1,509,428,732 and 1,918,698,012, respectively. We analyzed Hi-C using Juicer[98] with default settings and visualized with HiGlass[99]. Unless otherwise indicated, interaction frequencies were normalized using iterative matrix balancing. We generated virtual 4C tracks by extracting

normalized interaction frequencies from an anchor bin of interest from the contact matrix. We performed Aggregate Peak Analysis (APA) using juicer tools with settings "-u -r 10000".

### 2.5.23 Annotating fine-mapped diabetes risk variants

We annotated risk signals in compiled fine-mapping data for T2D. For each signal, we identified variants that were in the 99% credible set with PPA>0.01. We intersected these candidate variants with sites for each islet cell type and cell state, and then identified variants with predicted effects on the overlapping cell types/states. We annotated variants based on overlap with sites co-accessible to gene promoters. For target genes linked to diabetes risk variants we determined enriched gene sets using GSEA[94].

### 2.5.24 Analysis of *INS* promoter 4C data

We downloaded and re-analyzed published 4C of the *INS* promoter for EndoC-βH1[72] with 4C-ker[100]. We created a reduced genome using 25bp flanking sequences of BglII cutting sites. For the 3 replicates, we aligned reads to this reduced genome using bowtie2[101] (v.2.2.9; '-N 0 -5 20'). We extracted counts for each fragment after removing self-ligated and undigested fragments and input bedGraph files to R.4Cker. We generated normalized counts and called high interaction regions using 'nearBaitAnalysis' ('k=10').

### 2.5.25 CRISPR/Cas9-mediated genome editing in human embryonic stem cells

H1 hESCs (WA01; purchased from WiCell; NIH registration number: 0043) were seeded onto Matrigel®-coated six-well plates at a density of 50,000 cells/cm$^2$ and maintained in mTeSR1 media (StemCell Technologies) for 3-4 days with media changed daily. sgRNA oligos and genotyping primers are listed in **Table 2.4**. hESC research was approved by the University of California, San Diego, Institutional Review Board and Embryonic Stem Cell Research Oversight Committee.

### 2.5.26 Pancreatic differentiation of hESC clones

*KCNQ1* enhancer-deleted hESC lines were differentiated in a suspension-based format using rotational culture with modifications to a published protocol[74] *KCNQ1* base-edited hESC lines were differentiated in a suspension-based format using rotational culture with modifications to a published protocol[102]. Details on modifications are provided in the Supplementary Note. For the experimental analysis of clones, we performed flow cytometry, immunofluorescence staining, mRNA sequencing, quantitative PCR, and insulin content measurement.

## 2.6 Figures

**Figure 2.1. Pancreatic islet cell type accessible chromatin defined using snATAC-seq.**

(a) Clustering of accessible chromatin profiles from 15,298 pancreatic islet cells identifies 12 distinct clusters plotted on UMAP coordinates. The number of cells for each cluster is listed in parenthesis next to the cluster label. (b) Promoter accessibility in a 1 kb window around the TSS for selected marker genes. (c) Aggregate read density (counts per $1\times10^5$) at hormone marker genes: *GCG* (alpha), *INS-IGF2* (beta), *SST* (delta), and *PPY* (gamma). (d) Spearman correlation between t-statistics of cluster-specific genes based on promoter accessibility (snATAC-seq) and gene expression (scRNA-seq). (e) Row-normalized chromVAR motif enrichment z-scores for 141 TF sequence motifs with variable enrichment across clusters (left). Cell types with multiple clusters are collapsed into a single cluster (e.g. beta 1 + beta 2 into beta). Enrichment z-scores for FOXA1 and PDX1 motifs for each cell projected onto UMAP coordinates (right). (f) Pearson correlation of TF motif enrichment z-scores between endocrine and exocrine cell types (g) FDR-corrected p-values from two-sided two sample T-tests of differential chromVAR motif enrichment comparison between delta and beta cells for 366 TF motifs. (h) Enrichment z-scores for SCRT1 and MAFB motifs in 7,598 beta and 710 delta cells projected onto UMAP coordinates (top) and shown as violin distributions (bottom; lines represent median and quartiles).

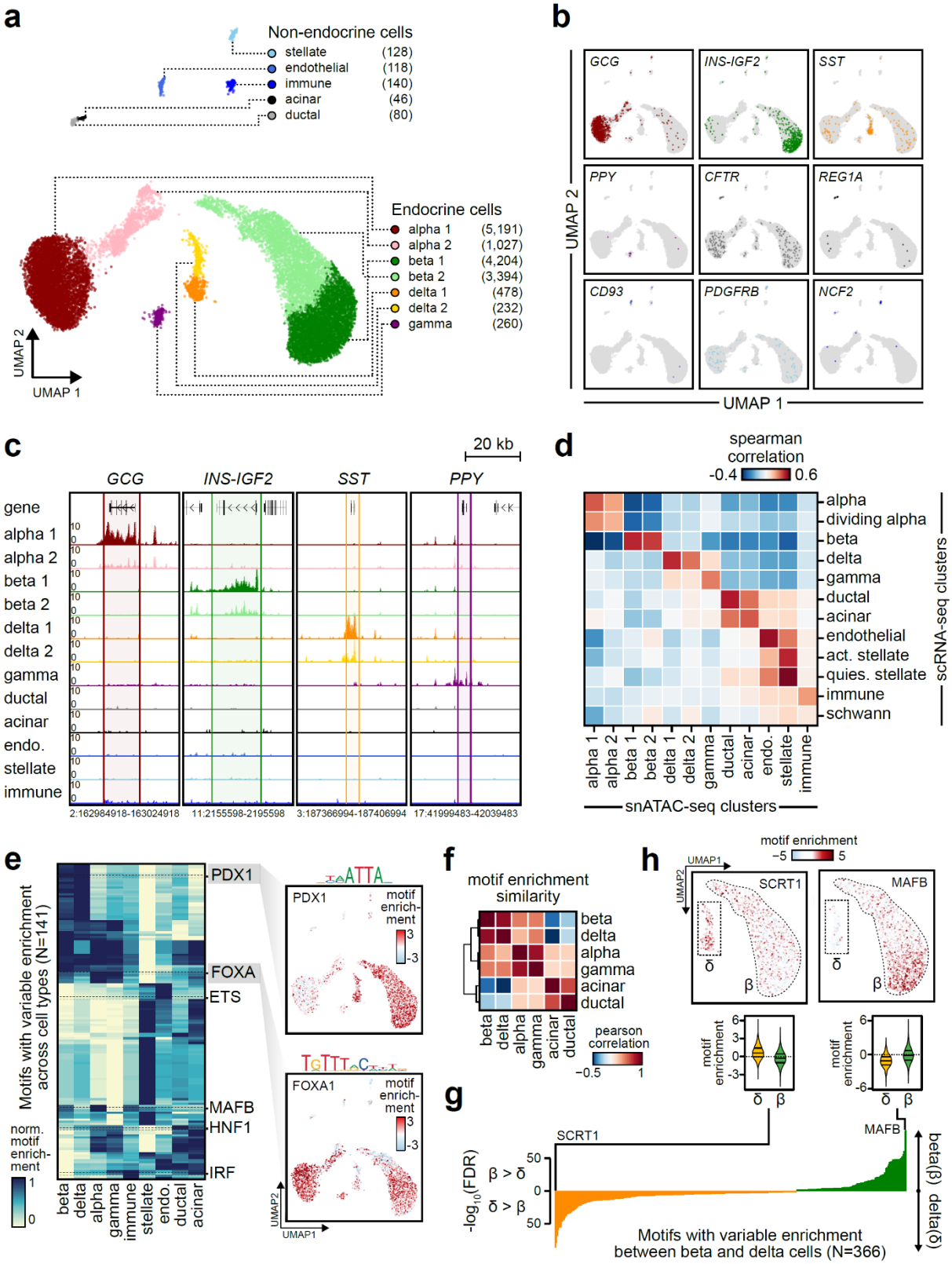

a
Non-endocrine cells
stellate (128)
endothelial (118)
immune (140)
acinar (46)
ductal (80)
Endocrine cells
alpha 1 (5,191)
alpha 2 (1,027)
beta 1 (4,204)
beta 2 (3,394)
delta 1 (478)
delta 2 (232)
gamma (260)
UMAP 2
UMAP 1
b
GCG
INS-IGF2
SST
PPY
CFTR
REG1A
CD93
PDGFRB
NCF2
UMAP 2
UMAP 1
c
20 kb
GCG
INS-IGF2
SST
PPY
gene
alpha 1
alpha 2
beta 1
beta 2
delta 1
delta 2
gamma
ductal
acinar
endo.
stellate
immune
2:162984918-163024918
11:2155598-2195598
3:187366994-187406994
17:41999483-42039483
d
spearman correlation
-0.4
0.6
alpha
dividing alpha
beta
delta
gamma
ductal
acinar
endothelial
act. stellate
quies. stellate
immune
schwann
scRNA-seq clusters
alpha 1
alpha 2
beta 1
beta 2
delta 1
delta 2
gamma
ductal
acinar
endo.
stellate
immune
snATAC-seq clusters
e
Motifs with variable enrichment across cell types (N=141)
norm. motif enrich-ment
1
0
PDX1
FOXA
ETS
MAFB
HNF1
IRF
beta
delta
alpha
gamma
immune
stellate
endo.
ductal
acinar
PDX1
motif enrich-ment
3
-3
FOXA1
motif enrich-ment
3
-3
UMAP2
UMAP1
f
motif enrichment similarity
beta
delta
alpha
gamma
acinar
ductal
pearson correlation
−0.5
1
h
motif enrichment
−5
5
UMAP1
UMAP2
SCRT1
MAFB
δ
β
motif enrichment
δ
β
g
SCRT1
MAFB
-log10(FDR)
β > δ
δ > β
50
0
50
beta(β) delta(δ)
Motifs with variable enrichment between beta and delta cells (N=366)

**Figure 2.2. Heterogeneity in endocrine cell accessible chromatin and regulatory programs.**

(a) Gene promoters with significantly differential chromatin accessibility between sub-clusters of alpha cells (left), beta cells (middle), and delta cells (right). (b) Enrichment of gene sets using ranked gene lists from the differential promoter analyses. Panels include genes with differential promoter accessibility between hormone-high or hormone-low states (first subpanel from left); genes expressed in beta cell sub-clusters from islet scRNA-seq (second and third subpanels); genes positively and negatively correlated with exocytosis from islet Patch-seq (fourth subpanel). (c) Enrichment of gene ontology terms related to glucose response, hormone secretion, stress response, and cell cycle among genes with differential promoter accessibility between endocrine cell states. (d) Row-normalized motif enrichments for 215 TF motifs with variable enrichment across endocrine cell states. Single cell motif enrichment z-scores for a representative RFX (RFX3) and FOS/JUN (FOS::JUN) motif are projected onto UMAP coordinates (right), and violin plots below show motif enrichment distribution within endocrine cell states (lines represent median and quartiles). (e) Ordering of alpha, beta and delta cells across pseudostate trajectories using high *GCG/INS-IGF2/SST* promoter accessibility as the reference point. Across each trajectory, the percentage of cells in the hormone-high state and the binary cluster call of individual cells are shown above the heatmaps, which show row-normalized motif enrichment for variable motifs between cell states. (f) Promoter accessibility for genes in the FOS/JUN motif family across pseudostate trajectories. Genes with matching promoter accessibility and motif enrichment patterns ($\rho>0.5$) are bolded.

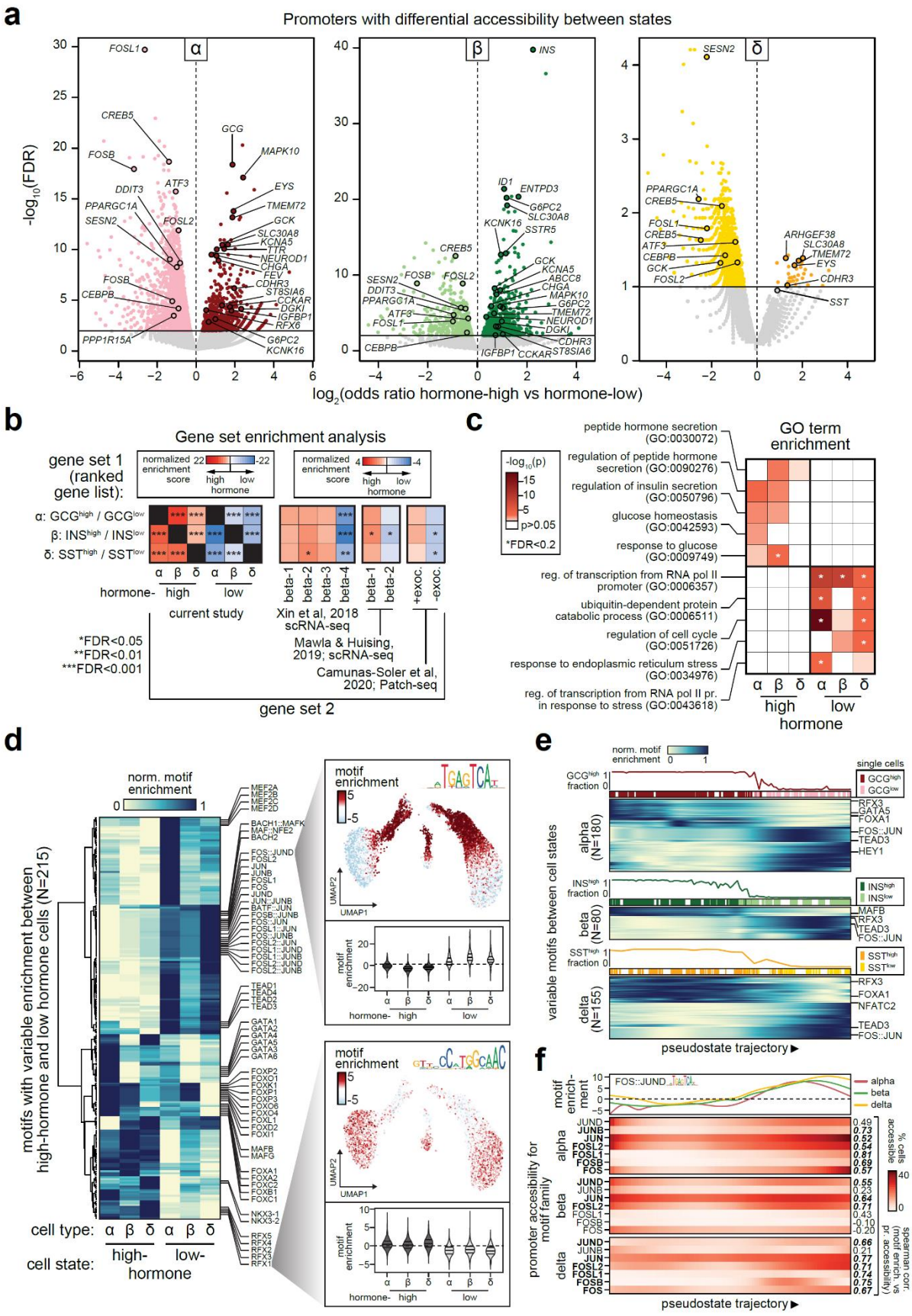

a
Promoters with differential accessibility between states
α
β
δ
FOSL1
INS
SESN2
-log10(FDR)
log2(odds ratio hormone-high vs hormone-low)
b
Gene set enrichment analysis
gene set 1 (ranked gene list):
normalized enrichment score
high hormone
low hormone
α: GCGhigh / GCGlow
β: INShigh / INSlow
δ: SSThigh / SSTlow
hormone-
high
low
current study
Xin et al, 2018 scRNA-seq
Mawla & Huising, 2019; scRNA-seq
Camunas-Soler et al, 2020; Patch-seq
*FDR<0.05
**FDR<0.01
***FDR<0.001
gene set 2
c
GO term enrichment
-log10(p)
p>0.05
*FDR<0.2
peptide hormone secretion (GO:0030072)
regulation of peptide hormone secretion (GO:0090276)
regulation of insulin secretion (GO:0050796)
glucose homeostasis (GO:0042593)
response to glucose (GO:0009749)
reg. of transcription from RNA pol II promoter (GO:0006357)
ubiquitin-dependent protein catabolic process (GO:0006511)
regulation of cell cycle (GO:0051726)
response to endoplasmic reticulum stress (GO:0034976)
reg. of transcription from RNA pol II pr. in response to stress (GO:0043618)
d
norm. motif enrichment
motifs with variable enrichment between high-hormone and low hormone cells (N=215)
cell type:
cell state:
high-hormone
low-hormone
motif enrichment
UMAP1
UMAP2
e
norm. motif enrichment
single cells
GCGhigh fraction
INShigh fraction
SSThigh fraction
variable motifs between cell states
alpha (N=180)
beta (N=80)
delta (N=155)
pseudostate trajectory
f
motif enrich-ment
promoter accessibility for motif family
% cells accessible
spearman corr. motif enrichment vs. promoter accessibility
pseudostate trajectory

**Figure 2.3. Enrichment of islet accessible chromatin for diabetes and fasting glycemia.**

(a) Stratified LD score regression enrichment z-scores for diabetes-related quantitative endophenotypes (top), type 1 and 2 diabetes (middle), and control traits (bottom) for islet cell types. **FDR<0.01 *FDR<0.1. (b) Single cell enrichment z-scores for fasting glucose level (FG) and T2D projected onto UMAP coordinates (left), enrichment per cell type (middle panels), and beta cell enrichment split into 10 trajectory bins (right). Boxplot center line, limits, and whiskers represent median, quartiles, and 1.5 interquartile range respectively. (c) Enrichment (estimate ± 95% CI by fgwas) of variants at loci associated with both T2D and FG (T2D/FG) within beta cell accessible chromatin. (d) Candidate causal T2D variant rs11708067 overlaps an enhancer active in $INS^{high}$ beta cells at the *ADCY5* locus, consistent with beta cell enrichment patterns for T2D/FG loci. (e) Enrichment (estimate ± 95% CI by fgwas) of variants at T2D loci in accessible chromatin for non-beta endocrine cells after removing beta accessible chromatin. (f) Candidate causal T2D variant rs1111875 overlaps a delta cell-specific site at the *HHEX* locus. (g) Correlation between single cell FG and TF motif enrichments across all 14.2k cells (left) and 7.2k beta cells (right). Across all cells, FG has positive correlations with beta-enriched TF families such as PDX, NKX6 and PAX. Within beta cells, FG has positive correlations with $INS^{high}$ beta-enriched TF families such as RFX, MAF/NRL, and FOXA. (h) Enrichment (effect±SE) of FG-associated variants directly overlapping sequence motifs for those either positively or negatively correlated with FG in beta cells.

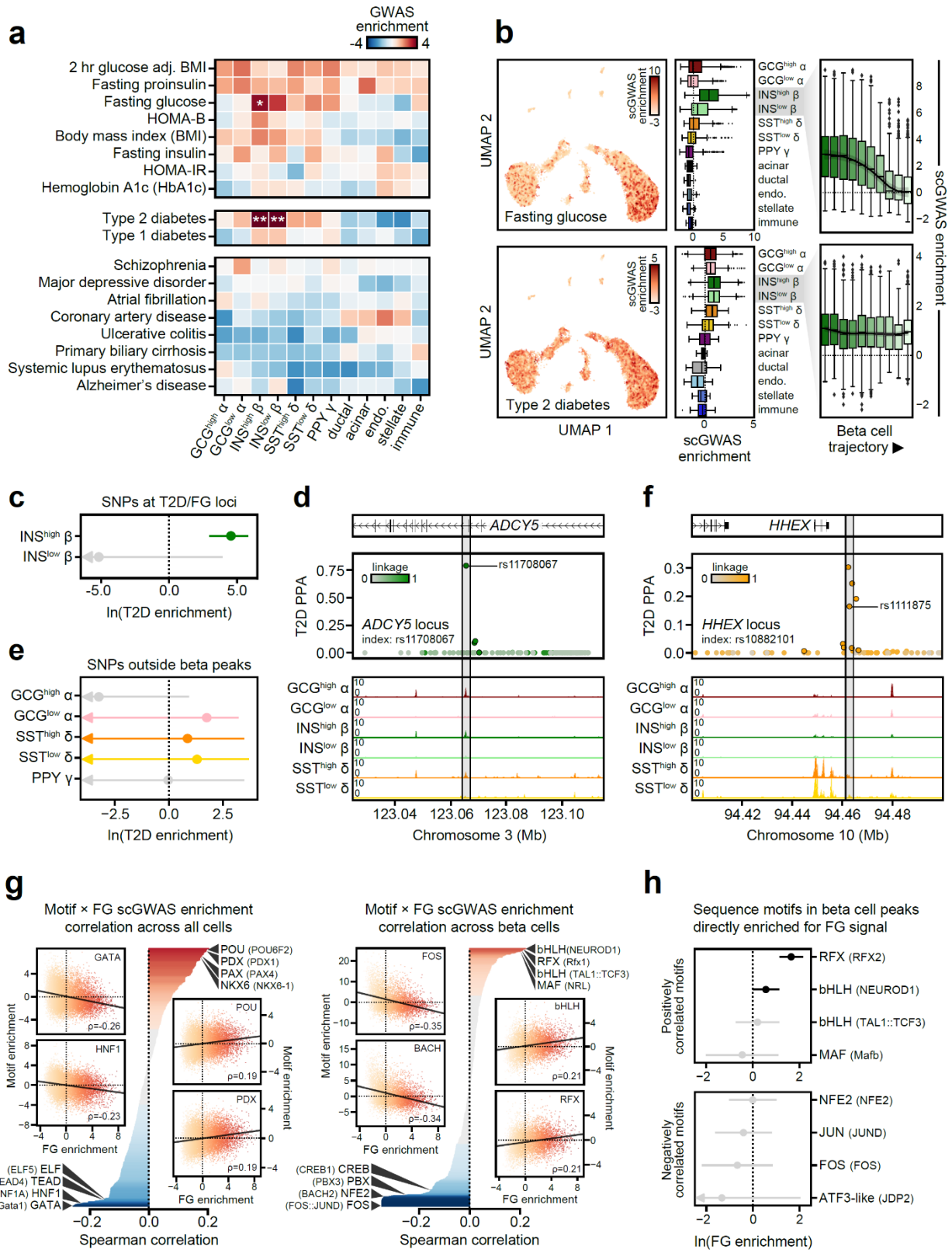

a
GWAS enrichment
-4
4
2 hr glucose adj. BMI
Fasting proinsulin
Fasting glucose
HOMA-B
Body mass index (BMI)
Fasting insulin
HOMA-IR
Hemoglobin A1c (HbA1c)
Type 2 diabetes
Type 1 diabetes
Schizophrenia
Major depressive disorder
Atrial fibrillation
Coronary artery disease
Ulcerative colitis
Primary biliary cirrhosis
Systemic lupus erythematosus
Alzheimer's disease
GCG^high α
GCG^low α
INS^high β
INS^low β
SST^high δ
SST^low δ
PPY γ
ductal
acinar
endo.
stellate
immune
b
UMAP 2
Fasting glucose
Type 2 diabetes
UMAP 1
scGWAS enrichment
Beta cell trajectory ▶
c
SNPs at T2D/FG loci
ln(T2D enrichment)
d
ADCY5
T2D PPA
linkage
rs11708067
ADCY5 locus
index: rs11708067
Chromosome 3 (Mb)
e
SNPs outside beta peaks
ln(T2D enrichment)
f
HHEX
rs1111875
HHEX locus
index: rs10882101
Chromosome 10 (Mb)
g
Motif × FG scGWAS enrichment correlation across all cells
Motif × FG scGWAS enrichment correlation across beta cells
Motif enrichment
FG enrichment
Spearman correlation
h
Sequence motifs in beta cell peaks directly enriched for FG signal
Positively correlated motifs
Negatively correlated motifs
RFX (RFX2)
bHLH (NEUROD1)
bHLH (TAL1::TCF3)
MAF (Mafb)
NFE2 (NFE2)
JUN (JUND)
FOS (FOS)
ATF3-like (JDP2)
ln(FG enrichment)

**Figure 2.4. Genetic variants with islet cell type- and state-specific effects on chromatin accessibility.**

(a) Percentage of HRC variants in any endocrine cell type peak (n=1,411,387 variants) that had significant deltaSVM effects (FDR<0.1) for the reference (ref.) or alternative (alt.) allele. (b) Spearman correlation between deltaSVM Z-score and chromatin accessibility allelic imbalance Z-scores for variants with predicted effects in alpha and beta states. (c) Relative luciferase reporter activity (mean ± 95% CI; n=3 replicates) for five T2D variants with predicted beta cell effects. The allele with predicted effect is circled. p-values by two-sided Student's T-tests. (d) Enrichment of islet caQTLs for variants with predicted effects in alpha and beta cells (left) and stratified based on shared, cell type- and state-specific effects (right). p-values by two-sided Fisher's exact test, ns, not significant. (e) Examples of variants with predicted effects in alpha and beta cells (left). TF motif families enriched in sequences surrounding the effect allele relative to the non-effect allele (middle). Promoter accessibility patterns for genes in enriched TF motif families (right). Genes with matching promoter accessibility and TF motif enrichment patterns are highlighted. (f) Enrichment (estimate ± 95% CI) of low frequency and rare variants with predicted effects on islet chromatin at different T2D association thresholds. p-values by two-sided binomial test. (g) Low-frequency T2D variant rs78840640 at the *IGF2BP3* locus with high causal probability (PPA=0.33) has predicted effects in beta cells,

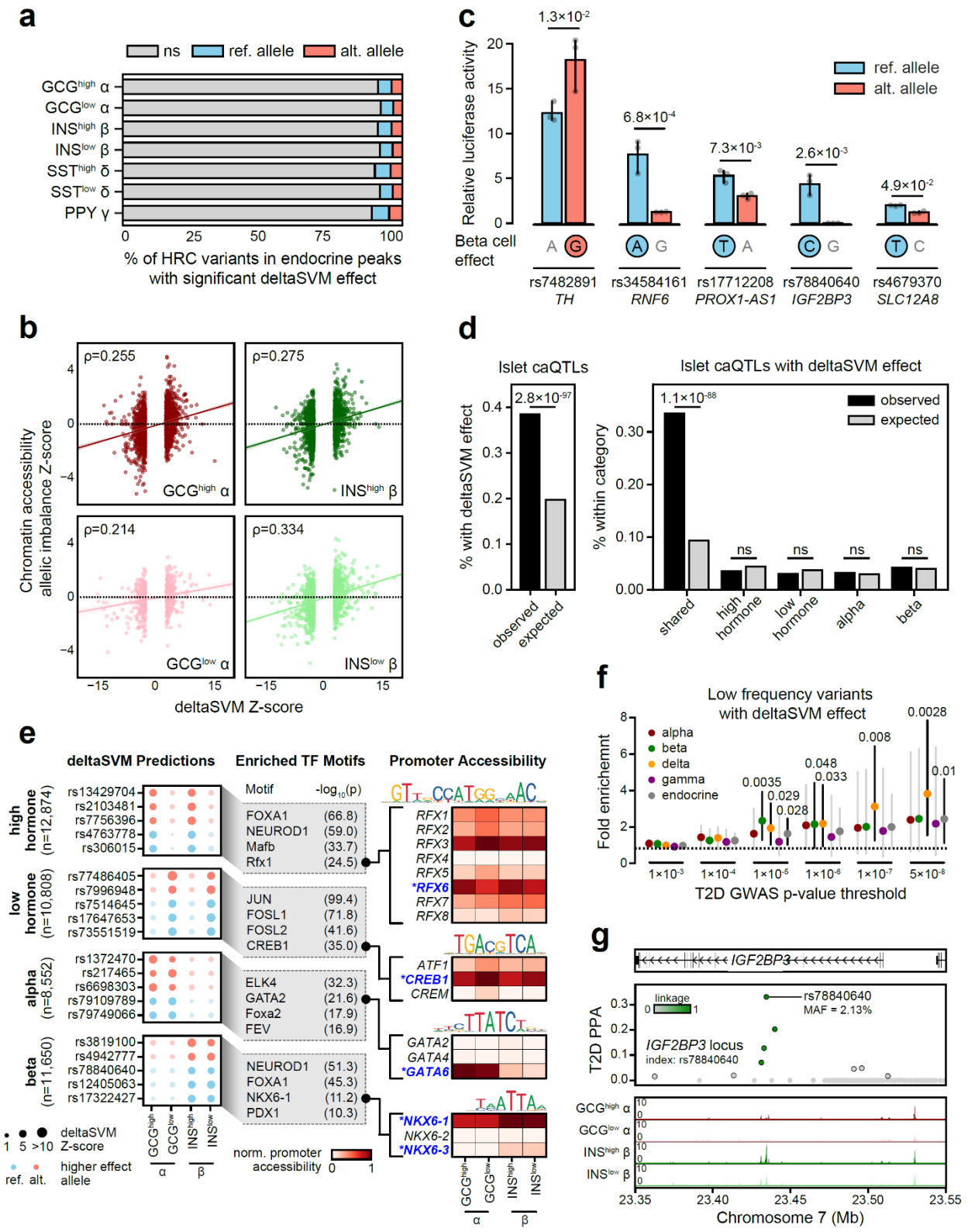
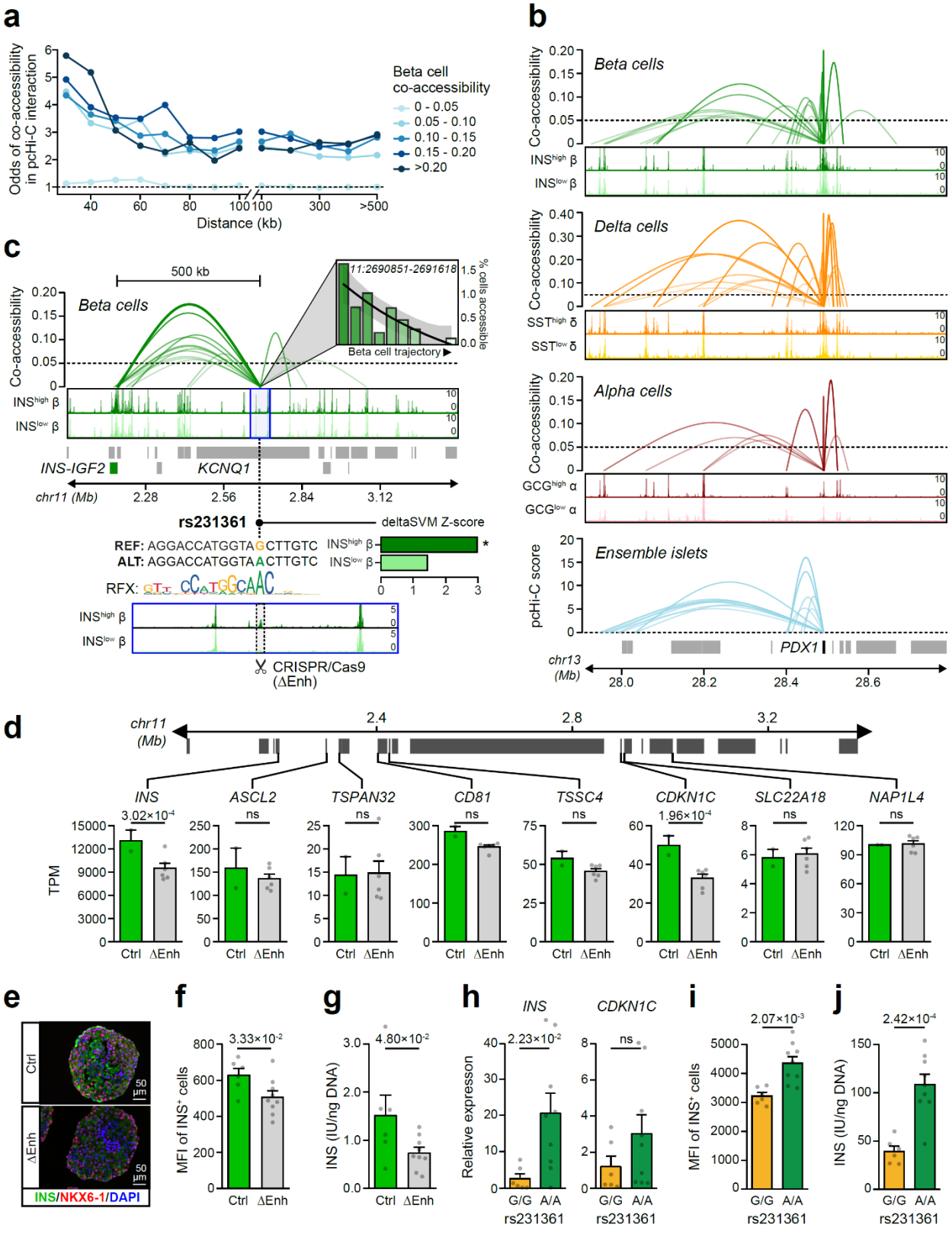

a
ns
ref. allele
alt. allele
GCG^high α
GCG^low α
INS^high β
INS^low β
SST^high δ
SST^low δ
PPY γ
0
25
50
75
100
% of HRC variants in endocrine peaks with significant deltaSVM effect
b
ρ=0.255
ρ=0.275
ρ=0.214
ρ=0.334
GCG^high α
INS^high β
GCG^low α
INS^low β
Chromatin accessibility allelic imbalance Z-score
deltaSVM Z-score
c
1.3×10^-2
6.8×10^-4
7.3×10^-3
2.6×10^-3
4.9×10^-2
Relative luciferase activity
ref. allele
alt. allele
Beta cell effect
rs7482891 TH
rs34584161 RNF6
rs17712208 PROX1-AS1
rs78840640 IGF2BP3
rs4679370 SLC12A8
d
Islet caQTLs
2.8×10^-97
% with deltaSVM effect
observed
expected
Islet caQTLs with deltaSVM effect
1.1×10^-88
observed
expected
% within category
ns
shared
high hormone
low hormone
alpha
beta
e
deltaSVM Predictions
Enriched TF Motifs
Promoter Accessibility
high hormone (n=12,874)
low hormone (n=10,808)
alpha (n=8,552)
beta (n=11,650)
Motif
-log10(p)
FOXA1 (66.8)
NEUROD1 (59.0)
Mafb (33.7)
Rfx1 (24.5)
JUN (99.4)
FOSL1 (71.8)
FOSL2 (41.6)
CREB1 (35.0)
ELK4 (32.3)
GATA2 (21.6)
Foxa2 (17.9)
FEV (16.9)
NEUROD1 (51.3)
FOXA1 (45.3)
NKX6-1 (11.2)
PDX1 (10.3)
RFX1
RFX2
RFX3
RFX4
RFX5
*RFX6
RFX7
RFX8
ATF1
*CREB1
CREM
GATA2
GATA4
*GATA6
*NKX6-1
NKX6-2
*NKX6-3
deltaSVM Z-score
higher effect allele
ref. alt.
norm. promoter accessibility
f
Low frequency variants with deltaSVM effect
alpha
beta
delta
gamma
endocrine
Fold enrichemnt
T2D GWAS p-value threshold
g
IGF2BP3
linkage
rs78840640
MAF = 2.13%
IGF2BP3 locus
index: rs78840640
T2D PPA
Chromosome 7 (Mb)

**Figure 2.5. Chromatin co-accessibility links diabetes risk variants to target genes.**

(a) Distance-matched odds that co-accessible sites in beta cells overlap islet pcHi-C interactions at different threshold bins. (b) Single cell co-accessibility and islet pcHi-C interactions at the *PDX1* promoter. (c) Enhancer harboring T2D variant rs231361 shows distal beta cell co-accessibility to the *INS* promoter and other non-promoter sites. The enhancer is accessible in $INS^{high}$ but not $INS^{low}$ beta and has decreasing accessibility across the beta cell trajectory. rs231361 disrupts an *RFX* motif and has predicted effects in $INS^{high}$ beta. *FDR<0.1. CRISPR/Cas9 deletion of a 2.6kb region (highlighted, ΔEnh) around the enhancer. (d) Expression (transcripts per million, TPM±SEM) of genes within 2Mb of the enhancer in beta cell stage cultures for ΔEnh (n=6; 3 clones × 2 differentiations) and control (n=2; 1 clone × 2 differentiations), p-values from DESeq2, ns not significant. (e) Representative immunofluorescence staining for INS (green), NKX6-1 (red), and DAPI (blue) in beta cell stage cultures for ΔEnh (n=3 independent experiments) and control (n=3 experiments) (f) Quantification of INS median fluorescence intensity (MFI) and (g) INS content in beta cell stage cultures for ΔEnh (n=9; 3 clones × 3 differentiations) and control (n=6; 2 clones × 3 differentiations). (h) Relative expression of *INS* and *CDKN1C*, (i) Quantification of INS MFI, and (j) INS content in beta cell stage cultures for $KCNQ1^{A/A}$ (n=9; 3 clones × 3 differentiations) and $KCNQ1^{G/G}$ (n=6; 2 clones × 3 differentiations). Data shown are mean±SEM, p-values by two-sided Student's T-test, ns, not significant.


a
Odds of co-accessibility in pcHi-C interaction
Distance (kb)
Beta cell co-accessibility
0 - 0.05
0.05 - 0.10
0.10 - 0.15
0.15 - 0.20
>0.20
b
Beta cells
Delta cells
Alpha cells
Ensemble islets
Co-accessibility
pcHi-C score
INS^high β
INS^low β
SST^high δ
SST^low δ
GCG^high α
GCG^low α
PDX1
chr13 (Mb)
c
500 kb
11:2690851-2691618
% cells accessible
Beta cell trajectory
INS-IGF2
KCNQ1
chr11 (Mb)
rs231361
deltaSVM Z-score
REF: AGGACCATGGTAGCTTGTC
ALT: AGGACCATGGTAACTTGTC
RFX:
CRISPR/Cas9 (ΔEnh)
d
chr11 (Mb)
INS
ASCL2
TSPAN32
CD81
TSSC4
CDKN1C
SLC22A18
NAP1L4
TPM
3.02×10-4
1.96×10-4
ns
Ctrl
ΔEnh
e
Ctrl
ΔEnh
50 μm
INS/NKX6-1/DAPI
f
MFI of INS+ cells
3.33×10-2
g
INS (IU/ng DNA)
4.80×10-2
h
INS
CDKN1C
Relative expresson
2.23×10-2
G/G
A/A
rs231361
i
MFI of INS+ cells
2.07×10-3
j
INS (IU/ng DNA)
2.42×10-4

## 2.7 Supplementary Figures

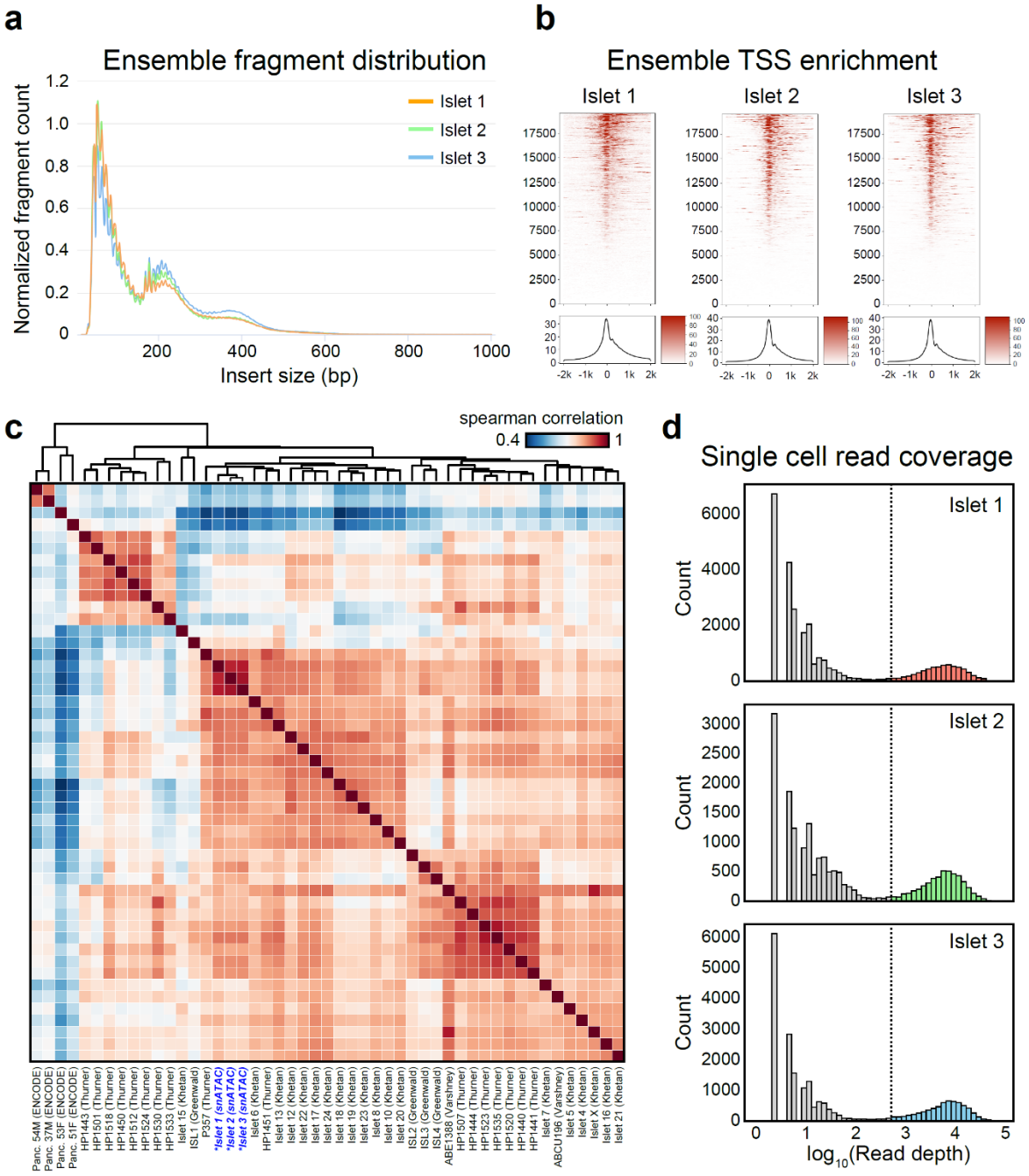



**Figure S2.1. Quality control metrics and aggregate comparison to bulk islet ATAC.**

(a) Insert size distribution for aggregate reads from each snATAC-seq experiment. (b) Aggregated read coverage from each snATAC-seq experiment in a ±2 kb window around individual promoters (top) and averaged across all promoters (bottom). (c) Spearman correlation between normalized read coverage within a merged set of peaks from 3 aggregated islet snATAC-seq, 42 bulk islet ATAC-seq, and 4 bulk pancreas ATAC-seq datasets. Names of samples are from the original sources of the data. (d) Binned $\log_{10}$ read depth distribution for each experiment.

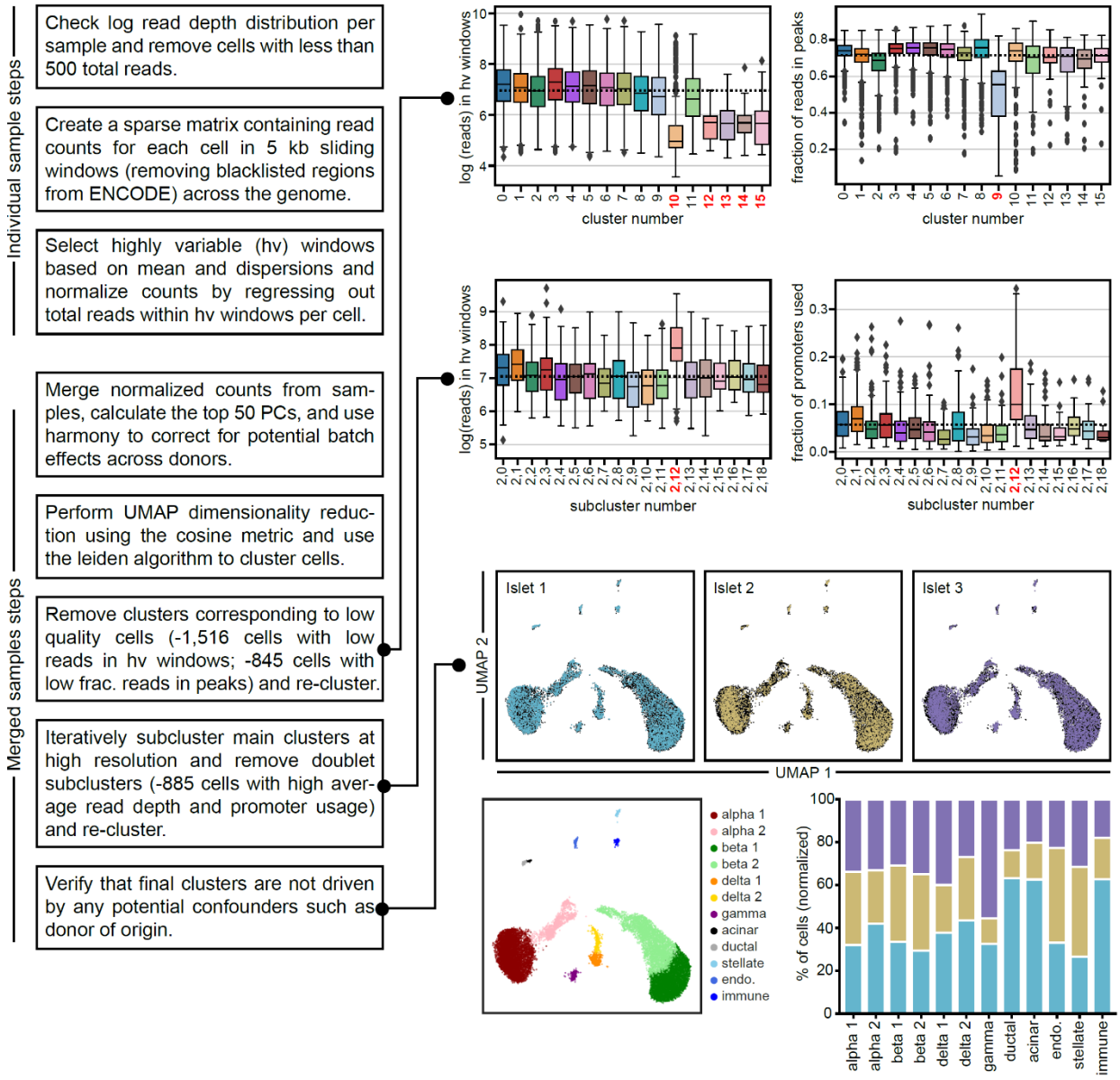



**Figure S2.2. Flowchart of the snATAC-seq data processing pipeline.**

(a) Flowchart summarizing key steps of the snATAC-seq processing pipeline, including the various steps where cells were filtered out. Samples were first processed individually. All samples were then combined using a batch correction method. Clusters corresponding to cells from low quality cells, including those with low read depth in highly variable windows and low fraction of reads in peaks were then removed. After re-clustering, iterative subclustering of the main clusters at high resolution was used to identify and remove doublet subclusters. The final clusters are not driven by potential confounders such as donor of origin. Boxplot center lines, limits, and whiskers represent median, quartiles, and 1.5 IQR respectively.

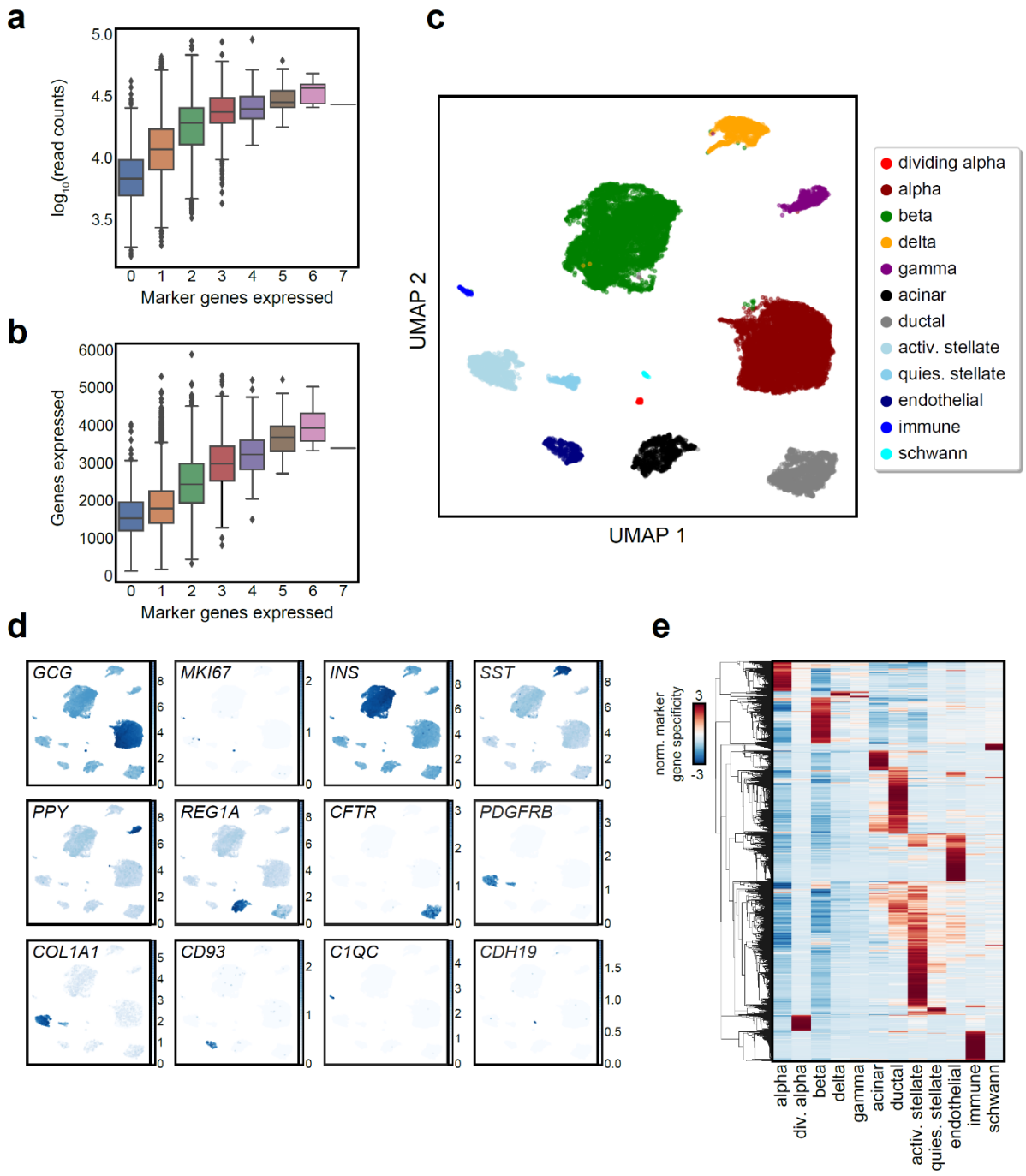



**Figure S2.3. Analysis of islet single cell gene expression data.**

(a) $\log_{10}$ transformed read depth or (b) total number of genes expressed compared with number of marker genes expressed per cell from scRNA-seq data. Boxplot center lines, limits, and whiskers represent median, quartiles, and 1.5 IQR respectively. Cells expressing more than one marker gene (defined by mixture models) were marked as doublets and filtered out. (c) Clusters of islet cells from single cell RNA-seq data plotted on UMAP coordinates. quies. stellate, quiescent stellate. activ. stellate, activated stellate. (d) Selected marker gene $\log_2$(expression) for each cluster plotted on UMAP coordinates. (e) Row-normalized t-statistics of marker gene specificity showing the most specific genes (t-statistic>20) for each cluster.

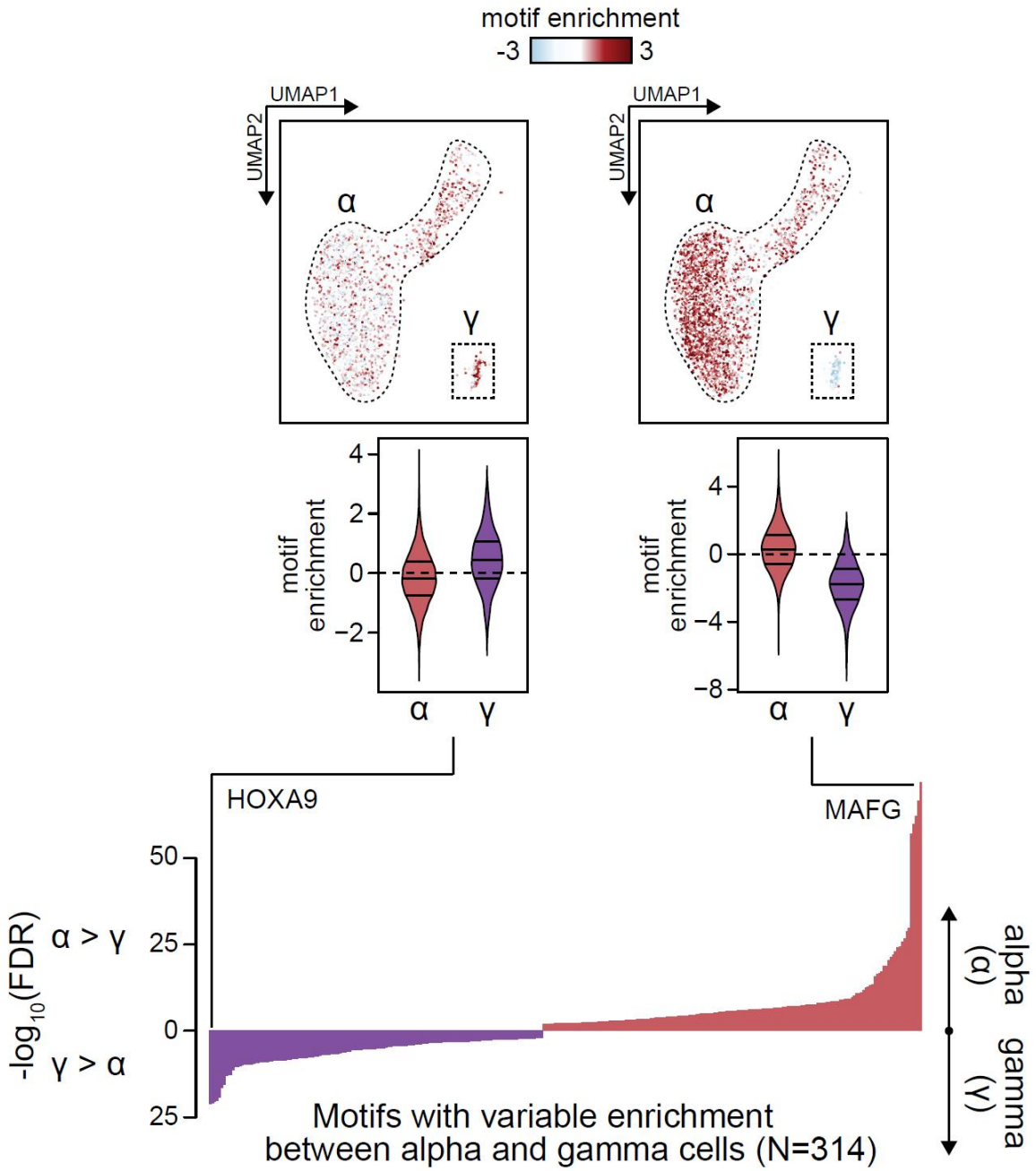
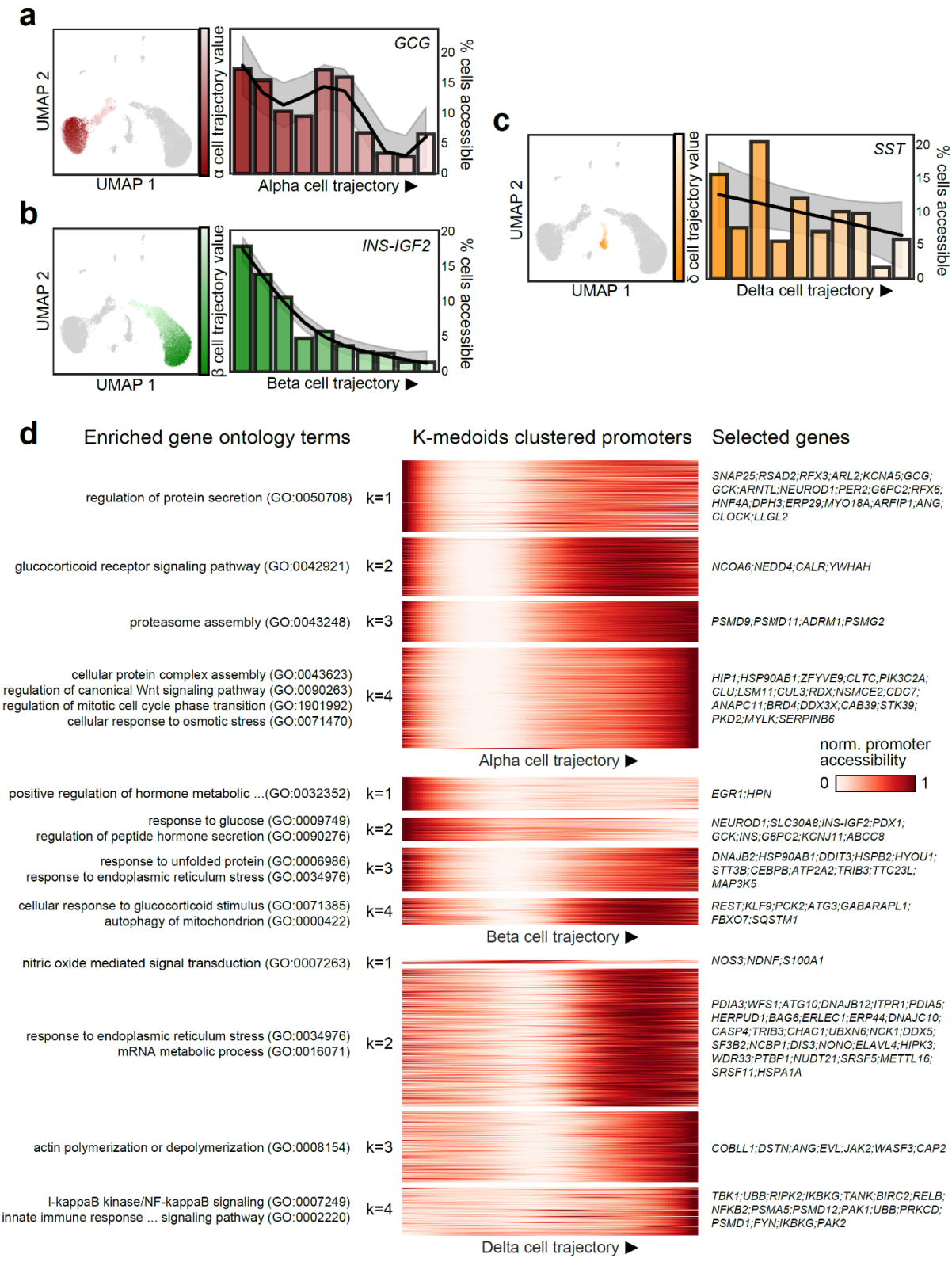



**Figure S2.4. Comparison of motif enrichment between alpha and gamma cells.**

Differential enrichment of motifs between alpha cell open chromatin regions and gamma cell open chromatin regions as measured by a 2-sided T-test, with FDR calculated by the Benjamini-Hochberg procedure. Examples are highlighted of motifs enriched in alpha cells and gamma cells, respectively (MAFG, HOXA9). UMAP plots show enrichment z-scores for the indicated motifs in alpha and gamma cells. Violin plots below show the distribution of enrichment z-scores across alpha or gamma cells, where the lines represent median and quartiles.

**Figure S2.5. Differentially accessible promoters across pseudo-states.**

(a) Pseudo-state (trajectory) values for alpha cells plotted on UMAP coordinates (left) and percentage of cells with *GCG* promoter accessibility decreases across 10 bins along the alpha (α) cell trajectory (right). (b) Pseudo-state (trajectory) values for beta (β) cells plotted on UMAP coordinates (left) and percentage of cells with *INS* promoter accessibility decreases across 10 bins along the beta cell trajectory (right). (c) Pseudo-state (trajectory) values for delta (δ) cells plotted on UMAP coordinates (left) and percentage of cells with *SST* promoter accessibility decreases across 10 bins along the beta cell trajectory (right). (d) Heatmaps showing promoters with dynamic accessibility across trajectories for alpha (top), beta (middle) and delta (bottom) cell trajectories. Gene promoters are clustered into 4 groups for each trajectory with k-medoids clustering. Enriched gene ontology for each k-mediod cluster (left) and selected genes present in at least one enriched gene ontology.


a
UMAP 2
UMAP 1
α cell trajectory value
GCG
% cells accessible
Alpha cell trajectory ▶
b
UMAP 2
UMAP 1
β cell trajectory value
INS-IGF2
% cells accessible
Beta cell trajectory ▶
c
UMAP 2
UMAP 1
δ cell trajectory value
SST
% cells accessible
Delta cell trajectory ▶
d
Enriched gene ontology terms
K-medoids clustered promoters
Selected genes
regulation of protein secretion (GO:0050708)
k=1
SNAP25;RSAD2;RFX3;ARL2;KCNA5;GCG;GCK;ARNTL;NEUROD1;PER2;G6PC2;RFX6;HNF4A;DPH3;ERP29;MYO18A;ARFIP1;ANG;CLOCK;LLGL2
glucocorticoid receptor signaling pathway (GO:0042921)
k=2
NCOA6;NEDD4;CALR;YWHAH
proteasome assembly (GO:0043248)
k=3
PSMD9;PSMD11;ADRM1;PSMG2
cellular protein complex assembly (GO:0043623)
regulation of canonical Wnt signaling pathway (GO:0090263)
regulation of mitotic cell cycle phase transition (GO:1901992)
cellular response to osmotic stress (GO:0071470)
k=4
HIP1;HSP90AB1;ZFYVE9;CLTC;PIK3C2A;CLU;LSM11;CUL3;RDX;NSMCE2;CDC7;ANAPC11;BRD4;DDX3X;CAB39;STK39;PKD2;MYLK;SERPINB6
Alpha cell trajectory ▶
norm. promoter accessibility
0
1
positive regulation of hormone metabolic ...(GO:0032352)
k=1
EGR1;HPN
response to glucose (GO:0009749)
regulation of peptide hormone secretion (GO:0090276)
k=2
NEUROD1;SLC30A8;INS-IGF2;PDX1;GCK;INS;G6PC2;KCNJ11;ABCC8
response to unfolded protein (GO:0006986)
response to endoplasmic reticulum stress (GO:0034976)
k=3
DNAJB2;HSP90AB1;DDIT3;HSPB2;HYOU1;STT3B;CEBPB;ATP2A2;TRIB3;TTC23L;MAP3K5
cellular response to glucocorticoid stimulus (GO:0071385)
autophagy of mitochondrion (GO:0000422)
k=4
REST;KLF9;PCK2;ATG3;GABARAPL1;FBXO7;SQSTM1
Beta cell trajectory ▶
nitric oxide mediated signal transduction (GO:0007263)
k=1
NOS3;NDNF;S100A1
response to endoplasmic reticulum stress (GO:0034976)
mRNA metabolic process (GO:0016071)
k=2
PDIA3;WFS1;ATG10;DNAJB12;ITPR1;PDIA5;HERPUD1;BAG6;ERLEC1;ERP44;DNAJC10;CASP4;TRIB3;CHAC1;UBXN6;NCK1;DDX5;SF3B2;NCBP1;DIS3;NONO;ELAVL4;HIPK3;WDR33;PTBP1;NUDT21;SRSF5;METTL16;SRSF11;HSPA1A
actin polymerization or depolymerization (GO:0008154)
k=3
COBLL1;DSTN;ANG;EVL;JAK2;WASF3;CAP2
I-kappaB kinase/NF-kappaB signaling (GO:0007249)
innate immune response ... signaling pathway (GO:0002220)
k=4
TBK1;UBB;RIPK2;IKBKG;TANK;BIRC2;RELB;NFKB2;PSMA5;PSMD12;PAK1;UBB;PRKCD;PSMD1;FYN;IKBKG;PAK2
Delta cell trajectory ▶

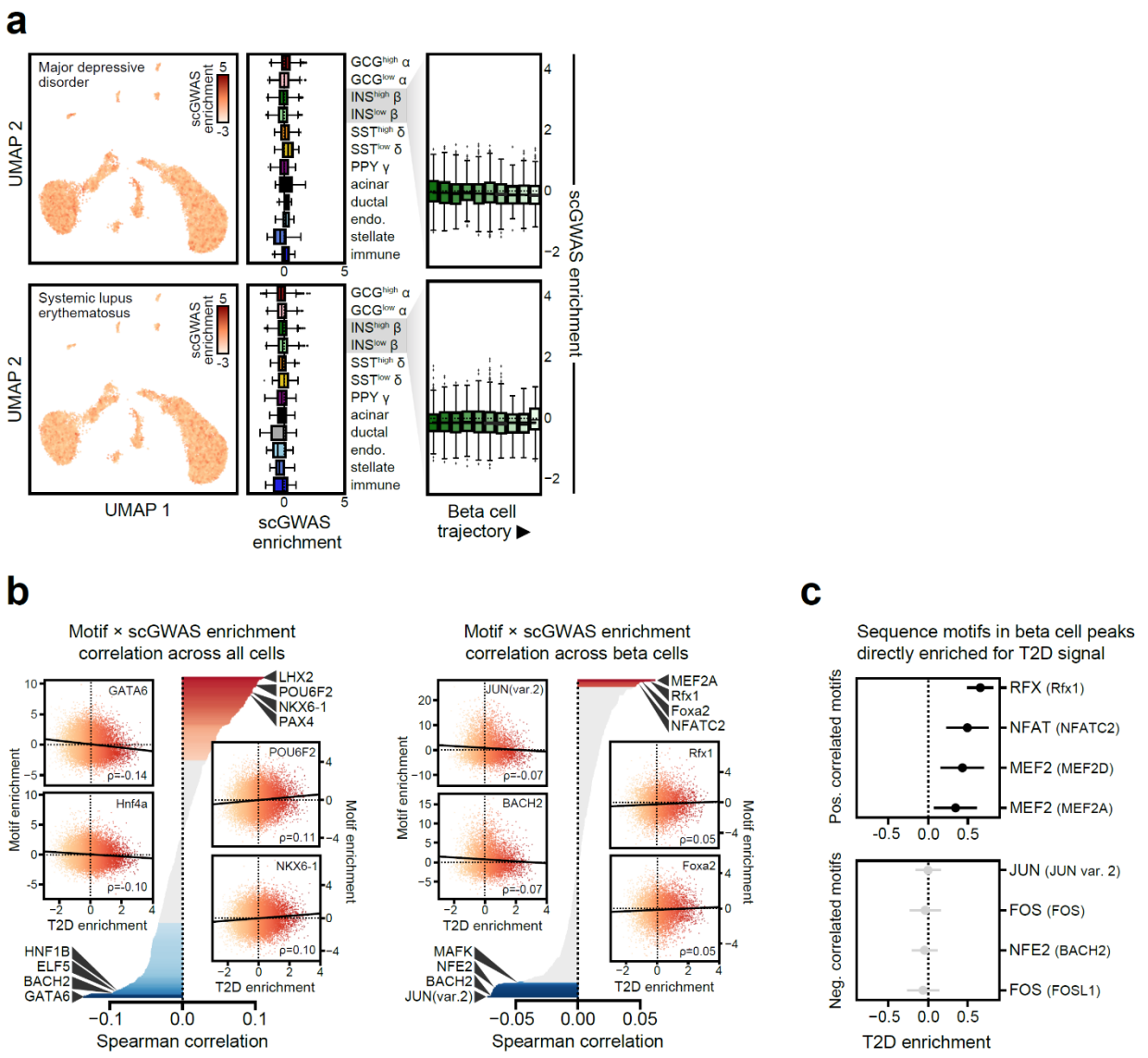



**Figure S2.6. Single cell GWAS enrichment and correlation with TF motifs.**

(a) Single cell GWAS enrichment z-scores for Major depressive disorder and Systemic lupus erythematosus projected onto UMAP coordinates (left panels), z-score enrichment distribution per cell type and state (middle panels) and z-score enrichment distribution split into 10 bins based on beta cell trajectory values (right panels). Boxplot center lines, limits, and whiskers represent median, quartiles, and 1.5 IQR respectively. (b) Correlation between single cell GWAS enrichment z-scores for Type 2 Diabetes and chromVAR TF motif enrichment z-scores across either all cells (left) or beta cells (right). Inset scatterplots highlight the top correlated motifs in either direction. (c) Variants mapping directly in sequence motifs positively correlated with T2D risk in beta cells are enriched for T2D association, whereas variants mapping in motifs negatively correlated with T2D risk in beta cells show no such enrichment.

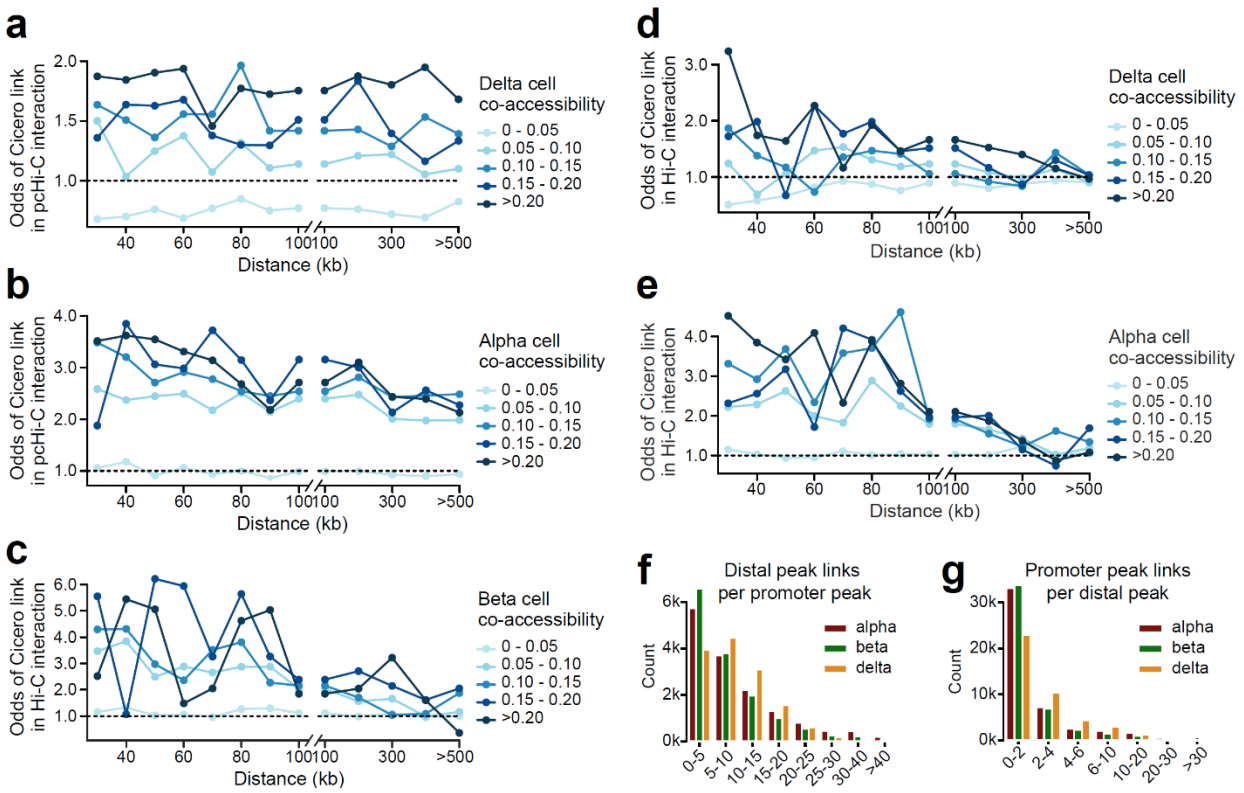



**Figure S2.7. Single cell co-accessibility analyses in islet cell types.**

(a) Distance-matched odds that delta cell co-accessibility links overlap islet pcHi-C chromatin loops at different co-accessibility threshold bins in 0.05 intervals demonstrate that co-accessible links are enriched for chromatin interactions. (b) Same analysis as in (a) but with alpha cell co-accessibility. (c) Same analysis as in (a) but with beta cell co-accessibility and Hi-C loops. (d) Same analysis as in (a) but with delta cell co-accessibility and Hi-C loops. (e) Same analysis as in (a) but with alpha cell co-accessibility and Hi-C loops. (f) Number of distal sites linked to each promoter peak for alpha, beta, and delta cells. (g) Number of promoters linked to each distal site for alpha, beta, and delta cells.

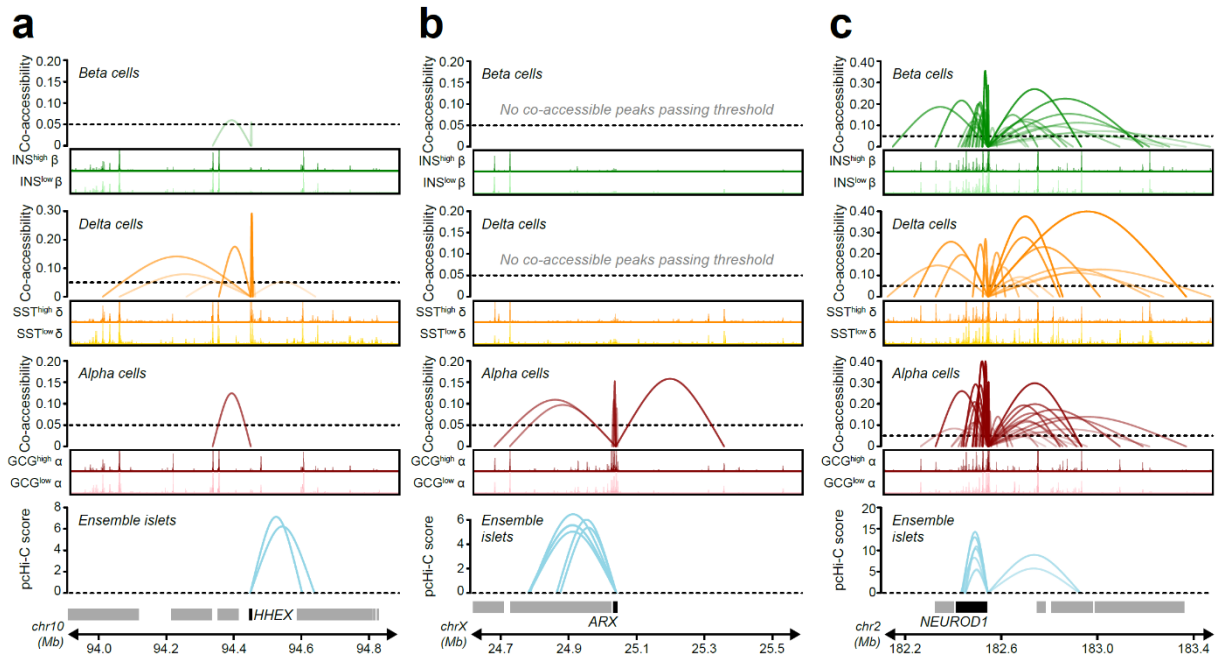



**Figure S2.8. Cell type-specific and shared co-accessible sites.**

(a) An example of co-accessibility anchored at the promoter for the delta cell identity TF *HHEX*. Co-accessibility for beta, delta, and alpha cells are shown compared to high-confidence pcHi-C loops from ensemble islets. Genome browser plots scale: 0-10. (b) An example of co-accessibility anchored at the promoter for the alpha cell identity TF *ARX*. (c) An example of shared co-accessibility anchored at the promoter for the shared islet identity TF *NEUROD1*.

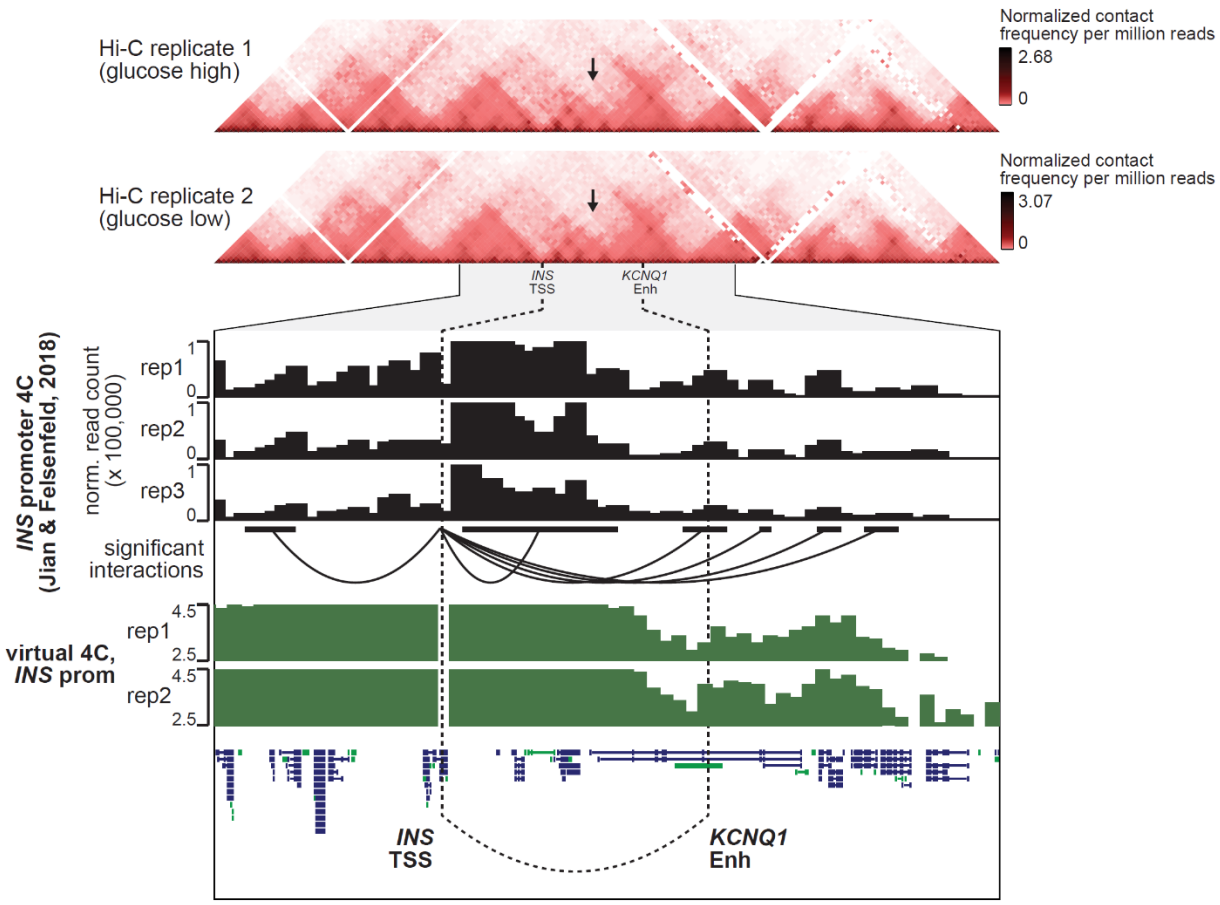



**Figure S2.9. 3D chromatin interactions at the T2D-associated KCNQ1 locus.**

Top panels show Hi-C contact matrices from hESC-derived beta cells, visualized at 25 kb resolution. Region shown is chr11:500,00-4,500,000, hg19. Black arrows indicate putative interaction point of *INS* TSS and *KCNQ1* enhancer. Genome browser plot below shows a zoomed view of chr11:1,750,000-3,250,000. Data from 4C-seq anchored on the INS promoter in EndoC-βH1 cells (Jian & Felsenfeld 2018) is shown, as analyzed with the 4C-ker package. Normalized read counts are shown in black from 3 biological replicates. Significant interactions from *INS* promoter are shown as arcs below read counts tracks. Interactions calls are from data pooled across 3 replicates are shown here. The region containing the *KCNQ1* enhancer was called as a significant interaction region with *INS* promotor independently in each 4C replicate. Virtual 4C plots in green show log(normalized Hi-C interaction frequency) from *INS* promoter.

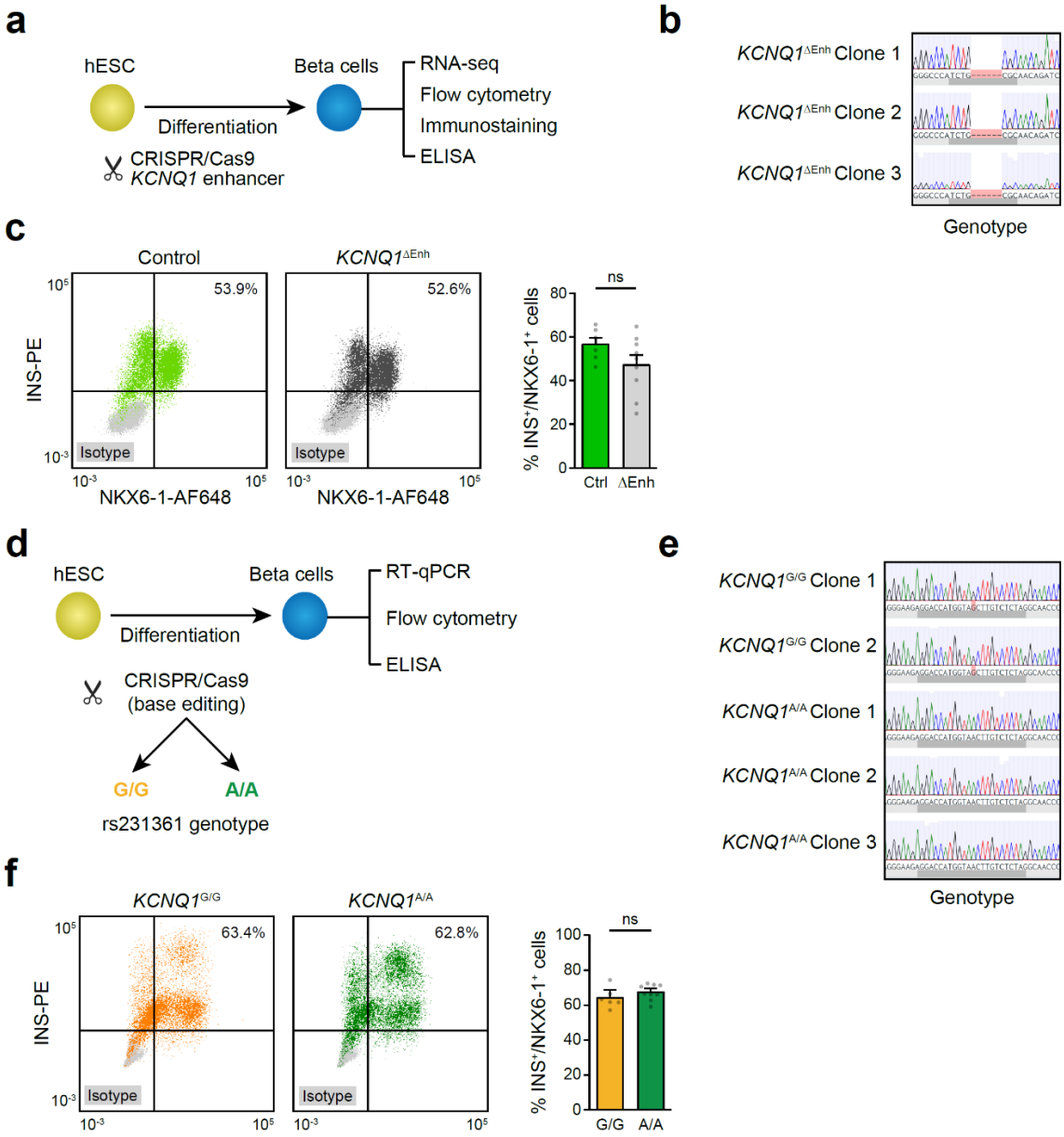



**Figure S2.10. Genome editing of the KCNQ1 locus in hESCs.**

(a) Schematic of the workflow and (b) Sanger sequencing for *KCNQ1* enhancer deletion in three independent hESC clones. (c) Representative figures of flow cytometry analysis for NKX6-1 and INS comparing control and *KCNQ1*$^{\Delta Enh}$ cells (left). Quantification of the percentage of NKX6-1$^+$/INS$^+$ cells in beta cell stage cultures from control (n=6; 2 clones × 3 differentiations) and *KCNQ1*$^{\Delta Enh}$ (n=9; 3 clones × 3 differentiations) cells (right). ns, not significant by two-sided Student's T-test without adjustment for multiple comparisons. (d) Schematic of the workflow and (e) Sanger sequencing for two independent *KCNQ1*$^{G/G}$ clones and three *KCNQ1*$^{A/A}$ clones. (f) Representative figures of flow cytometry analysis for NKX6-1 and INS comparing *KCNQ1*$^{G/G}$ and *KCNQ1*$^{A/A}$ clones (left). of the percentage of NKX6-1$^+$/INS$^+$ cells in beta cell stage cultures from *KCNQ1*$^{G/G}$ (n=6; 2 clones × 3 differentiations) and *KCNQ1*$^{A/A}$ (n=9; 3 clones × 3 differentiations) cells (right). ns, not significant by two-sided Student's T-test without adjustment for multiple comparisons. Error bars represent SEM.

# 2.8 Tables

## Table 2.1. Characteristics of donor samples used for snATAC-seq assays.

| Name | ID | Age (y) | Sex | Race | BMI | Blood glucose (mg/dl) | Diabetes status | Cause of death | Viability (%) | Purity (% Post culture) | snATAC assay type | Duplicate read % | Total unique reads | Mitochondrial read % | TSS read overlap % | TSS enrichment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Islet 1 | (UNOS) AFC2208 | 32 | M | Caucasian | 32.3 | 164.4 | Non-diabetic | Anoxia | 95 | 80 | combinatorial barcoding | 77.25% | 48,273,188 | 0.01% | 36.90% | 36.72 |
| Islet 2 | (UNOS) AFEA331 | 45 | M | African American | 29.3 | 155.4 | Non-diabetic | Stroke | 95 | 90 | combinatorial barcoding | 29.80% | 35,851,000 | 0.02% | 37.90% | 38.98 |
| Islet 3 | (UNOS) AFEP022 | 62 | M | African American | 36.1 | 178.4 | Non-diabetic | Anoxia | 95 | 90 | combinatorial barcoding | 45.87% | 54,268,192 | 2.15% | 37.90% | 38.24 |
| Pancreas 1 | (nPOD) 6004 | 33 | M | Caucasian | 30.9 | N/A | Non-diabetic | N/A | N/A | N/A | combinatorial barcoding | 75.28% | 20,090,168 | 5.78% | 30.90% | 44.89 |
| Pancreas 1 | (nPOD) 6004 | 33 | M | Caucasian | 30.9 | N/A | Non-diabetic | N/A | N/A | N/A | 10X | 37.37% | 228,259,258 | 1.66% | 35.10% | 42.21 |

## Table 2.2. Sequence motifs with variable enrichment across cell types.

| Motif | beta | delta | alpha | gamma | immune | stellate | endothelial | ductal | acinar |
|---|---|---|---|---|---|---|---|---|---|
| LHX6 | 0.87 | 1.00 | 0.12 | 0.36 | 0.32 | 0.00 | 0.15 | 0.37 | 0.08 |
| NOTO | 0.91 | 1.00 | 0.14 | 0.32 | 0.36 | 0.00 | 0.17 | 0.29 | 0.11 |
| PAX4 | 0.94 | 1.00 | 0.10 | 0.31 | 0.36 | 0.00 | 0.20 | 0.31 | 0.03 |
| Lhx8 | 0.94 | 1.00 | 0.10 | 0.31 | 0.36 | 0.00 | 0.20 | 0.31 | 0.03 |
| DLX6 | 0.98 | 1.00 | 0.21 | 0.46 | 0.43 | 0.07 | 0.29 | 0.41 | 0.00 |
| PDX1 | 0.97 | 1.00 | 0.16 | 0.41 | 0.49 | 0.12 | 0.23 | 0.38 | 0.00 |
| GSX1 | 0.92 | 1.00 | 0.14 | 0.35 | 0.38 | 0.03 | 0.22 | 0.47 | 0.00 |
| EMX1 | 0.87 | 1.00 | 0.19 | 0.42 | 0.38 | 0.00 | 0.17 | 0.49 | 0.04 |
| EMX2 | 0.98 | 1.00 | 0.21 | 0.46 | 0.37 | 0.00 | 0.23 | 0.49 | 0.09 |
| GBX1 | 0.93 | 1.00 | 0.17 | 0.41 | 0.38 | 0.00 | 0.32 | 0.44 | 0.12 |
| GBX2 | 0.93 | 1.00 | 0.30 | 0.52 | 0.38 | 0.01 | 0.38 | 0.24 | 0.00 |
| LHX2 | 0.98 | 1.00 | 0.16 | 0.39 | 0.36 | 0.00 | 0.29 | 0.34 | 0.12 |
| LBX1 | 0.98 | 1.00 | 0.22 | 0.37 | 0.41 | 0.00 | 0.34 | 0.35 | 0.15 |
| Alx4 | 0.80 | 1.00 | 0.27 | 0.36 | 0.37 | 0.00 | 0.16 | 0.40 | 0.13 |
| POU6F2 | 0.97 | 1.00 | 0.21 | 0.44 | 0.34 | 0.00 | 0.16 | 0.50 | 0.25 |
| mix-a | 0.81 | 1.00 | 0.30 | 0.50 | 0.38 | 0.00 | 0.17 | 0.53 | 0.27 |
| Alx1 | 0.79 | 1.00 | 0.25 | 0.46 | 0.35 | 0.00 | 0.16 | 0.49 | 0.21 |
| VAX1 | 0.86 | 1.00 | 0.25 | 0.49 | 0.47 | 0.02 | 0.33 | 0.45 | 0.00 |
| VAX2 | 0.86 | 1.00 | 0.25 | 0.49 | 0.47 | 0.02 | 0.33 | 0.45 | 0.00 |
| Hoxd3 | 0.85 | 1.00 | 0.32 | 0.49 | 0.42 | 0.00 | 0.32 | 0.52 | 0.12 |
| NKX6-1 | 0.87 | 1.00 | 0.26 | 0.47 | 0.43 | 0.00 | 0.33 | 0.45 | 0.09 |
| NKX6-2 | 0.87 | 1.00 | 0.26 | 0.47 | 0.43 | 0.00 | 0.33 | 0.45 | 0.09 |
| HOXA2 | 0.92 | 1.00 | 0.37 | 0.47 | 0.48 | 0.00 | 0.24 | 0.51 | 0.23 |
| MEOX1 | 0.91 | 1.00 | 0.35 | 0.51 | 0.47 | 0.00 | 0.37 | 0.47 | 0.16 |
| EVX2 | 0.99 | 1.00 | 0.36 | 0.48 | 0.53 | 0.00 | 0.37 | 0.48 | 0.20 |
| Rfx1 | 0.78 | 0.93 | 1.00 | 0.97 | 0.61 | 0.00 | 0.35 | 0.13 | 0.23 |
| RFX5 | 0.77 | 0.89 | 0.84 | 1.00 | 0.60 | 0.00 | 0.38 | 0.11 | 0.24 |
| RFX3 | 0.74 | 0.87 | 0.83 | 1.00 | 0.55 | 0.00 | 0.41 | 0.03 | 0.23 |
| RFX2 | 0.76 | 0.90 | 0.84 | 1.00 | 0.55 | 0.02 | 0.38 | 0.00 | 0.27 |
| RFX4 | 0.77 | 0.93 | 0.88 | 1.00 | 0.56 | 0.00 | 0.35 | 0.01 | 0.26 |
| Tcf12 | 1.00 | 0.80 | 0.71 | 0.71 | 0.40 | 0.18 | 0.00 | 0.05 | 0.25 |
| TAL1::TCF3 | 1.00 | 0.96 | 0.95 | 0.71 | 0.69 | 0.27 | 0.38 | 0.00 | 0.32 |
| NEUROD1 | 1.00 | 0.91 | 0.83 | 0.70 | 0.59 | 0.01 | 0.17 | 0.01 | 0.00 |
| Tcf21 | 1.00 | 0.99 | 0.94 | 0.82 | 0.62 | 0.15 | 0.28 | 0.00 | 0.00 |
| FOXP3 | 0.99 | 0.58 | 1.00 | 1.00 | 0.66 | 0.00 | 0.59 | 0.75 | 0.99 |
| FOXO6 | 0.99 | 0.58 | 1.00 | 1.00 | 0.66 | 0.00 | 0.59 | 0.75 | 0.99 |
| FOXO4 | 0.99 | 0.58 | 1.00 | 1.00 | 0.66 | 0.00 | 0.59 | 0.75 | 0.99 |
| FOXD2 | 0.99 | 0.58 | 1.00 | 1.00 | 0.66 | 0.00 | 0.59 | 0.75 | 0.99 |
| FOXI1 | 0.99 | 0.58 | 1.00 | 1.00 | 0.66 | 0.00 | 0.59 | 0.75 | 0.99 |
| FOXL1 | 0.99 | 0.58 | 1.00 | 1.00 | 0.66 | 0.00 | 0.59 | 0.75 | 0.99 |
| ZEB1 | 1.00 | 0.99 | 1.00 | 0.99 | 0.80 | 0.00 | 0.46 | 0.96 | 1.00 |
| TCF4 | 0.95 | 0.95 | 1.00 | 0.94 | 0.77 | 0.00 | 0.43 | 0.85 | 0.96 |
| TCF3 | 0.95 | 0.97 | 1.00 | 0.97 | 0.76 | 0.00 | 0.47 | 0.85 | 0.95 |
| SNAI2 | 0.95 | 0.99 | 1.00 | 0.97 | 0.79 | 0.00 | 0.46 | 0.82 | 0.93 |
| ID4 | 0.98 | 0.98 | 1.00 | 0.96 | 0.79 | 0.00 | 0.45 | 0.84 | 0.95 |
| Ascl2 | 1.00 | 0.87 | 0.77 | 0.77 | 0.58 | 0.00 | 0.15 | 0.31 | 0.36 |
| ASCL1 | 1.00 | 0.90 | 0.82 | 0.88 | 0.60 | 0.00 | 0.13 | 0.46 | 0.63 |
| Myod1 | 1.00 | 0.92 | 0.86 | 0.86 | 0.61 | 0.00 | 0.19 | 0.47 | 0.60 |
| Foxa2 | 0.82 | 0.60 | 1.00 | 0.97 | 0.51 | 0.00 | 0.35 | 0.56 | 0.67 |
| FOXA1 | 0.82 | 0.65 | 1.00 | 0.97 | 0.64 | 0.00 | 0.38 | 0.56 | 0.72 |
| FOXC2 | 0.77 | 0.37 | 1.00 | 0.93 | 0.40 | 0.00 | 0.44 | 0.53 | 0.84 |
| FOXB1 | 0.68 | 0.48 | 1.00 | 0.85 | 0.57 | 0.00 | 0.40 | 0.44 | 0.67 |
| FOXC1 | 0.68 | 0.36 | 1.00 | 0.86 | 0.48 | 0.00 | 0.30 | 0.36 | 0.64 |
| NFYB | 0.02 | 0.20 | 0.20 | 0.00 | 0.27 | 1.00 | 1.00 | 0.39 | 0.64 |
| NFYA | 0.00 | 0.19 | 0.26 | 0.05 | 0.37 | 0.83 | 1.00 | 0.31 | 0.55 |
| ETV6 | 0.00 | 0.18 | 0.15 | 0.25 | 0.53 | 0.58 | 1.00 | 0.47 | 0.40 |
| ETV2 | 0.00 | 0.15 | 0.12 | 0.24 | 0.31 | 0.51 | 1.00 | 0.35 | 0.29 |
| ELK4 | 0.00 | 0.11 | 0.10 | 0.19 | 0.34 | 0.69 | 1.00 | 0.35 | 0.32 |
| ERG | 0.00 | 0.08 | 0.07 | 0.14 | 0.27 | 0.58 | 1.00 | 0.24 | 0.25 |
| ETS1 | 0.00 | 0.09 | 0.07 | 0.15 | 0.26 | 0.63 | 1.00 | 0.26 | 0.26 |
| JUNB(var.2) | 0.00 | 0.18 | 0.01 | 0.06 | 0.02 | 0.81 | 0.79 | 0.29 | 1.00 |
| Creb5 | 0.00 | 0.18 | 0.02 | 0.05 | 0.07 | 1.00 | 0.77 | 0.28 | 0.90 |
| FOSL2::JUNB(var.2) | 0.01 | 0.19 | 0.00 | 0.02 | 0.10 | 1.00 | 0.76 | 0.42 | 0.95 |
| JUN::JUNB(var.2) | 0.04 | 0.20 | 0.02 | 0.00 | 0.15 | 1.00 | 0.79 | 0.45 | 0.95 |
| FOSB::JUNB(var.2) | 0.05 | 0.19 | 0.00 | 0.04 | 0.12 | 0.95 | 0.74 | 0.38 | 1.00 |
| FOSL1::JUN(var.2) | 0.04 | 0.19 | 0.00 | 0.02 | 0.13 | 0.99 | 0.69 | 0.35 | 1.00 |
| ATF7 | 0.00 | 0.18 | 0.02 | 0.00 | 0.13 | 1.00 | 0.73 | 0.36 | 1.00 |
| FOSB::JUN | 0.04 | 0.17 | 0.00 | 0.03 | 0.11 | 1.00 | 0.68 | 0.43 | 0.94 |
| JDP2(var.2) | 0.03 | 0.16 | 0.00 | 0.02 | 0.07 | 1.00 | 0.68 | 0.34 | 0.93 |
| BATF3 | 0.11 | 0.20 | 0.12 | 0.00 | 0.22 | 1.00 | 0.75 | 0.51 | 0.77 |
| FOSL1::JUND(var.2) | 0.07 | 0.18 | 0.00 | 0.00 | 0.15 | 1.00 | 0.71 | 0.42 | 0.71 |
| CREB1 | 0.14 | 0.22 | 0.08 | 0.00 | 0.18 | 1.00 | 0.53 | 0.47 | 0.87 |
| FOS::JUN(var.2) | 0.07 | 0.26 | 0.00 | 0.05 | 0.18 | 1.00 | 0.66 | 0.53 | 0.93 |
| FOSL2::JUN(var.2) | 0.12 | 0.23 | 0.00 | 0.05 | 0.13 | 1.00 | 0.69 | 0.42 | 0.92 |
| FOSL2::JUND(var.2) | 0.07 | 0.20 | 0.00 | 0.00 | 0.11 | 1.00 | 0.61 | 0.43 | 0.90 |
| RELA | 0.00 | 0.05 | 0.21 | 0.13 | 0.21 | 0.55 | 0.51 | 1.00 | 0.85 |
| REL | 0.00 | 0.03 | 0.21 | 0.10 | 0.22 | 0.53 | 0.55 | 1.00 | 0.80 |
| CEBPA | 0.10 | 0.04 | 0.08 | 0.08 | 0.57 | 1.00 | 0.00 | 0.65 | 0.64 |
| JUN | 0.20 | 0.25 | 0.01 | 0.00 | 0.20 | 1.00 | 0.35 | 0.61 | 0.81 |
| JUND(var.2) | 0.16 | 0.23 | 0.02 | 0.00 | 0.16 | 1.00 | 0.46 | 0.65 | 0.98 |
| Nfe2l2 | 0.36 | 0.42 | 0.15 | 0.00 | 0.33 | 1.00 | 0.42 | 0.54 | 0.52 |
| Bach1::Mafk | 0.38 | 0.41 | 0.17 | 0.00 | 0.35 | 1.00 | 0.45 | 0.54 | 0.47 |
| BACH2 | 0.35 | 0.41 | 0.15 | 0.00 | 0.33 | 1.00 | 0.47 | 0.53 | 0.46 |
| MAF::NFE2 | 0.36 | 0.40 | 0.15 | 0.00 | 0.34 | 1.00 | 0.44 | 0.52 | 0.47 |
| BATF::JUN | 0.28 | 0.41 | 0.08 | 0.00 | 0.29 | 1.00 | 0.54 | 0.53 | 0.46 |
| FOS | 0.30 | 0.44 | 0.10 | 0.00 | 0.28 | 1.00 | 0.51 | 0.55 | 0.48 |
| FOSL1::JUND | 0.29 | 0.44 | 0.09 | 0.00 | 0.29 | 1.00 | 0.52 | 0.54 | 0.49 |
| JUN(var.2) | 0.30 | 0.45 | 0.09 | 0.00 | 0.29 | 1.00 | 0.56 | 0.55 | 0.49 |
| JDP2 | 0.30 | 0.47 | 0.08 | 0.00 | 0.29 | 1.00 | 0.56 | 0.57 | 0.50 |
| FOSL1::JUNB | 0.31 | 0.45 | 0.09 | 0.00 | 0.30 | 1.00 | 0.51 | 0.55 | 0.51 |
| FOSB::JUNB | 0.31 | 0.47 | 0.09 | 0.00 | 0.30 | 1.00 | 0.52 | 0.56 | 0.51 |
| FOSL2::JUND | 0.31 | 0.46 | 0.09 | 0.00 | 0.30 | 1.00 | 0.52 | 0.56 | 0.51 |
| FOS::JUNB | 0.31 | 0.46 | 0.09 | 0.00 | 0.30 | 1.00 | 0.52 | 0.56 | 0.53 |
| FOSL2::JUNB | 0.30 | 0.46 | 0.09 | 0.00 | 0.29 | 1.00 | 0.53 | 0.56 | 0.52 |
| NFE2 | 0.32 | 0.46 | 0.09 | 0.00 | 0.31 | 1.00 | 0.54 | 0.56 | 0.51 |
| FOSL1 | 0.32 | 0.45 | 0.11 | 0.00 | 0.31 | 1.00 | 0.52 | 0.57 | 0.50 |
| JUNB | 0.31 | 0.45 | 0.09 | 0.00 | 0.29 | 1.00 | 0.53 | 0.57 | 0.49 |
| FOSL2 | 0.31 | 0.45 | 0.09 | 0.00 | 0.29 | 1.00 | 0.52 | 0.57 | 0.50 |
| JUND | 0.31 | 0.46 | 0.10 | 0.00 | 0.29 | 1.00 | 0.52 | 0.58 | 0.50 |
| FOS::JUN | 0.30 | 0.46 | 0.09 | 0.00 | 0.30 | 1.00 | 0.53 | 0.56 | 0.50 |
| JUN::JUNB | 0.31 | 0.45 | 0.09 | 0.00 | 0.30 | 1.00 | 0.54 | 0.56 | 0.51 |
| FOS::JUND | 0.31 | 0.46 | 0.09 | 0.00 | 0.29 | 1.00 | 0.52 | 0.57 | 0.51 |
| FOSL1::JUN | 0.31 | 0.46 | 0.09 | 0.00 | 0.29 | 1.00 | 0.52 | 0.57 | 0.51 |
| FOSL2::JUN | 0.31 | 0.46 | 0.09 | 0.00 | 0.29 | 1.00 | 0.53 | 0.57 | 0.52 |
| PBX3 | 0.22 | 0.39 | 0.34 | 0.00 | 0.28 | 1.00 | 0.36 | 0.80 | 0.51 |
| MAFG::NFE2L1 | 0.23 | 0.31 | 0.13 | 0.00 | 0.32 | 1.00 | 0.29 | 0.38 | 0.42 |
| ATF4 | 0.26 | 0.26 | 0.11 | 0.00 | 0.35 | 1.00 | 0.26 | 0.50 | 0.67 |

## Table 2.2. Sequence motifs with variable enrichment across cell types, continued.

| Motif | beta | delta | alpha | gamma | immune | stellate | endothelial | ductal | acinar |
|---|---|---|---|---|---|---|---|---|---|
| NRL | 0.74 | 0.19 | 1.00 | 0.00 | 0.79 | 0.43 | 0.41 | 0.28 | 0.21 |
| Mafb | 0.79 | 0.23 | 1.00 | 0.00 | 0.82 | 0.67 | 0.47 | 0.29 | 0.33 |
| MAFG | 0.80 | 0.41 | 1.00 | 0.00 | 0.81 | 0.66 | 0.50 | 0.49 | 0.40 |
| MEF2B | 0.66 | 0.35 | 0.54 | 0.65 | 0.52 | 1.00 | 0.75 | 0.00 | 0.20 |
| MAFK | 0.66 | 0.49 | 0.65 | 0.00 | 0.62 | 1.00 | 0.53 | 0.56 | 0.53 |
| MAFF | 0.63 | 0.47 | 0.62 | 0.00 | 0.61 | 1.00 | 0.49 | 0.57 | 0.55 |
| NR4A1 | 0.11 | 0.00 | 0.51 | 0.06 | 0.45 | 1.00 | 0.39 | 0.17 | 0.49 |
| USF1 | 0.44 | 0.35 | 0.35 | 0.00 | 0.60 | 1.00 | 0.63 | 0.33 | 0.89 |
| MITF | 0.28 | 0.25 | 0.20 | 0.00 | 0.48 | 1.00 | 0.58 | 0.25 | 0.62 |
| HNF1B | 0.28 | 0.58 | 0.73 | 0.75 | 0.31 | 0.00 | 0.18 | 1.00 | 0.80 |
| HNF1A | 0.26 | 0.54 | 0.71 | 0.71 | 0.31 | 0.00 | 0.17 | 1.00 | 0.76 |
| HNF4G | 0.00 | 0.26 | 0.72 | 0.77 | 0.20 | 0.08 | 0.03 | 0.69 | 1.00 |
| Hnf4a | 0.00 | 0.30 | 0.70 | 0.85 | 0.20 | 0.07 | 0.06 | 0.78 | 1.00 |
| CTCF | 0.22 | 0.26 | 0.67 | 0.98 | 0.67 | 0.00 | 0.56 | 0.46 | 1.00 |
| GATA1::TAL1 | 0.19 | 0.53 | 1.00 | 0.87 | 0.47 | 0.00 | 0.47 | 0.80 | 0.82 |
| GATA3 | 0.00 | 0.25 | 0.89 | 0.74 | 0.35 | 0.01 | 0.31 | 0.92 | 1.00 |
| GATA5 | 0.02 | 0.24 | 0.83 | 0.75 | 0.33 | 0.00 | 0.37 | 0.92 | 1.00 |
| GATA6 | 0.03 | 0.28 | 0.96 | 0.83 | 0.43 | 0.00 | 0.34 | 0.98 | 1.00 |
| GATA2 | 0.04 | 0.27 | 0.91 | 0.78 | 0.39 | 0.00 | 0.32 | 0.98 | 1.00 |
| Gata4 | 0.10 | 0.34 | 0.98 | 0.87 | 0.43 | 0.00 | 0.35 | 1.00 | 0.96 |
| Gata1 | 0.02 | 0.25 | 1.00 | 0.86 | 0.44 | 0.00 | 0.33 | 0.99 | 0.91 |
| STAT3 | 0.00 | 0.02 | 0.49 | 0.68 | 0.51 | 1.00 | 0.56 | 0.79 | 0.88 |
| STAT1 | 0.00 | 0.09 | 0.31 | 0.57 | 0.49 | 0.91 | 0.78 | 0.67 | 1.00 |
| Stat4 | 0.00 | 0.03 | 0.34 | 0.51 | 0.49 | 0.75 | 0.80 | 0.67 | 1.00 |
| ELF4 | 0.00 | 0.24 | 0.23 | 0.38 | 0.52 | 0.53 | 1.00 | 0.69 | 0.71 |
| ELF1 | 0.00 | 0.21 | 0.23 | 0.34 | 0.40 | 0.55 | 1.00 | 0.63 | 0.67 |
| ELF5 | 0.00 | 0.40 | 0.34 | 0.59 | 0.66 | 0.36 | 0.96 | 1.00 | 0.90 |
| ELF3 | 0.00 | 0.35 | 0.33 | 0.52 | 0.59 | 0.43 | 1.00 | 0.98 | 0.89 |
| EHF | 0.00 | 0.33 | 0.32 | 0.53 | 0.59 | 0.45 | 1.00 | 0.95 | 0.90 |
| IRF8 | 0.00 | 0.45 | 0.51 | 0.92 | 1.00 | 0.67 | 0.51 | 0.71 | 0.92 |
| IRF4 | 0.00 | 0.50 | 0.47 | 0.74 | 1.00 | 0.60 | 0.38 | 0.68 | 0.94 |
| SPI1 | 0.00 | 0.15 | 0.07 | 0.26 | 1.00 | 0.24 | 0.67 | 0.43 | 0.35 |
| SPIC | 0.00 | 0.27 | 0.13 | 0.43 | 1.00 | 0.48 | 0.98 | 0.63 | 0.51 |

## Table 2.3. Sequence motifs differentially enriched across endocrine cell states.

| Motif | GCG$^{high}$ alpha | INS$^{high}$ beta | SST$^{high}$ delta | GCG$^{low}$ alpha | INS$^{low}$ beta | SST$^{low}$ delta |
|---|---|---|---|---|---|---|
| NR2F1 | 0.44 | 0.35 | 0.00 | 1.00 | 0.48 | 0.41 |
| MEF2C | 0.21 | 0.35 | 0.00 | 1.00 | 0.67 | 0.37 |
| MEF2A | 0.21 | 0.29 | 0.00 | 1.00 | 0.61 | 0.35 |
| MEF2B | 0.21 | 0.33 | 0.00 | 1.00 | 0.65 | 0.42 |
| MEF2D | 0.22 | 0.30 | 0.00 | 1.00 | 0.61 | 0.41 |
| TFEC | 0.09 | 0.04 | 0.00 | 1.00 | 0.49 | 0.62 |
| MITF | 0.00 | 0.02 | 0.02 | 1.00 | 0.53 | 0.64 |
| USF1 | 0.05 | 0.07 | 0.00 | 1.00 | 0.57 | 0.64 |
| NRF1 | 0.00 | 0.02 | 0.03 | 1.00 | 0.37 | 0.49 |
| TFEB | 0.14 | 0.06 | 0.00 | 1.00 | 0.45 | 0.56 |
| TFE3 | 0.12 | 0.10 | 0.00 | 1.00 | 0.51 | 0.52 |
| USF2 | 0.06 | 0.05 | 0.00 | 1.00 | 0.47 | 0.53 |
| CREB3 | 0.27 | 0.22 | 0.00 | 1.00 | 0.37 | 0.71 |
| RARA | 0.28 | 0.17 | 0.00 | 1.00 | 0.44 | 0.55 |
| Srebf1(var.2) | 0.23 | 0.11 | 0.00 | 1.00 | 0.48 | 0.51 |
| SREBF2(var.2) | 0.19 | 0.06 | 0.00 | 1.00 | 0.47 | 0.48 |
| KLF16 | 0.24 | 0.00 | 0.26 | 1.00 | 0.40 | 0.61 |
| SP3 | 0.32 | 0.00 | 0.15 | 1.00 | 0.38 | 0.66 |
| NFYA | 0.38 | 0.00 | 0.24 | 1.00 | 0.30 | 0.70 |
| NFYB | 0.27 | 0.00 | 0.21 | 1.00 | 0.33 | 0.76 |
| Atf1 | 0.11 | 0.00 | 0.03 | 1.00 | 0.20 | 0.48 |
| Crem | 0.12 | 0.00 | 0.03 | 1.00 | 0.18 | 0.54 |
| Atf3 | 0.24 | 0.00 | 0.03 | 1.00 | 0.21 | 0.64 |
| Klf1 | 0.33 | 0.00 | 0.00 | 1.00 | 0.29 | 0.65 |
| Klf12 | 0.28 | 0.03 | 0.00 | 1.00 | 0.26 | 0.62 |
| KLF13 | 0.36 | 0.00 | 0.14 | 1.00 | 0.22 | 0.57 |
| SP4 | 0.25 | 0.00 | 0.09 | 1.00 | 0.34 | 0.55 |
| KLF14 | 0.32 | 0.00 | 0.10 | 1.00 | 0.31 | 0.55 |
| SP8 | 0.28 | 0.00 | 0.11 | 1.00 | 0.28 | 0.55 |
| PBX2 | 0.67 | 0.00 | 0.07 | 1.00 | 0.44 | 0.41 |
| HEY2 | 0.24 | 0.05 | 0.00 | 1.00 | 0.30 | 0.27 |
| MAX | 0.26 | 0.01 | 0.00 | 1.00 | 0.25 | 0.19 |
| MYC | 0.31 | 0.16 | 0.00 | 1.00 | 0.31 | 0.11 |
| HEY1 | 0.34 | 0.13 | 0.00 | 1.00 | 0.31 | 0.27 |
| ZNF384 | 0.33 | 0.15 | 0.00 | 1.00 | 0.35 | 0.32 |
| Arnt | 0.33 | 0.05 | 0.00 | 1.00 | 0.25 | 0.25 |
| Id2 | 0.36 | 0.14 | 0.00 | 1.00 | 0.25 | 0.22 |
| Hes1 | 0.30 | 0.04 | 0.00 | 1.00 | 0.27 | 0.46 |
| Nr2f6(var.2) | 0.38 | 0.02 | 0.00 | 1.00 | 0.26 | 0.50 |
| Rarg | 0.36 | 0.07 | 0.00 | 1.00 | 0.33 | 0.50 |
| Creb3l2 | 0.27 | 0.10 | 0.00 | 1.00 | 0.36 | 0.35 |
| MLX | 0.22 | 0.05 | 0.00 | 1.00 | 0.29 | 0.39 |
| BHLHE40 | 0.38 | 0.02 | 0.00 | 1.00 | 0.22 | 0.37 |
| Arntl | 0.31 | 0.03 | 0.00 | 1.00 | 0.24 | 0.39 |
| BHLHE41 | 0.31 | 0.04 | 0.00 | 1.00 | 0.25 | 0.41 |
| MLXIPL | 0.28 | 0.06 | 0.00 | 1.00 | 0.27 | 0.36 |
| Tcfl5 | 0.28 | 0.03 | 0.00 | 1.00 | 0.21 | 0.36 |
| DUX4 | 0.00 | 0.33 | 0.53 | 0.84 | 0.77 | 1.00 |
| DUXA | 0.00 | 0.57 | 0.64 | 0.63 | 0.84 | 1.00 |
| ATF4 | 0.03 | 0.00 | 0.10 | 0.74 | 0.88 | 1.00 |
| Bach1::Mafk | 0.00 | 0.19 | 0.19 | 0.76 | 0.71 | 1.00 |
| MAF::NFE2 | 0.00 | 0.16 | 0.18 | 0.77 | 0.70 | 1.00 |
| BACH2 | 0.00 | 0.15 | 0.20 | 0.78 | 0.68 | 1.00 |
| Nfe2l2 | 0.00 | 0.14 | 0.19 | 0.76 | 0.68 | 1.00 |
| FOS::JUND | 0.00 | 0.11 | 0.26 | 0.79 | 0.65 | 1.00 |
| FOSL2 | 0.00 | 0.11 | 0.26 | 0.78 | 0.65 | 1.00 |
| JUN(var.2) | 0.00 | 0.11 | 0.25 | 0.77 | 0.64 | 1.00 |
| JUNB | 0.00 | 0.11 | 0.26 | 0.77 | 0.64 | 1.00 |
| FOSL1 | 0.00 | 0.11 | 0.25 | 0.78 | 0.65 | 1.00 |
| FOS | 0.00 | 0.10 | 0.25 | 0.78 | 0.64 | 1.00 |
| JUND | 0.00 | 0.11 | 0.25 | 0.78 | 0.64 | 1.00 |
| Pax2 | 0.00 | 0.11 | 0.29 | 0.82 | 0.64 | 1.00 |
| JUN::JUNB | 0.00 | 0.10 | 0.25 | 0.82 | 0.66 | 1.00 |
| NFE2 | 0.00 | 0.12 | 0.25 | 0.80 | 0.67 | 1.00 |
| BATF::JUN | 0.00 | 0.10 | 0.23 | 0.79 | 0.64 | 1.00 |
| FOSB::JUNB | 0.00 | 0.11 | 0.26 | 0.80 | 0.65 | 1.00 |
| FOS::JUN | 0.00 | 0.10 | 0.25 | 0.80 | 0.64 | 1.00 |
| FOSL1::JUN | 0.00 | 0.11 | 0.25 | 0.79 | 0.65 | 1.00 |
| FOS::JUNB | 0.00 | 0.11 | 0.25 | 0.80 | 0.65 | 1.00 |
| FOSL2::JUN | 0.00 | 0.11 | 0.25 | 0.80 | 0.65 | 1.00 |
| FOSL1::JUND | 0.00 | 0.09 | 0.24 | 0.80 | 0.64 | 1.00 |
| FOSL1::JUNB | 0.00 | 0.10 | 0.24 | 0.80 | 0.64 | 1.00 |
| FOSL2::JUND | 0.00 | 0.10 | 0.25 | 0.81 | 0.65 | 1.00 |
| FOSL2::JUNB | 0.00 | 0.10 | 0.25 | 0.81 | 0.64 | 1.00 |
| JDP2 | 0.00 | 0.10 | 0.25 | 0.80 | 0.64 | 1.00 |
| KLF4 | 0.06 | 0.00 | 0.05 | 0.68 | 0.36 | 1.00 |
| PPARG | 0.00 | 0.07 | 0.23 | 0.51 | 0.49 | 1.00 |
| PBX3 | 0.24 | 0.00 | 0.15 | 1.00 | 0.64 | 0.95 |
| EGR2 | 0.12 | 0.00 | 0.09 | 0.82 | 0.46 | 1.00 |
| TBX1 | 0.21 | 0.00 | 0.11 | 0.86 | 0.40 | 1.00 |
| ZNF24 | 0.21 | 0.00 | 0.23 | 0.82 | 0.41 | 1.00 |
| EGR3 | 0.18 | 0.00 | 0.21 | 0.89 | 0.52 | 1.00 |
| MGA | 0.17 | 0.00 | 0.18 | 0.93 | 0.43 | 1.00 |
| TBX15 | 0.17 | 0.00 | 0.18 | 0.93 | 0.43 | 1.00 |
| FOSB::JUNB(var.2) | 0.00 | 0.05 | 0.12 | 1.00 | 0.43 | 0.88 |
| FOSL2::JUND(var.2) | 0.00 | 0.06 | 0.12 | 1.00 | 0.46 | 0.88 |
| CREB1 | 0.00 | 0.06 | 0.08 | 1.00 | 0.45 | 0.84 |
| FOSB::JUN | 0.00 | 0.04 | 0.09 | 1.00 | 0.43 | 0.87 |
| FOSL1::JUN(var.2) | 0.00 | 0.03 | 0.10 | 1.00 | 0.41 | 0.87 |
| JUNB(var.2) | 0.00 | 0.01 | 0.10 | 1.00 | 0.34 | 0.82 |
| ATF7 | 0.01 | 0.00 | 0.08 | 1.00 | 0.34 | 0.81 |
| Creb5 | 0.00 | 0.00 | 0.08 | 1.00 | 0.33 | 0.82 |
| JUN::JUNB(var.2) | 0.00 | 0.01 | 0.09 | 1.00 | 0.39 | 0.87 |
| FOSL2::JUNB(var.2) | 0.00 | 0.00 | 0.10 | 1.00 | 0.38 | 0.83 |
| JDP2(var.2) | 0.00 | 0.02 | 0.08 | 1.00 | 0.41 | 0.83 |
| BATF3 | 0.00 | 0.00 | 0.03 | 1.00 | 0.35 | 0.73 |
| HSF2 | 0.02 | 0.09 | 0.00 | 1.00 | 0.44 | 0.78 |
| EGR4 | 0.15 | 0.11 | 0.00 | 0.86 | 0.56 | 1.00 |
| HSF1 | 0.17 | 0.18 | 0.00 | 1.00 | 0.52 | 0.79 |
| HSF4 | 0.12 | 0.06 | 0.00 | 1.00 | 0.44 | 0.88 |
| XBP1 | 0.15 | 0.19 | 0.00 | 1.00 | 0.42 | 0.93 |
| FOS::JUN(var.2) | 0.00 | 0.05 | 0.19 | 1.00 | 0.48 | 0.94 |
| FOSL2::JUN(var.2) | 0.00 | 0.11 | 0.17 | 1.00 | 0.52 | 0.93 |
| SREBF1 | 0.05 | 0.10 | 0.00 | 1.00 | 0.54 | 0.93 |
| JUND(var.2) | 0.00 | 0.09 | 0.11 | 1.00 | 0.58 | 0.93 |
| FOSL1::JUND(var.2) | 0.00 | 0.03 | 0.11 | 1.00 | 0.50 | 0.92 |
| SREBF2 | 0.05 | 0.00 | 0.09 | 1.00 | 0.55 | 0.97 |

## Table 2.3. Sequence motifs differentially enriched across endocrine cell states, continued.

| Motif | GCG$^{high}$ alpha | INS$^{high}$ beta | SST$^{high}$ delta | GCG$^{low}$ alpha | INS$^{low}$ beta | SST$^{low}$ delta |
|---|---|---|---|---|---|---|
| JUN | 0.00 | 0.13 | 0.18 | 1.00 | 0.72 | 1.00 |
| MAFG::NFE2L1 | 0.00 | 0.06 | 0.13 | 0.87 | 0.60 | 1.00 |
| TEAD1 | 0.57 | 0.00 | 0.21 | 1.00 | 0.56 | 0.88 |
| TEAD4 | 0.54 | 0.00 | 0.19 | 1.00 | 0.53 | 0.85 |
| TEAD2 | 0.49 | 0.00 | 0.21 | 1.00 | 0.58 | 0.88 |
| TEAD3 | 0.48 | 0.00 | 0.17 | 1.00 | 0.64 | 0.90 |
| ESR1 | 0.34 | 0.36 | 0.00 | 1.00 | 0.79 | 0.90 |
| MAFF | 0.49 | 0.48 | 0.00 | 1.00 | 0.77 | 1.00 |
| MAFK | 0.51 | 0.50 | 0.00 | 1.00 | 0.74 | 0.98 |
| EHF | 0.94 | 0.00 | 1.00 | 0.38 | 0.10 | 0.61 |
| ELF3 | 0.90 | 0.00 | 1.00 | 0.27 | 0.09 | 0.52 |
| ELF5 | 0.82 | 0.00 | 1.00 | 0.32 | 0.14 | 0.53 |
| IRF9 | 0.74 | 0.04 | 1.00 | 0.15 | 0.00 | 0.16 |
| IRF4 | 0.84 | 0.10 | 1.00 | 0.26 | 0.00 | 0.31 |
| IRF8 | 0.95 | 0.12 | 1.00 | 0.23 | 0.00 | 0.22 |
| CTCF | 1.00 | 0.61 | 0.61 | 0.59 | 0.03 | 0.00 |
| Bhlha15 | 1.00 | 0.41 | 0.69 | 0.34 | 0.00 | 0.21 |
| NEUROD2 | 1.00 | 0.46 | 0.60 | 0.41 | 0.00 | 0.24 |
| BHLHE22 | 1.00 | 0.36 | 0.53 | 0.39 | 0.00 | 0.20 |
| BHLHE23 | 1.00 | 0.37 | 0.53 | 0.40 | 0.00 | 0.21 |
| Atoh1 | 1.00 | 0.38 | 0.56 | 0.39 | 0.00 | 0.18 |
| OLIG1 | 1.00 | 0.39 | 0.56 | 0.37 | 0.00 | 0.20 |
| OLIG2 | 1.00 | 0.36 | 0.44 | 0.39 | 0.00 | 0.24 |
| OLIG3 | 1.00 | 0.34 | 0.50 | 0.44 | 0.00 | 0.29 |
| HNF1A | 1.00 | 0.00 | 0.77 | 0.87 | 0.27 | 0.41 |
| HNF1B | 1.00 | 0.00 | 0.82 | 0.79 | 0.25 | 0.43 |
| Gata1 | 1.00 | 0.00 | 0.27 | 0.58 | 0.03 | 0.16 |
| GATA2 | 1.00 | 0.02 | 0.29 | 0.54 | 0.00 | 0.17 |
| Gata4 | 1.00 | 0.00 | 0.30 | 0.55 | 0.01 | 0.17 |
| GATA5 | 1.00 | 0.01 | 0.31 | 0.54 | 0.00 | 0.15 |
| GATA3 | 1.00 | 0.00 | 0.30 | 0.55 | 0.01 | 0.19 |
| GATA6 | 1.00 | 0.00 | 0.28 | 0.57 | 0.01 | 0.18 |
| GATA1::TAL1 | 1.00 | 0.08 | 0.45 | 0.48 | 0.00 | 0.33 |
| Hnf4a | 1.00 | 0.00 | 0.53 | 0.66 | 0.05 | 0.19 |
| HNF4G | 1.00 | 0.00 | 0.48 | 0.67 | 0.07 | 0.11 |
| FOXP2 | 0.81 | 0.86 | 0.48 | 0.51 | 1.00 | 0.00 |
| Foxo1 | 0.93 | 1.00 | 0.58 | 0.63 | 0.99 | 0.00 |
| FOXK1 | 1.00 | 0.91 | 0.45 | 0.69 | 0.96 | 0.00 |
| FOXP1 | 1.00 | 0.88 | 0.50 | 0.68 | 0.98 | 0.00 |
| FOXP3 | 1.00 | 0.90 | 0.49 | 0.67 | 0.96 | 0.00 |
| FOXO6 | 1.00 | 0.90 | 0.49 | 0.67 | 0.96 | 0.00 |
| FOXO4 | 1.00 | 0.90 | 0.49 | 0.67 | 0.96 | 0.00 |
| FOXL1 | 1.00 | 0.90 | 0.49 | 0.67 | 0.96 | 0.00 |
| FOXD2 | 1.00 | 0.90 | 0.49 | 0.67 | 0.96 | 0.00 |
| FOXI1 | 1.00 | 0.90 | 0.49 | 0.67 | 0.96 | 0.00 |
| NRL | 1.00 | 0.88 | 0.13 | 0.35 | 0.34 | 0.00 |
| Mafb | 1.00 | 0.89 | 0.00 | 0.55 | 0.41 | 0.06 |
| MAFG | 1.00 | 0.81 | 0.00 | 0.66 | 0.46 | 0.27 |
| FOXA1 | 1.00 | 0.72 | 0.61 | 0.42 | 0.56 | 0.00 |
| Foxa2 | 1.00 | 0.74 | 0.59 | 0.46 | 0.61 | 0.00 |
| FOXC2 | 1.00 | 0.69 | 0.43 | 0.65 | 0.71 | 0.00 |
| FOXB1 | 1.00 | 0.58 | 0.41 | 0.64 | 0.47 | 0.00 |
| FOXC1 | 1.00 | 0.64 | 0.41 | 0.65 | 0.56 | 0.00 |
| Alx1 | 0.00 | 0.68 | 1.00 | 0.41 | 0.76 | 0.91 |
| Hoxa9 | 0.00 | 0.64 | 1.00 | 0.44 | 0.83 | 0.96 |
| GSX1 | 0.00 | 0.90 | 1.00 | 0.32 | 0.82 | 0.86 |
| Lhx8 | 0.00 | 0.93 | 1.00 | 0.33 | 0.85 | 0.86 |
| PAX4 | 0.00 | 0.93 | 1.00 | 0.33 | 0.85 | 0.86 |
| EMX1 | 0.00 | 0.85 | 1.00 | 0.30 | 0.78 | 0.89 |
| GSX2 | 0.00 | 0.79 | 1.00 | 0.38 | 0.79 | 0.87 |
| NHLH1 | 0.70 | 1.00 | 0.93 | 0.00 | 0.57 | 0.28 |
| Myod1 | 0.75 | 1.00 | 0.91 | 0.00 | 0.54 | 0.25 |
| ZBTB18 | 0.75 | 1.00 | 0.93 | 0.00 | 0.58 | 0.25 |
| TFAP4 | 0.74 | 1.00 | 0.80 | 0.00 | 0.53 | 0.19 |
| ASCL1 | 0.70 | 1.00 | 0.89 | 0.00 | 0.50 | 0.24 |
| NEUROD1 | 0.72 | 1.00 | 0.88 | 0.00 | 0.47 | 0.26 |
| Myog | 0.66 | 1.00 | 0.87 | 0.00 | 0.55 | 0.25 |
| TWIST1 | 0.68 | 1.00 | 0.88 | 0.00 | 0.56 | 0.19 |
| Ascl2 | 0.68 | 1.00 | 0.88 | 0.00 | 0.52 | 0.22 |
| Tcf12 | 0.69 | 1.00 | 0.86 | 0.00 | 0.52 | 0.19 |
| NKX2-3 | 0.74 | 1.00 | 0.89 | 0.03 | 0.40 | 0.00 |
| MYF6 | 0.77 | 0.84 | 1.00 | 0.00 | 0.38 | 0.23 |
| MSC | 0.85 | 0.95 | 1.00 | 0.00 | 0.45 | 0.15 |
| Tcf21 | 0.83 | 0.98 | 1.00 | 0.00 | 0.44 | 0.18 |
| TAL1::TCF3 | 0.81 | 1.00 | 0.93 | 0.00 | 0.40 | 0.20 |
| ZEB1 | 0.91 | 1.00 | 0.92 | 0.00 | 0.37 | 0.16 |
| ID4 | 0.99 | 0.87 | 1.00 | 0.00 | 0.40 | 0.01 |
| NEUROG2 | 1.00 | 0.88 | 0.86 | 0.00 | 0.26 | 0.25 |
| NFATC1 | 0.89 | 1.00 | 0.76 | 0.32 | 0.53 | 0.00 |
| NFATC3 | 0.88 | 1.00 | 0.76 | 0.28 | 0.46 | 0.00 |
| MTF1 | 0.94 | 1.00 | 0.82 | 0.42 | 0.46 | 0.00 |
| NFAT5 | 1.00 | 0.95 | 0.75 | 0.34 | 0.48 | 0.00 |
| NFIC | 0.46 | 1.00 | 0.53 | 0.00 | 0.69 | 0.19 |
| NFATC2 | 0.82 | 1.00 | 0.58 | 0.20 | 0.56 | 0.00 |
| Nkx2-5(var.2) | 0.81 | 1.00 | 0.38 | 0.00 | 0.50 | 0.08 |
| FIGLA | 0.51 | 0.45 | 1.00 | 0.00 | 0.13 | 0.16 |
| ISL2 | 0.26 | 0.51 | 1.00 | 0.00 | 0.26 | 0.38 |
| Nkx3-1 | 0.42 | 0.44 | 1.00 | 0.00 | 0.12 | 0.31 |
| NKX3-2 | 0.38 | 0.52 | 1.00 | 0.00 | 0.20 | 0.35 |
| SPIB | 0.59 | 0.45 | 1.00 | 0.00 | 0.52 | 0.46 |
| EWSR1-FLI1 | 0.62 | 0.78 | 1.00 | 0.00 | 0.62 | 0.27 |
| SOX10 | 0.50 | 0.68 | 1.00 | 0.00 | 0.56 | 0.34 |
| Bcl6 | 1.00 | 0.54 | 0.67 | 0.00 | 0.53 | 0.22 |
| Stat5a::Stat5b | 1.00 | 0.65 | 0.58 | 0.00 | 0.41 | 0.26 |
| STAT1 | 1.00 | 0.62 | 0.71 | 0.00 | 0.26 | 0.28 |
| TCF4 | 1.00 | 0.64 | 0.68 | 0.11 | 0.20 | 0.00 |
| STAT3 | 1.00 | 0.67 | 0.72 | 0.00 | 0.27 | 0.08 |
| Stat4 | 1.00 | 0.63 | 0.68 | 0.00 | 0.26 | 0.14 |
| STAT1::STAT2 | 0.70 | 0.32 | 1.00 | 0.00 | 0.27 | 0.05 |
| IRF2 | 0.84 | 0.34 | 1.00 | 0.09 | 0.15 | 0.00 |
| IRF3 | 0.81 | 0.19 | 1.00 | 0.10 | 0.10 | 0.00 |
| RFX5 | 0.72 | 0.82 | 1.00 | 0.08 | 0.16 | 0.00 |
| RFX4 | 0.70 | 0.77 | 1.00 | 0.16 | 0.10 | 0.00 |
| RFX2 | 0.69 | 0.79 | 1.00 | 0.13 | 0.14 | 0.00 |
| RFX3 | 0.71 | 0.79 | 1.00 | 0.13 | 0.14 | 0.00 |
| Rfx1 | 0.88 | 0.80 | 1.00 | 0.04 | 0.14 | 0.00 |
| SNAI2 | 0.91 | 0.63 | 1.00 | 0.00 | 0.22 | 0.03 |

## Table 2.4. Primer sequences.

| Description | Sequence |
|---|---|
| rs7482891 | |
| construct primer left | AGAGGTCTGAGGAGCCCTTG |
| construct primer right | TAGACCCTGCAGAGCCACAG |
| rs34584161 | |
| construct primer left | AAGCTGACAGACAGAGGGTCA |
| construct primer right | GGGCTTCATAAACATCAGCA |
| rs17712208 | |
| construct primer left | AAGCCCACCTTCGTAAACAT |
| construct primer right | TGAAGTAGCTCCCAGTGAAGG |
| rs78840640 | |
| construct primer left | CACAATGAAGCCATGTCCTTT |
| construct primer right | TCAGCTTTCTATTTTGGGGAAA |
| rs4679370 | |
| construct primer left | TCAATGTCTACCTCAAAATTCTTTGT |
| construct primer right | CACTGCAGCCTTAAACTCCTG |
| rs7482891 | |
| SDM primer left | ACTGGTGCCCgCTCCCCGCGT |
| SDM primer right | CCCAGGGCACAGACGTGCTG |
| rs17712208 | |
| SDM primer left | GGAGCTATGGaTAATTATTGACTG |
| SDM primer right | ATTAACGATCCAGTCAGC |
| rs78840640 | |
| SDM primer left | ATCAGATTTGgTGAGAAAGAAGAAC |
| SDM primer right | GCCCATCAATTCTGAGCATG |
| rs4679370 | |
| SDM primer left | ATCAGTAAGCcCCTAAAGCCTG |
| SDM primer right | TAACTTGAGGCAATGGTG |
| Deletion | |
| sgRNA oligos: | |
| KCNQ1_sgRNA1-s: | ACTGTCGGGCCCATCTGCCA |
| KCNQ1_sgRNA1-as: | TGGTTGGATCTGTTGCGGGG |
| Genotyping primers: | |
| Span-F: | AGTGGGGCCATGAACAATAA |
| Span-R: | GCCTGAGTTTCCGTGACTGT |
| Base editing | |
| sgRNA oligos: | |
| KCNQ1_base editing_sgRNA1-s: | CACCGTGCCTAGAGACAAGCTACCA |
| KCNQ1_base editing_sgRNA1-as: | aaacTGGTAGCTTGTCTCTAGGCAC |
| ssODN: | ACCCTGCACATGACGGGCGAGGGAAGAGGACCATGGTAACTTGTCTCTAGGCAACCCATAGGTGCCCAATGGAGAGATAATCTCAAATATGGTAGCAGAGTTCAGACT CAGTGGCTTTCAGATGGTT |
| Genotyping primers: | |
| Span-F: | ATGTGCCTAGAGGCCTGAGA |
| Span-R: | TGAGGGACCTACCAAGGATG |
| qPCR | |
| hu-*TBP*-F: | ATTAAGGGAGGGAGTGGCAC |
| hu-*TBP*-R: | GCTTTGCTTCCCTTTCCCAA |
| hu-*INS*-F: | AAGAGGCCATCAAGCAGATCA |
| hu-*INS*-R: | CAGGAGGCGCATCCACA |
| hu-*CDKN1C*-F: | AGATCAGCGCCTGAGAAGTCGT |
| hu-*CDKN1C*-R: | TCGGGGCTCTTTGGGCTCTAAA |

## 2.9 Data and Code Availability

Raw sequencing data have been deposited into the NCBI Gene Expression Omnibus (GEO) with accession numbers GSE160472, GSE160473, and GSE163610. Processed data files and annotations for snATAC-seq are available through the Diabetes Epigenome Atlas (https://www.diabetesepigenome.org/). All other data are either contained within the article or available upon request to the corresponding author.

Code for processing and clustering the snATAC-seq datasets is available at: https://github.com/kjgaulton/pipelines/tree/master/islet_snATAC_pipeline.

## 2.10 Acknowledgements

This work was supported by NIH R01DK114650 and U01DK105554 (sub-award) to K.G., R01DK068471 and U01DK105541 to M.S., U01DK120429 to K.G. and M.S., and by the UC San Diego School of Medicine to the Center for Epigenomics. We thank the QB3 Macrolab at UC Berkeley for purification of the Tn5 transposase. We thank K. Jepsen, UCSD IGM Genomics Center, and S. Kuan for sequencing and B. Li for bioinformatics support. Data from the UK Biobank was accessed under application 24058. We thank I. Matta for preparation of RNA-seq libraries.

## 2.11 Author Contributions

K.J.G., D.U.G, and M.Sa. conceived of and supervised the research in the study; K.J.G., D.U.G., M.Sa., J.C., C.Z, and Z.C. wrote the manuscript; J.C. performed analyses of single cell and genetic data; C.Z., M.Sc and J.W. performed hESC experiments; Z.C. performed analyses of single cell and Hi-C data; J.Y.H. performed combinatorial barcoding single cell assays and genotyping; M.M performed 10x single cell assays; R.M. performed Hi-C experiments; S.H., A.D.

and M.O. performed reporter experiments; Y.Q. performed analyses of 4C data; Y.Sui performed analyses of hESC data; Y.Sun and P.K. developed and processed data for the epigenome database; R.F. contributed to analyses of single cell data; S.P. contributed to the development of single cell assays.

Chapter 2, in full, is a reformatted reprint of the material as it appears in “Single cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk“ in *Nature Genetics*, 2021. Joshua Chiou, Chun Zeng, Zhang Cheng, Jee Yun Han, Michael Schlichting, Michael Miller, Robert Mendez, Serina Huang, Jinzhao Wang, Yinghui Sui, Allison Deogaygay, Mei-Lin Okino, Yunjiang Qiu, Ying Sun, Parul Kudtarkar, Rongxin Fang, Sebastian Preissl, Maike Sander, David U Gorkin, Kyle J Gaulton. The dissertation author was the primary investigator and author of this paper.

## 2.12 References


1. Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., Bates, D., Hansen, R. S., Neph, S., Sabo, P. J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S. R., Kaul, R. & Stamatoyannopoulos, J. A. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* 337, 1190–1195 (2012).

2. Cusanovich, D. A., Daza, R., Adey, A., Pliner, H., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. & Shendure, J. Multiplex Single Cell Profiling of Chromatin Accessibility by Combinatorial Cellular Indexing. *Science* 348, 910–914 (2015).

3. Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y. & Greenleaf, W. J. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490 (2015).

4. Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D. U., Zhang, Y., Sos, B. C., Afzal, V., Dickel, D. E., Kuan, S., Visel, A., Pennacchio, L. A., Zhang, K. & Ren, B. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* 21, 432–439 (2018).

5. Litzenburger, U. M., Buenrostro, J. D., Wu, B., Shen, Y., Sheffield, N. C., Kathiria, A., Greenleaf, W. J. & Chang, H. Y. Single-cell epigenomic variability reveals functional cancer heterogeneity. *Genome Biol.* 18, 15 (2017).

6. Buenrostro, J. D., Corces, M. R., Lareau, C. A., Wu, B., Schep, A. N., Aryee, M. J., Majeti, R., Chang, H. Y. & Greenleaf, W. J. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* 173, 1535-1548.e16 (2018).

7. Ulirsch, J. C., Lareau, C. A., Bao, E. L., Ludwig, L. S., Guo, M. H., Benner, C., Satpathy, A. T., Kartha, V. K., Salem, R. M., Hirschhorn, J. N., Finucane, H. K., Aryee, M. J., Buenrostro, J. D. & Sankaran, V. G. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* 51, 683–693 (2019).

8. Pliner, H. A., Packer, J. S., McFaline-Figueroa, J. L., Cusanovich, D. A., Daza, R. M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., Adey, A. C., Steemers, F. J., Shendure, J. & Trapnell, C. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* 71, 858-871.e8 (2018).

9. Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M. H., Prins, B. P., Guo, X., Bielak, L. F., Below, J. E., Bowden, D. W., Chambers, J. C., Kim, Y. J., Ng, M. C. Y., Petty, L. E., Sim, X., Zhang, W., Bennett, A. J., Bork-Jensen, J., Brummett, C. M., Canouil, M., Ec Kardt, K.-U., Fischer, K., Kardia, S. L. R., Kronenberg, F., Läll, K., Liu, C.-T., Locke, A. E., Luan, J., Ntalla, I., Nylander, V., Schönherr, S., Schurmann, C., Yengo, L., Bottinger, E. P., Brandslund, I., Christensen, C., Dedoussis, G., Florez, J. C., Ford, I., Franco, O. H., Frayling, T. M., Giedraitis, V., Hackinger, S., Hattersley, A. T., Herder, C., Ikram, M. A., Ingelsson, M., Jørgensen, M. E., Jørgensen, T., Kriebel, J., Kuusisto, J., Ligthart, S., Lindgren,

C. M., Linneberg, A., Lyssenko, V., Mamakou, V., Meitinger, T., Mohlke, K. L., Morris, A. D., Nadkarni, G., Pankow, J. S., Peters, A., Sattar, N., Stančáková, A., Strauch, K., Taylor, K. D., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Witte, D. R., Dupuis, J., Peyser, P. A., Zeggini, E., Loos, R. J. F., Froguel, P., Ingelsson, E., Lind, L., Groop, L., Laakso, M., Collins, F. S., Jukema, J. W., Palmer, C. N. A., Grallert, H., Metspalu, A., Dehghan, A., Köttgen, A., Abecasis, G. R., Meigs, J. B., Rotter, J. I., Marchini, J., Pedersen, O., Hansen, T., Langenberg, C., Wareham, N. J., Stefansson, K., Gloyn, A. L., Morris, A. P., Boehnke, M. & McCarthy, M. I. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* 50, 1505–1513 (2018).

10. Wood, A. R., Jonsson, A., Jackson, A. U., Wang, N., van Leewen, N., Palmer, N. D., Kobes, S., Deelen, J., Boquete-Vilarino, L., Paananen, J., Stančáková, A., Boomsma, D. I., de Geus, E. J. C., Eekhoff, E. M. W., Fritsche, A., Kramer, M., Nijpels, G., Simonis-Bik, A., van Haeften, T. W., Mahajan, A., Boehnke, M., Bergman, R. N., Tuomilehto, J., Collins, F. S., Mohlke, K. L., Banasik, K., Groves, C. J., McCarthy, M. I., Diabetes Research on Patient Stratification (DIRECT), Pearson, E. R., Natali, A., Mari, A., Buchanan, T. A., Taylor, K. D., Xiang, A. H., Gjesing, A. P., Grarup, N., Eiberg, H., Pedersen, O., Chen, Y.-D., Laakso, M., Norris, J. M., Smith, U., Wagenknecht, L. E., Baier, L., Bowden, D. W., Hansen, T., Walker, M., Watanabe, R. M., 't Hart, L. M., Hanson, R. L. & Frayling, T. M. A Genome-Wide Association Study of IVGTT-Based Measures of First-Phase Insulin Secretion Refines the Underlying Physiology of Type 2 Diabetes Variants. *Diabetes* 66, 2296–2309 (2017).

11. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., Wheeler, E., Glazer, N. L., Bouatia-Naji, N., Gloyn, A. L., Lindgren, C. M., Mägi, R., Morris, A. P., Randall, J., Johnson, T., Elliott, P., Rybin, D., Thorleifsson, G., Steinthorsdottir, V., Henneman, P., Grallert, H., Dehghan, A., Hottenga, J. J., Franklin, C. S., Navarro, P., Song, K., Goel, A., Perry, J. R. B., Egan, J. M., Lajunen, T., Grarup, N., Sparsø, T., Doney, A., Voight, B. F., Stringham, H. M., Li, M., Kanoni, S., Shrader, P., Cavalcanti-Proença, C., Kumari, M., Qi, L., Timpson, N. J., Gieger, C., Zabena, C., Rocheleau, G., Ingelsson, E., An, P., O'Connell, J., Luan, J., Elliott, A., McCarroll, S. A., Payne, F., Roccasecca, R. M., Pattou, F., Sethupathy, P., Ardlie, K., Ariyurek, Y., Balkau, B., Barter, P., Beilby, J. P., Ben-Shlomo, Y., Benediktsson, R., Bennett, A. J., Bergmann, S., Bochud, M., Boerwinkle, E., Bonnefond, A., Bonnycastle, L. L., Borch-Johnsen, K., Böttcher, Y., Brunner, E., Bumpstead, S. J., Charpentier, G., Chen, Y.-D. I., Chines, P., Clarke, R., Coin, L. J. M., Cooper, M. N., Cornelis, M., Crawford, G., Crisponi, L., Day, I. N. M., Geus, E. J. C. de, Delplanque, J., Dina, C., Erdos, M. R., Fedson, A. C., Fischer-Rosinsky, A., Forouhi, N. G., Fox, C. S., Frants, R., Franzosi, M. G., Galan, P., Goodarzi, M. O., Graessler, J., Groves, C. J., Grundy, S., Gwilliam, R., Gyllensten, U., Hadjadj, S., Hallmans, G., Hammond, N., Han, X., Hartikainen, A.-L., Hassanali, N., Hayward, C., Heath, S. C., Hercberg, S., Herder, C., Hicks, A. A., Hillman, D. R., Hingorani, A. D., Hofman, A., Hui, J., Hung, J., Isomaa, B., Johnson, P. R. V., Jørgensen, T., Jula, A., Kaakinen, M., Kaprio, J., Kesaniemi, Y. A., Kivimaki, M., Knight, B., Koskinen, S., Kovacs, P., Kyvik, K. O., Lathrop, G. M., Lawlor, D. A., Le Bacquer, O., Lecoeur, C., Li, Y., Lyssenko, V., Mahley, R., Mangino, M., Manning, A. K., Martínez-Larrad, M. T., McAteer, J. B., McCulloch, L. J., McPherson, R., Meisinger, C., Melzer, D., Meyre, D., Mitchell, B. D., Morken, M. A., Mukherjee, S., Naitza, S., Narisu, N., Neville, M. J., Oostra, B. A., Orrù, M., Pakyz, R., Palmer, C. N. A., Paolisso, G., Pattaro, C., Pearson, D., Peden, J. F., Pedersen, N. L., Perola, M., Pfeiffer, A. F. H., Pichler, I., Polasek, O., Posthuma, D., Potter, S. C., Pouta, A., Province, M. A., Psaty, B. M., Rathmann, W., Rayner, N. W., Rice, K., Ripatti, S., Rivadeneira, F., Roden, M., Rolandsson, O., Sandbaek, A., Sandhu, M., Sanna, S., Sayer, A. A., Scheet, P., Scott, L. J., Seedorf, U., Sharp, S. J., Shields, B., Sigurðsson, G., Sijbrands, E. J. G., Silveira, A., Simpson, L., Singleton, A., Smith, N. L., Sovio, U., Swift, A., Syddall, H., Syvänen, A.-C., Tanaka, T., Thorand, B., Tichet, J.,

Tönjes, A., Tuomi, T., Uitterlinden, A. G., van Dijk, K. W., van Hoek, M., Varma, D., Visvikis-Siest, S., Vitart, V., Vogelzangs, N., Waeber, G., Wagner, P. J., Walley, A., Walters, G. B., Ward, K. L., Watkins, H., Weedon, M. N., Wild, S. H., Willemsen, G., Witteman, J. C. M., Yarnell, J. W. G., Zeggini, E., Zelenika, D., Zethelius, B., Zhai, G., Zhao, J. H., Zillikens, M. C., Consortium, D., Consortium, G., Consortium, G. Bp., Borecki, I. B., Loos, R. J. F., Meneton, P., Magnusson, P. K. E., Nathan, D. M., Williams, G. H., Hattersley, A. T., Silander, K., Salomaa, V., Smith, G. D., Bornstein, S. R., Schwarz, P., Spranger, J., Karpe, F., Shuldiner, A. R., Cooper, C., Dedoussis, G. V., Serrano-Ríos, M., Morris, A. D., Lind, L., Palmer, L. J., Hu, F. B., Franks, P. W., Ebrahim, S., Marmot, M., Kao, W. H. L., Pankow, J. S., Sampson, M. J., Kuusisto, J., Laakso, M., Hansen, T., Pedersen, O., Pramstaller, P. P., Wichmann, H. E., Illig, T., Rudan, I., Wright, A. F., Stumvoll, M., Campbell, H., Wilson, J. F., Consortium, A. H. on behalf of P. & Investigators, the M. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* 42, 105–116 (2010).

12. Manning, A. K., Hivert, M.-F., Scott, R. A., Grimsby, J. L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.-T., Bielak, L. F., Prokopenko, I., Amin, N., Barnes, D., Cadby, G., Hottenga, J.-J., Ingelsson, E., Jackson, A. U., Johnson, T., Kanoni, S., Ladenvall, C., Lagou, V., Lahti, J., Lecoeur, C., Liu, Y., Martinez-Larrad, M. T., Montasser, M. E., Navarro, P., Perry, J. R. B., Rasmussen-Torvik, L. J., Salo, P., Sattar, N., Shungin, D., Strawbridge, R. J., Tanaka, T., van Duijn, C. M., An, P., de Andrade, M., Andrews, J. S., Aspelund, T., Atalay, M., Aulchenko, Y., Balkau, B., Bandinelli, S., Beckmann, J. S., Beilby, J. P., Bellis, C., Bergman, R. N., Blangero, J., Boban, M., Boehnke, M., Boerwinkle, E., Bonnycastle, L. L., Boomsma, D. I., Borecki, I. B., Böttcher, Y., Bouchard, C., Brunner, E., Budimir, D., Campbell, H., Carlson, O., Chines, P. S., Clarke, R., Collins, F. S., Corbatón-Anchuelo, A., Couper, D., de Faire, U., Dedoussis, G. V., Deloukas, P., Dimitriou, M., Egan, J. M., Eiriksdottir, G., Erdos, M. R., Eriksson, J. G., Eury, E., Ferrucci, L., Ford, I., Forouhi, N. G., Fox, C. S., Franzosi, M. G., Franks, P. W., Frayling, T. M., Froguel, P., Galan, P., de Geus, E., Gigante, B., Glazer, N. L., Goel, A., Groop, L., Gudnason, V., Hallmans, G., Hamsten, A., Hansson, O., Harris, T. B., Hayward, C., Heath, S., Hercberg, S., Hicks, A. A., Hingorani, A., Hofman, A., Hui, J., Hung, J., Jarvelin, M.-R., Jhun, M. A., Johnson, P. C. D., Jukema, J. W., Jula, A., Kao, W. H., Kaprio, J., Kardia, S. L. R., Keinanen-Kiukaanniemi, S., Kivimaki, M., Kolcic, I., Kovacs, P., Kumari, M., Kuusisto, J., Kyvik, K. O., Laakso, M., Lakka, T., Lannfelt, L., Lathrop, G. M., Launer, L. J., Leander, K., Li, G., Lind, L., Lindstrom, J., Lobbens, S., Loos, R. J. F., Luan, J., Lyssenko, V., Mägi, R., Magnusson, P. K. E., Marmot, M., Meneton, P., Mohlke, K. L., Mooser, V., Morken, M. A., Miljkovic, I., Narisu, N., O'Connell, J., Ong, K. K., Oostra, B. A., Palmer, L. J., Palotie, A., Pankow, J. S., Peden, J. F., Pedersen, N. L., Pehlic, M., Peltonen, L., Penninx, B., Pericic, M., Perola, M., Perusse, L., Peyser, P. A., Polasek, O., Pramstaller, P. P., Province, M. A., Räikkönen, K., Rauramaa, R., Rehnberg, E., Rice, K., Rotter, J. I., Rudan, I., Ruokonen, A., Saaristo, T., Sabater-Lleal, M., Salomaa, V., Savage, D. B., Saxena, R., Schwarz, P., Seedorf, U., Sennblad, B., Serrano-Rios, M., Shuldiner, A. R., Sijbrands, E. J. G., Siscovick, D. S., Smit, J. H., Small, K. S., Smith, N. L., Smith, A. V., Stančáková, A., Stirrups, K., Stumvoll, M., Sun, Y. V., Swift, A. J., Tönjes, A., Tuomilehto, J., Trompet, S., Uitterlinden, A. G., Uusitupa, M., Vikström, M., Vitart, V., Vohl, M.-C., Voight, B. F., Vollenweider, P., Waeber, G., Waterworth, D. M., Watkins, H., Wheeler, E., Widen, E., Wild, S. H., Willems, S. M., Willemsen, G., Wilson, J. F., Witteman, J. C. M., Wright, A. F., Yaghootkar, H., Zelenika, D., Zemunik, T., Zgaga, L., Consortium, Dia. G. R. A. M. (DIAGRAM), Consortium, T. M. T. H. E. R. (MUTHER), Wareham, N. J., McCarthy, M. I., Barroso, I., Watanabe, R. M., Florez, J. C., Dupuis, J., Meigs, J. B. & Langenberg, C. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* 44, 659–669 (2012).

13. Scott, R. A., Lagou, V., Welch, R. P., Wheeler, E., Montasser, M. E., Luan, J., Mägi, R., Strawbridge, R. J., Rehnberg, E., Gustafsson, S., Kanoni, S., Rasmussen-Torvik, L. J., Yengo, L., Lecoeur, C., Shungin, D., Sanna, S., Sidore, C., Johnson, P. C. D., Jukema, J. W., Johnson, T., Mahajan, A., Verweij, N., Thorleifsson, G., Hottenga, J.-J., Shah, S., Smith, A. V., Sennblad, B., Gieger, C., Salo, P., Perola, M., Timpson, N. J., Evans, D. M., Pourcain, B. S., Wu, Y., Andrews, J. S., Hui, J., Bielak, L. F., Zhao, W., Horikoshi, M., Navarro, P., Isaacs, A., O'Connell, J. R., Stirrups, K., Vitart, V., Hayward, C., Esko, T., Mihailov, E., Fraser, R. M., Fall, T., Voight, B. F., Raychaudhuri, S., Chen, H., Lindgren, C. M., Morris, A. P., Rayner, N. W., Robertson, N., Rybin, D., Liu, C.-T., Beckmann, J. S., Willems, S. M., Chines, P. S., Jackson, A. U., Kang, H. M., Stringham, H. M., Song, K., Tanaka, T., Peden, J. F., Goel, A., Hicks, A. A., An, P., Müller-Nurasyid, M., Franco-Cereceda, A., Folkersen, L., Marullo, L., Jansen, H., Oldehinkel, A. J., Bruinenberg, M., Pankow, J. S., North, K. E., Forouhi, N. G., Loos, R. J. F., Edkins, S., Varga, T. V., Hallmans, G., Oksa, H., Antonella, M., Nagaraja, R., Trompet, S., Ford, I., Bakker, S. J. L., Kong, A., Kumari, M., Gigante, B., Herder, C., Munroe, P. B., Caulfield, M., Antti, J., Mangino, M., Small, K., Miljkovic, I., Liu, Y., Atalay, M., Kiess, W., James, A. L., Rivadeneira, F., Uitterlinden, A. G., Palmer, C. N. A., Doney, A. S. F., Willemsen, G., Smit, J. H., Campbell, S., Polasek, O., Bonnycastle, L. L., Hercberg, S., Dimitriou, M., Bolton, J. L., Fowkes, G. R., Kovacs, P., Lindström, J., Zemunik, T., Bandinelli, S., Wild, S. H., Basart, H. V., Rathmann, W., Grallert, H., DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, Maerz, W., Kleber, M. E., Boehm, B. O., Peters, A., Pramstaller, P. P., Province, M. A., Borecki, I. B., Hastie, N. D., Rudan, I., Campbell, H., Watkins, H., Farrall, M., Stumvoll, M., Ferrucci, L., Waterworth, D. M., Bergman, R. N., Collins, F. S., Tuomilehto, J., Watanabe, R. M., de Geus, E. J. C., Penninx, B. W., Hofman, A., Oostra, B. A., Psaty, B. M., Vollenweider, P., Wilson, J. F., Wright, A. F., Hovingh, G. K., Metspalu, A., Uusitupa, M., Magnusson, P. K. E., Kyvik, K. O., Kaprio, J., Price, J. F., Dedoussis, G. V., Deloukas, P., Meneton, P., Lind, L., Boehnke, M., Shuldiner, A. R., van Duijn, C. M., Morris, A. D., Toenjes, A., Peyser, P. A., Beilby, J. P., Körner, A., Kuusisto, J., Laakso, M., Bornstein, S. R., Schwarz, P. E. H., Lakka, T. A., Rauramaa, R., Adair, L. S., Smith, G. D., Spector, T. D., Illig, T., de Faire, U., Hamsten, A., Gudnason, V., Kivimaki, M., Hingorani, A., Keinanen-Kiukaanniemi, S. M., Saaristo, T. E., Boomsma, D. I., Stefansson, K., van der Harst, P., Dupuis, J., Pedersen, N. L., Sattar, N., Harris, T. B., Cucca, F., Ripatti, S., Salomaa, V., Mohlke, K. L., Balkau, B., Froguel, P., Pouta, A., Jarvelin, M.-R., Wareham, N. J., Bouatia-Naji, N., McCarthy, M. I., Franks, P. W., Meigs, J. B., Teslovich, T. M., Florez, J. C., Langenberg, C., Ingelsson, E., Prokopenko, I. & Barroso, I. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* 44, 991–1005 (2012).

14. Thurner, M., van de Bunt, M., Torres, J. M., Mahajan, A., Nylander, V., Bennett, A. J., Gaulton, K. J., Barrett, A., Burrows, C., Bell, C. G., Lowe, R., Beck, S., Rakyan, V. K., Gloyn, A. L. & McCarthy, M. I. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *eLife* 7, (2018).

15. Fuchsberger, C., Flannick, J., Teslovich, T. M., Mahajan, A., Agarwala, V., Gaulton, K. J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D. J., Rivas, M. A., Perry, J. R. B., Sim, X., Blackwell, T. W., Robertson, N. R., Rayner, N. W., Cingolani, P., Locke, A. E., Fernandez Tajes, J., Highland, H. M., Dupuis, J., Chines, P. S., Lindgren, C. M., Hartl, C., Jackson, A. U., Chen, H., Huyghe, J. R., van de Bunt, M., Pearson, R. D., Kumar, A., Müller-Nurasyid, M., Grarup, N., Stringham, H. M., Gamazon, E. R., Lee, J., Chen, Y., Scott, R. A., Below, J. E., Chen, P., Huang, J., Go, M. J., Stitzel, M. L., Pasko, D., Parker, S. C. J., Varga, T. V., Green, T., Beer, N. L., Day-Williams, A. G., Ferreira, T., Fingerlin, T., Horikoshi, M., Hu, C., Huh, I., Ikram, M. K., Kim, B.-J., Kim, Y., Kim, Y. J., Kwon, M.-S., Lee, J., Lee, S., Lin, K.-H., Maxwell, T. J., Nagai, Y., Wang, X., Welch, R. P., Yoon, J., Zhang, W., Barzilai, N., Voight, B. F., Han,

B.-G., Jenkinson, C. P., Kuulasmaa, T., Kuusisto, J., Manning, A., Ng, M. C. Y., Palmer, N. D., Balkau, B., Stancáková, A., Abboud, H. E., Boeing, H., Giedraitis, V., Prabhakaran, D., Gottesman, O., Scott, J., Carey, J., Kwan, P., Grant, G., Smith, J. D., Neale, B. M., Purcell, S., Butterworth, A. S., Howson, J. M. M., Lee, H. M., Lu, Y., Kwak, S.-H., Zhao, W., Danesh, J., Lam, V. K. L., Park, K. S., Saleheen, D., So, W. Y., Tam, C. H. T., Afzal, U., Aguilar, D., Arya, R., Aung, T., Chan, E., Navarro, C., Cheng, C.-Y., Palli, D., Correa, A., Curran, J. E., Rybin, D., Farook, V. S., Fowler, S. P., Freedman, B. I., Griswold, M., Hale, D. E., Hicks, P. J., Khor, C.-C., Kumar, S., Lehne, B., Thuillier, D., Lim, W. Y., Liu, J., van der Schouw, Y. T., Loh, M., Musani, S. K., Puppala, S., Scott, W. R., Yengo, L., Tan, S.-T., Taylor, H. A., Thameem, F., Wilson, G., Wong, T. Y., Njølstad, P. R., Levy, J. C., Mangino, M., Bonnycastle, L. L., Schwarzmayr, T., Fadista, J., Surdulescu, G. L., Herder, C., Groves, C. J., Wieland, T., Bork-Jensen, J., Brandslund, I., Christensen, C., Koistinen, H. A., Doney, A. S. F., Kinnunen, L., Esko, T., Farmer, A. J., Hakaste, L., Hodgkiss, D., Kravic, J., Lyssenko, V., Hollensted, M., Jørgensen, M. E., Jørgensen, T., Ladenvall, C., Justesen, J. M., Käräjämäki, A., Kriebel, J., Rathmann, W., Lannfelt, L., Lauritzen, T., Narisu, N., Linneberg, A., Melander, O., Milani, L., Neville, M., Orho-Melander, M., Qi, L., Qi, Q., Roden, M., Rolandsson, O., Swift, A., Rosengren, A. H., Stirrups, K., Wood, A. R., Mihailov, E., Blancher, C., Carneiro, M. O., Maguire, J., Poplin, R., Shakir, K., Fennell, T., DePristo, M., Hrabé de Angelis, M., Deloukas, P., Gjesing, A. P., Jun, G., Nilsson, P., Murphy, J., Onofrio, R., Thorand, B., Hansen, T., Meisinger, C., Hu, F. B., Isomaa, B., Karpe, F., Liang, L., Peters, A., Huth, C., O'Rahilly, S. P., Palmer, C. N. A., Pedersen, O., Rauramaa, R., Tuomilehto, J., Salomaa, V., Watanabe, R. M., Syvänen, A.-C., Bergman, R. N., Bharadwaj, D., Bottinger, E. P., Cho, Y. S., Chandak, G. R., Chan, J. C. N., Chia, K. S., Daly, M. J., Ebrahim, S. B., Langenberg, C., Elliott, P., Jablonski, K. A., Lehman, D. M., Jia, W., Ma, R. C. W., Pollin, T. I., Sandhu, M., Tandon, N., Froguel, P., Barroso, I., Teo, Y. Y., Zeggini, E., Loos, R. J. F., Small, K. S., Ried, J. S., DeFronzo, R. A., Grallert, H., Glaser, B., Metspalu, A., Wareham, N. J., Walker, M., Banks, E., Gieger, C., Ingelsson, E., Im, H. K., Illig, T., Franks, P. W., Buck, G., Trakalo, J., Buck, D., Prokopenko, I., Mägi, R., Lind, L., Farjoun, Y., Owen, K. R., Gloyn, A. L., Strauch, K., Tuomi, T., Kooner, J. S., Lee, J.-Y., Park, T., Donnelly, P., Morris, A. D., Hattersley, A. T., Bowden, D. W., Collins, F. S., Atzmon, G., Chambers, J. C., Spector, T. D., Laakso, M., Strom, T. M., Bell, G. I., Blangero, J., Duggirala, R., Tai, E. S., McVean, G., Hanis, C. L., Wilson, J. G., Seielstad, M., Frayling, T. M., Meigs, J. B., Cox, N. J., Sladek, R., Lander, E. S., Gabriel, S., Burtt, N. P., Mohlke, K. L., Meitinger, T., Groop, L., Abecasis, G., Florez, J. C., Scott, L. J., Morris, A. P., Kang, H. M., Boehnke, M., Altshuler, D. & McCarthy, M. I. The genetic architecture of type 2 diabetes. *Nature* 536, 41–47 (2016).

16. Gaulton, K. J., Nammo, T., Pasquali, L., Simon, J. M., Giresi, P. G., Fogarty, M. P., Panhuis, T. M., Mieczkowski, P., Secchi, A., Bosco, D., Berney, T., Montanya, E., Mohlke, K. L., Lieb, J. D. & Ferrer, J. A map of open chromatin in human pancreatic islets. *Nat. Genet.* 42, 255–259 (2010).

17. Gaulton, K. J., Ferreira, T., Lee, Y., Raimondo, A., Mägi, R., Reschen, M. E., Mahajan, A., Locke, A., Rayner, N. W., Robertson, N., Scott, R. A., Prokopenko, I., Scott, L. J., Green, T., Sparso, T., Thuillier, D., Yengo, L., Grallert, H., Wahl, S., Frånberg, M., Strawbridge, R. J., Kestler, H., Chheda, H., Eisele, L., Gustafsson, S., Steinthorsdottir, V., Thorleifsson, G., Qi, L., Karssen, L. C., van Leeuwen, E. M., Willems, S. M., Li, M., Chen, H., Fuchsberger, C., Kwan, P., Ma, C., Linderman, M., Lu, Y., Thomsen, S. K., Rundle, J. K., Beer, N. L., van de Bunt, M., Chalisey, A., Kang, H. M., Voight, B. F., Abecasis, G. R., Almgren, P., Baldassarre, D., Balkau, B., Benediktsson, R., Blüher, M., Boeing, H., Bonnycastle, L. L., Bottinger, E. P., Burtt, N. P., Carey, J., Charpentier, G., Chines, P. S., Cornelis, M. C., Couper, D. J., Crenshaw, A. T., van Dam, R. M., Doney, A. S. F., Dorkhan, M., Edkins, S., Eriksson, J. G., Esko, T., Eury, E.,

Fadista, J., Flannick, J., Fontanillas, P., Fox, C., Franks, P. W., Gertow, K., Gieger, C., Gigante, B., Gottesman, O., Grant, G. B., Grarup, N., Groves, C. J., Hassinen, M., Have, C. T., Herder, C., Holmen, O. L., Hreidarsson, A. B., Humphries, S. E., Hunter, D. J., Jackson, A. U., Jonsson, A., Jørgensen, M. E., Jørgensen, T., Kao, W.-H. L., Kerrison, N. D., Kinnunen, L., Klopp, N., Kong, A., Kovacs, P., Kraft, P., Kravic, J., Langford, C., Leander, K., Liang, L., Lichtner, P., Lindgren, C. M., Lindholm, E., Linneberg, A., Liu, C.-T., Lobbens, S., Luan, J., Lyssenko, V., Männistö, S., McLeod, O., Meyer, J., Mihailov, E., Mirza, G., Mühleisen, T. W., Müller-Nurasyid, M., Navarro, C., Nöthen, M. M., Oskolkov, N. N., Owen, K. R., Palli, D., Pechlivanis, S., Peltonen, L., Perry, J. R. B., Platou, C. G. P., Roden, M., Ruderfer, D., Rybin, D., van der Schouw, Y. T., Sennblad, B., Sigurðsson, G., Stančáková, A., Steinbach, G., Storm, P., Strauch, K., Stringham, H. M., Sun, Q., Thorand, B., Tikkanen, E., Tonjes, A., Trakalo, J., Tremoli, E., Tuomi, T., Wennauer, R., Wiltshire, S., Wood, A. R., Zeggini, E., Dunham, I., Birney, E., Pasquali, L., Ferrer, J., Loos, R. J. F., Dupuis, J., Florez, J. C., Boerwinkle, E., Pankow, J. S., van Duijn, C., Sijbrands, E., Meigs, J. B., Hu, F. B., Thorsteinsdottir, U., Stefansson, K., Lakka, T. A., Rauramaa, R., Stumvoll, M., Pedersen, N. L., Lind, L., Keinanen-Kiukaanniemi, S. M., Korpi-Hyövälti, E., Saaristo, T. E., Saltevo, J., Kuusisto, J., Laakso, M., Metspalu, A., Erbel, R., Jöcke, K.-H., Moebus, S., Ripatti, S., Salomaa, V., Ingelsson, E., Boehm, B. O., Bergman, R. N., Collins, F. S., Mohlke, K. L., Koistinen, H., Tuomilehto, J., Hveem, K., Njølstad, I., Deloukas, P., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., de Faire, U., Hamsten, A., Illig, T., Peters, A., Cauchi, S., Sladek, R., Froguel, P., Hansen, T., Pedersen, O., Morris, A. D., Palmer, C. N. A., Kathiresan, S., Melander, O., Nilsson, P. M., Groop, L. C., Barroso, I., Langenberg, C., Wareham, N. J., O'Callaghan, C. A., Gloyn, A. L., Altshuler, D., Boehnke, M., Teslovich, T. M., McCarthy, M. I., Morris, A. P., & DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* 47, 1415–1425 (2015).

18. Pasquali, L., Gaulton, K. J., Rodríguez-Seguí, S. A., Mularoni, L., Miguel-Escalada, I., Akerman, I., Tena, J. J., Morán, I., Gómez-Marín, C., van de Bunt, M., Ponsa-Cobas, J., Castro, N., Nammo, T., Cebola, I., García-Hurtado, J., Maestro, M. A., Pattou, F., Piemonti, L., Berney, T., Gloyn, A. L., Ravassard, P., Gómez-Skarmeta, J. L., Müller, F., McCarthy, M. I. & Ferrer, J. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* 46, 136–143 (2014).

19. van der Meulen, T., Donaldson, C. J., Cáceres, E., Hunter, A. E., Cowing–Zitron, C., Pound, L. D., Adams, M. W., Zembrzycki, A., Grove, K. L. & Huising, M. O. Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. *Nat. Med.* 21, 769–776 (2015).

20. Caicedo, A. Paracrine and autocrine interactions in the human islet: more than meets the eye. *Semin. Cell Dev. Biol.* 24, 11–21 (2013).

21. DiGruccio, M. R., Mawla, A. M., Donaldson, C. J., Noguchi, G. M., Vaughan, J., Cowing-Zitron, C., van der Meulen, T. & Huising, M. O. 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 (2016).

22. Dorrell, C., Schug, J., Canaday, P. S., Russ, H. A., Tarlow, B. D., Grompe, M. T., Horton, T., Hebrok, M., Streeter, P. R., Kaestner, K. H. & Grompe, M. Human islets contain four distinct subtypes of β cells. *Nat. Commun.* 7, ncomms11756 (2016).

23. Xin, Y., Gutierrez, G. D., Okamoto, H., Kim, J., Lee, A.-H., Adler, C., Ni, M., Yancopoulos, G. D., Murphy, A. J. & Gromada, J. Pseudotime Ordering of Single Human β-Cells Reveals States of Insulin Production and Unfolded Protein Response. *Diabetes* db180365 (2018) doi:10.2337/db18-0365.

24. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. & Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16, 1289–1296 (2019).

25. Bader, E., Migliorini, A., Gegg, M., Moruzzi, N., Gerdes, J., Roscioni, S. S., Bakhti, M., Brandl, E., Irmler, M., Beckers, J., Aichler, M., Feuchtinger, A., Leitzinger, C., Zischka, H., Wang-Sattler, R., Jastroch, M., Tschöp, M., Machicao, F., Staiger, H., Häring, H.-U., Chmelova, H., Chouinard, J. A., Oskolkov, N., Korsgren, O., Speier, S. & Lickert, H. Identification of proliferative and mature β-cells in the islets of Langerhans. *Nature* 535, 430–434 (2016).

26. Liu, J. S. E. & Hebrok, M. All mixed up: defining roles for β-cell subtypes in mature islets. *Genes Dev.* 31, 228–240 (2017).

27. Greenwald, W. W., Chiou, J., Yan, J., Qiu, Y., Dai, N., Wang, A., Nariai, N., Aylward, A., Han, J. Y., Kadakia, N., Regue, L., Okino, M.-L., Drees, F., Kramer, D., Vinckier, N., Minichiello, L., Gorkin, D., Avruch, J., Frazer, K. A., Sander, M., Ren, B. & Gaulton, K. J. Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nat. Commun.* 10, 1–12 (2019).

28. Khetan, S., Kursawe, R., Youn, A., Lawlor, N., Jillette, A., Marquez, E. J., Ucar, D. & Stitzel, M. L. Type 2 Diabetes-Associated Genetic Variants Regulate Chromatin Accessibility in Human Islets. *Diabetes* 67, 2466–2477 (2018).

29. Varshney, A., Scott, L. J., Welch, R. P., Erdos, M. R., Chines, P. S., Narisu, N., Albanus, R. D., Orchard, P., Wolford, B. N., Kursawe, R., Vadlamudi, S., Cannon, M. E., Didion, J. P., Hensley, J., Kirilusha, A., NISC Comparative Sequencing Program, Bonnycastle, L. L., Taylor, D. L., Watanabe, R., Mohlke, K. L., Boehnke, M., Collins, F. S., Parker, S. C. J. & Stitzel, M. L. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl. Acad. Sci. U. S. A.* 114, 2301–2306 (2017).

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

31. Galvagni, F., Nardi, F., Maida, M., Bernardini, G., Vannuccini, S., Petraglia, F., Santucci, A. & Orlandini, M. CD93 and dystroglycan cooperation in human endothelial cell adhesion and migration adhesion and migration. *Oncotarget* 7, 10090–10103 (2016).

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

33. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* 14, 975–978 (2017).

34. Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S. R., Tan, G., Baranasic, D., Arenillas, D. J., Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W. W., Parcy, F. & Mathelier, A. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 46, D1284 (2018).

35. Wilson, M. E., Scheel, D. & German, M. S. Gene expression cascades in pancreatic development. *Mech. Dev.* 120, 65–80 (2003).

36. Conrad, E., Dai, C., Spaeth, J., Guo, M., Cyphert, H. A., Scoville, D., Carroll, J., Yu, W.-M., Goodrich, L. V., Harlan, D. M., Grove, K. L., Roberts, C. T., Powers, A. C., Gu, G. & Stein, R. The MAFB transcription factor impacts islet α-cell function in rodents and represents a unique signature of primate islet β-cells. *Am. J. Physiol. Endocrinol. Metab.* 310, E91–E102 (2016).

37. Katoh, M. C., Jung, Y., Ugboma, C. M., Shimbo, M., Kuno, A., Basha, W. A., Kudo, T., Oishi, H. & Takahashi, S. MafB Is Critical for Glucagon Production and Secretion in Mouse Pancreatic α Cells In Vivo. *Mol. Cell. Biol.* 38, (2018).

38. Nishimura, W., Takahashi, S. & Yasuda, K. MafA is critical for maintenance of the mature beta cell phenotype in mice. *Diabetologia* 58, 566–574 (2015).

39. Ozato, K., Tailor, P. & Kubota, T. The interferon regulatory factor family in host defense: mechanism of action. *J. Biol. Chem.* 282, 20065–20069 (2007).

40. De Val, S. & Black, B. L. Transcriptional Control of Endothelial Cell Development. *Dev. Cell* 16, 180–195 (2009).

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

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

43. Mawla, A. M. & Huising, M. O. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes. *Diabetes* 68, 1380–1393 (2019).

44. Camunas-Soler, J., Dai, X.-Q., Hang, Y., Bautista, A., Lyon, J., Suzuki, K., Kim, S. K., Quake, S. R. & MacDonald, P. E. Patch-Seq Links Single-Cell Transcriptomes to Human Islet Dysfunction in Diabetes. *Cell Metab.* 31, 1017-1031.e4 (2020).

45. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. & Rinn, J. L. Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat. Biotechnol.* 32, 381 (2014).

46. Parker, S. C. J., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., van Bueren, K. L., Chines, P. S., Narisu, N., NISC Comparative Sequencing Program, Black, B. L., Visel, A., Pennacchio, L. A., Collins, F. S., National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors, & NISC Comparative Sequencing Program Authors. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U. S. A.* 110, 17921–17926 (2013).

47. Aylward, A., Chiou, J., Okino, M.-L., Kadakia, N. & Gaulton, K. J. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Hum. Mol. Genet.* (2018) doi:10.1093/hmg/ddy314.

48. Strawbridge, R. J., Dupuis, J., Prokopenko, I., Barker, A., Ahlqvist, E., Rybin, D., Petrie, J. R., Travers, M. E., Bouatia-Naji, N., Dimas, A. S., Nica, A., Wheeler, E., Chen, H., Voight, B. F., Taneera, J., Kanoni, S., Peden, J. F., Turrini, F., Gustafsson, S., Zabena, C., Almgren, P., Barker, D. J. P., Barnes, D., Dennison, E. M., Eriksson, J. G., Eriksson, P., Eury, E., Folkersen, L., Fox, C. S., Frayling, T. M., Goel, A., Gu, H. F., Horikoshi, M., Isomaa, B., Jackson, A. U., Jameson, K. A., Kajantie, E., Kerr-Conte, J., Kuulasmaa, T., Kuusisto, J., Loos, R. J. F., Luan, J., Makrilakis, K., Manning, A. K., Martínez-Larrad, M. T., Narisu, N., Nastase Mannila, M., Ohrvik, J., Osmond, C., Pascoe, L., Payne, F., Sayer, A. A., Sennblad, B., Silveira, A., Stancáková, A., Stirrups, K., Swift, A. J., Syvänen, A.-C., Tuomi, T., van 't Hooft, F. M., Walker, M., Weedon, M. N., Xie, W., Zethelius, B., DIAGRAM Consortium, GIANT Consortium, MuTHER Consortium, CARDIoGRAM Consortium, C4D Consortium, Ongen, H., Mälarstig, A., Hopewell, J. C., Saleheen, D., Chambers, J., Parish, S., Danesh, J., Kooner, J., Ostenson, C.-G., Lind, L., Cooper, C. C., Serrano-Ríos, M., Ferrannini, E., Forsen, T. J., Clarke, R., Franzosi, M. G., Seedorf, U., Watkins, H., Froguel, P., Johnson, P., Deloukas, P., Collins, F. S., Laakso, M., Dermitzakis, E. T., Boehnke, M., McCarthy, M. I., Wareham, N. J., Groop, L., Pattou, F., Gloyn, A. L., Dedoussis, G. V., Lyssenko, V., Meigs, J. B., Barroso, I., Watanabe, R. M., Ingelsson, E., Langenberg, C., Hamsten, A. & Florez, J. C. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 60, 2624–2634 (2011).

49. Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., Croteau-Chonka, D. C., Esko, T., Fall, T., Ferreira, T., Gustafsson, S., Kutalik, Z., Luan, J., Mägi, R., Randall, J. C., Winkler, T. W., Wood, A. R., Workalemahu, T., Faul, J. D., Smith, J. A., Zhao, J. H., Zhao, W., Chen, J., Fehrmann, R., Hedman, Å. K., Karjalainen, J., Schmidt, E. M., Absher, D., Amin, N., Anderson, D., Beekman, M., Bolton, J. L., Bragg-Gresham, J. L., Buyske, S., Demirkan, A., Deng, G., Ehret, G. B., Feenstra, B., Feitosa, M. F., Fischer, K., Goel, A., Gong, J., Jackson, A. U., Kanoni, S., Kleber, M. E., Kristiansson, K., Lim, U., Lotay, V., Mangino, M., Leach, I. M., Medina-Gomez, C., Medland, S. E., Nalls, M. A., Palmer, C. D., Pasko, D., Pechlivanis, S., Peters, M. J., Prokopenko, I., Shungin, D., Stančáková, A., Strawbridge, R. J., Sung, Y. J., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S. W., van Setten, J., Van Vliet-Ostaptchouk, J. V., Wang, Z., Yengo, L., Zhang, W., Isaacs, A., Albrecht, E., Ärnlöv, J., Arscott, G. M., Attwood, A. P., Bandinelli, S., Barrett, A., Bas, I. N., Bellis, C., Bennett, A. J., Berne, C., Blagieva, R., Blüher, M., Böhringer, S., Bonnycastle, L. L., Böttcher, Y., Boyd, H. A., Bruinenberg, M., Caspersen, I. H., Chen, Y.-D. I., Clarke, R., Daw, E. W., de Craen, A. J. M., Delgado, G.,

Dimitriou, M., Doney, A. S. F., Eklund, N., Estrada, K., Eury, E., Folkersen, L., Fraser, R. M., Garcia, M. E., Geller, F., Giedraitis, V., Gigante, B., Go, A. S., Golay, A., Goodall, A. H., Gordon, S. D., Gorski, M., Grabe, H.-J., Grallert, H., Grammer, T. B., Gräßler, J., Grönberg, H., Groves, C. J., Gusto, G., Haessler, J., Hall, P., Haller, T., Hallmans, G., Hartman, C. A., Hassinen, M., Hayward, C., Heard-Costa, N. L., Helmer, Q., Hengstenberg, C., Holmen, O., Hottenga, J.-J., James, A. L., Jeff, J. M., Johansson, Å., Jolley, J., Juliusdottir, T., Kinnunen, L., Koenig, W., Koskenvuo, M., Kratzer, W., Laitinen, J., Lamina, C., Leander, K., Lee, N. R., Lichtner, P., Lind, L., Lindström, J., Lo, K. S., Lobbens, S., Lorbeer, R., Lu, Y., Mach, F., Magnusson, P. K. E., Mahajan, A., McArdle, W. L., McLachlan, S., Menni, C., Merger, S., Mihailov, E., Milani, L., Moayyeri, A., Monda, K. L., Morken, M. A., Mulas, A., Müller, G., Müller-Nurasyid, M., Musk, A. W., Nagaraja, R., Nöthen, M. M., Nolte, I. M., Pilz, S., Rayner, N. W., Renstrom, F., Rettig, R., Ried, J. S., Ripke, S., Robertson, N. R., Rose, L. M., Sanna, S., Scharnagl, H., Scholtens, S., Schumacher, F. R., Scott, W. R., Seufferlein, T., Shi, J., Smith, A. V., Smolonska, J., Stanton, A. V., Steinthorsdottir, V., Stirrups, K., Stringham, H. M., Sundström, J., Swertz, M. A., Swift, A. J., Syvänen, A.-C., Tan, S.-T., Tayo, B. O., Thorand, B., Thorleifsson, G., Tyrer, J. P., Uh, H.-W., Vandenput, L., Verhulst, F. C., Vermeulen, S. H., Verweij, N., Vonk, J. M., Waite, L. L., Warren, H. R., Waterworth, D., Weedon, M. N., Wilkens, L. R., Willenborg, C., Wilsgaard, T., Wojczynski, M. K., Wong, A., Wright, A. F., Zhang, Q., LifeLines Cohort Study, Brennan, E. P., Choi, M., Dastani, Z., Drong, A. W., Eriksson, P., Franco-Cereceda, A., Gådin, J. R., Gharavi, A. G., Goddard, M. E., Handsaker, R. E., Huang, J., Karpe, F., Kathiresan, S., Keildson, S., Kiryluk, K., Kubo, M., Lee, J.-Y., Liang, L., Lifton, R. P., Ma, B., McCarroll, S. A., McKnight, A. J., Min, J. L., Moffatt, M. F., Montgomery, G. W., Murabito, J. M., Nicholson, G., Nyholt, D. R., Okada, Y., Perry, J. R. B., Dorajoo, R., Reinmaa, E., Salem, R. M., Sandholm, N., Scott, R. A., Stolk, L., Takahashi, A., Tanaka, T., van 't Hooft, F. M., Vinkhuyzen, A. A. E., Westra, H.-J., Zheng, W., Zondervan, K. T., ADIPOGen Consortium, AGEN-BMI Working Group, CARDIOGRAMplusC4D Consortium, CKDGen Consortium, GLGC, ICBP, MAGIC Investigators, MuTHER Consortium, MIGen Consortium, PAGE Consortium, ReproGen Consortium, GENIE Consortium, International Endogene Consortium, Heath, A. C., Arveiler, D., Bakker, S. J. L., Beilby, J., Bergman, R. N., Blangero, J., Bovet, P., Campbell, H., Caulfield, M. J., Cesana, G., Chakravarti, A., Chasman, D. I., Chines, P. S., Collins, F. S., Crawford, D. C., Cupples, L. A., Cusi, D., Danesh, J., de Faire, U., den Ruijter, H. M., Dominiczak, A. F., Erbel, R., Erdmann, J., Eriksson, J. G., Farrall, M., Felix, S. B., Ferrannini, E., Ferrières, J., Ford, I., Forouhi, N. G., Forrester, T., Franco, O. H., Gansevoort, R. T., Gejman, P. V., Gieger, C., Gottesman, O., Gudnason, V., Gyllensten, U., Hall, A. S., Harris, T. B., Hattersley, A. T., Hicks, A. A., Hindorff, L. A., Hingorani, A. D., Hofman, A., Homuth, G., Hovingh, G. K., Humphries, S. E., Hunt, S. C., Hyppönen, E., Illig, T., Jacobs, K. B., Jarvelin, M.-R., Jöckel, K.-H., Johansen, B., Jousilahti, P., Jukema, J. W., Jula, A. M., Kaprio, J., Kastelein, J. J. P., Keinanen-Kiukaanniemi, S. M., Kiemeney, L. A., Knekt, P., Kooner, J. S., Kooperberg, C., Kovacs, P., Kraja, A. T., Kumari, M., Kuusisto, J., Lakka, T. A., Langenberg, C., Marchand, L. L., Lehtimäki, T., Lyssenko, V., Männistö, S., Marette, A., Matise, T. C., McKenzie, C. A., McKnight, B., Moll, F. L., Morris, A. D., Morris, A. P., Murray, J. C., Nelis, M., Ohlsson, C., Oldehinkel, A. J., Ong, K. K., Madden, P. A. F., Pasterkamp, G., Peden, J. F., Peters, A., Postma, D. S., Pramstaller, P. P., Price, J. F., Qi, L., Raitakari, O. T., Rankinen, T., Rao, D. C., Rice, T. K., Ridker, P. M., Rioux, J. D., Ritchie, M. D., Rudan, I., Salomaa, V., Samani, N. J., Saramies, J., Sarzynski, M. A., Schunkert, H., Schwarz, P. E. H., Sever, P., Shuldiner, A. R., Sinisalo, J., Stolk, R. P., Strauch, K., Tönjes, A., Trégouët, D.-A., Tremblay, A., Tremoli, E., Virtamo, J., Vohl, M.-C., Völker, U., Waeber, G., Willemsen, G., Witteman, J. C., Zillikens, M. C., Adair, L. S., Amouyel, P., Asselbergs, F. W., Assimes, T. L., Bochud, M., Boehm, B. O., Boerwinkle, E., Bornstein, S. R., Bottinger, E. P., Bouchard, C., Cauchi, S., Chambers, J. C., Chanock, S. J., Cooper, R. S., de Bakker, P. I. W., Dedoussis, G., Ferrucci, L., Franks, P. W., Froguel, P., Groop, L. C., Haiman, C. A., Hamsten, A., Hui, J.,

Hunter, D. J., Hveem, K., Kaplan, R. C., Kivimaki, M., Kuh, D., Laakso, M., Liu, Y., Martin, N. G., März, W., Melbye, M., Metspalu, A., Moebus, S., Munroe, P. B., Njølstad, I., Oostra, B. A., Palmer, C. N. A., Pedersen, N. L., Perola, M., Pérusse, L., Peters, U., Power, C., Quertermous, T., Rauramaa, R., Rivadeneira, F., Saaristo, T. E., Saleheen, D., Sattar, N., Schadt, E. E., Schlessinger, D., Slagboom, P. E., Snieder, H., Spector, T. D., Thorsteinsdottir, U., Stumvoll, M., Tuomilehto, J., Uitterlinden, A. G., Uusitupa, M., van der Harst, P., Walker, M., Wallaschofski, H., Wareham, N. J., Watkins, H., Weir, D. R., Wichmann, H.-E., Wilson, J. F., Zanen, P., Borecki, I. B., Deloukas, P., Fox, C. S., Heid, I. M., O'Connell, J. R., Strachan, D. P., Stefansson, K., van Duijn, C. M., Abecasis, G. R., Franke, L., Frayling, T. M., McCarthy, M. I., Visscher, P. M., Scherag, A., Willer, C. J., Boehnke, M., Mohlke, K. L., Lindgren, C. M., Beckmann, J. S., Barroso, I., North, K. E., Ingelsson, E., Hirschhorn, J. N., Loos, R. J. F. & Speliotes, E. K. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197–206 (2015).

50. Saxena, R., Hivert, M.-F., Langenberg, C., Tanaka, T., Pankow, J. S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A. U., Kao, W. H. L., Li, M., Glazer, N. L., Manning, A. K., Luan, J., Stringham, H. M., Prokopenko, I., Johnson, T., Grarup, N., Boesgaard, T. W., Lecoeur, C., Shrader, P., O'Connell, J., Ingelsson, E., Couper, D. J., Rice, K., Song, K., Andreasen, C. H., Dina, C., Köttgen, A., Le Bacquer, O., Pattou, F., Taneera, J., Steinthorsdottir, V., Rybin, D., Ardlie, K., Sampson, M., Qi, L., van Hoek, M., Weedon, M. N., Aulchenko, Y. S., Voight, B. F., Grallert, H., Balkau, B., Bergman, R. N., Bielinski, S. J., Bonnefond, A., Bonnycastle, L. L., Borch-Johnsen, K., Böttcher, Y., Brunner, E., Buchanan, T. A., Bumpstead, S. J., Cavalcanti-Proença, C., Charpentier, G., Chen, Y.-D. I., Chines, P. S., Collins, F. S., Cornelis, M., Crawford, G. J., Delplanque, J., Doney, A., Egan, J. M., Erdos, M. R., Firmann, M., Forouhi, N. G., Fox, C. S., Goodarzi, M. O., Graessler, J., Hingorani, A., Isomaa, B., Jørgensen, T., Kivimaki, M., Kovacs, P., Krohn, K., Kumari, M., Lauritzen, T., Lévy-Marchal, C., Mayor, V., McAteer, J. B., Meyre, D., Mitchell, B. D., Mohlke, K. L., Morken, M. A., Narisu, N., Palmer, C. N. A., Pakyz, R., Pascoe, L., Payne, F., Pearson, D., Rathmann, W., Sandbaek, A., Sayer, A. A., Scott, L. J., Sharp, S. J., Sijbrands, E., Singleton, A., Siscovick, D. S., Smith, N. L., Sparsø, T., Swift, A. J., Syddall, H., Thorleifsson, G., Tönjes, A., Tuomi, T., Tuomilehto, J., Valle, T. T., Waeber, G., Walley, A., Waterworth, D. M., Zeggini, E., Zhao, J. H., Consortium, G. & Investigators, the M. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* 42, 142–148 (2010).

51. Wheeler, E., Leong, A., Liu, C.-T., Hivert, M.-F., Strawbridge, R. J., Podmore, C., Li, M., Yao, J., Sim, X., Hong, J., Chu, A. Y., Zhang, W., Wang, X., Chen, P., Maruthur, N. M., Porneala, B. C., Sharp, S. J., Jia, Y., Kabagambe, E. K., Chang, L.-C., Chen, W.-M., Elks, C. E., Evans, D. S., Fan, Q., Giulianini, F., Go, M. J., Hottenga, J.-J., Hu, Y., Jackson, A. U., Kanoni, S., Kim, Y. J., Kleber, M. E., Ladenvall, C., Lecoeur, C., Lim, S.-H., Lu, Y., Mahajan, A., Marzi, C., Nalls, M. A., Navarro, P., Nolte, I. M., Rose, L. M., Rybin, D. V., Sanna, S., Shi, Y., Stram, D. O., Takeuchi, F., Tan, S. P., van der Most, P. J., Van Vliet-Ostaptchouk, J. V., Wong, A., Yengo, L., Zhao, W., Goel, A., Martinez Larrad, M. T., Radke, D., Salo, P., Tanaka, T., van Iperen, E. P. A., Abecasis, G., Afaq, S., Alizadeh, B. Z., Bertoni, A. G., Bonnefond, A., Böttcher, Y., Bottinger, E. P., Campbell, H., Carlson, O. D., Chen, C.-H., Cho, Y. S., Garvey, W. T., Gieger, C., Goodarzi, M. O., Grallert, H., Hamsten, A., Hartman, C. A., Herder, C., Hsiung, C. A., Huang, J., Igase, M., Isono, M., Katsuya, T., Khor, C.-C., Kiess, W., Kohara, K., Kovacs, P., Lee, J., Lee, W.-J., Lehne, B., Li, H., Liu, J., Lobbens, S., Luan, J., Lyssenko, V., Meitinger, T., Miki, T., Miljkovic, I., Moon, S., Mulas, A., Müller, G., Müller-Nurasyid, M., Nagaraja, R., Nauck, M., Pankow, J. S., Polasek, O., Prokopenko, I., Ramos, P. S., Rasmussen-Torvik, L., Rathmann, W., Rich, S. S., Robertson, N. R., Roden, M., Roussel, R., Rudan, I., Scott, R. A., Scott, W. R., Sennblad, B., Siscovick, D. S., Strauch, K., Sun, L., Swertz,

M., Tajuddin, S. M., Taylor, K. D., Teo, Y.-Y., Tham, Y. C., Tönjes, A., Wareham, N. J., Willemsen, G., Wilsgaard, T., Hingorani, A. D., EPIC-CVD Consortium, EPIC-InterAct Consortium, Lifelines Cohort Study, Egan, J., Ferrucci, L., Hovingh, G. K., Jula, A., Kivimaki, M., Kumari, M., Njølstad, I., Palmer, C. N. A., Serrano Ríos, M., Stumvoll, M., Watkins, H., Aung, T., Blüher, M., Boehnke, M., Boomsma, D. I., Bornstein, S. R., Chambers, J. C., Chasman, D. I., Chen, Y.-D. I., Chen, Y.-T., Cheng, C.-Y., Cucca, F., de Geus, E. J. C., Deloukas, P., Evans, M. K., Fornage, M., Friedlander, Y., Froguel, P., Groop, L., Gross, M. D., Harris, T. B., Hayward, C., Heng, C.-K., Ingelsson, E., Kato, N., Kim, B.-J., Koh, W.-P., Kooner, J. S., Körner, A., Kuh, D., Kuusisto, J., Laakso, M., Lin, X., Liu, Y., Loos, R. J. F., Magnusson, P. K. E., März, W., McCarthy, M. I., Oldehinkel, A. J., Ong, K. K., Pedersen, N. L., Pereira, M. A., Peters, A., Ridker, P. M., Sabanayagam, C., Sale, M., Saleheen, D., Saltevo, J., Schwarz, P. E., Sheu, W. H. H., Snieder, H., Spector, T. D., Tabara, Y., Tuomilehto, J., van Dam, R. M., Wilson, J. G., Wilson, J. F., Wolffenbuttel, B. H. R., Wong, T. Y., Wu, J.-Y., Yuan, J.-M., Zonderman, A. B., Soranzo, N., Guo, X., Roberts, D. J., Florez, J. C., Sladek, R., Dupuis, J., Morris, A. P., Tai, E.-S., Selvin, E., Rotter, J. I., Langenberg, C., Barroso, I. & Meigs, J. B. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med.* 14, e1002383 (2017).

52. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427 (2014).

53. de Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., Jostins, L., Rice, D. L., Gutierrez-Achury, J., Ji, S.-G., Heap, G., Nimmo, E. R., Edwards, C., Henderson, P., Mowat, C., Sanderson, J., Satsangi, J., Simmons, A., Wilson, D. C., Tremelling, M., Hart, A., Mathew, C. G., Newman, W. G., Parkes, M., Lees, C. W., Uhlig, H., Hawkey, C., Prescott, N. J., Ahmad, T., Mansfield, J. C., Anderson, C. A. & Barrett, J. C. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* 49, 256–261 (2017).

54. Bentham, J., Morris, D. L., Graham, D. S. C., Pinder, C. L., Tombleson, P., Behrens, T. W., Martín, J., Fairfax, B. P., Knight, J. C., Chen, L., Replogle, J., Syvänen, A.-C., Rönnblom, L., Graham, R. R., Wither, J. E., Rioux, J. D., Alarcón-Riquelme, M. E. & Vyse, T. J. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* 47, 1457–1464 (2015).

55. Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., DeStafano, A. L., Bis, J. C., Beecham, G. W., Grenier-Boley, B., Russo, G., Thorton-Wells, T. A., Jones, N., Smith, A. V., Chouraki, V., Thomas, C., Ikram, M. A., Zelenika, D., Vardarajan, B. N., Kamatani, Y., Lin, C. F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M. L., Ruiz, A., Bihoreau, M. T., Choi, S. H., Reitz, C., Pasquier, F., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O. L., De Jager, P. L., Deramecourt, V., Johnston, J. A., Evans, D., Lovestone, S., Letenneur, L., Morón, F. J., Rubinsztein, D. C., Eiriksdottir, G., Sleegers, K., Goate, A. M., Fiévet, N., Huentelman, M. W., Gill, M., Brown, K., Kamboh, M. I., Keller, L., Barberger-Gateau, P., McGuiness, B., Larson, E. B., Green, R., Myers, A. J., Dufouil, C., Todd, S., Wallon, D., Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J., Lleo, A., Bayer, A., Tsuang, D. W., Yu, L., Tsolaki, M., Bossù, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S., Sanchez-Garcia, F., Fox, N. C., Hardy, J., Deniz Naranjo, M. C., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., European Alzheimer's Disease Initiative (EADI), Genetic and Environmental Risk in Alzheimer's Disease, Alzheimer's Disease Genetic Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology, Moebus, S.,

Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., Nacmias, B., Gilbert, J. R., Mayhaus, M., Lannefelt, L., Hakonarson, H., Pichler, S., Carrasquillo, M. M., Ingelsson, M., Beekly, D., Alvarez, V., Zou, F., Valladares, O., Younkin, S. G., Coto, E., Hamilton-Nelson, K. L., Gu, W., Razquin, C., Pastor, P., Mateo, I., Owen, M. J., Faber, K. M., Jonsson, P. V., Combarros, O., O'Donovan, M. C., Cantwell, L. B., Soininen, H., Blacker, D., Mead, S., Mosley, T. H., Bennett, D. A., Harris, T. B., Fratiglioni, L., Holmes, C., de Bruijn, R. F., Passmore, P., Montine, T. J., Bettens, K., Rotter, J. I., Brice, A., Morgan, K., Foroud, T. M., Kukull, W. A., Hannequin, D., Powell, J. F., Nalls, M. A., Ritchie, K., Lunetta, K. L., Kauwe, J. S., Boerwinkle, E., Riemenschneider, M., Boada, M., Hiltuenen, M., Martin, E. R., Schmidt, R., Rujescu, D., Wang, L. S., Dartigues, J. F., Mayeux, R., Tzourio, C., Hofman, A., Nöthen, M. M., Graff, C., Psaty, B. M., Jones, L., Haines, J. L., Holmans, P. A., Lathrop, M., Pericak-Vance, M. A., Launer, L. J., Farrer, L. A., van Duijn, C. M., Van Broeckhoven, C., Moskvina, V., Seshadri, S., Williams, J., Schellenberg, G. D. & Amouyel, P. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* 45, 1452–1458 (2013).

56. Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J. H., Clarke, T.-K., Coleman, J. I. R., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G. E., Crowley, C. A., Dashti, H. S., Davies, G., Deary, I. J., Degenhardt, F., Derks, E. M., Direk, N., Dolan, C. V., Dunn, E. C., Eley, T. C., Eriksson, N., Escott-Price, V., Kiadeh, F. H. F., Finucane, H. K., Forstner, A. J., Frank, J., Gaspar, H. A., Gill, M., Giusti-Rodríguez, P., Goes, F. S., Gordon, S. D., Grove, J., Hall, L. S., Hannon, E., Hansen, C. S., Hansen, T. F., Herms, S., Hickie, I. B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.-J., Hougaard, D. M., Hu, M., Hyde, C. L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles, J. A., Kohane, I. S., Kraft, J., Kretzschmar, W. W., Krogh, J., Kutalik, Z., Lane, J. M., Li, Y., Li, Y., Lind, P. A., Liu, X., Lu, L., MacIntyre, D. J., MacKinnon, D. F., Maier, R. M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S. E., Mehta, D., Middeldorp, C. M., Mihailov, E., Milaneschi, Y., Milani, L., Mill, J., Mondimore, F. M., Montgomery, G. W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M. G., Nyholt, D. R., O'Reilly, P. F., Oskarsson, H., Owen, M. J., Painter, J. N., Pedersen, C. B., Pedersen, M. G., Peterson, R. E., Pettersson, E., Peyrot, W. J., Pistis, G., Posthuma, D., Purcell, S. M., Quiroz, J. A., Qvist, P., Rice, J. P., Riley, B. P., Rivera, M., Saeed Mirza, S., Saxena, R., Schoevers, R., Schulte, E. C., Shen, L., Shi, J., Shyn, S. I., Sigurdsson, E., Sinnamon, G. B. C., Smit, J. H., Smith, D. J., Stefansson, H., Steinberg, S., Stockmeier, C. A., Streit, F., Strohmaier, J., Tansey, K. E., Teismann, H., Teumer, A., Thompson, W., Thomson, P. A., Thorgeirsson, T. E., Tian, C., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A. G., Umbricht, D., Van der Auwera, S., van Hemert, A. M., Viktorin, A., Visscher, P. M., Wang, Y., Webb, B. T., Weinsheimer, S. M., Wellmann, J., Willemsen, G., Witt, S. H., Wu, Y., Xi, H. S., Yang, J., Zhang, F., eQTLGen, 23andMe, Arolt, V., Baune, B. T., Berger, K., Boomsma, D. I., Cichon, S., Dannlowski, U., de Geus, E. C. J., DePaulo, J. R., Domenici, E., Domschke, K., Esko, T., Grabe, H. J., Hamilton, S. P., Hayward, C., Heath, A. C., Hinds, D. A., Kendler, K. S., Kloiber, S., Lewis, G., Li, Q. S., Lucae, S., Madden, P. F. A., Magnusson, P. K., Martin, N. G., McIntosh, A. M., Metspalu, A., Mors, O., Mortensen, P. B., Müller-Myhsok, B., Nordentoft, M., Nöthen, M. M., O'Donovan, M. C., Paciga, S. A., Pedersen, N. L., Penninx, B. W. J. H., Perlis, R. H., Porteous, D. J., Potash, J. B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T. G., Smoller, J. W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M. M., Werge, T., Winslow, A. R., Lewis, C. M., Levinson, D. F., Breen, G., Børglum, A. D., Sullivan, P. F., & Major Depressive Disorder Working Group of the Psychiatric Genomics

Consortium. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* 50, 668–681 (2018).

57. Nelson, C. P., Goel, A., Butterworth, A. S., Kanoni, S., Webb, T. R., Marouli, E., Zeng, L., Ntalla, I., Lai, F. Y., Hopewell, J. C., Giannakopoulou, O., Jiang, T., Hamby, S. E., Di Angelantonio, E., Assimes, T. L., Bottinger, E. P., Chambers, J. C., Clarke, R., Palmer, C. N. A., Cubbon, R. M., Ellinor, P., Ermel, R., Evangelou, E., Franks, P. W., Grace, C., Gu, D., Hingorani, A. D., Howson, J. M. M., Ingelsson, E., Kastrati, A., Kessler, T., Kyriakou, T., Lehtimäki, T., Lu, X., Lu, Y., März, W., McPherson, R., Metspalu, A., Pujades-Rodriguez, M., Ruusalepp, A., Schadt, E. E., Schmidt, A. F., Sweeting, M. J., Zalloua, P. A., AlGhalayini, K., Keavney, B. D., Kooner, J. S., Loos, R. J. F., Patel, R. S., Rutter, M. K., Tomaszewski, M., Tzoulaki, I., Zeggini, E., Erdmann, J., Dedoussis, G., Björkegren, J. L. M., EPIC-CVD Consortium, CARDIoGRAMplusC4D, UK Biobank CardioMetabolic Consortium CHD working group, Schunkert, H., Farrall, M., Danesh, J., Samani, N. J., Watkins, H. & Deloukas, P. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* 49, 1385–1391 (2017).

58. Nielsen, J. B., Thorolfsdottir, R. B., Fritsche, L. G., Zhou, W., Skov, M. W., Graham, S. E., Herron, T. J., McCarthy, S., Schmidt, E. M., Sveinbjornsson, G., Surakka, I., Mathis, M. R., Yamazaki, M., Crawford, R. D., Gabrielsen, M. E., Skogholt, A. H., Holmen, O. L., Lin, M., Wolford, B. N., Dey, R., Dalen, H., Sulem, P., Chung, J. H., Backman, J. D., Arnar, D. O., Thorsteinsdottir, U., Baras, A., O'Dushlaine, C., Holst, A. G., Wen, X., Hornsby, W., Dewey, F. E., Boehnke, M., Kheterpal, S., Mukherjee, B., Lee, S., Kang, H. M., Holm, H., Kitzman, J., Shavit, J. A., Jalife, J., Brummett, C. M., Teslovich, T. M., Carey, D. J., Gudbjartsson, D. F., Stefansson, K., Abecasis, G. R., Hveem, K. & Willer, C. J. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* 50, 1234–1239 (2018).

59. Cordell, H. J., Han, Y., Mells, G. F., Li, Y., Hirschfield, G. M., Greene, C. S., Xie, G., Juran, B. D., Zhu, D., Qian, D. C., Floyd, J. A. B., Morley, K. I., Prati, D., Lleo, A., Cusi, D., Canadian-US PBC Consortium, Italian PBC Genetics Study Group, UK-PBC Consortium, Gershwin, M. E., Anderson, C. A., Lazaridis, K. N., Invernizzi, P., Seldin, M. F., Sandford, R. N., Amos, C. I. & Siminovitch, K. A. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat. Commun.* 6, 8019 (2015).

60. Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L. & Neale, B. M. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47, 1236–1241 (2015).

61. Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M. J., Price, A. L. & Neale, B. M. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47, 291–295 (2015).

62. Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, RACI Consortium, Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R. B., Okada, Y., Raychaudhuri, S., Daly, M. J., Patterson, N., Neale, B. M. & Price, A. L. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235 (2015).

63. Scott, R. A., Scott, L. J., Mägi, R., Marullo, L., Gaulton, K. J., Kaakinen, M., Pervjakova, N., Pers, T. H., Johnson, A. D., Eicher, J. D., Jackson, A. U., Ferreira, T., Lee, Y., Ma, C., Steinthorsdottir, V., Thorleifsson, G., Qi, L., Van Zuydam, N. R., Mahajan, A., Chen, H., Almgren, P., Voight, B. F., Grallert, H., Müller-Nurasyid, M., Ried, J. S., Rayner, W. N., Robertson, N., Karssen, L. C., van Leeuwen, E. M., Willems, S. M., Fuchsberger, C., Kwan, P., Teslovich, T. M., Chanda, P., Li, M., Lu, Y., Dina, C., Thuillier, D., Yengo, L., Jiang, L., Sparso, T., Kestler, H. A., Chheda, H., Eisele, L., Gustafsson, S., Frånberg, M., Strawbridge, R. J., Benediktsson, R., Hreidarsson, A. B., Kong, A., Sigurðsson, G., Kerrison, N. D., Luan, J., Liang, L., Meitinger, T., Roden, M., Thorand, B., Esko, T., Mihailov, E., Fox, C., Liu, C.-T., Rybin, D., Isomaa, B., Lyssenko, V., Tuomi, T., Couper, D. J., Pankow, J. S., Grarup, N., Have, C. T., Jørgensen, M. E., Jørgensen, T., Linneberg, A., Cornelis, M. C., van Dam, R. M., Hunter, D. J., Kraft, P., Sun, Q., Edkins, S., Owen, K. R., Perry, J. R., Wood, A. R., Zeggini, E., Tajes-Fernandes, J., Abecasis, G. R., Bonnycastle, L. L., Chines, P. S., Stringham, H. M., Koistinen, H. A., Kinnunen, L., Sennblad, B., Mühleisen, T. W., Nöthen, M. M., Pechlivanis, S., Baldassarre, D., Gertow, K., Humphries, S. E., Tremoli, E., Klopp, N., Meyer, J., Steinbach, G., Wennauer, R., Eriksson, J. G., Männistö, S., Peltonen, L., Tikkanen, E., Charpentier, G., Eury, E., Lobbens, S., Gigante, B., Leander, K., McLeod, O., Bottinger, E. P., Gottesman, O., Ruderfer, D., Blüher, M., Kovacs, P., Tonjes, A., Maruthur, N. M., Scapoli, C., Erbel, R., Jöckel, K.-H., Moebus, S., de Faire, U., Hamsten, A., Stumvoll, M., Deloukas, P., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., Ripatti, S., Salomaa, V., Pedersen, N. L., Boehm, B. O., Bergman, R. N., Collins, F. S., Mohlke, K. L., Tuomilehto, J., Hansen, T., Pedersen, O., Barroso, I., Lannfelt, L., Ingelsson, E., Lind, L., Lindgren, C. M., Cauchi, S., Froguel, P., Loos, R. J., Balkau, B., Boeing, H., Franks, P. W., Gurrea, A. B., Palli, D., van der Schouw, Y. T., Altshuler, D., Groop, L. C., Langenberg, C., Wareham, N. J., Sijbrands, E., van Duijn, C. M., Florez, J. C., Meigs, J. B., Boerwinkle, E., Gieger, C., Strauch, K., Metspalu, A., Morris, A. D., Palmer, C. N., Hu, F. B., Thorsteinsdottir, U., Stefansson, K., Dupuis, J., Morris, A. P., Boehnke, M., McCarthy, M. I., Prokopenko, I., & DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* (2017) doi:10.2337/db16-1253.

64. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* 94, 559–573 (2014).

65. Lee, D., Gorkin, D. U., Baker, M., Strober, B. J., Asoni, A. L., McCallion, A. S. & Beer, M. A. A method to predict the impact of regulatory variants from DNA sequence. *Nat. Genet.* 47, 955–961 (2015).

66. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., van Duijn, C. M., Gillies, C. E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J., Boomsma, D. I., Branham, K., Breen, G., Brummet, C., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F. S., Corbin, L., Davey Smith, G., Dedoussis, G., Dorr, M., Farmaki, A.-E., Ferrucci, L., Forer, L., Fraser, R. M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O. L., Hveem, K., Kretzler, M., Lee, J., McGue, M., Meitinger, T., Melzer, D., Min, J., Mohlke, K. L., Vincent, J., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, B., Sala, C., Salomaa, V., Schlessinger, D., Schoenheer, S., Slagboom, P. E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van den Berg, L., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M. G., Wilson, J. F., Frayling, T., de Bakker,

P., Swertz, M. A., McCarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., Soranzo, N., Walter, K., Swaroop, A., Cucca, F., Anderson, C., Boehnke, M., McCarthy, M. I., Durbin, R., Abecasis, G. & Marchini, J. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48, 1279–1283 (2016).

67. Wingender, E., Schoeps, T., Haubrock, M. & Dönitz, J. TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res.* 43, D97-102 (2015).

68. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272–286 (2014).

69. Schmitt, A. D., Hu, M. & Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.* 17, 743–755 (2016).

70. Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., Frum, T., Giste, E., Johnson, A. K., Johnson, E. M., Kutyavin, T., Lajoie, B., Lee, B.-K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E. D., Qu, H., Reynolds, A. P., Roach, V., Safi, A., Sanchez, M. E., Sanyal, A., Shafer, A., Simon, J. M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M. O., Hansen, R. S., Navas, P. A., Stamatoyannopoulos, G., Iyer, V. R., Lieb, J. D., Sunyaev, S. R., Akey, J. M., Sabo, P. J., Kaul, R., Furey, T. S., Dekker, J., Crawford, G. E. & Stamatoyannopoulos, J. A. The accessible chromatin landscape of the human genome. *Nature* 489, 75–82 (2012).

71. Miguel-Escalada, I., Bonàs-Guarch, S., Cebola, I., Ponsa-Cobas, J., Mendieta-Esteban, J., Atla, G., Javierre, B. M., Rolando, D. M. Y., Farabella, I., Morgan, C. C., García-Hurtado, J., Beucher, A., Morán, I., Pasquali, L., Ramos-Rodríguez, M., Appel, E. V. R., Linneberg, A., Gjesing, A. P., Witte, D. R., Pedersen, O., Grarup, N., Ravassard, P., Torrents, D., Mercader, J. M., Piemonti, L., Berney, T., de Koning, E. J. P., Kerr-Conte, J., Pattou, F., Fedko, I. O., Groop, L., Prokopenko, I., Hansen, T., Marti-Renom, M. A., Fraser, P. & Ferrer, J. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat. Genet.* 51, 1137–1148 (2019).

72. Jian, X. & Felsenfeld, G. Insulin promoter in human pancreatic β cells contacts diabetes susceptibility loci and regulates genes affecting insulin metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 115, E4633–E4641 (2018).

73. Lawlor, N., Márquez, E. J., Orchard, P., Narisu, N., Shamim, M. S., Thibodeau, A., Varshney, A., Kursawe, R., Erdos, M. R., Kanke, M., Gu, H., Pak, E., Dutra, A., Russell, S., Li, X., Piecuch, E., Luo, O., Chines, P. S., Fuchbserger, C., NIH Intramural Sequencing Center, Sethupathy, P., Aiden, A. P., Ruan, Y., Aiden, E. L., Collins, F. S., Ucar, D., Parker, S. C. J. & Stitzel, M. L. Multiomic Profiling Identifies cis-Regulatory Networks Underlying Human Pancreatic β Cell Identity and Function. *Cell Rep.* 26, 788-801.e6 (2019).

74. Rezania, A., Bruin, J. E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., Yang, Y. H. C., Johnson, J. D. & Kieffer, T. J. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* 32, 1121–1133 (2014).

75. Claussnitzer, M., Dankel, S. N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I. S., Beaudry, J. L., Puviindran, V., Abdennur, N. A., Liu, J., Svensson, P.-A., Hsu, Y.-H., Drucker, D. J., Mellgren, G., Hui, C.-C., Hauner, H. & Kellis, M. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* 373, 895–907 (2015).

76. Fogarty, M. P., Cannon, M. E., Vadlamudi, S., Gaulton, K. J. & Mohlke, K. L. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet.* 10, e1004633 (2014).

77. Rusu, V., Hoch, E., Mercader, J. M., Tenen, D. E., Gymrek, M., Hartigan, C. R., DeRan, M., von Grotthuss, M., Fontanillas, P., Spooner, A., Guzman, G., Deik, A. A., Pierce, K. A., Dennis, C., Clish, C. B., Carr, S. A., Wagner, B. K., Schenone, M., Ng, M. C. Y., Chen, B. H., MEDIA Consortium, SIGMA T2D Consortium, Centeno-Cruz, F., Zerrweck, C., Orozco, L., Altshuler, D. M., Schreiber, S. L., Florez, J. C., Jacobs, S. B. R. & Lander, E. S. Type 2 Diabetes Variants Disrupt Function of SLC16A11 through Two Distinct Mechanisms. *Cell* 170, 199-212.e20 (2017).

78. Carrat, G. R., Hu, M., Nguyen-Tu, M.-S., Chabosseau, P., Gaulton, K. J., van de Bunt, M., Siddiq, A., Falchi, M., Thurner, M., Canouil, M., Pattou, F., Leclerc, I., Pullen, T. J., Cane, M. C., Prabhala, P., Greenwald, W., Schulte, A., Marchetti, P., Ibberson, M., MacDonald, P. E., Manning Fox, J. E., Gloyn, A. L., Froguel, P., Solimena, M., McCarthy, M. I. & Rutter, G. A. Decreased STARD10 Expression Is Associated with Defective Insulin Secretion in Humans and Mice. *Am. J. Hum. Genet.* 100, 238–256 (2017).

79. Claussnitzer, M., Dankel, S. N., Klocke, B., Grallert, H., Glunk, V., Berulava, T., Lee, H., Oskolkov, N., Fadista, J., Ehlers, K., Wahl, S., Hoffmann, C., Qian, K., Rönn, T., Riess, H., Müller-Nurasyid, M., Bretschneider, N., Schroeder, T., Skurk, T., Horsthemke, B., DIAGRAM+Consortium, Spieler, D., Klingenspor, M., Seifert, M., Kern, M. J., Mejhert, N., Dahlman, I., Hansson, O., Hauck, S. M., Blüher, M., Arner, P., Groop, L., Illig, T., Suhre, K., Hsu, Y.-H., Mellgren, G., Hauner, H. & Laumen, H. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* 156, 343–358 (2014).

80. Roman, T. S., Cannon, M. E., Vadlamudi, S., Buchkovich, M. L., Wolford, B. N., Welch, R. P., Morken, M. A., Kwon, G. J., Varshney, A., Kursawe, R., Wu, Y., Jackson, A. U., Comparative, S. P. N., Erdos, M. R., Kuusisto, J., Laakso, M., Scott, L. J., Boehnke, M., Collins, F. S., Parker, S. C. J., Stitzel, M. L. & Mohlke, K. L. A Type 2 Diabetes-Associated Functional Regulatory Variant in a Pancreatic Islet Enhancer at the Adcy5 Locus. *Diabetes* db170464 (2017) doi:10.2337/db17-0464.

81. Kycia, I., Wolford, B. N., Huyghe, J. R., Fuchsberger, C., Vadlamudi, S., Kursawe, R., Welch, R. P., Albanus, R. d'Oliveira, Uyar, A., Khetan, S., Lawlor, N., Bolisetty, M., Mathur, A., Kuusisto, J., Laakso, M., Ucar, D., Mohlke, K. L., Boehnke, M., Collins, F. S., Parker, S. C. J. & Stitzel, M. L. A Common Type 2 Diabetes Risk Variant Potentiates Activity of an Evolutionarily Conserved Islet Stretch Enhancer and Increases C2CD4A and C2CD4B Expression. *Am. J. Hum. Genet.* 102, 620–635 (2018).

82. Avrahami, D., Klochendler, A., Dor, Y. & Glaser, B. Beta cell heterogeneity: an evolving concept. *Diabetologia* 60, 1363–1369 (2017).

83. Modi, H., Skovsø, S., Ellis, C., Krentz, N. A. J., Zhao, Y. B., Cen, H., Noursadeghi, N., Panzhinskiy, E., Hu, X., Dionne, D. A., Xuan, S., Huising, M. O., Kieffer, T. J., Lynn, F. C. & Johnson, J. D. Ins2 gene bursting activity defines a mature β-cell state. *bioRxiv* 702589 (2019) doi:10.1101/702589.

84. Farack, L., Golan, M., Egozi, A., Dezorella, N., Bahar Halpern, K., Ben-Moshe, S., Garzilli, I., Tóth, B., Roitman, L., Krizhanovsky, V. & Itzkovitz, S. Transcriptional Heterogeneity of Beta Cells in the Intact Pancreas. *Dev. Cell* 48, 115-125.e4 (2019).

85. Rai, V., Quang, D. X., Erdos, M. R., Cusanovich, D. A., Daza, R. M., Narisu, N., Zou, L. S., Didion, J. P., Guan, Y., Shendure, J., Parker, S. C. J. & Collins, F. S. 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.* (2019) doi:10.1016/j.molmet.2019.12.006.

86. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).

87. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* 9, 1–5 (2019).

88. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15 (2018).

89. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* 9, 1–12 (2019).

90. Arda, H. E., Tsai, J., Rosli, Y. R., Giresi, P., Bottino, R., Greenleaf, W. J., Chang, H. Y. & Kim, S. K. A Chromatin Basis for Cell Lineage and Disease Risk in the Human Pancreas. *Cell Syst.* 7, 310-322.e4 (2018).

91. Ackermann, A. M., Wang, Z., Schug, J., Naji, A. & Kaestner, K. H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* 5, 233–244 (2016).

92. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842 (2010).

93. Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W. & Ma'ayan, A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90-97 (2016).

94. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550 (2005).

95. Li, Y. I., van de Geijn, B., Raj, A., Knowles, D. A., Petti, A. A., Golan, D., Gilad, Y. & Pritchard, J. K. RNA splicing is a primary link between genetic variation and disease. *Science* 352, 600–604 (2016).

96. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W. & Noble, W. S. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202-208 (2009).

97. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* 12, 1061–1063 (2015).

98. Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S. & Aiden, E. L. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* 3, 95–98 (2016).

99. Kerpedjiev, P., Abdennur, N., Lekschas, F., McCallum, C., Dinkla, K., Strobelt, H., Luber, J. M., Ouellette, S. B., Azhir, A., Kumar, N., Hwang, J., Lee, S., Alver, B. H., Pfister, H., Mirny, L. A., Park, P. J. & Gehlenborg, N. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* 19, 125 (2018).

100. Raviram, R., Rocha, P. P., Müller, C. L., Miraldi, E. R., Badri, S., Fu, Y., Swanzey, E., Proudhon, C., Snetkova, V., Bonneau, R. & Skok, J. A. 4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments. *PLoS Comput. Biol.* 12, e1004780 (2016).

101. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359 (2012).

102. Velazco-Cruz, L., Song, J., Maxwell, K. G., Goedegebuure, M. M., Augsornworawat, P., Hogrebe, N. J. & Millman, J. R. Acquisition of Dynamic Function in Human Stem Cell-Derived β Cells. *Stem Cell Rep.* 12, 351–365 (2019).

# Chapter 3: Cell type-specific genetic mechanisms of type 1 diabetes risk

## 3.1 Abstract

Genetic risk variants identified in genome-wide association studies (GWAS) of complex disease are primarily non-coding, and translating risk variants into mechanistic insight requires detailed gene regulatory maps in disease-relevant cell types. Here, we combined a GWAS of type 1 diabetes (T1D) in 520,580 samples with candidate *cis*-regulatory elements (cCREs) in pancreas and peripheral blood mononuclear cell types defined using single nucleus ATAC-seq (snATAC-seq) of 131,554 nuclei. T1D risk variants were enriched in cCREs active in T-cells and additional cell types including acinar and ductal cells of the exocrine pancreas. Risk variants at multiple T1D signals overlapped exocrine-specific cCREs linked to genes with exocrine-specific expression. At the *CFTR* locus, T1D risk variant rs7795896 mapped in a ductal-specific cCRE which regulated *CFTR*, and the risk allele reduced ductal cell transcription factor binding, enhancer activity and *CFTR* expression. These findings support a role for the exocrine pancreas in T1D pathogenesis and highlight the power of large-scale GWAS and single cell epigenomics in understanding the cellular origins of complex disease.

## 3.2 Introduction

Type 1 diabetes (T1D) is a complex autoimmune disease characterized by the loss of insulin-producing pancreatic beta cells[1], where the triggers of autoimmunity and disease onset remain poorly understood. T1D has a strong genetic component, most prominently at the major histocompatibility complex (MHC) locus, but also from 59 additional risk loci[2–4]. T1D risk variants are largely non-coding, and intersection of risk variants with epigenomic data has identified enrichment within lymphoid enhancers[2]. However, due to limited sample sizes, incomplete variant coverage, and limited cell type resolution of existing epigenomic maps, the causal variants and cellular mechanisms of action of T1D risk loci are largely unresolved.

## 3.3 Results

### 3.3.1 Comprehensive discovery and fine mapping of T1D risk signals

We performed a GWAS of 18,942 T1D cases and 501,638 controls of European ancestry from 9 cohorts (**Table 3.1**). After applying uniform quality-control (**Figure S3.1**), we imputed genotypes into the TOPMed r2 panel and tested for T1D association[5]. Through meta-analysis, we combined association results for 61,947,369 variants and observed 81 loci reaching genome-wide significance ($P<5\times10^{-8}$), including 48 of 59 known loci and 33 previously unreported loci (**Figure 3.1a**, **Figure S3.2, Table 3.2**). At 92 total loci (59 known and 33 novel), we discovered 44 independent signals using FINEMAP, of which 36 were previously unreported (**Figure S3.3, Table 3.2**). Nearly a third (32%; 29/92) of loci contained more than one signal; for example, the *PTPN2* and *BCL11A* loci each had three signals (**Figure 3.1b, Figure 3.1c**).

We next fine-mapped causal variants for 136 T1D signals including 92 main and 44 independent signals (**Figure 3.1c**). We used FINEMAP to obtain the posterior probability of association (PPA) for tested variants and define 99% credible sets for each signal. (**Table 3.3**). Compared to a previous study[6], fine-mapping resolution was improved based on credible set size and maximum posterior probability (**Figure 3.1d, Figure S3.4**). The median credible set size was

31 variants, where nearly a quarter (24%; 32/136) contained 5 or fewer variants, and 28% (38/136) contained a single variant with >0.50 PPA (**Figure 1c**). Credible sets at 15% (21/136) of signals contained a nonsynonymous variant with PPA>0.01, including novel loci *AIRE* p.Arg471Cys (PPA=0.99), *BATF3* p.Val11Ile (PPA=0.078), *PRF1* p.Ala91Val (PPA=0.28), and *INPP5B* p.Gly250Cys (PPA=0.055) (**Table 3.4**).

The TOPMed r2 panel enables more accurate imputation of rare variants. We identified four novel variants with minor allele frequency (MAF)<0.005 and large effects on T1D (**Figure S3.5a**). Among these, rs541856133 (MAF=0.0015, OR=3.01, 95% CI=2.33-3.89) mapped directly upstream of *CEL,* a gene implicated as causal for maturity-onset diabetes of the young (MODY8)[7]. We also identified a novel protein-coding protective variant at *IFIH1* (p.Asn160Asp, rs75671397, MAF=0.002, OR=0.35, 95% CI=0.22-0.55) independent of known signals in this gene. Two additional non-coding risk variants mapped to *SH2B3* (rs570074821, MAF=0.0019, OR=1.89, 95% CI=1.37-2.61) and *CAMK4* (rs72663304, MAF=0.0013, OR=2.54, 95% CI=1.72-3.76) (**Figure S3.5b**).

We characterized genetic correlations between T1D and other complex traits and diseases with LD score regression. In line with previous reports[2,8], T1D had significant (FDR<0.10) positive correlations with autoimmune diseases such as rheumatoid arthritis ($r_g$=0.44, FDR=$7.52\times10^{-5}$) and systemic lupus erythematosus ($r_g$=0.35, FDR=$5.05\times10^{-7}$), and negative correlation with ulcerative colitis ($r_g$=-0.18, FDR=$1.95\times10^{-3}$) (**Figure S3.6**). We also observed significant positive correlations with metabolic traits such as fasting insulin level ($r_g$=0.18, FDR=$4.04\times10^{-3}$), coronary artery disease ($r_g$=0.12, FDR=$1.23\times10^{-2}$), and type 2 diabetes ($r_g$=0.10, FDR=$1.95\times10^{-3}$). These results demonstrate relationships between genetic effects on T1D and autoimmune and metabolic disease.

### 3.3.2 Defining cell type-specific *cis*-regulatory programs in T1D-relevant tissues

The majority of T1D risk likely affects gene regulation[2]. To annotate T1D risk variants, we generated an accessible chromatin reference map using snATAC-seq of peripheral blood and pancreas from non-diabetic donors (**Table 3.5**). We used a modified pipeline[9] to group chromatin accessibility profiles from 131,554 cells into 28 clusters (**Figure 3.2a, Figure S3.7**) and assigned cell type identities using chromatin accessibility at known marker genes (**Table 3.6**). For example, chromatin accessibility at *C1QB* marked pancreas tissue-resident macrophages, *REG1A* marked acinar cells, and *CFTR* marked ductal cells (**Figure 3.2b**). We also observed patterns of chromatin accessibility at marker genes for distinct cell sub-types, such as *FOXP3* for regulatory T-cells (**Figure 3.2b**). To relate cell type-resolved accessible chromatin to gene expression, we created a single cell RNA-seq (scRNA-seq) reference map in peripheral blood and pancreas. We assigned cell type identities for 90,495 cells in 29 clusters, which largely identified similar cell types and proportions as snATAC-seq (**Figure S3.8**).

To characterize *cis*-regulatory programs, we aggregated reads from cells within each snATAC-seq cluster and identified accessible chromatin peaks representing cCREs. There were 448,142 cCREs across all 28 clusters and an average of 77,812 cCREs per cluster. We also aggregated reads from cells within each snRNA-seq cluster to derive normalized expression (transcripts per million, TPM). To delineate regulatory programs specifying each cell type, we identified 25,436 cCREs with accessibility patterns most specific to each cluster (**Figure 3.2c**)**.** Genes within 100 kb of cell type-specific cCREs had more cell type-specific expression relative to other cCREs (**Figure S3.9**). Cell type-specific cCREs were also strongly enriched for GO terms representing highly specialized cellular processes (**Figure 3.2c**).

We defined transcriptional regulators of cCRE activity by assessing transcription factor (TF) motif enrichment using chromVAR. Enriched TF motifs included those with lineage, cell type, and cell state specificity (**Figure 3.2d**). As TFs within subfamilies often have similar motifs, we

grouped TFs into subfamilies to identify TFs with matching cell type expression and motif enrichment patterns (**Table 3.7**). For example, FOXA subfamily TFs *FOXA2* and *FOXA3* were specifically expressed in pancreatic endocrine and exocrine cells, HNF1 subfamily TF *HNF1B* was specifically expressed in ductal cells, and ROR subfamily TF *RORC* was specifically expressed in memory CD8+ T-cells (**Figure 3.2d, Table 3.7**).

As the target genes of cCRE activity are largely unknown, we identified cell type-resolved co-accessibility links between distal (non-promoter) cCREs and putative target gene promoters using Cicero. Across all cell types, we observed a total of 1,028,428 links (co-accessibility>0.05) between distal cCREs and gene promoters. Co-accessible links were often cell type-specific; for example, multiple distal cCREs were co-accessible with the *AQP1* promoter in ductal cells and the *CEL* promoter in acinar cells (**Figure 3.2e**). In nearly every cell type, target genes co-accessible with distal cCREs were more likely to be expressed in the cell type compared to matched non-co-accessible genes (**Figure S3.10**).

### 3.3.3 Annotating T1D risk variants with cell type-specific regulatory programs

We determined enrichment of variants associated with T1D and other complex traits and diseases for cell type cCREs using stratified LD score regression. For T1D, the most significant enrichment was in T cell cCREs (naïve T Z=5.57, FDR=$2.26\times10^{-5}$; memory CD8+ T Z=4.80, FDR=$4.67\times10^{-4}$; activated CD4+ T Z=4.62, FDR=$6.74\times10^{-4}$; cytotoxic CD8+ T Z=4.49 FDR=$1.09\times10^{-3}$; regulatory T Z=3.26, FDR=$7.23\times10^{-3}$) and adaptive NK cells (Z=3.50, FDR=$9.93\times10^{-3}$) (**Figure 3.3a**). Notably, we did not observe enrichment in pancreatic resident immune cells (CD8+ T Z=0.65, FDR=1.0; macrophage Z=-0.56, FDR=1.0). Other immune-related diseases were primarily enriched within lymphocyte cCREs, while type 2 diabetes and glycemic traits were enriched in pancreatic endocrine, acinar, and ductal cCREs (**Figure 3.3a**). These results demonstrate that T1D variants are broadly enriched for T cell cCREs and highlight other traits enriched for pancreatic and immune cell cCREs.

Despite the strong enrichment of T1D-associated variants in T-cells, many T1D signals did not overlap a T-cell cCRE suggesting that additional cell types contribute to T1D risk. To identify additional disease-relevant cell types, we used an orthogonal approach to test for enrichment of T1D variants within the subset of cell type-specific cCREs with fgwas. As expected, T1D-associated variants were enriched in cCREs specific to T-cells and beta cells (activated CD4+ T ln(enrich)=4.25, 95% CI=1.11-5.43; cytotoxic CD8+ T ln(enrich)=4.04, 95% CI=0.20-5.20); $INS^{high}$ beta cells (ln(enrich)=3.58, 95% CI=0.95-4.84) (**Figure 3.3b**). Interestingly, T1D variants were also enriched in cCREs specific to plasmacytoid dendritic (pDC) (ln(enrich)=4.00, 95% CI=1.96-5.10), classical monocytes (ln(enrich)=3.78, 95% CI=2.23-4.74), acinar (ln(enrich)=3.35, 95% CI=1.59-4.46) and ductal cells (ln(enrich)=3.28, 95% CI=0.18-4.69) (**Figure 3.3b**).

Given insight into key T1D-relevant cell types, we next annotated T1D signals in cCREs for these cell types. Over 75% (103/136) of T1D signals contained at least one variant (PPA>0.01) overlapping a cCRE, and at 65% (67/103) of these signals the cCRE was co-accessible with a gene promoter. Variants with high probabilities (PPA>0.50) were significantly more likely to map in a cCRE compared to other credible set variants (OR=3.9, 95% CI 1.9-7.8, $P=1.9\times10^{-4}$), and these cCREs were more likely to be co-accessible with a promoter (OR=6.1, 95% CI 1.3-55.9, $P=7.1\times10^{-3}$). For each signal, we calculated the cumulative posterior probability (cPPA) of credible set variants overlapping distal cCREs in each disease-enriched cell type. Numerous T1D signals had high cPPA in T cell cCREs and not in other disease-relevant cell types (**Figure 3.3c**). We also observed T1D signals with high cPPA in acinar and ductal (exocrine), beta cell, monocyte and pDC cCREs, several of which were highly cell type-specific (**Figure 3.3c**). For each signal, we further annotated genes within 1 Mb expressed in the same cell type and co-accessible with cCREs (**Figure 3.3c**).

Multiple T1D signals had high cPPA specifically in pancreatic exocrine cells and were linked to genes with exocrine-specific expression. At the *GP2* locus, three variants accounted for 0.951 PPA and mapped in an acinar-specific cCRE co-accessible with the promoter of *GP2*, which had acinar-specific expression (**Figure 3.3c,d**). Similarly, rs72802342 at the *BCAR1* locus (PPA=0.30) mapped in an acinar-specific cCRE co-accessible with the promoters of *CTRB1* and *CTRB2*, both of which had acinar-specific expression (**Figure 3.3c,e**). Other signals such as *CEL* had similar exocrine-specific profiles (**Figure S3.11a-c**). Exocrine cCREs at T1D loci were also largely specific relative to stimulated immune cell and islet accessible chromatin (**Table 3.8**).

### 3.3.4 Risk variant at novel T1D locus has pancreatic ductal cell-specific effects on *CFTR*

The *CFTR* locus contained a fine-mapped variant rs7795896 (PPA=0.63) in a distal cCRE specific to ductal cells and co-accessible with the *CFTR* promoter in addition to other genes (**Figure 3.4a**). Recessive mutations in *CFTR* cause cystic fibrosis (CF), which is often comorbid with exocrine pancreas insufficiency and CF-related diabetes (CFRD)[10]. Furthermore, carriers of *CFTR* mutations often develop chronic pancreatitis[11]. As *CFTR* has not been previously implicated in T1D, we sought to validate the mechanism of this locus. The T1D risk allele of rs7795896 significantly reduced enhancer activity (594bp sequence $P=1.15\times10^{-2}$, **Figure 3.4b**; 180bp sequence $P=3.35\times10^{-2}$, **Figure S3.12a**) and reduced protein binding (bound fraction rs7795896-C=0.007, rs7795896-T=0.081; **Figure 3.4c, Figure S3.12b**) in Capan-1 cells. The variant mapped in a sequence motif for the ductal TF HNF1B, albeit in a position predicted to have minimal impact, and overlapped a HNF1B ChIP-seq site previously identified in ductal cells[12] (**Figure S3.12c**).

To determine whether the enhancer harboring rs7795896 regulated *CFTR* in ductal cells, we used CRISPR interference (CRISPRi) to inactivate enhancer activity ($CFTR^{Enh}$) in Capan-1 cells (**Table 3.9**). As positive and negative controls, we inactivated the *CFTR* promoter ($CFTR^{Prom}$) and used a non-targeting guide RNA, respectively. Quantitative PCR revealed a significant

reduction in *CFTR* expression after enhancer inactivation (ANCOVA $P=9.10\times10^{-5}$), whereas expression of other genes at the locus was unchanged (**Figure 3.4d, Figure S3.12d**). We determined whether risk variants affected *CFTR* expression using pancreas eQTL data from GTEx[13]. Out of 13 genes tested for association, only *CFTR* had evidence for an eQTL ($P=4.31\times10^{-4}$), and this eQTL was colocalized with the T1D signal ($PP_{shared}=91.8\%$) (**Figure 3.4e**). We identified four candidate variants with evidence for driving the shared signal using eCAVIAR (CLPP>0.01), and only rs7795896 mapped in a cCRE. The T1D risk allele of rs7795896 was associated with decreased *CFTR* expression, consistent with effects on enhancer activity and TF binding. To evaluate whether the *CFTR* eQTL in pancreas was driven by ductal cells, we re-calculated eQTL associations including estimated cell type proportion as an interaction term, and only ductal cells had significant association ($P=2.37\times10^{-4}$) (**Figure 3.4f-g, Figure S3.12e**).

As *CFTR* has been implicated in pancreatic cancer[14] and pancreatitis[15], we asked whether rs7795896 was associated with these phenotypes in UK biobank and FinnGen. The T1D risk allele was associated with increased risk of pancreatitis (chronic pancreatitis OR=1.15, $P=3.18\times10^{-3}$; acute pancreatitis OR=1.07, $P=1.15\times10^{-2}$) and other pancreatic diseases (OR=1.13, $P=4.72\times10^{-5}$) (**Figure 3.4h**). In contrast, rs7795896 was not associated with other autoimmune diseases (all P>0.05). T1D signals associated with increased risk of pancreatic disease had significantly higher cPPA in exocrine cCREs compared to other signals (two-sided Student's t-test P=0.027) and no difference for T cell cCREs (P=0.36). Together, our findings support a model in which variants regulating *CFTR* and other genes in the exocrine pancreas increase risk of T1D and pancreatic diseases (**Figure 3.4i**).

## 3.4 Discussion

High-resolution mapping of both genetic variants influencing T1D risk and cell type-specific *cis*-regulatory programs in T1D-relevant tissues enabled new insight into disease mechanisms. Risk variants at multiple novel loci mapped to genes with specialized function in

exocrine cells. While our results support variants in exocrine cCREs mediating T1D risk, fine-mapping has not resolved a single variant at most loci. Risk variants in exocrine-specific cCREs may also function in other cell types in the context of development, environmental changes, or disease progression. Continued fine-mapping in trans-ethnic cohorts with systematic evaluation of variant function across a breadth of conditions in relevant cell types will help clarify risk mechanisms. Furthermore, while we linked cCREs to putative target genes using co-accessibility, these links represent correlations that require both sites to vary in their accessibility. Future studies will benefit from linking changes in chromatin to gene expression directly through single cell multi-omics.

Observational studies have reported exocrine pancreas abnormalities at T1D onset[16], but it was unknown whether this was causing disease[17]. Genomic studies have also identified changes in exocrine cell proportions in T1D[18]. Exocrine pancreas abnormalities in T1D have been considered secondary to other disease processes, such as beta cell loss causing reduced insulinotropic effects on exocrine cells or viral infection leading to exocrine inflammation. In contrast, our findings provide evidence that exocrine cells intrinsically contribute to T1D pathogenesis. Reduced *CFTR* leads to CFRD via intra-islet inflammation and immune infiltration, and immune infiltration in the exocrine pancreas has been suggested to contribute to T1D[19–21]. Other implicated genes encode proteins secreted from acinar cells linked to risk of pancreatic disease[22–24], and may contribute to an inflammatory state. We therefore hypothesize a causal role for pancreatic exocrine gene regulation in T1D, which may provide novel avenues for therapeutic discovery.

## 3.5 Methods

### 3.5.1 Genotype quality control and imputation

We compiled individual-level genotype data and summary statistics of 18,942 T1D cases and 501,638 controls of European ancestry from public sources (**Table 3.1**), where T1D case

cohorts were matched to population control cohorts based on genotyping array (Affymetrix, Illumina Infinium, Illumina Omni, and Immunochip) and country of origin where possible (US, British, and Ireland). For the GENIE-UK cohort, because we were unable to find a matched country of origin control cohort, we used individuals of British ancestry (defined by individuals within 1.5 interquartile range of CEU/GBR subpopulations on the first 4 PCs from PCA with European 1000 Genomes Project samples) from the University of Michigan Health and Retirement study (HRS). For non-UK Biobank cohorts, we first applied individual and variant exclusion lists (where available) to remove low quality, duplicate, or non-European ancestry samples and failed genotype calls for each cohort. For control cohorts, we also used phenotype files (where available) to remove individuals with type 2 diabetes or autoimmune diseases.

We then applied a uniform processing pipeline and used PLINK[25] (version 1.90b6.7) to remove variants based on (i) low frequency (MAF<1%), (ii) missing genotypes (missing>5%), (iii) violation of Hardy-Weinberg equilibrium (HWE $P<1\times10^{-5}$ in control cohorts and HWE $P<1\times10^{-10}$ in case cohorts), (iv) difference in allele frequency >0.2 compared to the Haplotype Reference Consortium r1.1 reference panel[26], and (v) allele ambiguity defined as AT/GC variants with MAF>40%[35]. We further removed individuals based on (i) missing genotypes (missing>5%), (ii) sex mismatch with phenotype records ($hom_{chrX}$>0.2 for females and $hom_{chrX}$<0.8 for males), (iii) cryptic relatedness through identity-by-descent (IBD>0.2), and (iv) non-European ancestry through PCA with 1000 Genomes Project[27] (>3 interquartile range from 25$^{th}$ and 75$^{th}$ percentiles of European 1KGP samples on the first 4 PCs) (**Figure S3.1**). Lists of independent variants for IBD and PCA calculations were generated using PLINK ('--indep 50 5 2'). For the affected sib-pair (ASP) cohort genotyped on the Immunochip, we retained only one T1D sample from each family selected at random. For the GRID case and 1958 Birth control cohorts genotyped on the Immunochip, a portion of the cases overlapped the T1DGC or 1958 Birth cohorts genotyped on a genome-wide array. We thus used sample IDs from the phenotype files to remove these

samples from the GRID and 1958 Birth cohorts and verified that no samples were duplicated between the Immunochip and genome-wide array datasets by checking IBD. We combined data for matched case and control cohorts based on genotyping array and country of origin for imputation. We used the TOPMed Imputation Server[28] to impute genotypes into the TOPMed r2 panel[5] and removed variants based on low imputation quality ($R^2$<0.3). Following imputation, we implemented post-imputation filters to remove variants based on potential genotyping or imputation artifacts based on empirical $R^2$ (genotyped variants with empirical $R^2$<0.5 and all imputed variants in at least low linkage disequilibrium; LD, $r^2$>0.3).

For the UK Biobank cohort, we downloaded imputed genotype data from the UK Biobank v3 release which were imputed using a combination of the HRC and UK10K + 1000 Genomes reference panels. We removed individuals who had withdrawn participation from the UK biobank. We used phenotype data to remove individuals of non-European descent. To resolve duplicate samples represented in both the UK biobank and other cohorts on different genotyping arrays, we calculated IBD between samples in the UK biobank and cohorts of UK origin, removing duplicated samples from the UK biobank (IBD>0.9). Following these filters, we then used a combination of ICD10 codes to define 1,445 T1D cases (T1D diagnosis, insulin treatment within a year of diagnosis, no T2D diagnosis). We defined controls as 362,050 individuals without diabetes (no T1D, T2D, or gestational diabetes diagnosis) or other autoimmune diseases (systemic lupus erythematosus, rheumatoid arthritis, juvenile arthritis, Sjögren syndrome, alopecia areata, multiple sclerosis, autoimmune thyroiditis, vitiligo, celiac disease, primary biliary cirrhosis, psoriasis, or ulcerative colitis). We removed variants with low imputation quality ($R^2$<0.3).

For the FinnGen cohort, we downloaded GWAS summary statistics for type 1 diabetes (T1D_STRICT) from FinnGen freeze 3 (http://r3.finngen.fi/). This phenotype definition excluded individuals with type 2 diabetes from both cases and controls.

### 3.5.2 Association testing and meta-analysis

We tested variants with MAF>$1\times10^{-5}$ for association to T1D with firth bias reduced logistic regression using EPACTS (https://genome.sph.umich.edu/wiki/EPACTS) for non-UK Biobank cohorts or SAIGE[29] (version 0.38) for the UK Biobank, using genotype dosages adjusted for sex and the first four ancestry PCs. For the UK Biobank we used SAIGE as it is designed to run on biobank-scale cohorts and with highly imbalanced ratios of cases vs controls. For FinnGen, we used association results from the freeze 3 release that were generated using SAIGE. Prior to meta-analysis, we used liftOver to convert GRCh37/hg19 into GRCh38/hg38 coordinates for the UK biobank. We then combined association results across matched cohorts through inverse-variance weighted meta-analysis. We used liftOver to convert GRCh38/hg38 back into GRCh37/hg19 coordinates for the meta-analysis. We removed variants that were unable to be converted, were duplicated after coordinate conversion, or were located on different chromosomes after conversion. In total, our association data contained summary statistics for 61,947,369 variants. To evaluate the extent to which genomic inflation was driven by the polygenic nature of T1D or population stratification, we used LD score regression[30] to compare the LDSC intercept to lambda genomic control (GC). We observed an intercept of 1.07 (SE=0.03) compared to a lambda GC of 1.20, suggesting that the majority of the observed inflation was driven by polygenicity rather than population stratification.

### 3.5.3 Stochastic search and fine mapping of independent signals

We identified 59 loci (excluding the MHC locus) with T1D risk variants reported in previous genetic studies of T1D[2–4,31], and considered a locus in our study known if the most associated variant mapped within 500 kb of a previously reported T1D variant. We defined 33 novel loci where a variant reached genome-wide significance ($P<5\times10^{-8}$), and both mapped at least 500 kb away and was not in LD ($r^2<0.01$) with a previously reported T1D variant. At 92 (59 known and 33

novel) loci, we defined the 'index' variant as the variant with strongest T1D association at the locus.

For all 92 loci, we used a 1 Mb window around the index variant as the region for fine-mapping using FINEMAP[32] (version 1.4). For each region, we first filtered for variants with MAF>0.0005 and constructed pairwise LD matrices with PLINK[25] ( '--r --square --keep-allele-order') using the TOPMed2-imputed cohorts with genome-wide coverage (DCCT-EDIC, GENIE-ROI, GENIE-UK, GoKinD, T1DGC, WTCCC1-T1D and their respective control cohorts). We then applied FINEMAP using these matrices to conduct shotgun stochastic search and Bayesian fine-mapping using the default prior ('--sss --n-causal-snps 10 --prob-cred-set 0.99 --prior-std 0.05'). We selected the number of independent signals (causal variants) for each region based on the configuration with the highest FINEMAP posterior probability and used 99% credible sets from the FINEMAP output for the resulting signals. We calculated the effective sample size for all credible set variants, and no credible set variant with PPA>0.01 had <50% of the maximum effective sample size. We compared fine-mapping results to a previous fine-mapping dataset[6]. At 56 signals in common to both studies, we calculated the number of variants in the 99% credible set and the probability of the most likely causal variant.

### 3.5.4 GWAS correlation analyses

We used LD score regression[30,33] (version 1.0.1) to estimate genome-wide genetic correlations between T1D and immune diseases[34–42], other diseases[43–53], and non-disease traits[54–72], using European subsets of GWAS where applicable. For acute pancreatitis, chronic pancreatitis, and pancreatic cancer, we used inverse variance weighted meta-analysis to combine SAIGE analysis results from the UK biobank[29] (PheCodes 577.1, 577.2, and 157) and FinnGen r3 (K11_ACUTPANC, K11_CHRONPANC, C3_PANCREAS_EXALLC). We used pre-computed European 1000 Genomes LD scores to calculate correlation estimates ($r_g$) and standard errors. We then corrected p-values for multiple tests using FDR correction and considered FDR<0.1 as

significant. We also performed genetic correlation analyses using a version of the T1D meta-analysis excluding the Immunochip cohorts and observed highly similar results.

### 3.5.5 Generation of snATAC-seq libraries

Combinatorial indexing single cell ATAC-seq (snATAC-seq/sci-ATAC-seq). snATAC-seq was performed as described previously[9,73,74] with several modifications as described below. For the islet samples, approximately 3,000 islet equivalents (IEQ, roughly 1,000 cells each) were resuspended in 1 mL nuclei permeabilization buffer (10mM Tris-HCL (pH 7.5), 10mM NaCl, 3mM $MgCl_2$, 0.1% Tween-20 (Sigma), 0.1% IGEPAL-CA630 (Sigma) and 0.01% Digitonin (Promega) in water) and homogenized using 1mL glass dounce homogenizer with a tight-fitting pestle for 15 strokes. Homogenized islets were incubated for 10 min at 4°C and filtered with 30 µm filter (CellTrics). For the pancreas samples, frozen tissue was pulverized with a mortar and pestle while frozen and immersed in liquid nitrogen. Approximately 22 mg of pulverized tissue was then transferred to an Eppendorf tube and resuspended in 1 mL of cold permeabilization buffer for 10 minutes on a rotator at 4°C. Permeabilized sample was filtered with a 30µm filter (CellTrics), and the filter was washed with 300 µL of permeabilization buffer to increase nuclei recovery.

Once permeabilized and filtered, nuclei were pelleted with a swinging bucket centrifuge (500×g, 5 min, 4°C; 5920R, Eppendorf) and resuspended in 500 µL high salt tagmentation buffer (36.3 mM Tris-acetate (pH = 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and counted using a hemocytometer. Concentration was adjusted to 4500 nuclei/9 µL, and 4,500 nuclei were dispensed into each well of a 96-well plate. Glycerol was added to the leftover nuclei suspension for a final concentration of 25 % and nuclei were stored at -80°C. For tagmentation, 1 µL barcoded Tn5 transposomes were added using a BenchSmart™ 96 (Mettler Toledo), mixed five times and incubated for 60 min at 37°C with shaking (500 rpm). To inhibit the Tn5 reaction, 10 µL of 40 mM EDTA were added to each well with a BenchSmart™ 96 (Mettler Toledo) and the plate was incubated at 37°C for 15 min with shaking (500 rpm). Next, 20 µL 2 x sort buffer (2 %

BSA, 2 mM EDTA in PBS) were added using a BenchSmart™ 96 (Mettler Toledo). All wells were combined into a FACS tube and stained with 3 µM Draq7 (Cell Signaling). Using a SH800 (Sony), 20 nuclei were sorted per well into eight 96-well plates (total of 768 wells) containing 10.5 µL EB (25 pmol primer i7, 25 pmol primer i5, 200 ng BSA (Sigma)). Preparation of sort plates and all downstream pipetting steps were performed on a Biomek i7 Automated Workstation (Beckman Coulter). After addition of 1 µL 0.2% SDS, samples were incubated at 55°C for 7 min with shaking (500 rpm). We added 1 µL 12.5% Triton-X to each well to quench the SDS and 12.5 µL NEBNext High-Fidelity 2× PCR Master Mix (NEB). Samples were PCR-amplified (72°C 5 min, 98°C 30 s, (98°C 10 s, 63°C 30 s, 72°C 60 s) × 12 cycles, held at 12°C). After PCR, all wells were combined. Libraries were purified according to the MinElute PCR Purification Kit manual (Qiagen) using a vacuum manifold (QIAvac 24 plus, Qiagen) and size selection was performed with SPRI Beads (Beckmann Coulter, 0.55x and 1.5x). Libraries were purified one more time with SPRI Beads (Beckmann Coulter, 1.5x). Libraries were quantified using a Qubit fluorimeter (Life technologies) and the nucleosomal pattern was verified using a TapeStation (High Sensitivity D1000, Agilent). The library was sequenced on a HiSeq2500 sequencer (Illumina) using custom sequencing primers, 25% spike-in library and following read lengths: 50+43+40+50 (Read1+Index1+Index2+Read2).

Droplet-based 10x single cell ATAC-seq (scATAC-seq). 10x scATAC-seq protocol from 10x Genomics was followed: Chromium SingleCell ATAC ReagentKits UserGuide (CG000209, Rev A). Cryopreserved PBMC samples were thawed in 37°C water bath for 2 min and followed ‘PBMC thawing protocol’ in the UserGuide. After thawing cells, the pellets were resuspended again in 1 mL chilled PBS (with 0.04% PBS) and filtered with 50 µm CellTrics (04-0042-2317, Sysmex). The cells were centrifuged (300×g, 5 min, 4°C) and permeabilized with 100 µl of chilled lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 0.1% IGEPAL-CA630, 0.01% digitonin and 1% BSA). The samples were incubated on ice for 3 min and

resuspended with 1mL chilled wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20 and 1% BSA). After centrifugation (500×g, 5 min, 4°C), the pellets were resuspended in 100 µL of chilled Nuclei buffer (2000153, 10x Genomics). The nuclei concentration was adjusted between 3,000 to 7,000 per µl and 15,300 nuclei which targets 10,000 nuclei was used for the experiment. For pancreas tissue (pulverized as described above), approximately 31.7 mg of pulverized tissue was transferred to a LoBind tube (Eppendorf) and resuspended in 1 mL of cold permeabilization buffer (10mM Tris-HCL (pH 7.5), 10mM NaCl, 3mM $MgCl_2$, 0.1% Tween-20 (Sigma), 0.1% IGEPAL-CA630 (Sigma), 0.01% Digitonin (Promega) and 1% BSA (Proliant 7500804) in water) for 10 min on a rotator at 4°C. Permeabilized nuclei were filtered with 30 µm filter (CellTrics). Filtered nuclei were pelleted with a swinging bucket centrifuge (500×g, 5 min, 4°C; 5920R, Eppendorf) and resuspended in 1 mL Wash buffer (10mM Tris-HCL (pH 7.5), 10mM NaCl, 3mM $MgCl_2$, 0.1% Tween-20, and 1% BSA (Proliant 7500804) in molecular biology-grade water). Nuclei wash was repeated once. Next, washed nuclei were resuspended in 30 µL of 1X Nuclei Buffer (10X Genomics). Nuclei were counted using a hemocytometer, and finally the nuclei concentration was adjusted to 3,000 nuclei/µL. 15,360 nuclei were used as input for tagmentation.

Nuclei were diluted to 5 µl with 1X Nuclei buffer (10x Genomics) and, mixed with ATAC buffer (10x Genomics) and ATAC enzyme (10x Genomics) for tagmentation (60 min, 37°C). Single cell ATAC-seq libraries were generated using the (Chromium Chip E Single Cell ATAC kit (10x Genomics, 1000086) and indexes (Chromium i7 Multiplex Kit N, Set A, 10x Genomics, 1000084) following manufacturer instructions. Final libraries were quantified using a Qubit fluorimeter (Life technologies) and the nucleosomal pattern was verified using a TapeStation (High Sensitivity D1000, Agilent). Libraries were sequenced on a NextSeq 500 and HiSeq 4000 sequencer (Illumina) with following read lengths: 50+8+16+50 (Read1+Index1+Index2+Read2).

### 3.5.6 Single cell chromatin accessibility data processing

Prior to read alignment, we used trim_galore (version 0.4.4) to remove adapter sequences from reads using default parameters. We aligned reads to the hg19 reference genome using bwa mem[75] (version 0.7.17-r1188; '-M -C') and removed low mapping quality (MAPQ<30), secondary, unmapped, and mitochondrial reads using samtools[76] (version 1.10). To remove duplicate sequences on a per-barcode level, we used the MarkDuplicates tool from picard ('BARCODE_TAG'). For each tissue and snATAC-seq technology, we used log-transformed read depth distributions from each experiment to determine a threshold separating real cell barcodes from background noise. We used >500 total reads for combinatorial barcoding snATAC-seq and >2,300-4,000 total reads, as well as >0.3 fraction of reads in peaks for 10x snATAC-seq experiments (**Figure S3.7a**).

### 3.5.7 Single cell chromatin accessibility clustering

We identified snATAC-seq clusters using a previously described pipeline with a few modifications[9]. For each experiment, we first constructed a counts matrix consisting of read counts in 5 kb windows for each cell. Using scanpy[77] (version 1.4.4.post1), we normalized cells to a uniform read depth and log-transformed counts. We extracted highly variable (*hv*) windows ('min_mean=0.01, min_disp=0.25') and regressed out the total log-transformed read depth within *hv* windows (usable counts). We then merged datasets from the same tissue and performed PCA to extract the top 50 PCs. We used Harmony[78] (version 1.0) to correct the PCs for batch effects across experiments, using categorical covariates including donor-of-origin, biological sex, and snATAC-seq assay technology. We used the corrected components to construct a 30 nearest neighbor graph using the cosine metric, which we used for UMAP dimensionality reduction ('min_dist=0.3') and clustering with the Leiden algorithm[79] ('resolution=1.5').

Prior to combining cells across all tissues, we performed iterative clustering to identify and remove cells with low fraction of reads in peaks (using preliminary peaks called from data in bulk)

or low usable counts (islets: 948, pancreas: 2,588, PBMCs: 5,268 cells removed in total). Next, after removing low-quality cells and repeating the previous clustering steps, we sub-clustered the resulting main clusters at high resolution ('resolution=3.0') to identify sub-clusters containing potential doublets (islets: 886, pancreas: 4,495, PBMCs: 5,844 cells removed in total). We noted that these sub-clusters tended to have higher average usable counts, promoter usage, and accessibility at more than one marker gene promoter. After removing 20,029 low-quality or potential doublet cells, we performed a final round of clustering using experiments from all tissues, including tissue-of-origin as another covariate. We further removed 672 cells mapping to improbable cluster assignments (islet or pancreatic cells in PBMC clusters or vice versa). After all filters, we ended up with 131,554 cells mapping to 28 distinct clusters with consistent representation across samples from the same tissue (**Figure S3.7b**). We cataloged known marker genes for each cell type using a combination of literature search and PanglaoDB[80] (**Table 3.6**) and assessed gene accessibility (sum of read counts across each gene body) to assign labels to each cluster.

### 3.5.8 Single cell gene expression clustering

We compiled publicly available scRNA-seq datasets of peripheral blood (10x Genomics; v1 Chemistry – 3k, 6k, and 33k; v2 Chemistry – 4k and 8k, v3 Chemistry – 5k and 10k, v3.1 Chemistry – 5k, 10k single indexed, and 10k dual indexed) and pancreatic islets[81]. We re-processed each dataset using Cell Ranger (version 4.0.0) with the GRCh37 reference genome and removed cells with <500 genes expressed (non-zero counts). We extracted *hv* genes for PBMCs and pancreatic islets separately and merged both lists to obtain a single set of *hv* genes. For each sample, we used count matrices as input for scanpy[77] (version 1.4.4.post1), normalized counts for each cell to uniform read depth, log-transformed the normalized counts, and regressed out the log total counts for *hv* genes. We then merged all datasets and extracted the top 100 PCs using PCA. We used Harmony[78] (version 1.0) to correct PCs for covariates including the

experiment, donor, tissue, and biological sex. We constructed a 30 nearest neighbor graph using the cosine metric, performed UMAP dimensionality reduction ('min_dist=0.3'), and clustered with the Leiden algorithm[79] ('resolution=1.25'). We performed iterative clustering to remove 10,014 low quality and 5,286 potential doublet cells, leaving 90,495 cells for the cell type-resolved expression reference map. We used a combination of literature search and PanglaoDB[80] (**Table 3.6**) to assign labels to each cluster. For each cell type, we normalized aggregated reads from individual cells to derive TPM for each gene.

### 3.5.9 Comparing single cell chromatin accessibility and gene expression clusters

To compare cell types from snATAC-seq and scRNA-seq, we first derived gene expression t-statistics for each gene using linear regression models of log-transformed read count as a function of cluster assignment, donor-of-origin, and log sequencing depth, treating cells as individual data points. For each gene, we also derived chromatin accessibility t-statistics for promoter cCREs (see 3.5.11 Defining cell type-specific cCREs). For each scRNA-seq cluster, we extracted the top 100 most specific genes based on the gene expression t-statistic. Using a merged list of the most specific genes across all clusters, we compared gene expression and promoter accessibility t-statistics using Pearson correlation.

### 3.5.10 Cataloging cell type-resolved cCREs

We identified chromatin accessibility peaks with MACS2[82] (version 2.1.2) by calling peaks on aggregated reads from each cluster. In brief, we extracted reads from all cells within a given cluster, shifted reads aligned to the positive strand by +4 bp and reads aligned to the negative strand by -5 bp, and centered the reads. We then used MACS2 to call peaks ('--nomodel --keep-dup-all') and removed peaks overlapping ENCODE blacklisted regions[83,84]. We then merged peaks from all 28 clusters with bedtools[85] (version 2.26.0) to create a consistent set of 448,142 cCREs for subsequent analyses.

To compare accessible chromatin profiles from snATAC-seq to those from bulk ATAC-seq on FACS purified cell types, we reprocessed published ATAC-seq data from sorted pancreatic[86] and unstimulated immune cells[87]. We created pseudobulk profiles from the snATAC-seq data for each donor and cluster, retaining those that contained information from >50 cells. We then extracted read counts in the 448,142 cCREs for all sorted and pseudobulk profiles. We used PCA to extract the top 20 principal components and used UMAP for dimensionality reduction and visualization ('min_dist=0.5, n_neighbors=80').

### 3.5.11 Defining cell type-specific cCREs

To identify cCREs with accessibility levels most specific to each cluster, we used logistic regression models for each cCRE treating each cell as an individual data point. We performed separate regressions for each cluster, with binary cluster assignment and the covariates donor-of-origin and the log usable count as predictors and binary accessibility of the peak as the outcome, to calculate chromatin accessibility (CA) t-statistics. For a given cluster, we defined cCREs with activity most specific to that cluster by taking the top 1000 cCREs with the highest CA t-statistics, after first filtering out cCREs which also had high CA t-statistics for other clusters (cCRE cell type CA t-statistics>90$^{th}$ percentile in >2 other cell types). The cCREs were all significant after Bonferroni correction for the number of peaks ($P<1.1\times10^{-7}$) except for pancreatic CD8+ T (n=428 after correction), regulatory T (n=347) and memory CD8+ T (n=175). We then used GREAT[88] (version 3) to annotate gene ontology terms enriched in each set of cell type-specific cCREs compared to a background of all cCREs.

To assess whether cell type-specific cCREs tended to be close in proximity to genes with cell type-specific expression, we defined 100 kb windows around the midpoint of each cell type-specific cCRE and annotated genes with overlapping TSS. For each cell type that had a corresponding cluster in scRNA-seq, we compared whether genes around cell type-specific cCREs for that cell type had higher gene expression specificity scores than the rest of the cell

type-specific cCREs using two-sided Welch's t-tests. We collapsed cell type-specific cCREs for cell types with more than one state in snATAC-seq but only one state in scRNA-seq.

### 3.5.12 Single cell motif enrichment

We estimated TF motif enrichment z-scores for each cell using chromVAR[89] (version 1.5.0) by following the steps outlined in the user manual. First, we constructed a sparse binary matrix encoding read overlap with merged peaks for each cell. For each merged peak, we estimated the GC content bias to obtain a set of matched background peaks. To ensure a motif enrichment value for each cell, we did not apply any additional filters based on total reads or the fraction of reads in peaks. Next, using 580 TF motifs within the JASPAR 2018 CORE vertebrate (non-redundant) set[90], we computed GC bias-corrected enrichment z-scores (chromVAR deviation scores) for each cell. For each cell type, we considered a TF motif enriched if the average z-score across cells was greater than 2. We used the TFClass database[91] (http://tfclass.bioinf.med.uni-goettingen.de/) to group enriched TF motifs into structural sub-families. We determined the expression of all TFs within the subfamily in each cell type identified in scRNA-seq and considered TFs expressed in a cell type with TPM>1.

### 3.5.13 Single cell co-accessibility

We used Cicero[92] (version 1.3.3) to calculate co-accessibility scores between pairs of peaks for each cluster. As in the single cell motif enrichment analysis, we started from a sparse binary matrix. For each cluster, we only retained merged peaks that overlapped peaks from the cluster. Within each cluster, we aggregated cells based on the 50 nearest neighbors and used cicero to calculate co-accessibility scores, using a 1 Mb window size and a distance constraint of 500 kb. We then defined promoters as ±500 bp from the TSS of protein coding transcripts from GENCODE v19[93] to annotate co-accessibility links between gene promoters and distal cCREs (non-promoter cCREs).

To assess whether genes with co-accessible links between the promoter and distal cCREs (co-accessible genes; co-accessibility score>0.05) were expressed more often than non-co-accessible genes (co-accessibility score<0) within each cell type, we separated co-accessible links into bins based on the distance between the gene promoter and distal cCRE. Within each bin, we then compared the fraction of genes expressed in the cell type (TPM>1 from scRNA-seq) between co-accessible and non-co-accessible genes using 2-sided Fisher's exact tests. We collapsed co-accessible links for cell types with more than one state in snATAC-seq but only one state in scRNA-seq (alpha, beta, and delta cells). No comparison was made for pancreatic CD8+ T-cells, which did not have a corresponding cluster in scRNA-seq.

### 3.5.14 GWAS enrichment analyses

We used LD score regression[30,94,95] (version 1.0.1) to calculate genome-wide enrichment z-scores for 32 diseases and traits including T1D. We obtained GWAS summary statistics for autoimmune and inflammatory diseases (immune-related)[34–42], other diseases[43–51], and quantitative endophenotypes[54–63], and where necessary, we filled in variant IDs and alleles. Using 'munge_sumstats.py', we converted summary statistics to the LD score regression standard format. For each cluster, we considered overlap with chromatin accessibility peaks as a binary annotation for variants. Then, we computed annotation-specific LD scores by following the instructions for creating partitioned LD scores. We estimated enrichment coefficient z-scores for each annotation relative to the background annotations in the baseline-LD model (version 2.2). Using the enrichment z-scores, we computed two-sided p-values to assess significance and corrected for multiple tests using the Benjamini-Hochberg procedure. We also calculated GWAS enrichment z-scores for T1D using a version of the meta-analysis excluding the Immunochip cohorts and observed highly similar enrichment results.

From the full GWAS summary statistics, we first extracted variants with MAF>0.05 and calculated approximate Bayes factors[96] for each variant, assuming prior variance in allelic effects

= 0.04. We then used fgwas[97] (version 0.3.6) to estimate T1D enrichment for common variants (MAF>0.05) within cell type-specific cCREs using an average window size of 1 Mb also including annotations for coding exons, 3'/5'UTR regions and 1 kb upstream of the transcription start site (TSS) from GENCODE in each model. We considered cell type annotations enriched where ln(95% CI lower bound) >0 and depleted where ln(95% CI upper bound) <0.

### 3.5.15 Annotating cell type mechanisms of variants at fine mapped signals

We compared the proportion of credible set variants with PPA>0.50 overlapping a cCRE compared to other credible set variants using a Chi-square test. Among credible set variants in cCREs, we further compared the proportion of credible set variants with PPA>0.50 in a cCRE co-accessible with a gene promoter compared to other credible set variants using a two-sided Fisher's exact test.

For each T1D signal, we calculated the cumulative posterior probability of all credible set variants overlapping cCREs active in T-cells, monocytes, plasmacytoid dendritic cells, beta cells, acinar cells and ductal cells. For each signal overlapping cCREs, we annotated genes within 1 Mb of the index variant that were (i) expressed in the same cell type(s) (TPM>1 from scRNA-seq) and (ii) co-accessible with a cCRE harboring a credible set variant with PPA>0.01.

### 3.5.16 Luciferase reporter assays

We tested for allelic differences in enhancer activity at rs7795896 using multiple constructs. First, we cloned a 180 bp sequence of human DNA (Coriell) containing the reference or alternate allele into the luciferase reporter vector pGL4.23 (Promega) in the forward direction using the restriction enzymes SacI and KpnI. Second, we cloned a larger 594 bp sequence of human DNA (Coriell) containing the rs7795896 reference allele corresponding to the coordinates of the ductal-specific cCRE into pGL4.23 in the forward direction using the restriction enzymes SacI and KpnI. We introduced the alternate allele via SDM using the NEB Q5 Site Directed Mutagenesis kit (New England Biolabs) on 1ng plasmid containing the reference allele and

primers designed using the NEBaseChanger v.1.2.8 software. Sequence identity for all plasmids was confirmed with Sanger sequencing using the RV3 primer. Cloning primers were designed using Primer3 version 0.4.0. Primer sequences for cloning and SDM are listed in **Table 3.9**.

We grew Capan-1 cells (ATCC), a model for ductal cells[98], to approximately 70% confluency according to ATCC culture recommendations in 6-well or 24-well plates and fed complete growth media the day before transfection. For the 180bp construct, 2500 ng experimental or empty (pGL4.23) vector was co-transfected with 50 ng pRL-SV40 per sample using Lipofectamine 3000 (Invitrogen) into Capan-1 cells grown in a 6-well plate. For the 594 bp construct, 500 ng experimental or empty vector was co-transfected with 10 ng pRL-TK per sample using Lipofectamine 3000 (Invitrogen) into Capan-1 cells grown in a 24-well plate. For all experiments, 48 hours post-transfection samples were assayed using the Dual-Glo Luciferase Assay System (Promega). We normalized Firefly:Renilla ratios with respect to the empty vector and used two-sided Student's T-tests to compare luciferase activity between the two alleles.

### 3.5.17 Electrophoretic mobility shift assay

We ordered double-stranded 5' biotinylated and corresponding unlabeled (cold) oligonucleotides of 16 bp centered on rs7795896 with the reference and alternate alleles from Integrated DNA Technologies. Oligo sequences are listed in Supplementary Table 11. We performed EMSA using the LightShift Chemiluminescent EMSA kit (Thermo Fisher) according to manufacturer's instructions with the following adjustments: 100 fmol of biotinylated duplex probe per reaction, and 20 pmol of the same-allele non-biotinylated duplex "cold" probe in competition reactions (200× molar excess of the biotin probe). We used the NE-PER Nuclear Protein and Cytoplasmic Extraction Reagents (Thermo Fisher) kit to extract nuclear protein from Capan-1 cells and used 2uL of nuclear extract per binding reaction, corresponding to approximately 5-15ug of nuclear protein per reaction. We quantified bound and free probe (unbound) band intensity using ImageJ (v.1.53) and calculated the ratio of bound to unbound intensity. We then averaged bound ratios for replicates of each allele and compared ratios between alleles.

### 3.5.18 CRISPR inactivation of enhancer element

We maintained HEK293T cells in DMEM containing 100 units/mL penicillin and 100 mg/mL streptomycin sulfate supplemented with 10% fetal bovine serum. To generate CRISPRi lentiviral expression vectors, we designed guide RNA sequences to target the enhancer containing rs7795896 or the *CFTR* promoter. These guide RNAs, as well as a non-targeting control guide RNA, were placed downstream of the human U6 promoter in the pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-Puro backbone (Addgene, plasmid #71236). Targeting guide RNAs were designed using Benchling and selected to maximize both on-target binding[99] and guide specificity[100]. The non-targeting control guide RNA was selected from a previously validated genome-wide library[101]. Guide RNA sequences and targeted regions are listed in **Table 3.9**. Higher scores indicate greater on-target binding and specificity.

We generated high-titer lentiviral supernatants by co-transfection of the resulting plasmid and lentiviral packaging constructs into HEK293T cells. Specifically, we co-transfected CRISPRi vectors with the pCMV-R8.74 (Addgene, #22036) and pMD2.G (Addgene, #12259) expression plasmids into HEK293T cells using a 1mg/ml PEI solution (Polysciences). We collected lentiviral supernatants at 48 and 72 hours after transfection and concentrated lentiviruses by ultracentrifugation for 120 min at 19,500 rpm using a Beckman SW28 ultracentrifuge rotor at 4°C. Lentiviral titers were subsequently determined using a qPCR Lentivirus Titer Kit (Abm Bio), and aliquots were stored at -80°C.

We obtained Capan-1 pancreatic ductal adenocarcinoma cell lines from ATCC and cultured using Iscove's Modified Dulbecco's Media with 20% fetal bovine serum, 100 units/mL penicillin, and 100 mg/mL streptomycin sulfate. 24 hours prior to transduction, we passaged cells into a 12-well plate at a density of 100,000 cells per well. The following day, we added fresh medium containing 8µg/mL polybrene and concentrated CRISPRi lentivirus at an MOI of 40 to each well. For each condition (1 non-targeting guide RNA, 3 enhancer-targeting guide RNAs, and

1 promoter-targeting guide RNA) we transduced 3 wells for a total of 15 wells. We additionally included 3 wells of mock-transduced cells without lentivirus. We incubated the cells at 37°C for 30 minutes and then spun them in a centrifuge for 1 hour at 30°C at 950×g. 6 hours later, we replaced viral medium with fresh base culture medium for cell recovery. After 48 hours, we replaced medium daily with the addition of 1µg/mL puromycin for an additional 72 hours, at which point all mock-transduced cells were killed. We reduced the concentration of puromycin to 0.5 µg/mL and cultured cells with daily medium changes for an additional week before passaging each cell line into a 48-well plate at a density of approximately 100,000 cells per well. The following morning, we harvested cells from each condition and isolated RNA using the RNeasy® Micro Kit (Qiagen) according to the manufacturer's instructions.

For qRT-PCR, we performed cDNA synthesis using the iScript™ cDNA Synthesis Kit (Bio-Rad) and 250 ng of isolated RNA per reaction. We performed qRT-PCR reactions in triplicate with 5 ng of template cDNA per reaction using a CFX96™ Real-Time PCR Detection System and the iQ™ SYBR® Green Supermix (Bio-Rad). We used PCR of the TATA binding protein (TBP) coding sequence as an internal control, quantified relative expression via double delta CT analysis, and compared relative expression using ANCOVA (enhancer inactivation versus non-targeting control, enhancer sgRNA as a covariate) or a 2-sided Student's t-test (promoter inactivation versus non-targeting control). Genes with CT values greater than 34 were considered as not expressed. We also evaluated changes in expression of the puromycin resistance gene and the dCAS9 gene as additional controls. For eukaryotic genes, each primer pair was designed to span an exon-exon junction. Primers used for qPCR are listed in **Table 3.9**.

### 3.5.19 Colocalization and deconvolution of the pancreas *CFTR* eQTL

We obtained GTEx v7[13] eQTL summary statistics for pancreas tissue from 220 samples and used effect size (beta) and standard error estimates from the regression model for *CFTR* expression to calculate approximate Bayes factors[96] for each variant, assuming prior variance in

allelic effects = 0.04. We considered all variants in a 500 kb window around the T1D index variant at *CFTR* (rs7795896) tested in both the GWAS and eQTL datasets and used coloc[102] (version 4.0.4) to calculate the probability that the variants driving T1D and eQTL signals were shared, using prior probabilities $PP_{T1D}=1\times10^{-4}$, $PP_{eQTL}=1\times10^{-4}$, and $PP_{shared}=1\times10^{-5}$. We considered the T1D and *CFTR* eQTL signals colocalized based on the probability that they were shared ($PP_{shared}$) >0.9. We applied eCAVIAR[103] (version 2.2) using variants in a 500 kb window tested for both T1D and eQTL association using LD calculated from EUR samples in 1000 Genomes[27] and considered variants with CLPP>0.01 to be candidate causal variants for a shared signal.

We used MuSiC[104] (version 0.1.1) to estimate the proportions of major pancreatic cell types (acinar, duct, stellate, alpha, beta, delta, gamma) in each GTEx v7 pancreas sample . As input, we used raw count matrices from scRNA-seq of pancreatic cell types with labels from the gene expression reference map and GTEx v7 pancreas samples. For each cell type, we used the proportion as an interaction term and constructed linear models of TMM-normalized *CFTR* expression as a function of the interaction between genotype dosage and cell type proportion, accounting for covariates used by GTEx including sex, sequencing platform, 3 genotype PCs, and 28 inferred PCs from the expression data. From the original 30 inferred PCs, we excluded inferred PCs 2 and 3 because they were highly correlated with acinar cell proportion (Spearman's $\rho>0.7$). No remaining PCs were highly correlated (Spearman's $\rho<0.3$) with the proportions of other cell types.

### 3.5.20 Phenotype associations at T1D signals

We tested association of the T1D index variant at *CFTR* (rs7795896) for pancreatic and autoimmune disease phenotypes. For acute pancreatitis, chronic pancreatitis, and pancreatic cancer, we used inverse variance weighted meta-analysis to combine SAIGE analysis results from the UK biobank[29] (PheCodes 577.1, 577.2, and 157) and FinnGen (K11_ACUTPANC, K11_CHRONPANC, C3_PANCREAS_EXALLC). As mutations that cause cystic fibrosis (CF)

map to this locus, which are risk factors for pancreatitis and pancreatic cancer, we determined the impact of the most common CF mutation F508del/rs199826652 on the association results for rs7795896. For T1D, we tested for association of rs7795896 conditional on F508del/rs199826652 in all cohorts except for FinnGen and observed no evidence for a difference in T1D association. For pancreatitis and pancreatic cancer, we identified F508del/rs199826652 carriers in UK Biobank and repeated the association analysis for these phenotypes in UK biobank data after removing these individuals and observed no evidence of a change in the effect of rs7795896.

We identified T1D signals where the risk allele had at least nominal association ($P<0.05$) with increased risk of acute pancreatitis, chronic pancreatitis, or pancreatic cancer. We then tested whether these T1D signals had a difference in cPPA in exocrine cell cCREs or T cell cCREs compared to other T1D signals using a two-sided Student's T-test.

## 3.6 Figures

**Figure 3.1. Genome-wide association and fine mapping identifies novel signals for T1D risk.**

(a) Genome-wide T1D association (two-sided -$\log_{10}$ transformed p-values from logistic regression meta-analysis, unadjusted for multiple comparisons). Novel loci are colored (±250 kb of the index variant) and labeled based on the nearest gene. Dotted line indicates genome-wide significance ($P=5\times10^{-8}$). (b) Bayes factors (natural log-transformed) for independent association signals at the known *PTPN2* locus (left) and the novel *BCL11A* locus (right). Variants are colored based on linkage disequilibrium ($r^2$) with the index variant for each signal. (c) Breakdown of 136 T1D risk signals, including 92 main signals (59 known and 33 novel), and 44 independent signals (38 at known and 6 at novel loci). (d) Number of signals per locus (top), 99% credible set variants from fine mapping (middle), and variants with posterior probability of association (PPA) at various thresholds (bottom).

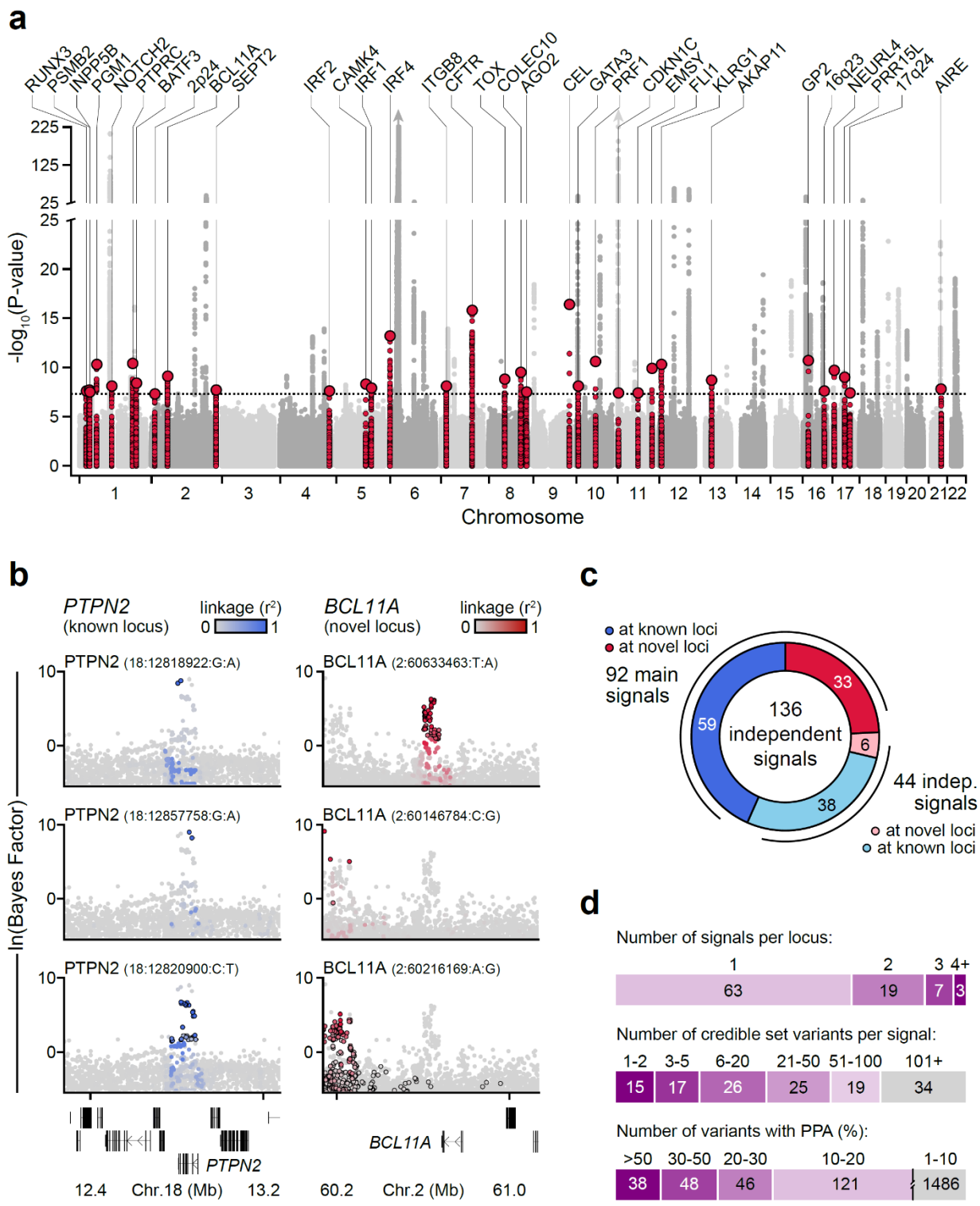

a
RUNX3
PSMB2
INPP5B
PGM1
NOTCH2
PTPRC
BATF3
2p24
BCL11A
SEPT2
IRF2
CAMK4
IRF1
IRF4
ITGB8
CFTR
TOX
COLEC10
AGO2
CEL
GATA3
PRF1
CDKN1C
EMSY
FLI1
KLRG1
AKAP11
GP2
16q23
NEURL4
PRR15L
17q24
AIRE
-log10(P-value)
Chromosome
b
PTPN2
(known locus)
linkage (r²)
BCL11A
(novel locus)
ln(Bayes Factor)
PTPN2 (18:12818922:G:A)
BCL11A (2:60633463:T:A)
PTPN2 (18:12857758:G:A)
BCL11A (2:60146784:C:G)
PTPN2 (18:12820900:C:T)
BCL11A (2:60216169:A:G)
12.4
Chr.18 (Mb)
13.2
60.2
Chr.2 (Mb)
61.0
c
at known loci
at novel loci
92 main signals
136 independent signals
44 indep. signals
d
Number of signals per locus:
Number of credible set variants per signal:
Number of variants with PPA (%):

**Figure 3.2. Reference map of 131,554 single cell chromatin accessibility profiles from T1D-relevant tissues.**

(a) Leiden clustering of single cell accessible chromatin profiles from 131,554 cells. Cells are plotted on the first two UMAP components, clusters are grouped into categories of cell types, and the number of cells per cluster are in parentheses. (b) Relative gene accessibility (column-normalized chromatin accessibility reads in gene bodies) showing examples of marker genes used to identify cluster labels. Aggregated chromatin accessibility profiles in a 50 kb window around selected marker genes (bottom). (c) Relative accessibility (row-normalized) for 25,436 cCREs most specific to each cluster (left), and enriched gene ontology terms for cCREs specific to activated stellate, pancreatic macrophages, ductal, and acinar cells (right). (d) Single cell motif enrichment z-scores (left) and expression of motif subfamily members (right) for examples of TFs with lineage-, cell type-, or cell state-specific motif enrichment and expression. TFs with matching motif enrichment and expression are highlighted. (e) Co-accessibility between *AQP1* and cCREs in ductal cells (left) and *CEL* and cCREs in acinar cells (right).

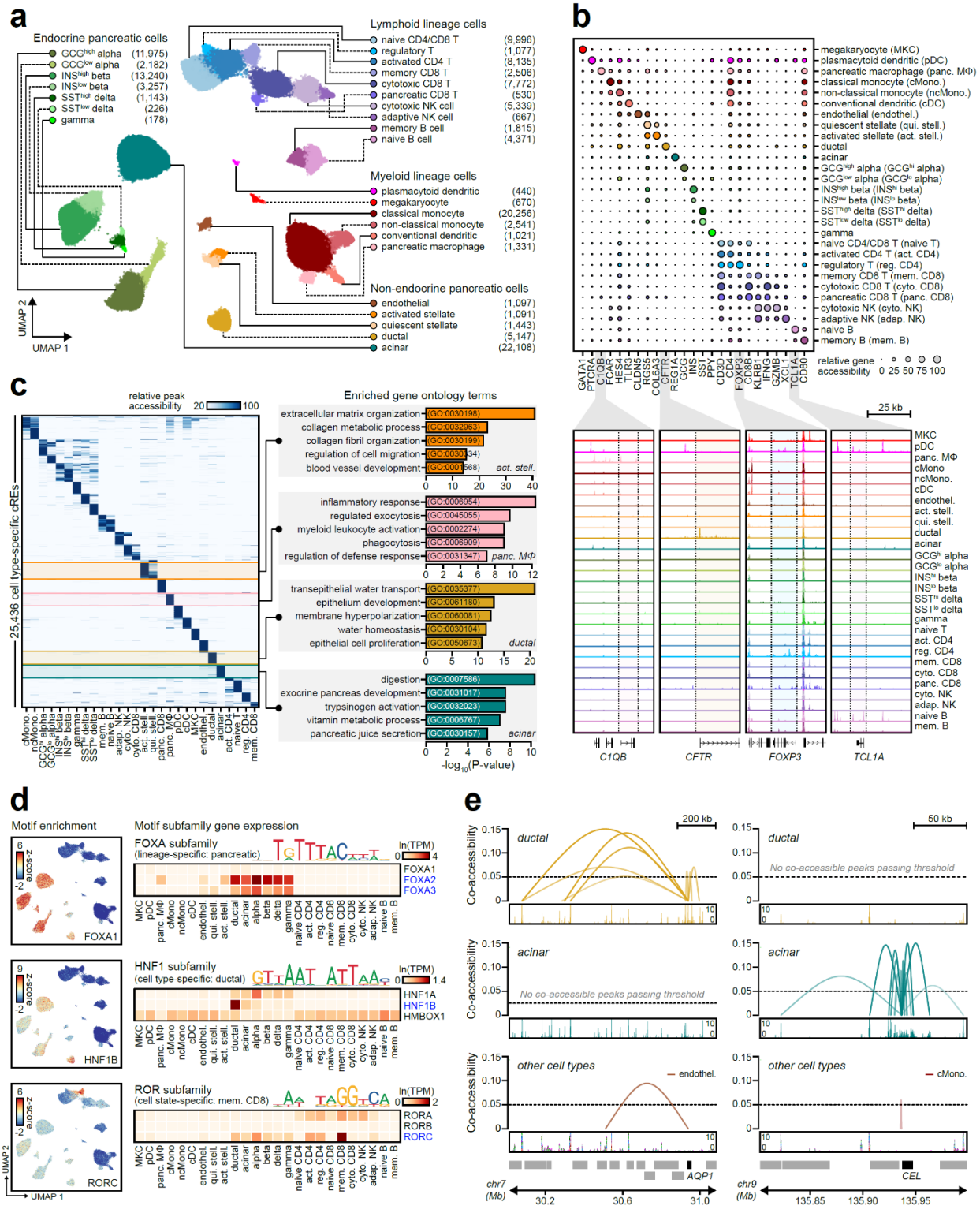

a
Endocrine pancreatic cells
GCG^high alpha (11,975)
GCG^low alpha (2,182)
INS^high beta (13,240)
INS^low beta (3,257)
SST^high delta (1,143)
SST^low delta (226)
gamma (178)
Lymphoid lineage cells
naive CD4/CD8 T (9,996)
regulatory T (1,077)
activated CD4 T (8,135)
memory CD8 T (2,506)
cytotoxic CD8 T (7,772)
pancreatic CD8 T (530)
cytotoxic NK cell (5,339)
adaptive NK cell (667)
memory B cell (1,815)
naive B cell (4,371)
Myeloid lineage cells
plasmacytoid dendritic (440)
megakaryocyte (670)
classical monocyte (20,256)
non-classical monocyte (2,541)
conventional dendritic (1,021)
pancreatic macrophage (1,331)
Non-endocrine pancreatic cells
endothelial (1,097)
activated stellate (1,091)
quiescent stellate (1,443)
ductal (5,147)
acinar (22,108)
UMAP 2
UMAP 1
b
megakaryocyte (MKC)
plasmacytoid dendritic (pDC)
pancreatic macrophage (panc. MΦ)
classical monocyte (cMono.)
non-classical monocyte (ncMono.)
conventional dendritic (cDC)
endothelial (endothel.)
quiescent stellate (qui. stell.)
activated stellate (act. stell.)
ductal
acinar
GCG^high alpha (GCG^hi alpha)
GCG^low alpha (GCG^lo alpha)
INS^high beta (INS^hi beta)
INS^low beta (INS^lo beta)
SST^high delta (SST^hi delta)
SST^low delta (SST^lo delta)
gamma
naive CD4/CD8 T (naive T)
activated CD4 T (act. CD4)
regulatory T (reg. CD4)
memory CD8 T (mem. CD8)
cytotoxic CD8 T (cyto. CD8)
pancreatic CD8 T (panc. CD8)
cytotoxic NK (cyto. NK)
adaptive NK (adap. NK)
naive B
memory B (mem. B)
GATA1 PTCRA C1QB FCER1A HES4 TLR3 CLDN5 RGS5 COL6A3 CFTR REG1A GCG INS SST PPY CD3D CD4 FOXP3 CD8B KLRB1 IFNG GZMB XCL1 TCL1A CD60
relative gene accessibility 0 25 50 75 100
25 kb
C1QB
CFTR
FOXP3
TCL1A
c
relative peak accessibility 20 100
25,436 cell type-specific cREs
Enriched gene ontology terms
extracellular matrix organization (GO:0030198)
collagen metabolic process (GO:0032963)
collagen fibril organization (GO:0030199)
regulation of cell migration (GO:0030334)
blood vessel development (GO:0001568)
act. stell.
inflammatory response (GO:0006954)
regulated exocytosis (GO:0045055)
myeloid leukocyte activation (GO:0002274)
phagocytosis (GO:0006909)
regulation of defense response (GO:0031347)
panc. MΦ
transepithelial water transport (GO:0035377)
epithelium development (GO:0061180)
membrane hyperpolarization (GO:0060081)
water homeostasis (GO:0030104)
epithelial cell proliferation (GO:0050673)
ductal
digestion (GO:0007586)
exocrine pancreas development (GO:0031017)
trypsinogen activation (GO:0032023)
vitamin metabolic process (GO:0006767)
pancreatic juice secretion (GO:0030157)
acinar
-log10(P-value)
d
Motif enrichment
Motif subfamily gene expression
FOXA subfamily
(lineage-specific: pancreatic)
ln(TPM)
FOXA1
FOXA2
FOXA3
HNF1 subfamily
(cell type-specific: ductal)
HNF1A
HNF1B
HMBOX1
ROR subfamily
(cell state-specific: mem. CD8)
RORA
RORB
RORC
z-score
UMAP 2
UMAP 1
e
ductal
acinar
other cell types
endothel.
cMono.
No co-accessible peaks passing threshold
Co-accessibility
200 kb
50 kb
AQP1
CEL
chr7 (Mb)
chr9 (Mb)

**Figure 3.3. Cell type-specific enrichment and mechanisms of T1D risk variants.**

(a) Stratified LD score regression coefficient z-scores for autoimmune and inflammatory diseases (top), other diseases (middle), and non-disease quantitative endophenotypes (bottom) for cCREs active in immune and pancreatic cell types. ***FDR<0.001 **FDR<0.01 *FDR<0.1. (b) T1D enrichment within cell type-specific cCREs. Labeled cell types have positive enrichment and 95% CI lower bound >0. Data are natural log enrichment ± 95% CI from fgwas. (c) T1D signals with highest cumulative PPA (cPPA) in cCREs for disease-enriched cell types (>0.20 cPPA for T-cells and monocytes, >0.10 cPPA for other groups), and >0.01 cPPA away from the next closest group (top). Column-normalized expression for genes with TPM>1 in the highlighted cell type(s) and within ±500 kb of the index variant. Genes co-accessible with cCREs containing risk variants are annotated in rectangles (bottom). (d) The *GP2* locus contains three variants in a distal cCRE co-accessible with the GP2 promoter in acinar cells which account for the majority of the causal probability (cPPA=0.98). Chromatin accessibility at both the distal cCRE and the *GP2* promoter is highly specific to acinar cells. (e) Variant rs72802342 at the *CTRB1*/*2*/*BCAR1* locus overlaps a distal cCRE co-accessible with the *CTRB2* and *CTRB1* promoters in acinar cells. Chromatin accessibility at the *CTRB1* and *CTRB2* promoters is highly specific to acinar cells. Variants contained in the 99% credible set are circled in black.

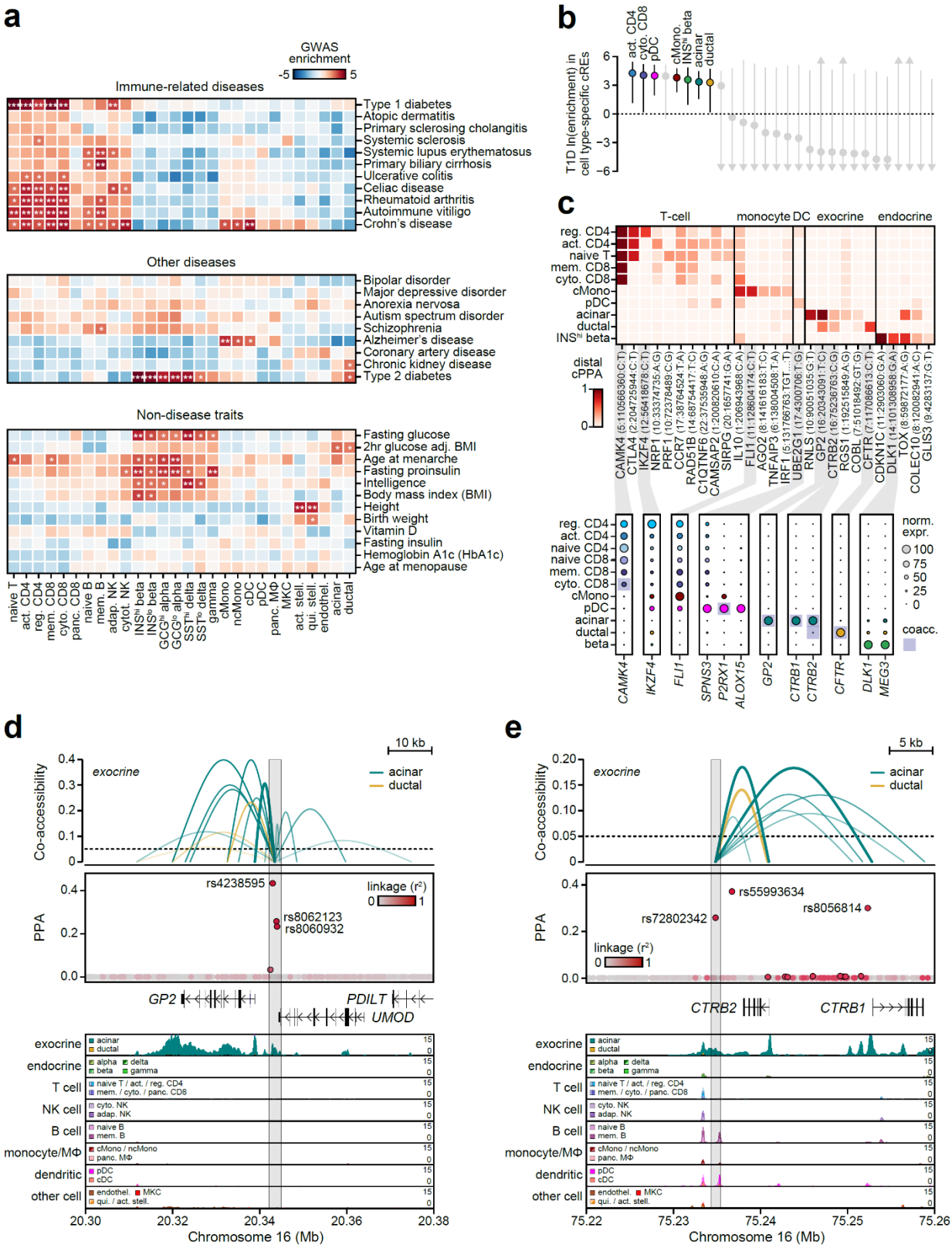
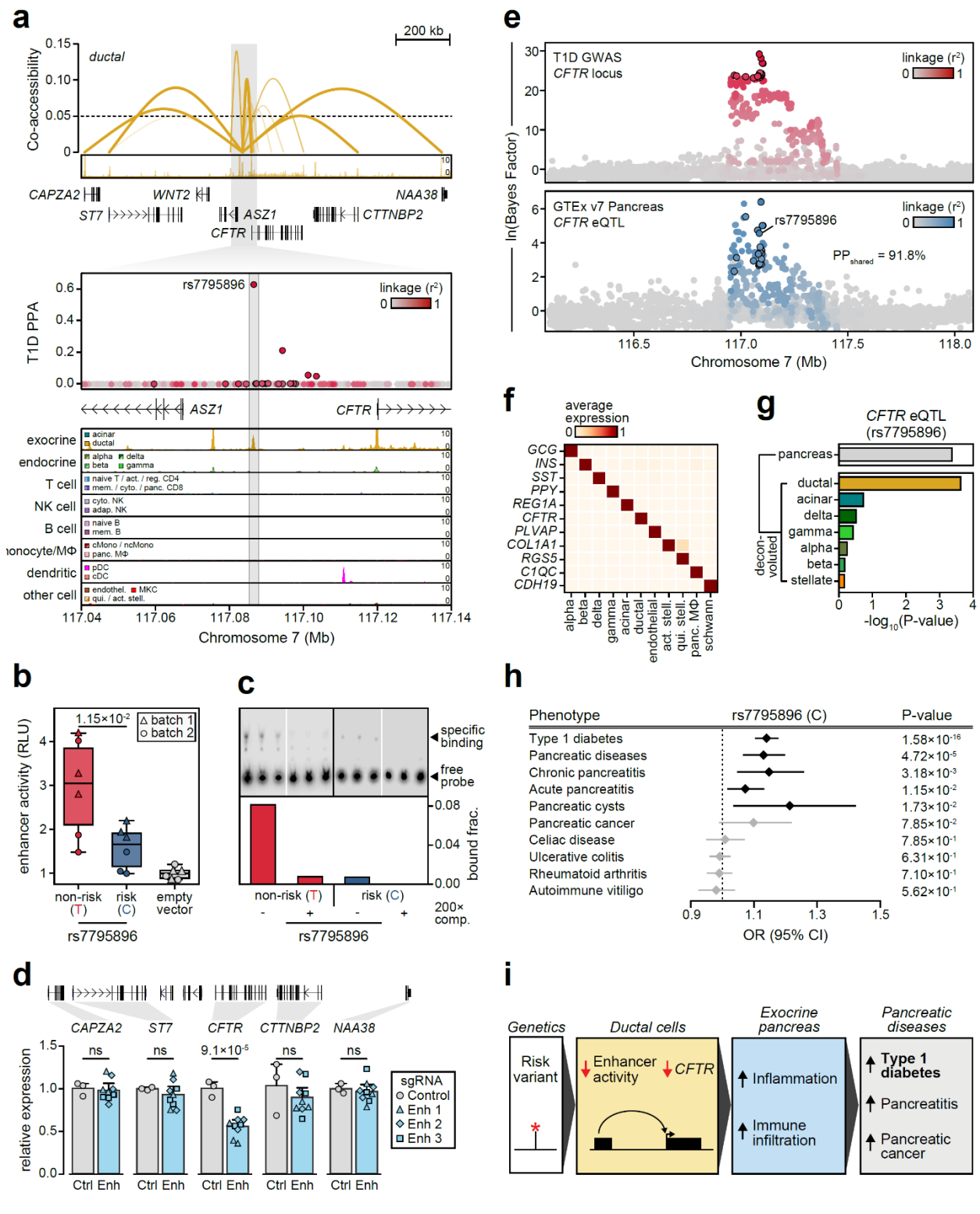

a
GWAS enrichment
-5
5
Immune-related diseases
Type 1 diabetes
Atopic dermatitis
Primary sclerosing cholangitis
Systemic sclerosis
Systemic lupus erythematosus
Primary biliary cirrhosis
Ulcerative colitis
Celiac disease
Rheumatoid arthritis
Autoimmune vitiligo
Crohn's disease
Other diseases
Bipolar disorder
Major depressive disorder
Anorexia nervosa
Autism spectrum disorder
Schizophrenia
Alzheimer's disease
Coronary artery disease
Chronic kidney disease
Type 2 diabetes
Non-disease traits
Fasting glucose
2hr glucose adj. BMI
Age at menarche
Fasting proinsulin
Intelligence
Body mass index (BMI)
Height
Birth weight
Vitamin D
Fasting insulin
Hemoglobin A1c (HbA1c)
Age at menopause
naive T
act. CD4
reg. CD4
mem. CD8
cyto. CD8
panc. CD8
naive B
mem. B
adap. NK
cytot. NK
INShi beta
INSlo beta
GCGhi alpha
GCGlo alpha
SSThi delta
SSTlo delta
gamma
cMono
ncMono
cDC
pDC
panc. MΦ
MKC
act. stell.
qui. stell.
endothel.
acinar
ductal
b
T1D ln(enrichment) in cell type-specific cREs
6
3
0
−3
−6
act. CD4
cyto. CD8
pDC
cMono
INShi beta
acinar
ductal
c
T-cell
monocyte
DC
exocrine
endocrine
reg. CD4
act. CD4
naive T
mem. CD8
cyto. CD8
cMono
pDC
acinar
ductal
INShi beta
distal cPPA
1
0
CAMK4 (5:110566360:C:T)
CTLA4 (2:204725944:C:T)
IKZF4 (12:56418678:C:T)
NRP1 (10:33374735:A:G)
PRF1 (10:72378489:C:G)
CCR7 (17:38764524:T:A)
RAD51B (14:68754417:T:C)
C1QTNF6 (22:37535948:A:G)
CAMSAP2 (1:200820610:C:A)
SIRPG (20:1657741:G:A)
IL10 (1:206943968:C:A)
FLI1 (11:128604174:C:T)
AGO2 (8:141616183:T:C)
TNFAIP3 (6:138004508:T:A)
IRF1 (5:131766763:TGT...:T)
UBE2G1 (17:43000706:T:G)
RNLS (10:90051035:G:T)
GP2 (16:20343091:T:C)
CTRB2 (16:75236763:C:G)
RGS1 (1:192515849:A:G)
COBL (7:51018492:GT:G)
CFTR (7:117086613:C:T)
CDKN1C (11:2903060:G:A)
DLK1 (14:101308958:G:A)
TOX (8:59872177:A:G)
COLEC10 (8:120082941:A:C)
GLIS3 (9:4283137:G:T)
reg. CD4
act. CD4
naive CD4
naive CD8
mem. CD8
cyto. CD8
cMono
pDC
acinar
ductal
beta
norm. expr.
100
75
50
25
0
coacc.
CAMK4
IKZF4
FLI1
SPNS3
P2RX1
ALOX15
GP2
CTRB1
CTRB2
CFTR
DLK1
MEG3
d
10 kb
Co-accessibility
exocrine
acinar
ductal
PPA
rs4238595
rs8062123
rs8060932
linkage (r²)
0
1
GP2
PDILT
UMOD
exocrine
endocrine
T cell
NK cell
B cell
monocyte/MΦ
dendritic
other cell
acinar
ductal
alpha
beta
delta
gamma
naive T / act. / reg. CD4
mem. / cyto. / panc. CD8
cyto. NK
adap. NK
naive B
mem. B
cMono / ncMono
panc. MΦ
pDC
cDC
endothel.
MKC
qui. / act. stell.
20.30
20.32
20.34
20.36
20.38
Chromosome 16 (Mb)
e
5 kb
Co-accessibility
exocrine
acinar
ductal
PPA
rs55993634
rs8056814
rs72802342
linkage (r²)
0
1
CTRB2
CTRB1
exocrine
endocrine
T cell
NK cell
B cell
monocyte/MΦ
dendritic
other cell
75.22
75.23
75.24
75.25
75.26
Chromosome 16 (Mb)

**Figure 3.4. Fine-mapped T1D variant regulates CFTR in pancreatic ductal cells.**

(a) Variant rs7795896 at the *CFTR* locus mapped in a cCRE co-accessible with *CFTR* and other genes. Zoomed-in view shows the cCRE is ductal-specific. (b) Relative luciferase units (RLU) for reporter containing 594 bp sequence surrounding rs7795896 in Capan-1 (n=6; 2 batches × 3 transfections). P-value by ANCOVA with batch as a covariate. (c) Electrophoretic mobility shift assay with nuclear extract from Capan-1 using probes for rs7795896 alleles, with or without 200× unlabeled competitor probe (200× comp.). Quantification of the bound fraction (specific binding / free probe) averaged across three technical replicates for each allele. (b) Expression of genes co-accessible with the distal cCRE in CRISPR-inactivated enhancer (Enh; n=9, 3 sgRNAs × 3 biological replicates) compared to non-targeting control (Ctrl; n=3 biological replicates) in Capan-1. Data are mean ± 95% CI. P-values by two-sided ANCOVA with batch as a covariate; ns, not significant. (e) Bayesian colocalization of T1D signal and *CFTR* pancreas eQTL. Variants in the T1D credible set are circled. (f) Expression of pancreatic cell type marker genes from scRNA-seq. (g) Two-sided cell type deconvolution p-values ($-\log_{10}$ transformed; from linear regression interaction between dosage and cell type proportion) of the *CFTR* pancreas eQTL. (h) rs7795896 GWAS association for T1D (from full meta-analysis), pancreatic disease, and autoimmune disease. Points and lines represent odds ratio estimates and 95% CI. Two-sided p-values from GWAS meta-analysis are unadjusted for multiple comparisons. (i) Variants regulating genes with specialized exocrine pancreas function influence T1D risk, and we hypothesize these effects are mediated through inflammation and immune infiltration.


a
200 kb
Co-accessibility
ductal
CAPZA2
WNT2
NAA38
ST7
ASZ1
CTTNBP2
CFTR
T1D PPA
rs7795896
linkage (r²)
exocrine
endocrine
T cell
NK cell
B cell
dendritic
other cell
Chromosome 7 (Mb)
e
T1D GWAS
CFTR locus
GTEx v7 Pancreas
CFTR eQTL
PP shared = 91.8%
ln(Bayes Factor)
Chromosome 7 (Mb)
f
average expression
GCG
INS
SST
PPY
REG1A
CFTR
PLVAP
COL1A1
RGS5
C1QC
CDH19
g
CFTR eQTL (rs7795896)
pancreas
ductal
acinar
delta
gamma
alpha
beta
stellate
deconvoluted
-log10(P-value)
b
1.15×10-2
batch 1
batch 2
enhancer activity (RLU)
non-risk (T)
risk (C)
empty vector
rs7795896
c
specific binding
free probe
bound frac.
200× comp.
h
Phenotype
rs7795896 (C)
P-value
Type 1 diabetes
Pancreatic diseases
Chronic pancreatitis
Acute pancreatitis
Pancreatic cysts
Pancreatic cancer
Celiac disease
Ulcerative colitis
Rheumatoid arthritis
Autoimmune vitiligo
OR (95% CI)
d
relative expression
ns
9.1×10-5
sgRNA
Control
Enh 1
Enh 2
Enh 3
Ctrl Enh
i
Genetics
Ductal cells
Exocrine pancreas
Pancreatic diseases
Risk variant
Enhancer activity
CFTR
Inflammation
Immune infiltration
Type 1 diabetes
Pancreatitis
Pancreatic cancer

## 3.7 Supplementary Figures

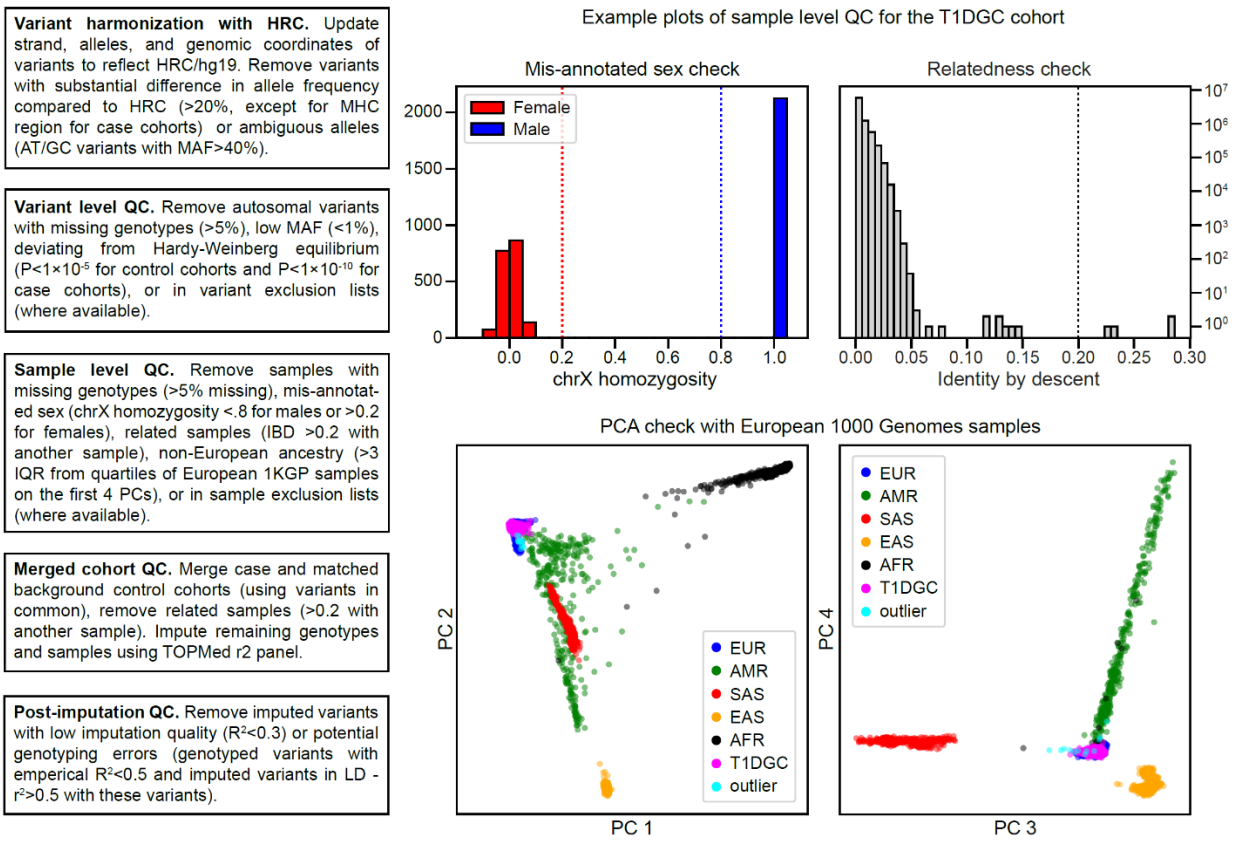



**Figure S3.1. Flowchart of quality control steps for genotype imputation.**

Flowchart (left) showing the steps of cohort-level, merged cohort, and post-imputation QC steps taken to filter out potential low quality variants and samples for non-biobank (non-UK biobank and non-FinnGen) samples. Example plots (right) showing sample level QC steps for the T1DGC cohorts. Samples with ambiguous sex (dotted red and blue lines marks thresholds for females and males respectively), relatedness with another sample (dotted line marks threshold for relatedness), and non-European ancestry outliers (cyan points) are filtered out.

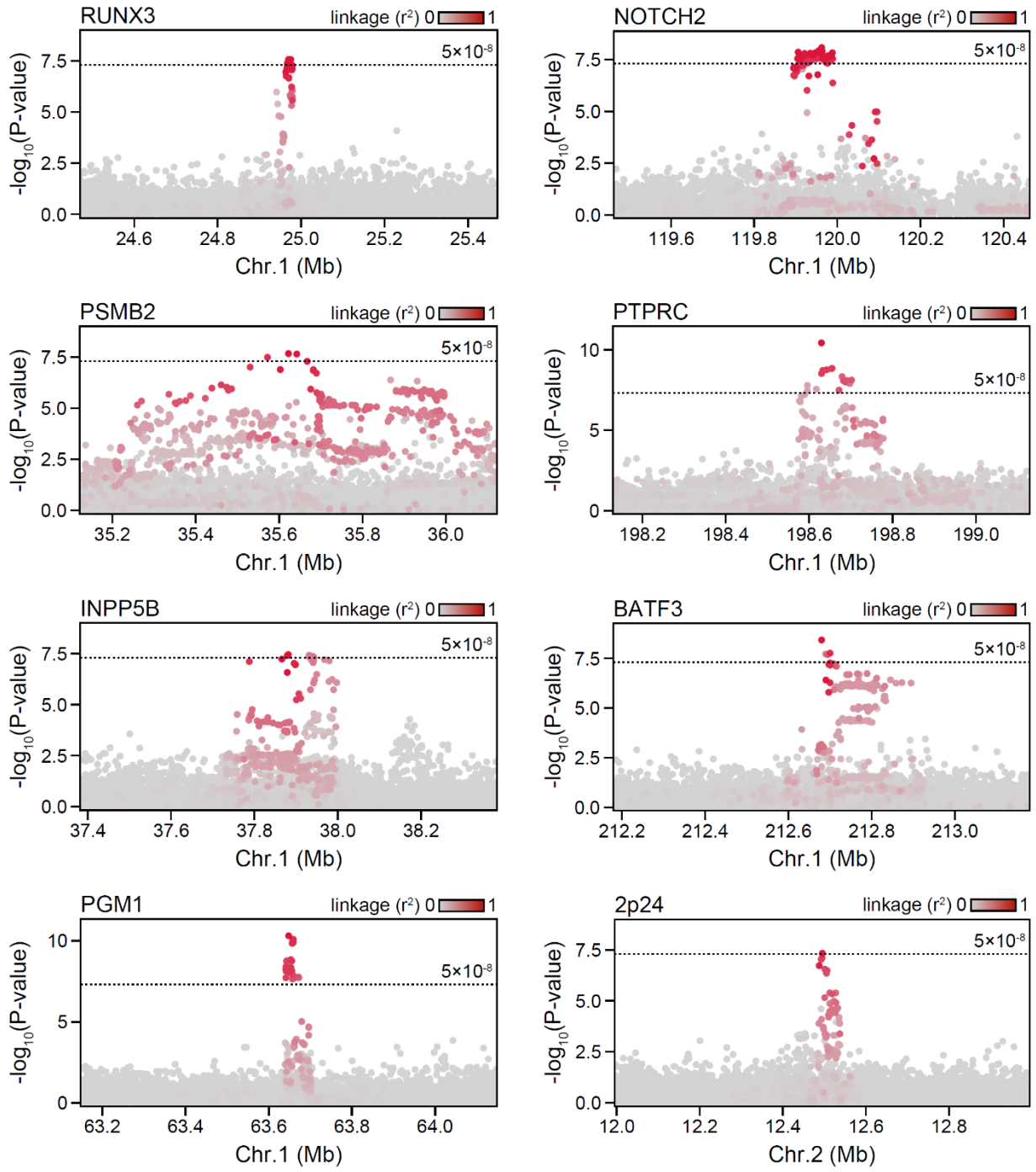



**Figure S3.2. Locus plots of novel T1D loci.**

Locus plots showing T1D GWAS meta-analysis association p-values for all variants in a ±500 kb window around the index variant for 33 genome-wide significant novel T1D loci. Variants are colored based on linkage disequilibrium with the index variant.

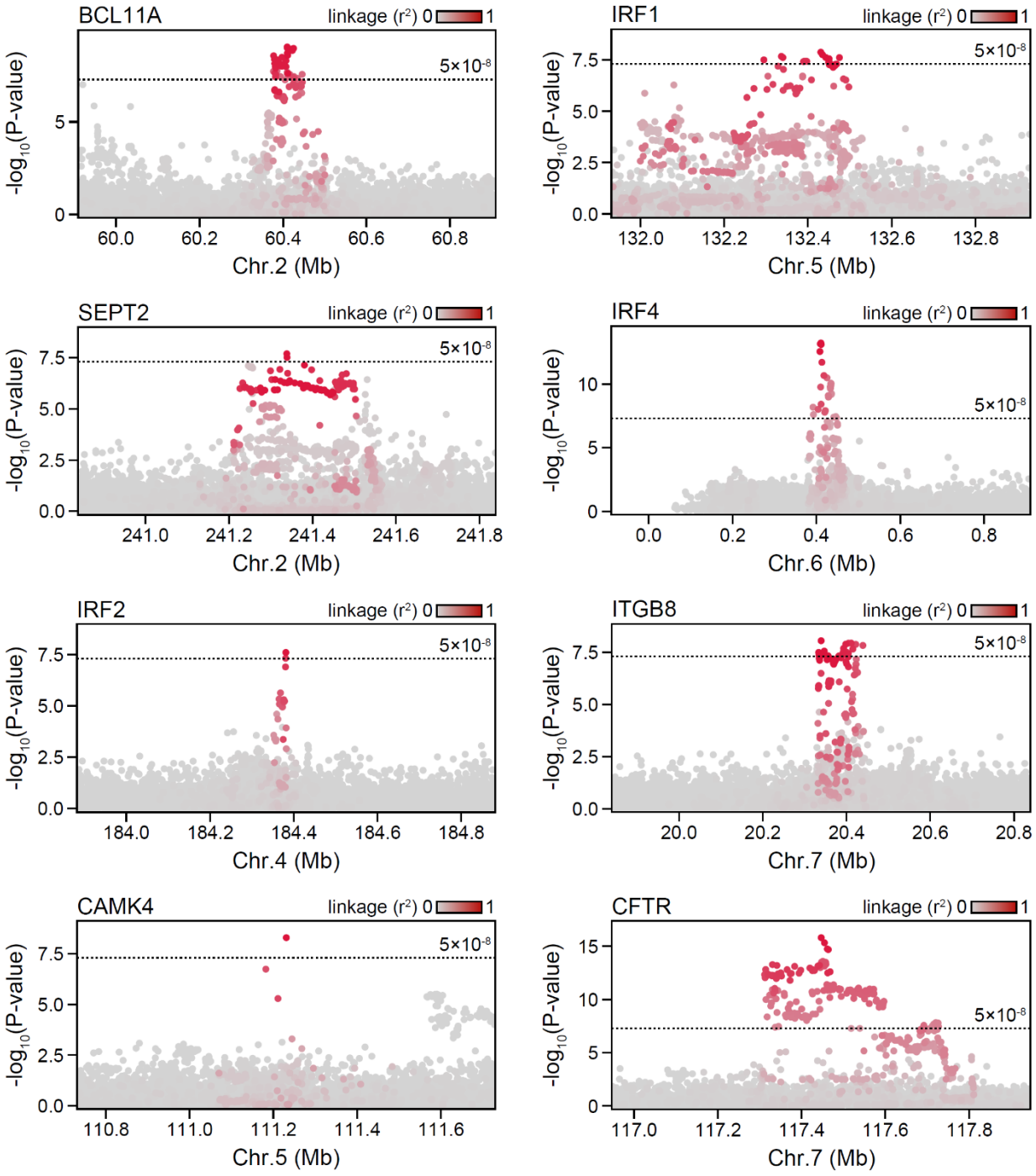



**Figure S3.2. Locus plots of novel T1D loci, continued.**

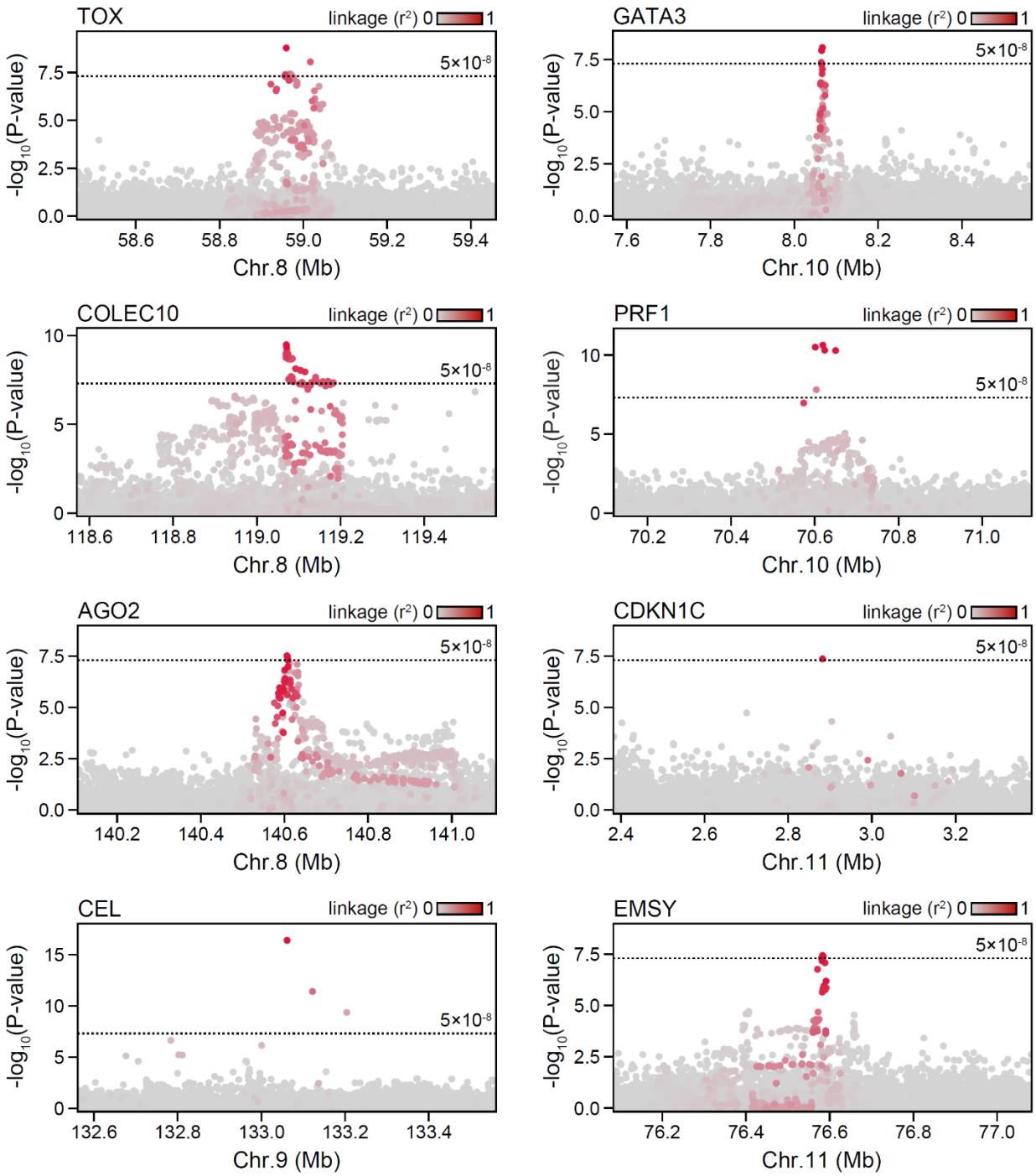



**Figure S3.2. Locus plots of novel T1D loci, continued.**

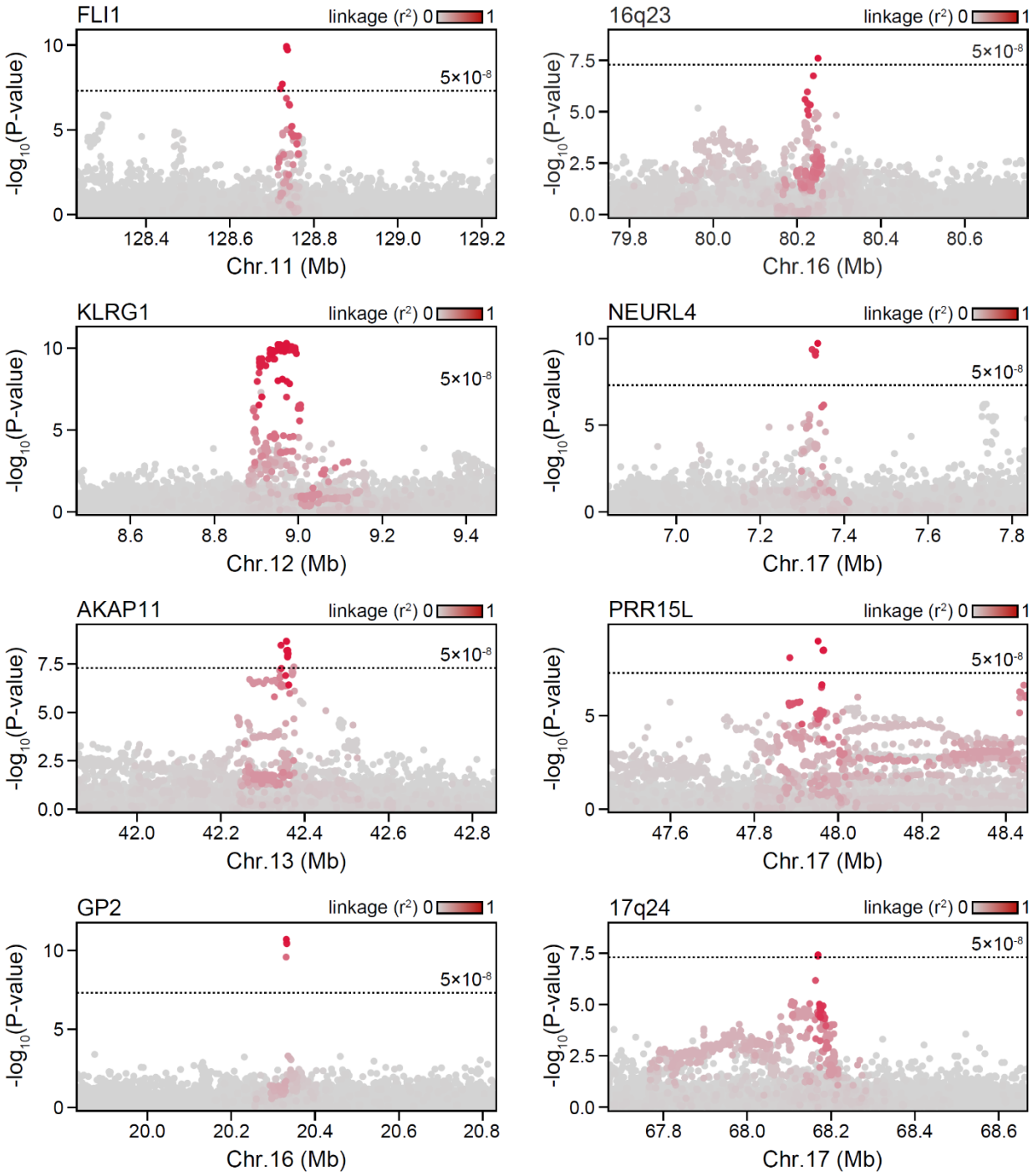



**Figure S3.2. Locus plots of novel T1D loci, continued.**

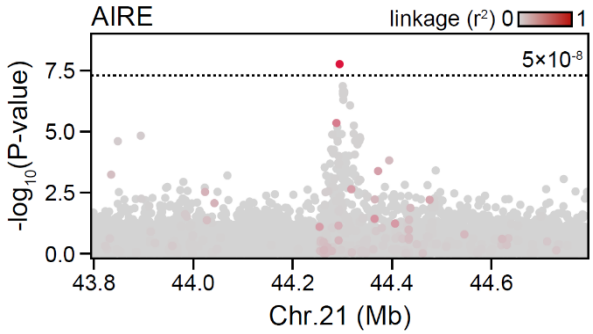



**Figure S3.2. Locus plots of novel T1D loci, continued.**

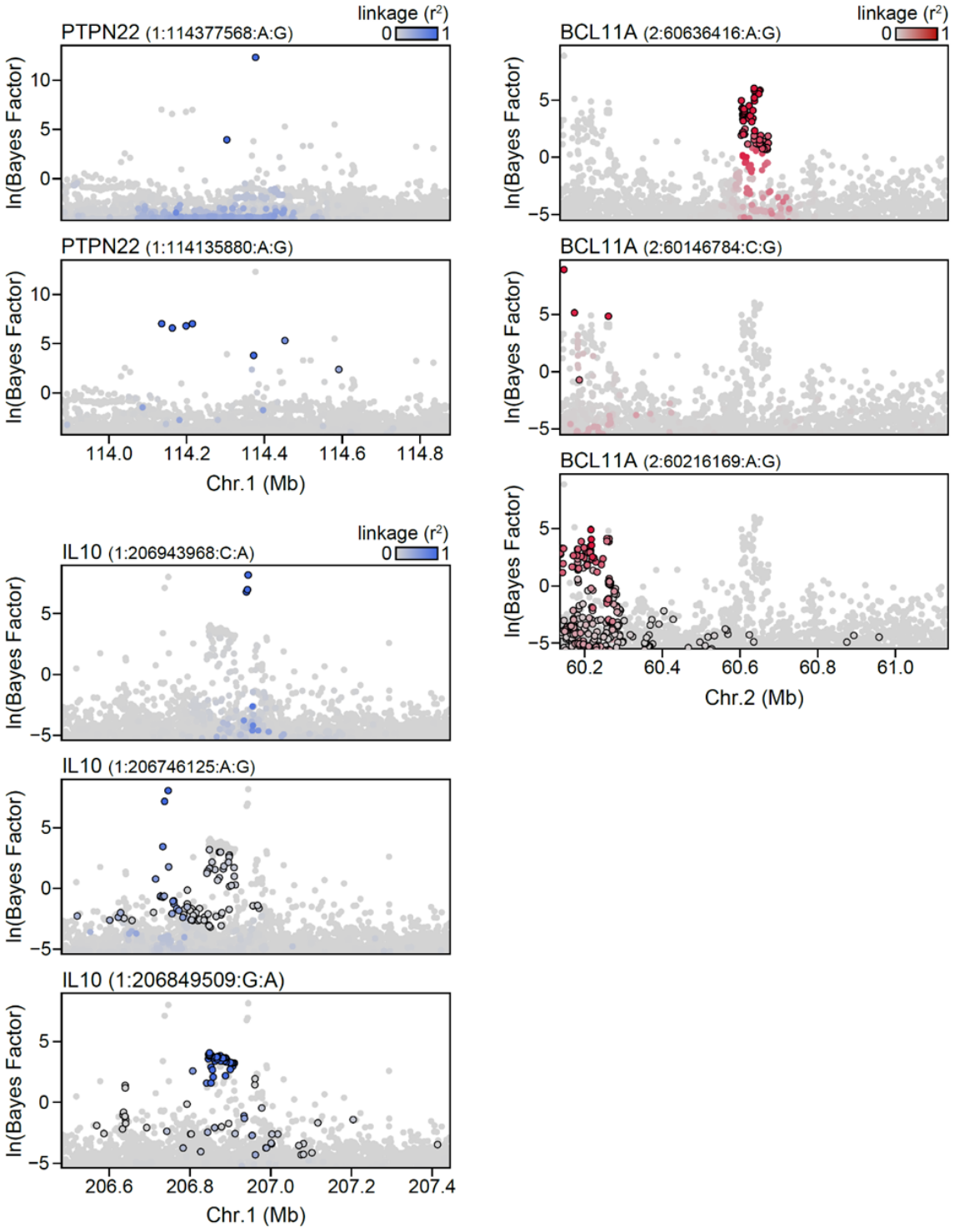



**Figure S3.3. Locus plots of independent T1D signals.**

Locus plots showing independent association signals for loci where FINEMAP identified more than one causal variant. Log-transformed Bayes factors from FINEMAP are plotted for each variant. Variants in a ±500 kb window around the search region are colored by linkage disequilibrium ($r^2$) with the most probable variant from fine mapping (blue for known loci and red for novel loci) and are circled if they are included in the 99% credible set for each signal.

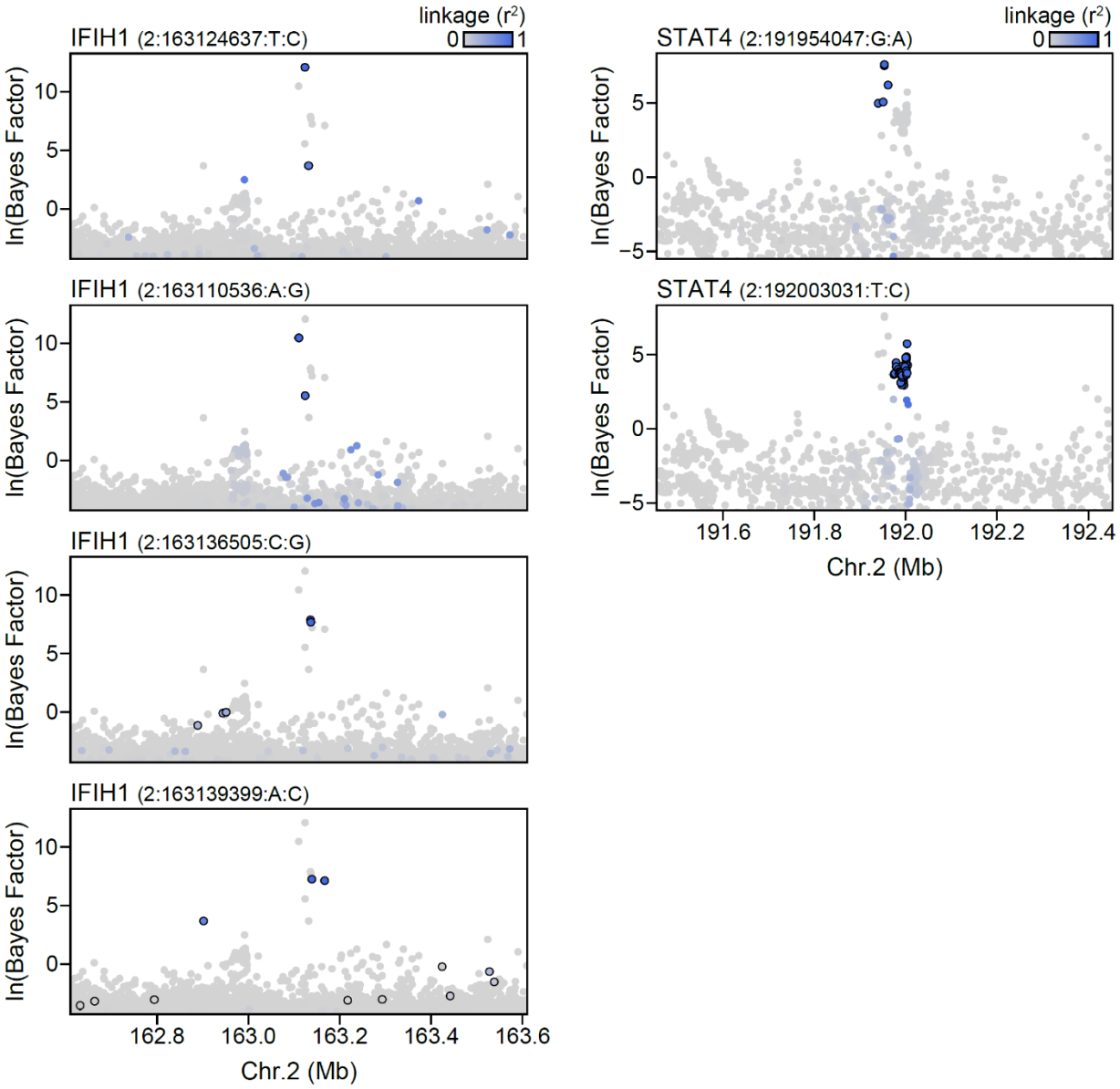



**Figure S3.3. Locus plots of independent T1D signals, continued.**

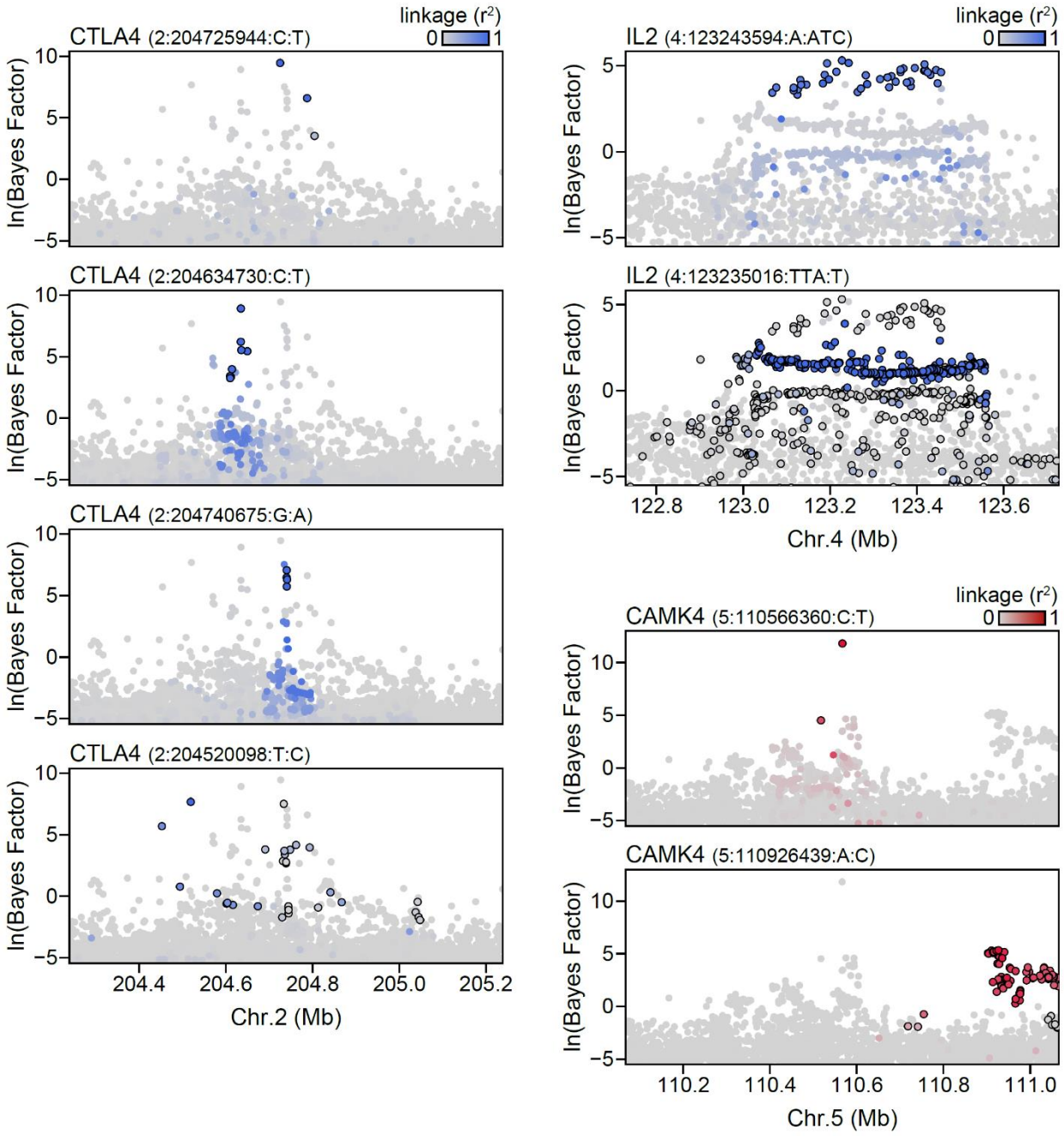



**Figure S3.3. Locus plots of independent T1D signals, continued.**

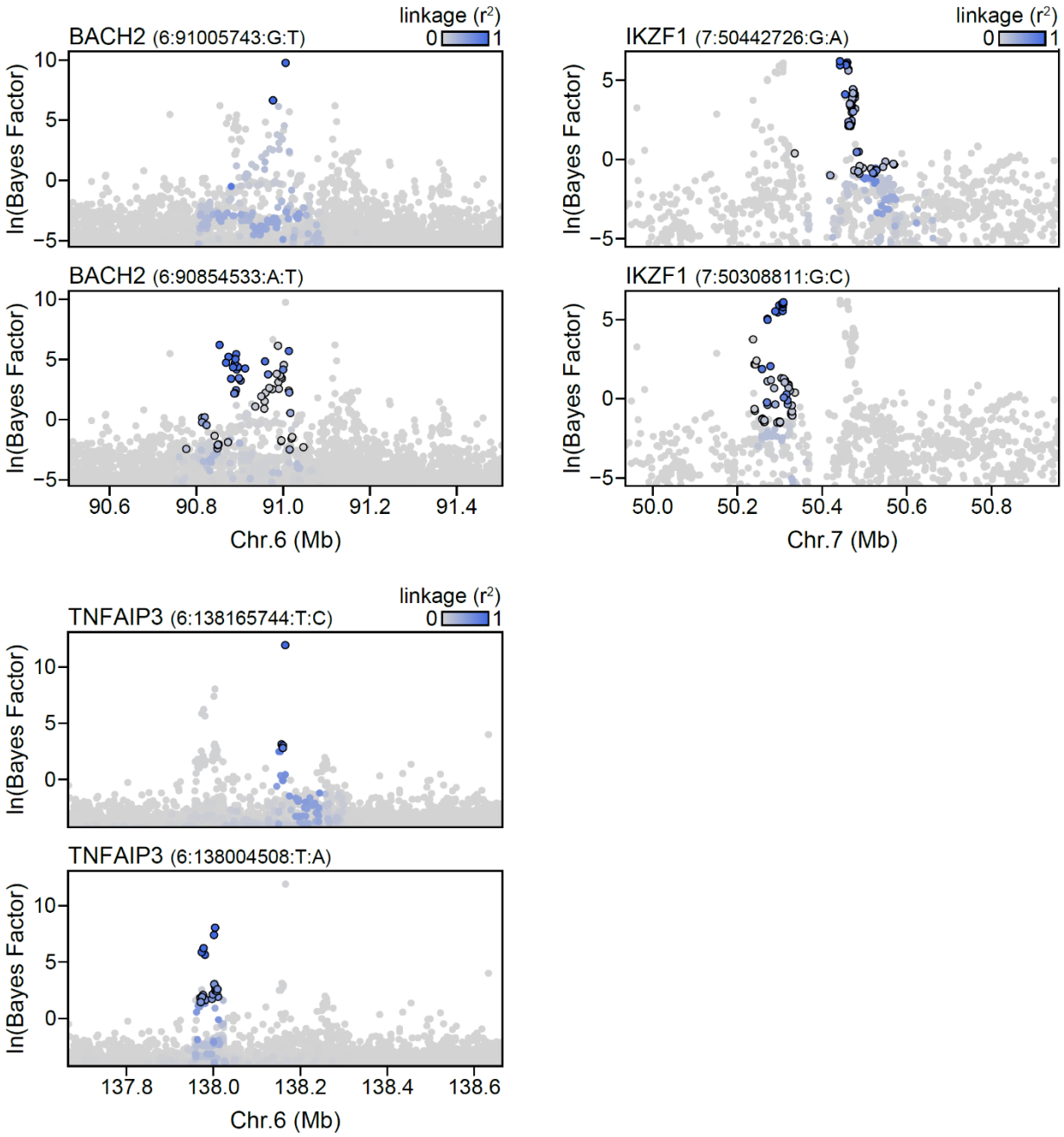



**Figure S3.3. Locus plots of independent T1D signals, continued.**

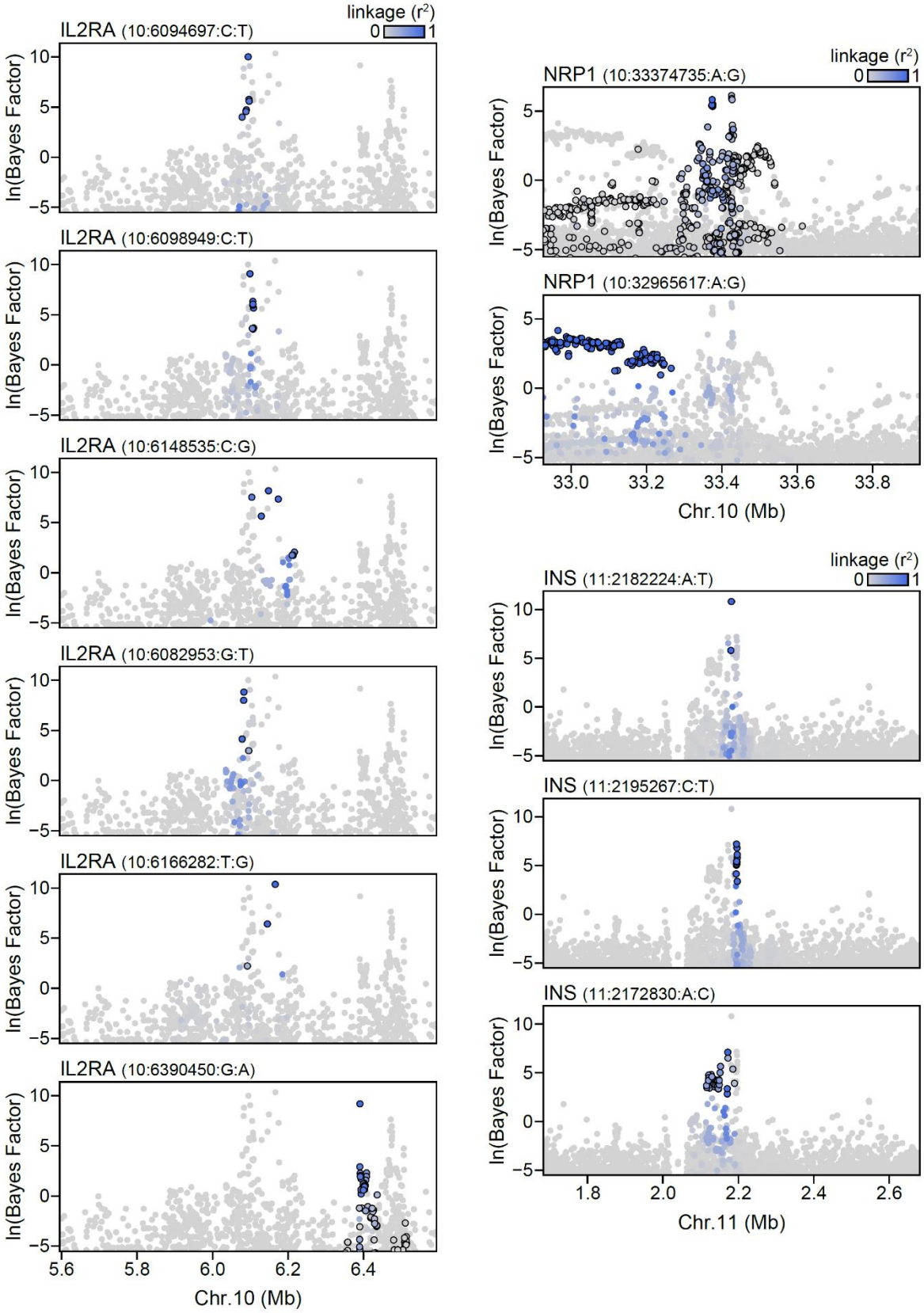



**Figure S3.3. Locus plots of independent T1D signals, continued.**

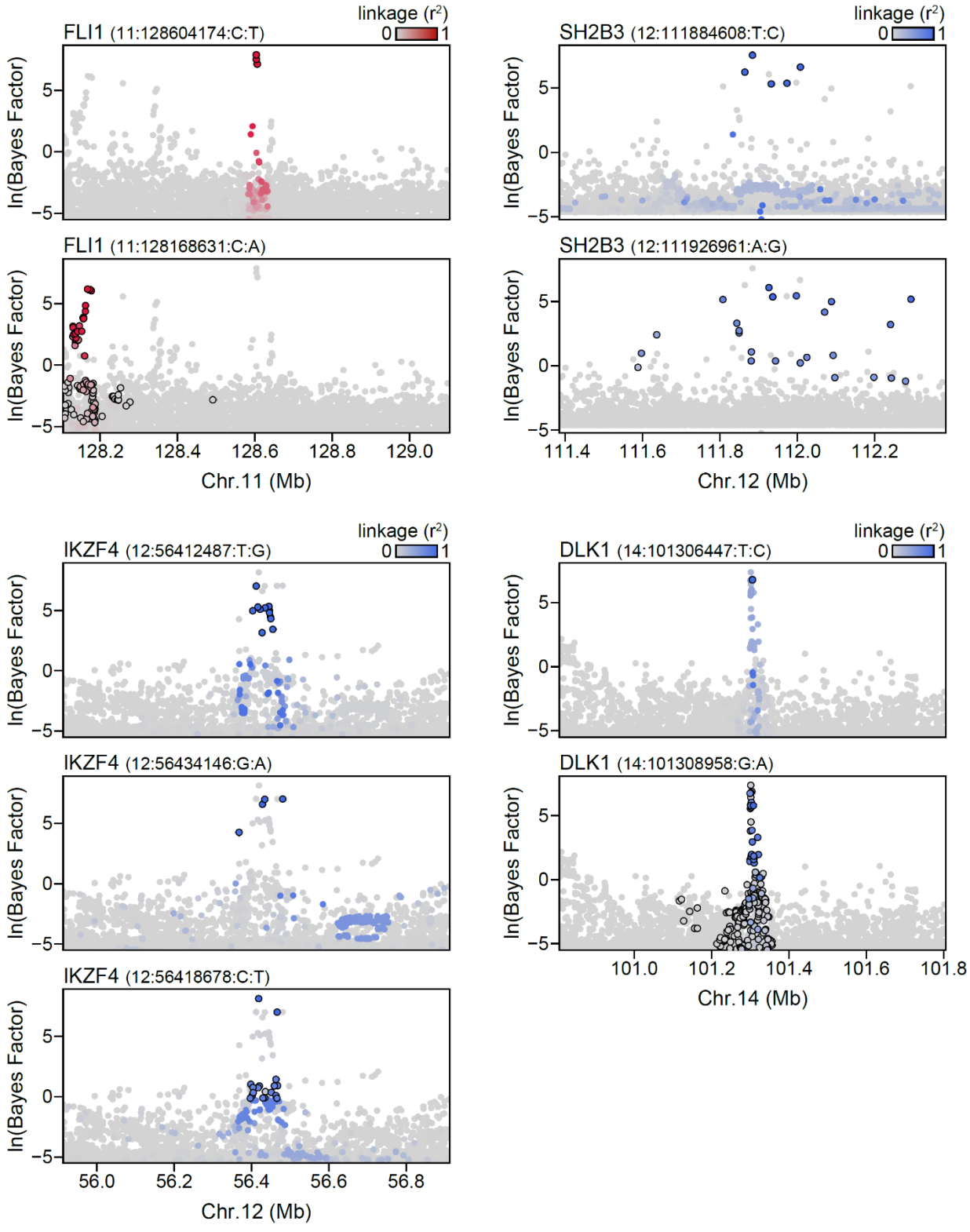



**Figure S3.3. Locus plots of independent T1D signals, continued.**

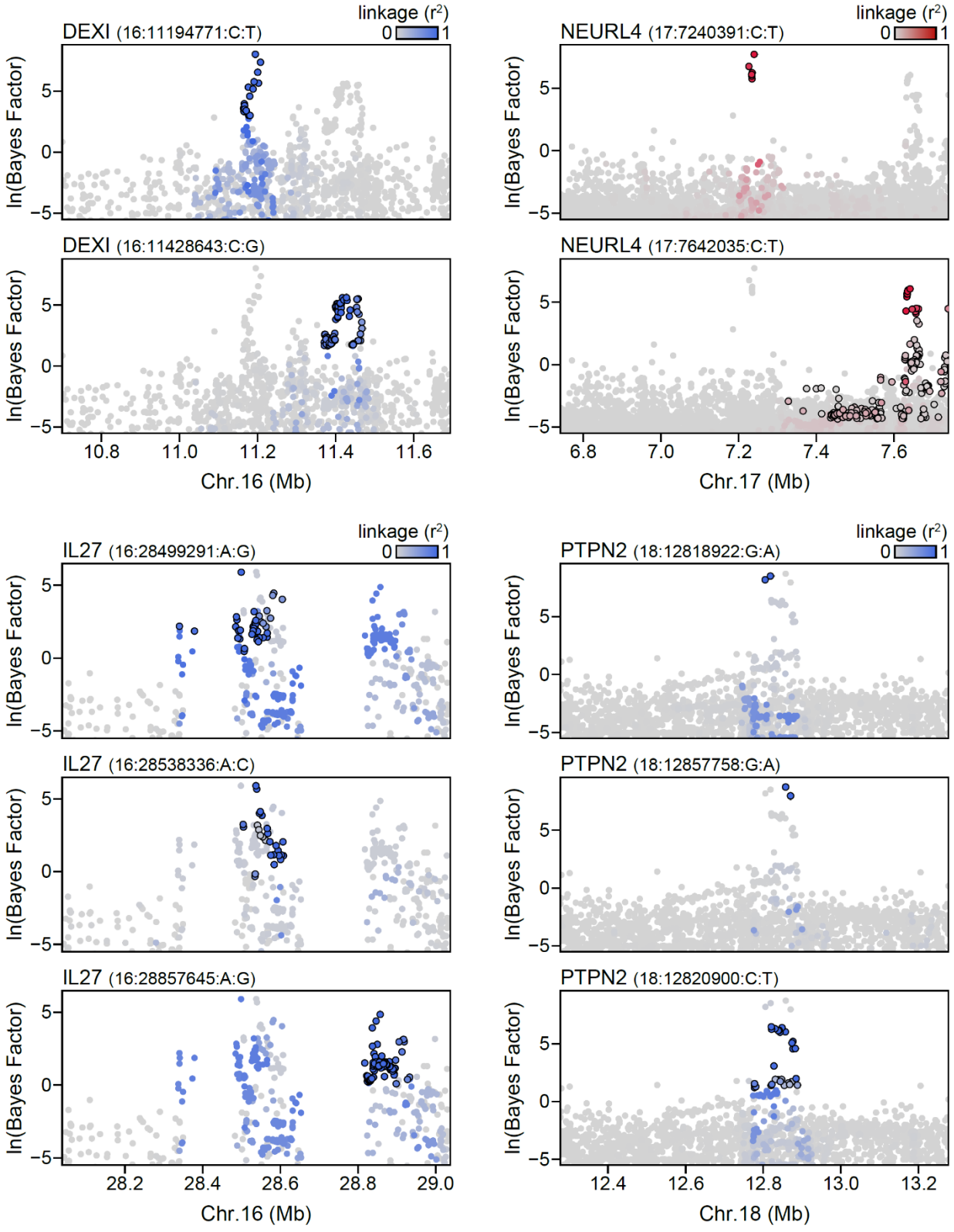



**Figure S3.3. Locus plots of independent T1D signals, continued.**

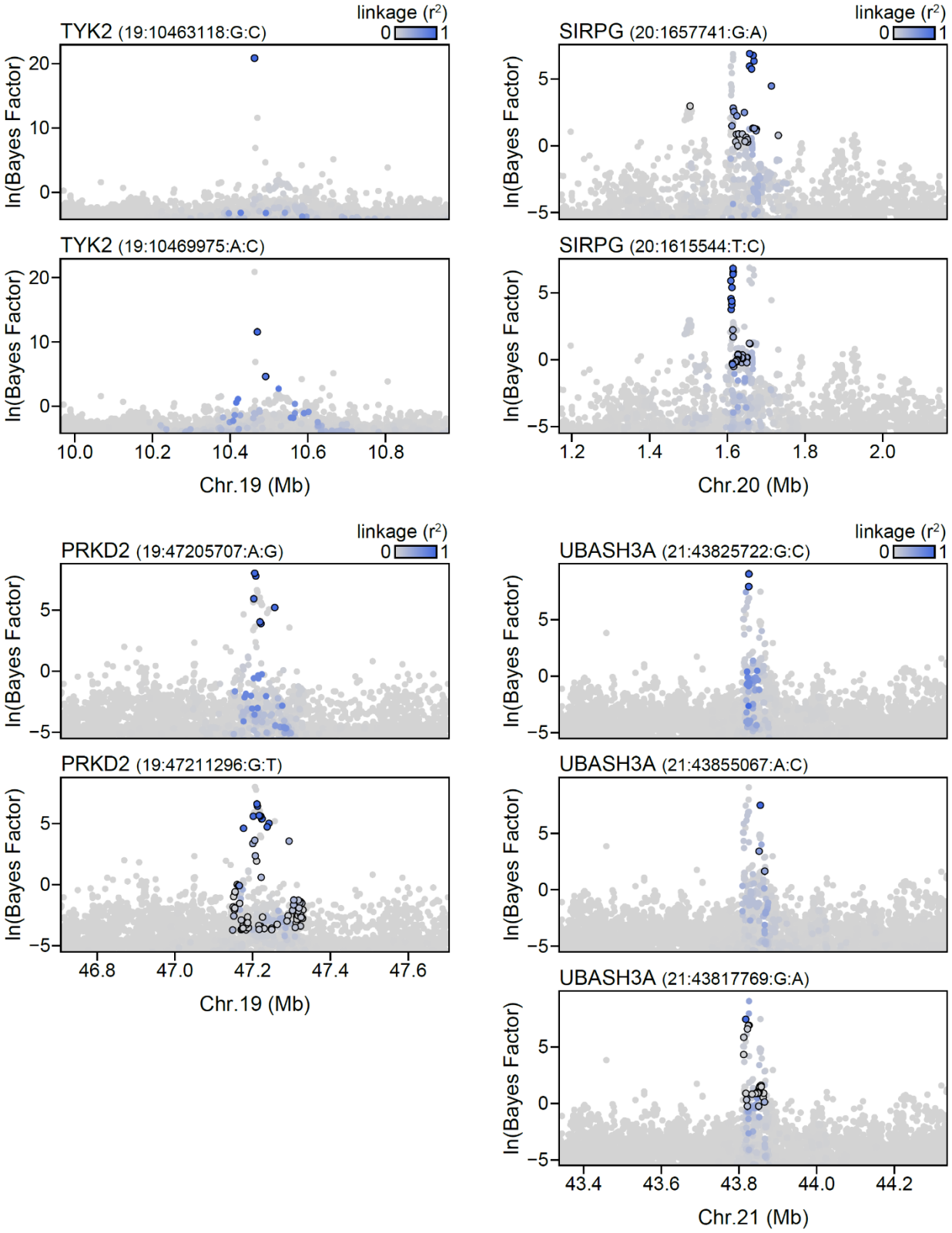



**Figure S3.3. Locus plots of independent T1D signals, continued.**

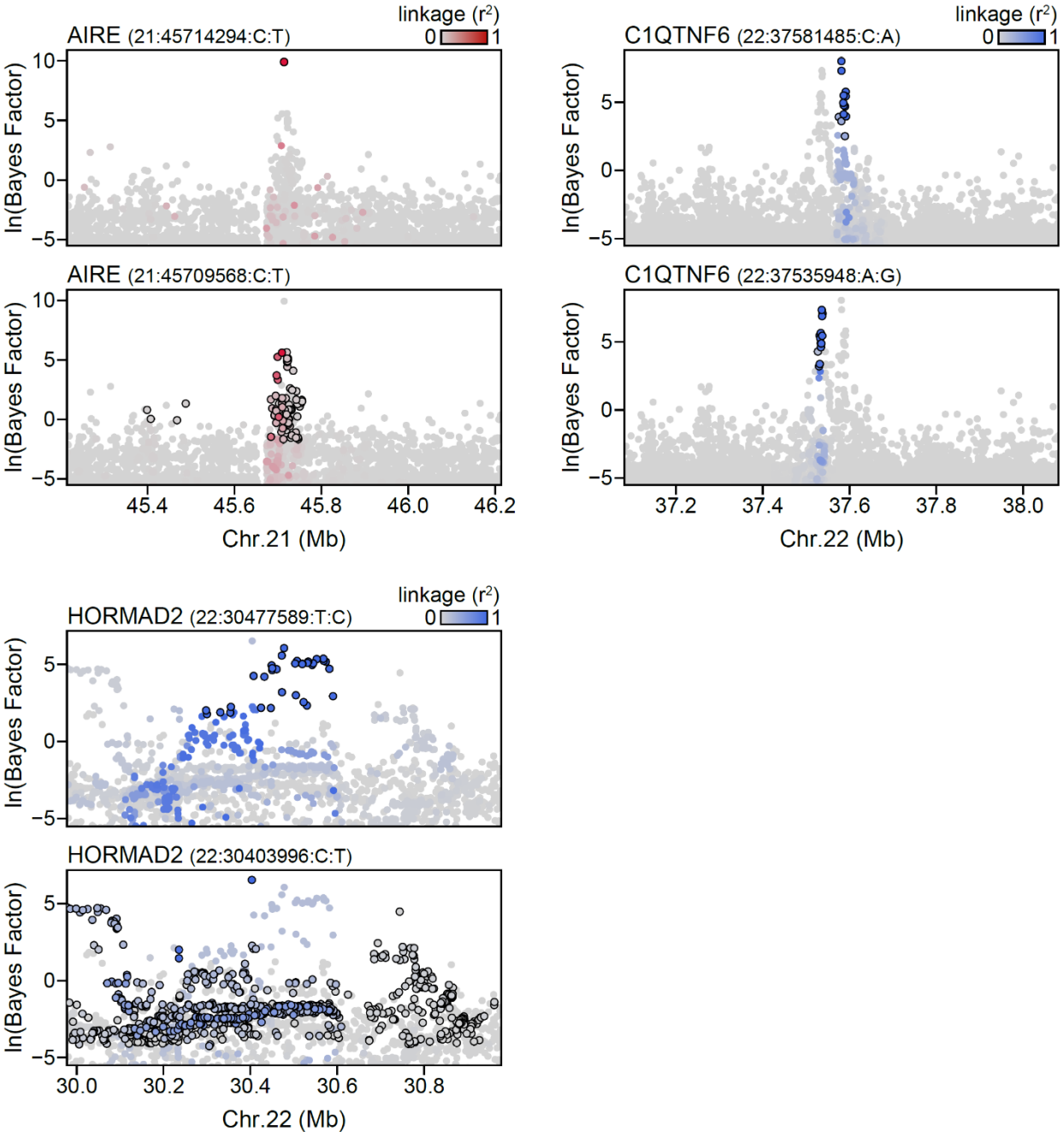



**Figure S3.3. Locus plots of independent T1D signals, continued.**

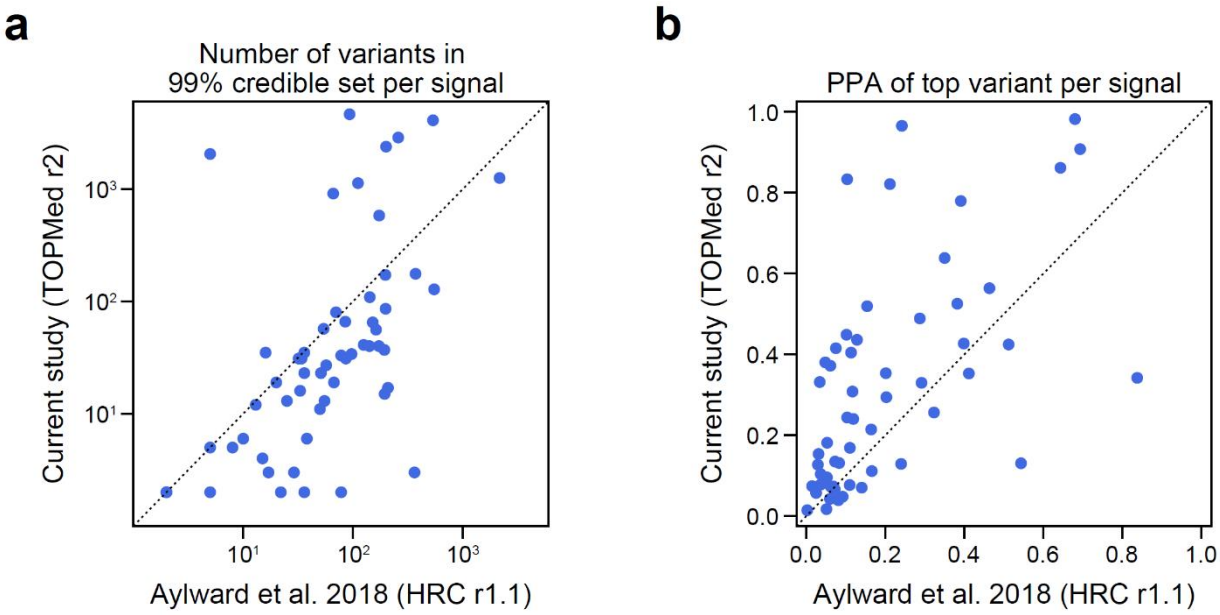



**Figure S3.4. Comparison of fine mapping resolution with a previous study.**

(a) Comparison of the number of variants in the 99% credible set or (b) posterior probability of association (PPA) of the top variant for 56 main T1D risk signals in common between the current study and Aylward et al., 2018, which were imputed using the TOPMed r2 and HRC r1.1 reference panels respectively. Lower numbers of 99% credible set variants and higher PPA of top variants generally indicate higher fine mapping resolution.

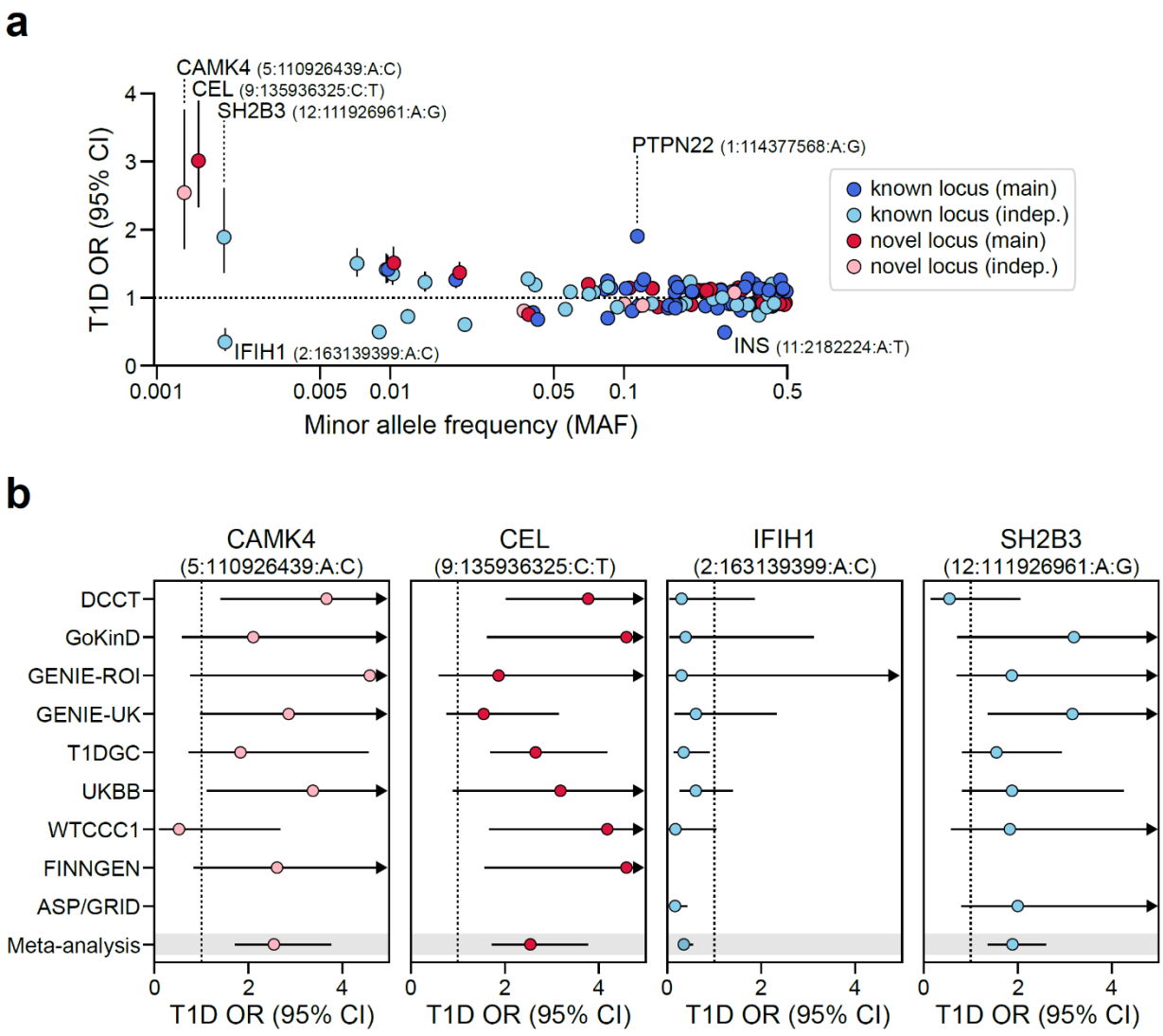



**Figure S3.5. Rare variants with large effects on T1D risk.**

(a) The relationship between minor allele frequency and T1D odds ratios (OR) for index variants at 136 T1D signals. Signals with common index variants and large effect size estimates (PTPN22 1:114377568:A:G and INS 11:2182224:A:T) or rare index variants (MAF<0.005) are labeled. Points and lines represent estimates for OR and 95% CI respectively. (b) Comparison of OR across cohorts for rare variants. Missing values indicate that the variant was not tested in the cohort.

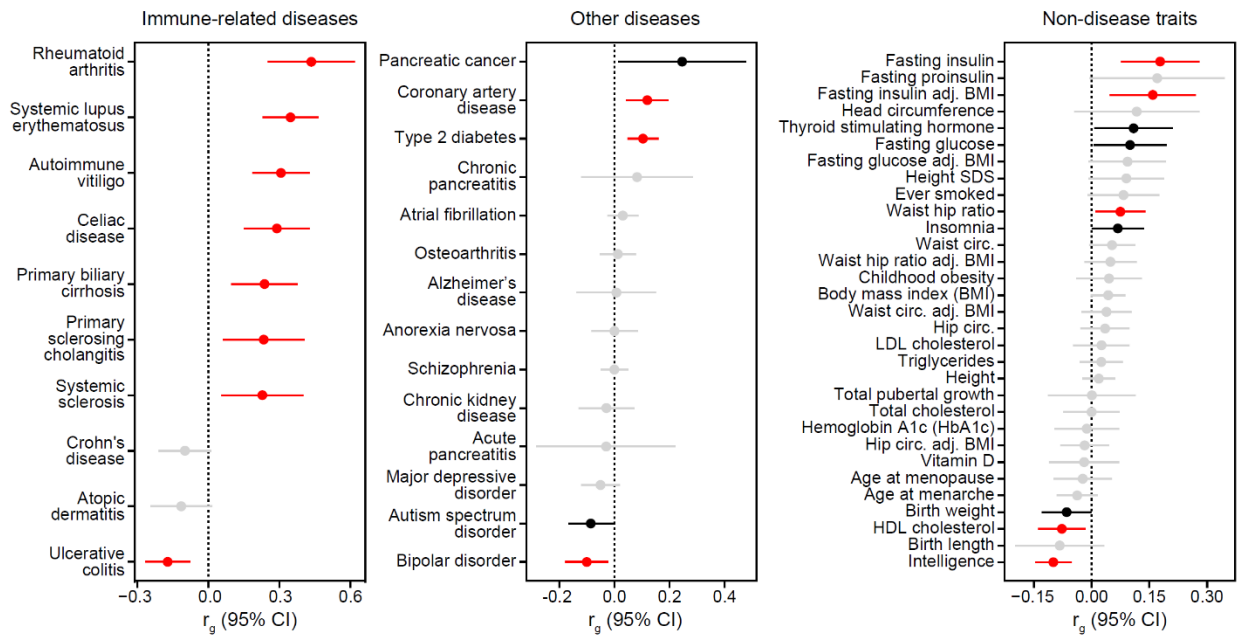



**Figure S3.6. Type 1 diabetes genetic correlations.**

Genetic correlations between T1D and other traits, including immune-related diseases (left), other diseases (middle), and non-disease traits (right), adj.=adjusted, circ.=circumference. Points represent genetic correlation estimates and lines represent 95% confidence intervals. Colors indicate significance: red indicates that the correlation is significant after FDR correction (FDR<0.1), black indicates that the correlation is nominally significant (P<0.05) but not significant after FDR correction, and grey indicates that the correlation is not significant.

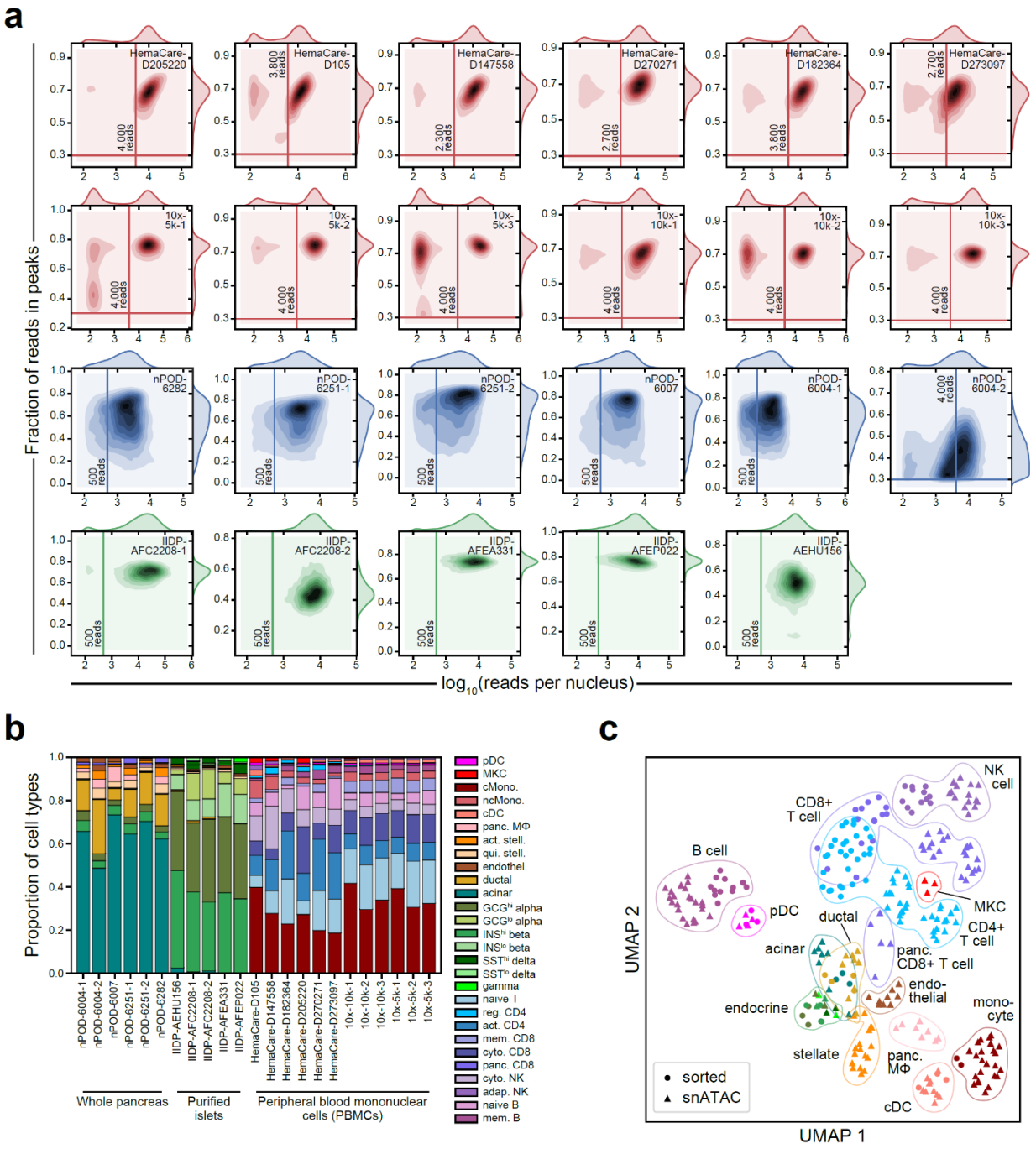



**Figure S3.7. snATAC-seq quality control metrics and comparison to sorted datasets.**

(a) Kernel density of the number of unique reads (log-transformed) compared to the fraction of reads in bulk peaks per cell. Thresholds for unique reads range between 500-4000 as indicated on each plot, and an additional threshold of fraction of reads in peaks > 0.3 was set for datasets generated using 10x Genomics technology. (b) Proportion of cell types within each experiment for cells passing QC. (c) Comparison of pseudobulk snATAC-seq profiles (within donor, at least 50 cells) to FACS purified ATAC-seq for major pancreatic or unstimulated immune cell types using PCA+ UMAP dimensionality reduction on read counts within 448,142 merged cREs from snATAC-seq clusters.

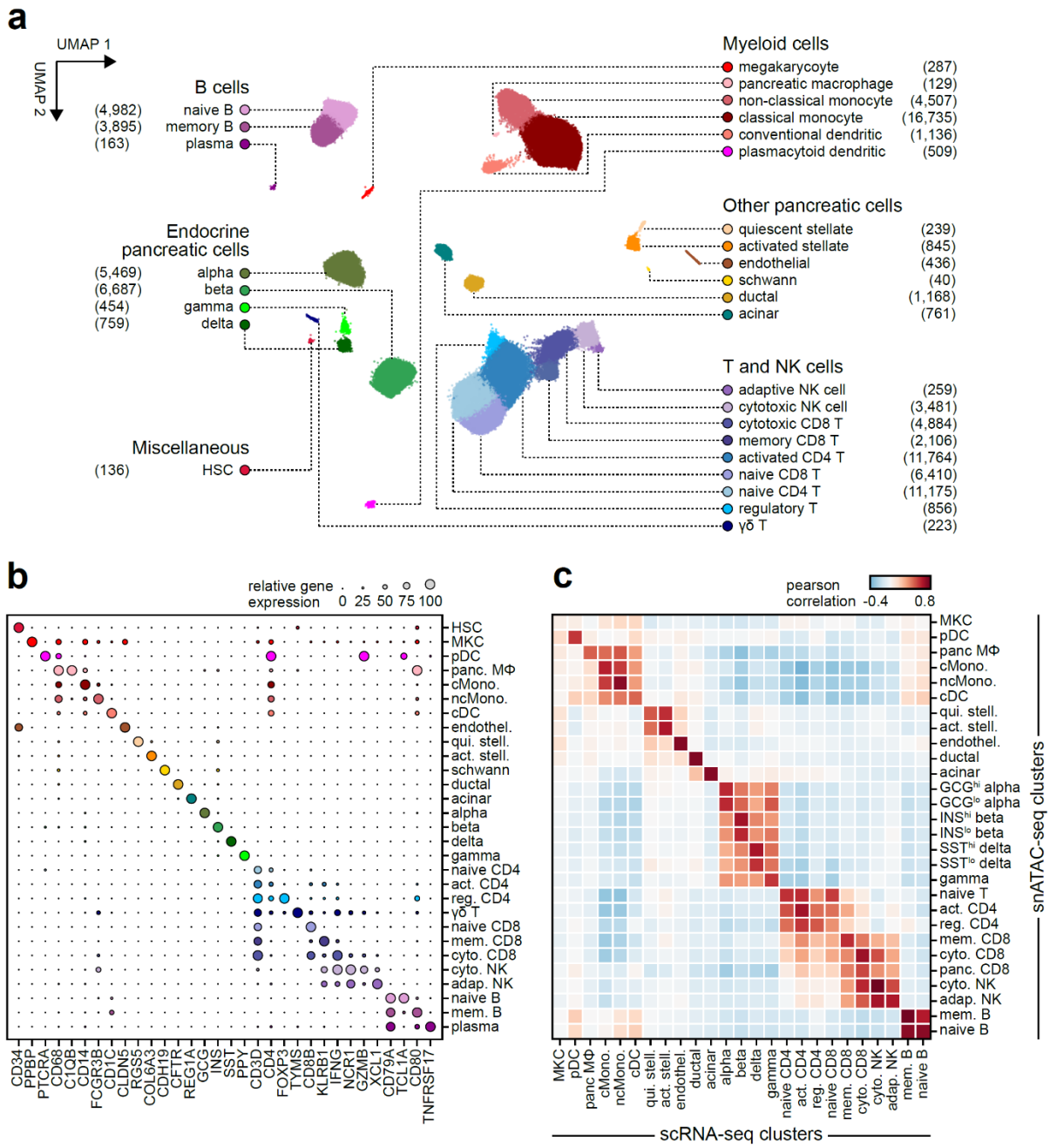



**Figure S3.8. Single cell RNA-seq reference map of PBMCs and pancreatic islets.**

(a) Clustering of 90,495 expression profiles from single cell RNA-seq experiments of peripheral blood mononuclear cells and pancreatic islets from published studies. Cells are plotted on the first two UMAP components and colored based on cluster assignment. The number of cells in each cluster is shown next to its corresponding label. HSC, hematopoietic stem cell. γδ T, gamma delta T. pDC, plasmacytoid dendritic. (b) Relative gene expression (average expression for all cells within a cluster and scaled from 0-100 across clusters) showing examples of marker genes used to assign cluster labels. (c) Pearson correlation coefficient between gene expression and promoter accessibility specificity scores using a list containing the top 100 most specific genes for each scRNA-seq cluster found in snATAC-seq.

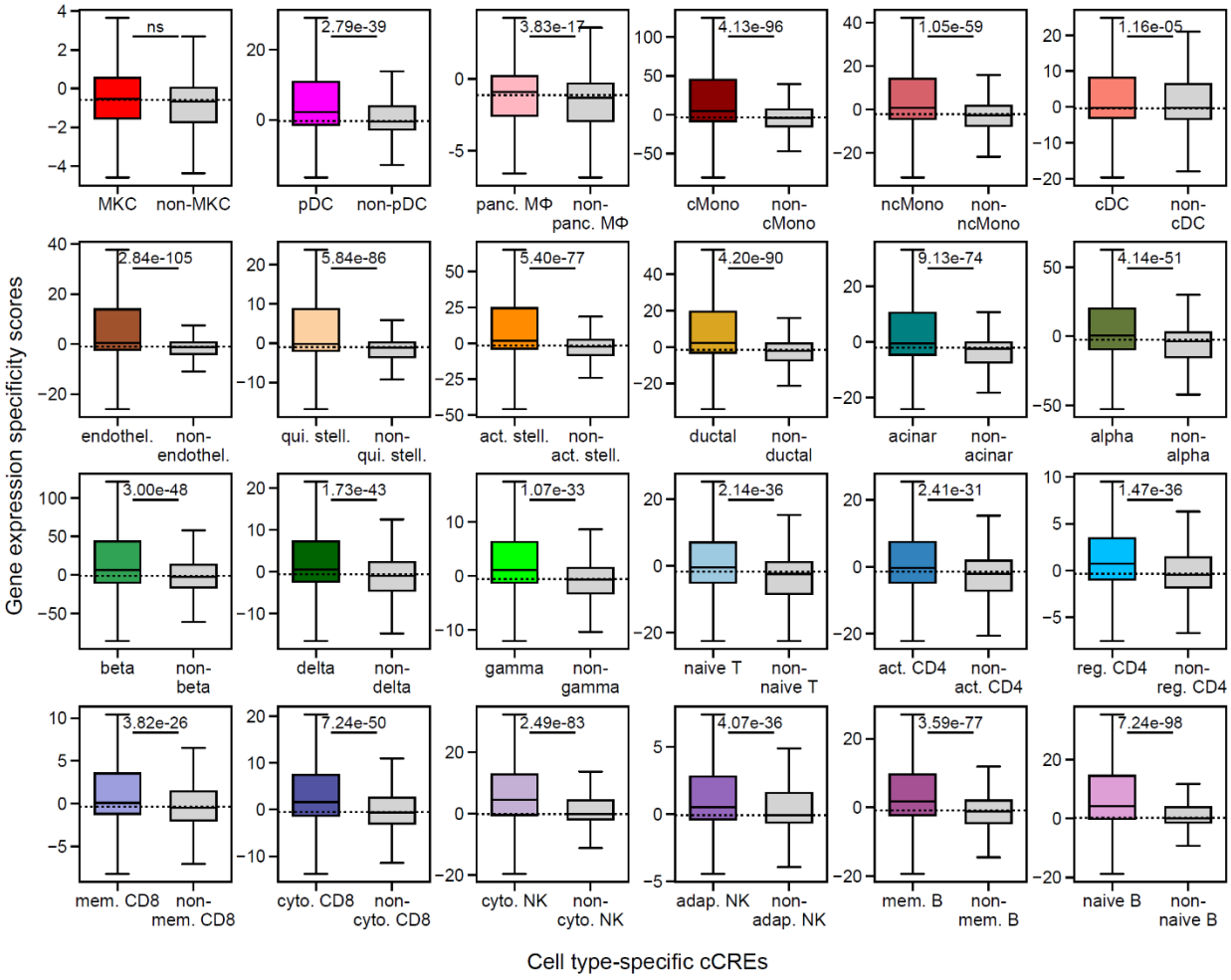



**Figure S3.9. Cell type-specific cCREs are in proximity to genes with cell type-specific expression.**

Gene expression specificity scores (t-statistics from linear regression models) for genes within 100 kb windows around cell type-specific cCREs for each cell type, as compared to cell type-specific cCREs for other cell types. Boxplot center line, limits, and whiskers represent median, quartiles, and 1.5 IQR, respectively. Outlier points (>1.5 IQR) are not shown for clarity of scale. ns, not significant by 2-sided Welch's t-test.

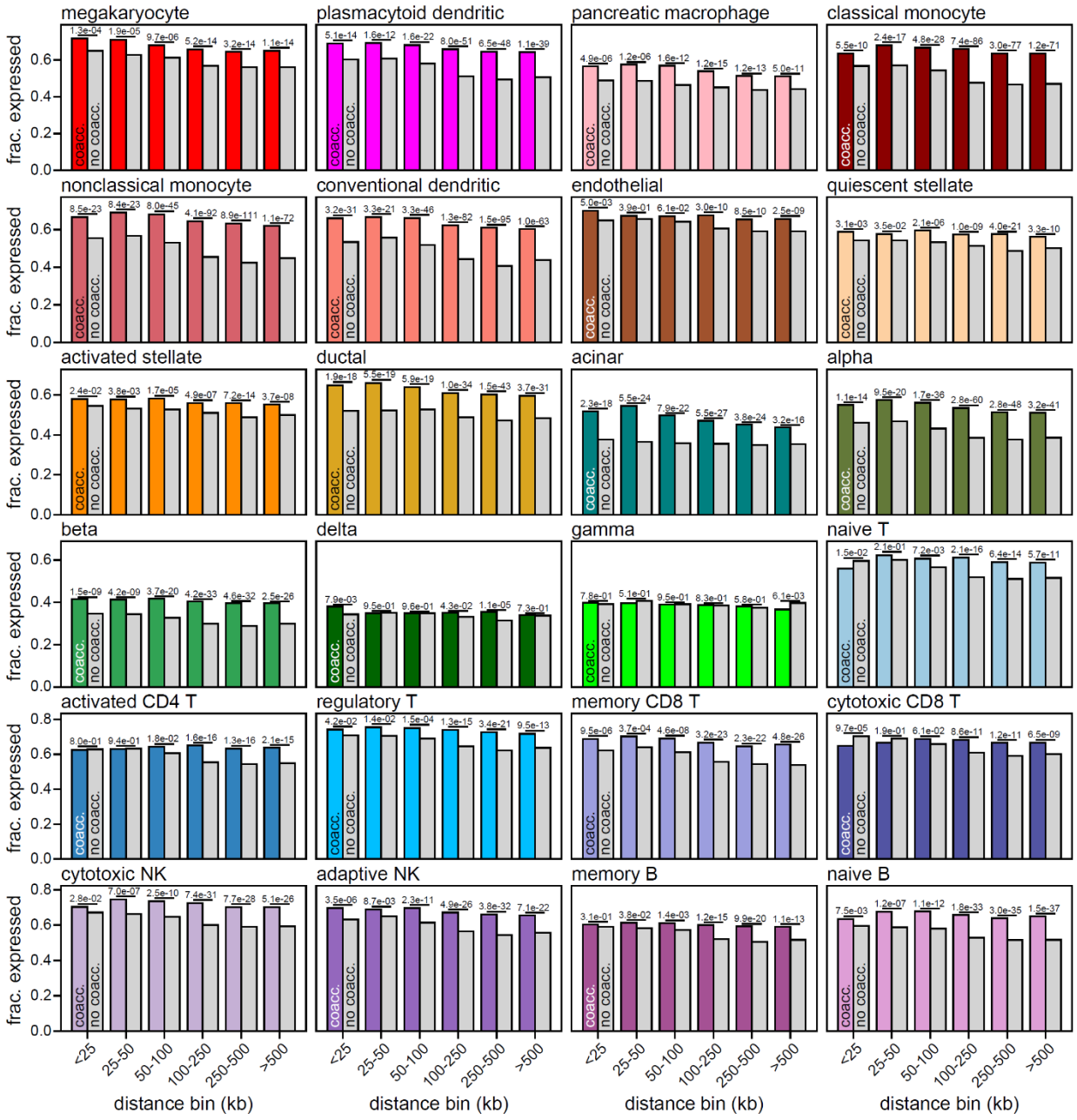



**Figure S3.10. Co-accessible genes are expressed more often than non-co-accessible genes within the same cell type.**

Distance-matched comparisons of the fraction of expressed genes (scRNA-seq pseudobulk TPM>1) between those with co-accessible links to distal regulatory elements (coacc.) and those without co-accessible links (no coacc.), defined by co-accessibility<0, show that genes with co-accessible links are generally enriched for expressed genes (2-sided P-values, Fisher's exact test).

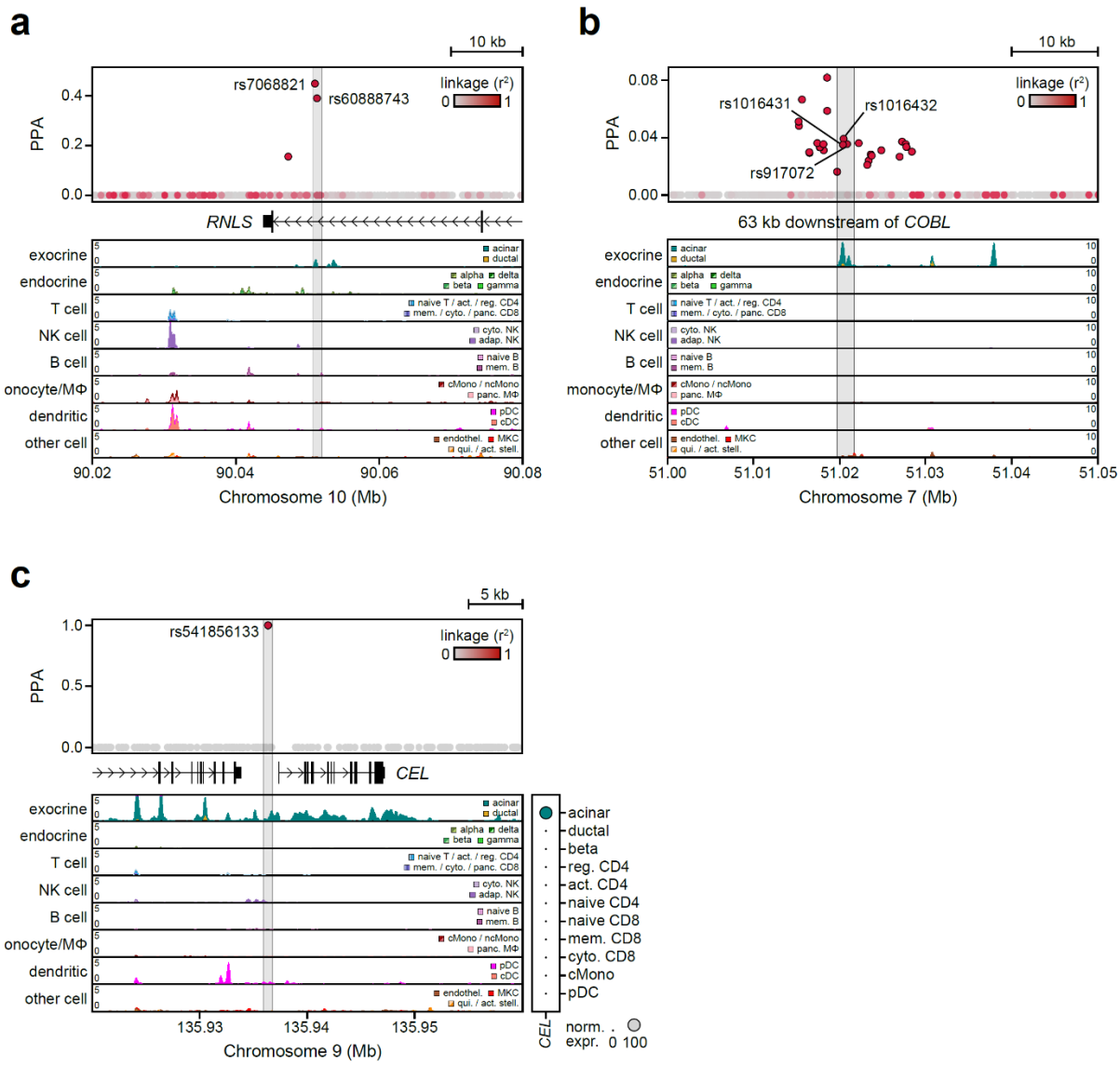



**Figure S3.11. Fine mapped variants in acinar regulatory elements.**

(a) The RNLS (10:90051035:G:T) signal contains two variants (rs7068821, PPA=0.448; rs60888743, PPA=0.389) located in an acinar-specific intronic peak within *RNLS*. (b) The COBL (7:51018492:GT:G) signal contains three variants (rs917072, PPA=0.036; rs1016432, PPA=0.039; rs1016431, PPA=0.035) located in an acinar-specific distal peak 63 kb downstream of *COBL*. (c) The CEL (9:135936325:C:T) signal contains a single fine mapped rare variant (rs541856133, PPA=1.0) in an acinar-specific broad region of chromatin accessibility directly upstream of the *CEL* promoter (left). CEL has acinar-specific gene expression (right).

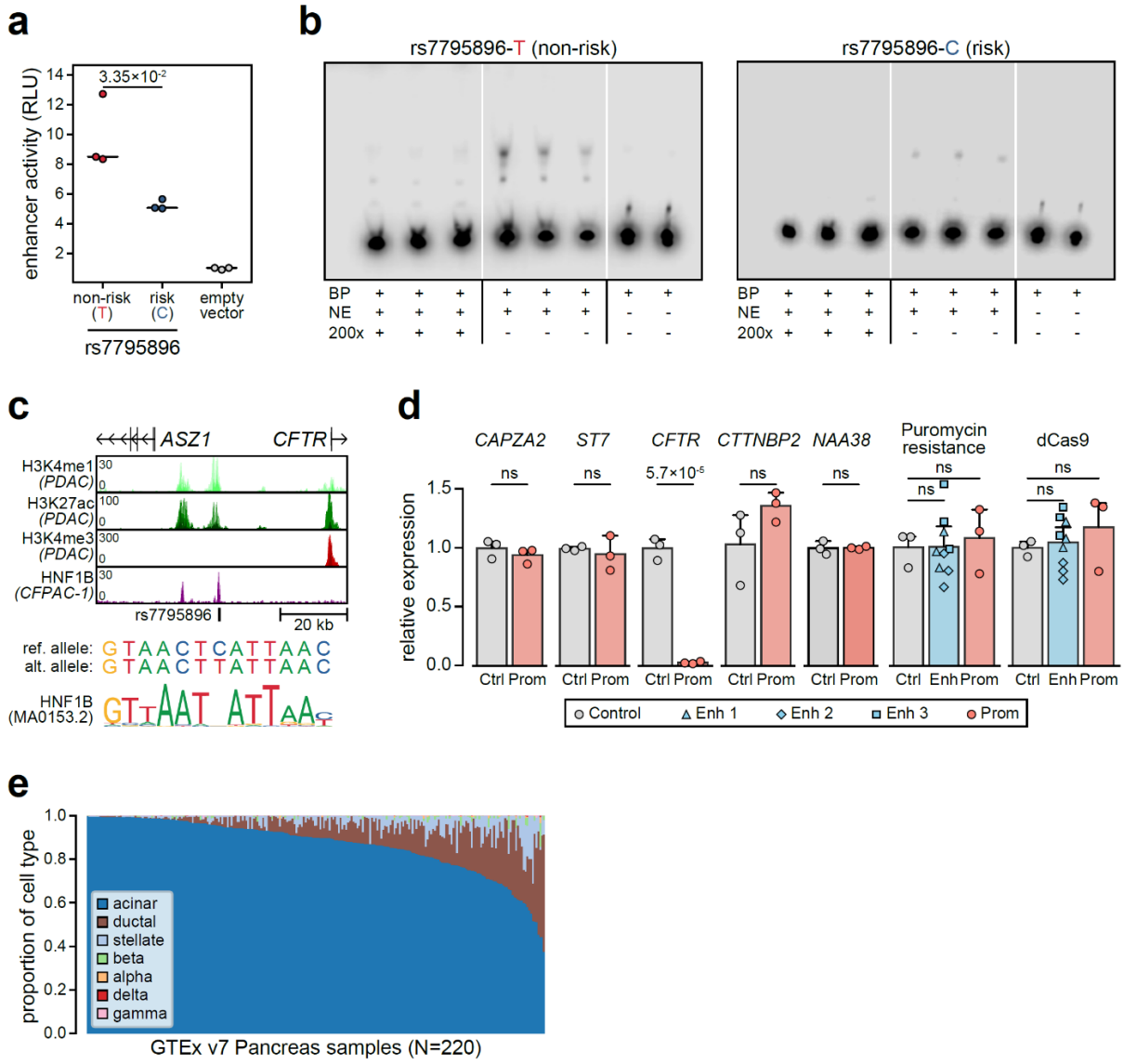



**Figure S3.12. rs7795896 has allelic effects on ductal enhancer activity.**

(a) Luciferase reporter assay in Capan-1 cells transfected with pGL4.23 minimal promoter plasmids containing rs7795896. Relative luciferase units (RLU) represent Firefly:Renilla ratios normalized to the empty pGL4.23 vector. p-value by two-tailed Student's t-test. (b) Full gels for electrophoretic mobility shift assay (EMSA) with nuclear extract from Capan-1 cells, using probes containing the T allele (left) or the C allele (right) of rs7795896. (c) rs7795896 overlaps histone marks of active enhancers but not promoters in pancreatic ductal adenocarcinoma cell lines (Capan-1, Capan-2, and CFPAC-1). rs7795896 overlaps a ChIP-seq peak for HNF1B in CFPAC-1 cells and a predicted HNF1B motif. (d) Relative expression for genes around rs7795896 with non-zero expression and the puromycin resistance and dCas9 genes. Ctrl n=3 biological replicates; Enh n=9, 3 sgRNAs × 3 biological replicates; Prom n=3 biological replicates. ns, not significant by two-tailed Student's t-test (Prom vs Ctrl) or ANCOVA with sgRNA as a covariate (Enh vs Ctrl). (e) Proportions of selected pancreatic cell types estimated by MuSiC for 220 bulk pancreas RNA-seq samples from the GTEx v7 release using single cell expression profiles.

# 3.8 Tables

## Table 3.1. Matched ancestry T1D case and control cohorts.

| Cohort | N (cases) | N (controls) | N (total) | % female | Ancestry (country) | Genotyping array | Data accession |
|---|---|---|---|---|---|---|---|
| T1DGC | 3,955 | 0 | 3,955 | 46.57% | UK | Illumina HumanHap 550 | phs000180.v3.p2 |
| 1958 Birth Cohort (WTCCC 2) | 0 | 2,694 | 2,694 | 48.44% | UK | Illumina Human 1.2M | EGAD00000000022 |
| UK National Blood Service (WTCCC 2) | 0 | 2,494 | 2,494 | 50.56% | UK | Illumina Human 1.2M | EGAD00000000024 |
| GENIE (ROI) | 455 | 0 | 455 | 52.09% | Ireland | Illumina Omni1 | phs000389.v1.p1 |
| Collaborative Study of Genes, Nutrients and Metabolites (CSGNM) | 0 | 2,317 | 2,317 | 58.74% | Ireland | Illumina Omni1 | phs000789.v1.p1 |
| GENIE (UK) | 1,242 | 0 | 1,242 | 48.07% | UK | Illumina Omni1 | phs000389.v1.p1 |
| Health and Retirement Study (HRS) | 22 | 5,393 | 5,415 | 57.65% | USA* | Illumina Omni2.5 | phs000428.v2.p2 |
| GoKinD | 1,682 | 0 | 1,682 | 53.92% | USA | Affymetrix 500K | phs000018.v2.p1 |
| NIMH Schizophrenia Controls | 0 | 1,361 | 1,361 | 54.08% | USA | Affymetrix 6.0 | phs000021.v3.p2 |
| DCCT-EDIC | 1,288 | 0 | 1,288 | 46.82% | USA | Illumina HumanHap 550 | phs000086.v3.p1 |
| Neurodevelopmental Genomics | 0 | 1,317 | 1,317 | 47.30% | USA | Illumina HumanHap 550 | phs000607.v3.p2 |
| eMERGE Network | 0 | 4,275 | 4,275 | 55.46% | USA | Illumina Human 660W | phs000360.v3.p1 |
| T1D (WTCCC 1) | 1,918 | 0 | 1,918 | 49.22% | UK | Affymetrix 500K | EGAD00000000008 |
| 1958 Birth Cohort (WTCCC 1) | 0 | 90 | 90 | 60.00% | UK | Affymetrix 500K | EGAD00000000001 |
| UK National Blood Service (WTCCC 1) | 0 | 158 | 158 | 54.43% | UK | Affymetrix 500K | EGAD00000000002 |
| Bipolar Disorder (WTCCC 1) | 0 | 1,857 | 1,857 | 62.63% | UK | Affymetrix 500K | EGAD00000000003 |
| T1DGC (ASP) | 2,663 | 0 | 2,663 | 48.37% | UK | Illumina Infinium ImmunoArray | phs000911.v1.p1 |
| T1DGC (UKGRID) | 2,853 | 0 | 2,853 | 51.88% | UK | Illumina Infinium ImmunoArray | phs000911.v1.p1 |
| 1958 Birth Cohort (Immunochip) | 0 | 4,008 | 4,008 | 50.50% | UK | Illumina Infinium ImmunoArray | EGAD00010000248 |
| UK Biobank | 1,445 | 362,050 | 363,495 | 54.73% | UK | Affymetrix UK Biobank Axiom | https://www.ukbiobank.ac.uk/ |
| FinnGen (freeze 3) - Type 1 diabetes, strict | 1,419 | 113,624 | 115,043 | - | Finland | ThermoFisher Axiom custom | https://www.finngen.fi/en |

## Table 3.2. Index variants for 33 novel T1D risk loci.

| Marker | Chrom. | Position (hg19) | Position (hg38) | Allele | | Locus name | Alt. AF | Beta | SE | P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Ref. | Alt. | | | | | |
| rs10751776 | 1 | 25296743 | 24970252 | A | C | RUNX3 | 0.5099 | 0.0781 | 0.0141 | 2.67E-08 |
| rs574384 | 1 | 36087661 | 35622060 | C | A | PSMB2 | 0.8947 | -0.1336 | 0.0239 | 2.20E-08 |
| rs12742756 | 1 | 38347417 | 37881745 | A | G | INPP5B | 0.4282 | -0.0831 | 0.0151 | 3.54E-08 |
| rs855330 | 1 | 64113889 | 63648218 | T | C | PGM1 | 0.2594 | 0.1112 | 0.0169 | 4.89E-11 |
| rs1493696 | 1 | 120505532 | 119962909 | G | A | NOTCH2 | 0.1323 | 0.1283 | 0.0223 | 8.45E-09 |
| rs570794153 | 1 | 198598389 | 198629259 | G | GA | PTPRC | 0.1397 | -0.1454 | 0.0220 | 3.89E-11 |
| rs12128789 | 1 | 212852832 | 212679490 | T | C | BATF3 | 0.1318 | 0.1270 | 0.0215 | 3.73E-09 |
| rs1881146 | 2 | 12634794 | 12494668 | A | T | 2p24 | 0.3105 | -0.0952 | 0.0174 | 4.57E-08 |
| rs12464462 | 2 | 60636416 | 60409281 | A | G | BCL11A | 0.4104 | -0.0880 | 0.0143 | 8.61E-10 |
| rs13018977 | 2 | 242278007 | 241338592 | T | A | SEPT2 | 0.2246 | 0.1003 | 0.0179 | 2.01E-08 |
| rs12644686 | 4 | 185302902 | 184381748 | C | G | IRF2 | 0.1938 | -0.1077 | 0.0193 | 2.44E-08 |
| rs114378220 | 5 | 110566360 | 111230662 | C | T | CAMK4 | 0.0703 | 0.1779 | 0.0304 | 5.11E-09 |
| rs766751473 | 5 | 131766763 | 132431071 | TGTGATACCCCAA | T | IRF1 | 0.4039 | 0.0802 | 0.0141 | 1.31E-08 |
| rs1050979 | 6 | 410417 | 410417 | A | G | IRF4 | 0.5151 | 0.1062 | 0.0141 | 5.65E-14 |
| rs7805218 | 7 | 20378801 | 20339178 | G | A | ITGB8 | 0.3077 | 0.0954 | 0.0166 | 8.67E-09 |
| rs7795896 | 7 | 117086613 | 117446559 | C | T | CFTR | 0.6916 | -0.1354 | 0.0164 | 1.58E-16 |
| rs1947178 | 8 | 59872177 | 58959618 | A | G | TOX | 0.7919 | -0.1033 | 0.0171 | 1.67E-09 |
| rs13259300 | 8 | 120082941 | 119070702 | A | C | COLEC10 | 0.5983 | -0.0922 | 0.0147 | 3.28E-10 |
| rs3802214 | 8 | 141616183 | 140606084 | T | C | AGO2 | 0.7988 | -0.1066 | 0.0192 | 2.96E-08 |
| rs541856133 | 9 | 135936325 | 133060938 | C | T | CEL | 0.0015 | 1.1022 | 0.1311 | 4.08E-17 |
| rs537544 | 10 | 8108382 | 8066419 | C | T | GATA3 | 0.6156 | -0.0864 | 0.0150 | 8.77E-09 |
| rs78325861 | 10 | 72378489 | 70618733 | C | G | PRF1 | 0.0390 | -0.2821 | 0.0422 | 2.31E-11 |
| rs140215710 | 11 | 2903060 | 2881830 | G | A | CDKN1C | 0.0103 | 0.4126 | 0.0753 | 4.19E-08 |
| rs7936434 | 11 | 76293805 | 76582761 | G | C | EMSY | 0.4643 | 0.0769 | 0.0140 | 3.58E-08 |
| rs607703 | 11 | 128604174 | 128734279 | C | T | FLI1 | 0.4844 | 0.0920 | 0.0143 | 1.17E-10 |
| rs7301381 | 12 | 9123932 | 8971336 | T | C | KLRG1 | 0.4637 | -0.0936 | 0.0143 | 5.25E-11 |
| rs238265 | 13 | 42930659 | 42356523 | T | G | AKAP11 | 0.6952 | -0.0908 | 0.0152 | 2.08E-09 |
| rs4238595 | 16 | 20343091 | 20331769 | T | C | GP2 | 0.6805 | -0.1121 | 0.0167 | 1.93E-11 |
| rs8046043 | 16 | 80284024 | 80250127 | G | C | 16q23 | 0.3916 | -0.0846 | 0.0152 | 2.49E-08 |
| rs61759532 | 17 | 7240391 | 7337072 | C | T | NEURL4 | 0.2346 | 0.1184 | 0.0186 | 1.91E-10 |
| rs11651753 | 17 | 46029089 | 47951723 | C | T | PRR15L | 0.4198 | -0.0869 | 0.0142 | 9.96E-10 |
| rs57209021 | 17 | 66164918 | 68168777 | C | T | 17q24 | 0.2256 | 0.1007 | 0.0183 | 3.73E-08 |
| rs74203920 | 21 | 45714294 | 44294411 | C | T | AIRE | 0.0198 | 0.3139 | 0.0557 | 1.77E-08 |

## Table 3.3. Most probable variants for 136 fine mapped T1D risk signals.

| Marker | Chrom. | Position (hg19) | Position (hg38) | Allele Ref. | Allele Alt. | Signal name | Alt. AF | PPA | Previous studies Index ($r^2$) | Previous studies PMID |
|---|---|---|---|---|---|---|---|---|---|---|
| rs10751776 | 1 | 25296743 | 24970252 | A | C | RUNX3 (1:25296743:A:C) | 0.5099 | 0.0406 | | |
| rs574384 | 1 | 36087661 | 35622060 | C | A | PSMB2 (1:36087661:C:A) | 0.8947 | 0.1953 | | |
| rs12742756 | 1 | 38347417 | 37881745 | A | G | INPP5B (1:38347417:A:G) | 0.4282 | 0.0705 | | |
| rs855330 | 1 | 64113889 | 63648218 | T | C | PGM1 (1:64113889:T:C) | 0.2594 | 0.1612 | | |
| rs150709401 | 1 | 114135880 | 113593258 | A | G | PTPN22 (1:114135880:A:G) | 0.0096 | 0.2576 | | |
| rs2476601 | 1 | 114377568 | 113834946 | A | G | PTPN22 (1:114377568:A:G) | 0.8860 | 0.9820 | rs2476601 (1.00) | 25751624 |
| rs1493696 | 1 | 120505532 | 119962909 | G | A | NOTCH2 (1:120505532:G:A) | 0.1323 | 0.0467 | | |
| rs10801128 | 1 | 192515849 | 192546719 | A | G | RGS1 (1:192515849:A:G) | 0.7174 | 0.1116 | rs2816316 (0.50) | 19430480 |
| rs570794153 | 1 | 198598389 | 198629259 | G | GA | PTPRC (1:198598389:G:GA) | 0.1397 | 0.8525 | | |
| rs10920052 | 1 | 200820610 | 200851482 | C | A | CAMSAP2 (1:200820610:C:A) | 0.2202 | 0.2400 | rs6691977 (0.94) | 25751624 |
| rs58579536 | 1 | 206746125 | 206572797 | A | G | IL10 (1:206746125:A:G) | 0.1318 | 0.5692 | | |
| rs77599401 | 1 | 206849509 | 206676164 | G | A | IL10 (1:206849509:G:A) | 0.0102 | 0.0273 | | |
| rs3024493 | 1 | 206943968 | 206770623 | C | A | IL10 (1:206943968:C:A) | 0.1543 | 0.5250 | rs3024505 (1.00) | 25751624 |
| rs12128789 | 1 | 212852832 | 212679490 | T | C | BATF3 (1:212852832:T:C) | 0.1318 | 0.3463 | | |
| rs1881146 | 2 | 12634794 | 12494668 | A | T | 2p24 (2:12634794:A:T) | 0.3105 | 0.3425 | | |
| rs55893453 | 2 | 25139367 | 24916498 | A | G | ADCY3 (2:25139367:A:G) | 0.2018 | 0.1268 | rs478222 (0.13) | 21980299 |
| rs77146844 | 2 | 60146784 | 59919649 | C | G | BCL11A (2:60146784:C:G) | 0.0373 | 0.8602 | | |
| rs13034691 | 2 | 60216169 | 59989034 | A | G | BCL11A (2:60216169:A:G) | 0.1004 | 0.0811 | | |
| rs12464462 | 2 | 60636416 | 60409281 | A | G | BCL11A (2:60636416:A:G) | 0.4104 | 0.0996 | | |
| rs4490209 | 2 | 100766711 | 100150249 | C | G | AFF3 (2:100766711:C:G) | 0.3599 | 0.0569 | rs9653442 (0.40) | 17554260 |
| rs567088138 | 2 | 112020548 | 111262971 | G | A | ACOXL (2:112020548:G:A) | 0.0098 | 0.5189 | rs4849135 (0) | 25751624 |
| rs2111485 | 2 | 163110536 | 162254026 | A | G | IFIH1 (2:163110536:A:G) | 0.6042 | 0.9076 | rs2111485 (1.00) | 25751624 |
| rs35667974 | 2 | 163124637 | 162268127 | T | C | IFIH1 (2:163124637:T:C) | 0.0208 | 0.9784 | rs35667974 (1.00) | 25751624 |
| rs35337543 | 2 | 163136505 | 162279995 | C | G | IFIH1 (2:163136505:C:G) | 0.0119 | 0.5204 | rs72871627 (1.00) | 25751624 |
| rs75671397 | 2 | 163139399 | 162282889 | A | C | IFIH1 (2:163139399:A:C) | 0.0020 | 0.4749 | | |
| rs10931480 | 2 | 191954047 | 191089321 | G | A | STAT4 (2:191954047:G:A) | 0.8372 | 0.3805 | | |
| rs7423529 | 2 | 192003031 | 191138305 | T | C | STAT4 (2:192003031:T:C) | 0.6187 | 0.0849 | rs7574865 (0.08) | 21980299 |
| rs143829908 | 2 | 204520098 | 203655375 | T | C | CTLA4 (2:204520098:T:C) | 0.0141 | 0.4778 | | |
| rs231389 | 2 | 204634730 | 203770007 | C | T | CTLA4 (2:204634730:C:T) | 0.7786 | 0.5200 | | |
| rs78847176 | 2 | 204725944 | 203861221 | C | T | CTLA4 (2:204725944:C:T) | 0.0072 | 0.7462 | | |
| rs231725 | 2 | 204740675 | 203875952 | G | A | CTLA4 (2:204740675:G:A) | 0.3611 | 0.3416 | rs3087243 (0.40) | 25751624 |
| rs13018977 | 2 | 242278007 | 241338592 | T | A | SEPT2 (2:242278007:T:A) | 0.2246 | 0.1535 | | |
| rs28607988 | 3 | 45913338 | 45871846 | A | C | CCR5 (3:45913338:A:C) | 0.0829 | 0.1036 | rs113010081 (0.01) | 25751624 |
| rs7668577 | 4 | 26123319 | 26121697 | A | C | RBPJ (4:26123319:A:C) | 0.3117 | 0.0953 | rs10517086 (0.94) | 19430480 |
| 4:123235016:TTA:T | 4 | 123235016 | 122313861 | TTA | T | IL2 (4:123235016:TTA:T) | 0.0742 | 0.0362 | | |
| rs77516441 | 4 | 123243594 | 122322439 | A | ATC | IL2 (4:123243594:A:ATC) | 0.6117 | 0.0765 | rs4505848 (0.84) | 19430480 |
| rs2611211 | 4 | 166575439 | 165654287 | C | T | CPE (4:166575439:C:T) | 0.8237 | 0.0722 | rs2611215 (1.00) | 25751624 |
| rs12644686 | 4 | 185302902 | 184381748 | C | G | IRF2 (4:185302902:C:G) | 0.1938 | 0.5525 | | |
| rs2303137 | 5 | 35895725 | 35895623 | A | T | IL7R (5:35895725:A:T) | 0.4419 | 0.0659 | rs11954020 (0.37) | 25751624 |
| rs114378220 | 5 | 110566360 | 111230662 | C | T | CAMK4 (5:110566360:C:T) | 0.0703 | 0.9690 | | |
| rs72663304 | 5 | 110926439 | 111590742 | A | C | CAMK4 (5:110926439:A:C) | 0.0013 | 0.0533 | | |
| rs766751473 | 5 | 131766763 | 132431071 | TGTGATACCCCAA | T | IRF1 (5:131766763:TGTGATACCCCAA:T) | 0.4039 | 0.0892 | | |
| rs1050979 | 6 | 410417 | 410417 | A | G | IRF4 (6:410417:A:G) | 0.5151 | 0.3473 | | |
| rs60649859 | 6 | 90854533 | 90144814 | A | T | BACH2 (6:90854533:A:T) | 0.0561 | 0.1359 | | |
| rs6908626 | 6 | 91005743 | 90296024 | G | T | BACH2 (6:91005743:G:T) | 0.1656 | 0.8210 | rs72928038 (0.99) | 25751624 |
| 6:126690257:AT:A | 6 | 126690257 | 126369111 | AT | A | CENPW (6:126690257:AT:A) | 0.4439 | 0.1533 | rs1538171 (0.98) | 25751624 |
| rs142761146 | 6 | 138004508 | 137683371 | T | A | TNFAIP3 (6:138004508:T:A) | 0.1700 | 0.4266 | rs10499194 (0.60) | 21980299 |
| rs4548024 | 6 | 138165744 | 137844607 | T | C | TNFAIP3 (6:138165744:T:C) | 0.2336 | 0.9702 | | |
| rs182429 | 6 | 159469574 | 159048542 | A | G | TAGAP (6:159469574:A:G) | 0.5693 | 0.1303 | rs1738074 (1.00) | 21980299 |
| rs11756073 | 6 | 170363580 | 170048356 | G | A | 6q27 (6:170363580:G:A) | 0.1679 | 0.0389 | rs924043 (0.60) | 21980299 |
| rs7805218 | 7 | 20378801 | 20339178 | G | A | ITGB8 (7:20378801:G:A) | 0.3077 | 0.0730 | | |
| rs17323934 | 7 | 26904330 | 26864711 | C | G | SKAP2 (7:26904330:C:G) | 0.2231 | 0.0738 | rs7804356 (0.95) | 19430480 |
| rs876039 | 7 | 50308811 | 50269215 | G | C | IKZF1 (7:50308811:G:C) | 0.3125 | 0.1065 | | |
| rs7809377 | 7 | 50442726 | 50375028 | G | A | IKZF1 (7:50442726:G:A) | 0.1157 | 0.1109 | rs62447205 (0.32) | 25751624 |
| rs34414436 | 7 | 51018492 | 50950795 | GT | G | COBL (7:51018492:GT:G) | 0.9592 | 0.0817 | rs10277986 (0.86) | 25751624 |
| rs7795896 | 7 | 117086613 | 117446559 | C | T | CFTR (7:117086613:C:T) | 0.6916 | 0.6287 | | |
| rs1947178 | 8 | 59872177 | 58959618 | A | G | TOX (8:59872177:A:G) | 0.7919 | 0.5298 | | |
| rs13259300 | 8 | 120082941 | 119070702 | A | C | COLEC10 (8:120082941:A:C) | 0.5983 | 0.1982 | | |
| rs3802214 | 8 | 141616183 | 140606084 | T | C | AGO2 (8:141616183:T:C) | 0.7988 | 0.1426 | | |
| rs1574285 | 9 | 4283137 | 4283137 | G | T | GLIS3 (9:4283137:G:T) | 0.5911 | 0.2140 | rs6476839 (0.95) | 25751624 |
| rs541856133 | 9 | 135936325 | 133060938 | C | T | CEL (9:135936325:C:T) | 0.0015 | 1.0000 | | |
| rs6602398 | 10 | 6082953 | 6040990 | G | T | IL2RA (10:6082953:G:T) | 0.3102 | 0.5468 | | |
| rs61839660 | 10 | 6094697 | 6052734 | C | T | IL2RA (10:6094697:C:T) | 0.0851 | 0.7793 | rs61839660 (1.00) | 25751624 |
| rs706778 | 10 | 6098949 | 6056986 | C | T | IL2RA (10:6098949:C:T) | 0.4302 | 0.6141 | rs10795791 (0.92) | 25751624 |
| rs41295159 | 10 | 6148535 | 6106572 | C | G | IL2RA (10:6148535:C:G) | 0.0090 | 0.3950 | rs41295121 (0.95) | 25751624 |
| rs77710246 | 10 | 6166282 | 6124319 | T | G | IL2RA (10:6166282:T:G) | 0.0416 | 0.8068 | | |
| rs947474 | 10 | 6390450 | 6348488 | G | A | IL2RA (10:6390450:G:A) | 0.8164 | 0.8791 | | |
| rs11596750 | 10 | 6471816 | 6429854 | T | C | PRKCQ (10:6471816:T:C) | 0.7093 | 0.2436 | rs11258747 (0.71) | 21980299 |
| rs537544 | 10 | 8108382 | 8066419 | C | T | GATA3 (10:8108382:C:T) | 0.6156 | 0.2562 | | |
| rs58763752 | 10 | 32965617 | 32676689 | A | G | NRP1 (10:32965617:A:G) | 0.1019 | 0.0206 | | |
| rs11009222 | 10 | 33374735 | 33085807 | A | G | NRP1 (10:33374735:A:G) | 0.1951 | 0.1288 | rs722988 (0.30) | 25371288 |
| rs78325861 | 10 | 72378489 | 70618733 | C | G | PRF1 (10:72378489:C:G) | 0.0390 | 0.3697 | | |
| rs7068821 | 10 | 90051035 | 88291278 | G | T | RNLS (10:90051035:G:T) | 0.2511 | 0.4483 | rs12416116 (0.82) | 25751624 |
| rs7948458 | 11 | 2172830 | 2151600 | A | C | INS (11:2172830:A:C) | 0.8083 | 0.2868 | | |

## Table 3.3. Most probable variants for 136 fine mapped T1D risk signals, continued.

| Marker | Chrom. | Position (hg19) | Position (hg38) | Allele Ref. | Allele Alt. | Signal name | Alt. AF | PPA | Previous studies Index ($r^2$) | Previous studies PMID |
|---|---|---|---|---|---|---|---|---|---|---|
| rs689 | 11 | 2182224 | 2160994 | A | T | INS (11:2182224:A:T) | 0.7306 | 0.8613 | rs689 (1.00) | 25751624 |
| rs10770143 | 11 | 2195267 | 2174037 | C | T | INS (11:2195267:C:T) | 0.6232 | 0.3048 | rs72853903 (0.67) | 25751624 |
| rs140215710 | 11 | 2903060 | 2881830 | G | A | CDKN1C (11:2903060:G:A) | 0.0103 | 0.9874 | | |
| rs663743 | 11 | 64107735 | 64340263 | G | A | BAD (11:64107735:G:A) | 0.3490 | 0.4887 | rs694739 (0.83) | 25371288 |
| rs7936434 | 11 | 76293805 | 76582761 | G | C | EMSY (11:76293805:G:C) | 0.4643 | 0.1645 | | |
| rs55836957 | 11 | 128168631 | 128298736 | C | A | FLI1 (11:128168631:C:A) | 0.2253 | 0.1309 | | |
| rs607703 | 11 | 128604174 | 128734279 | C | T | FLI1 (11:128604174:C:T) | 0.4844 | 0.4068 | | |
| rs7301381 | 12 | 9123932 | 8971336 | T | C | KLRG1 (12:9123932:T:C) | 0.4637 | 0.0254 | | |
| rs35954981 | 12 | 9878144 | 9725548 | TA | T | CD69 (12:9878144:TA:T) | 0.5060 | 0.0477 | rs917911 (0.48) | 25751624 |
| rs12232003 | 12 | 53593632 | 53199848 | T | C | ITGB7 (12:53593632:T:C) | 0.0879 | 0.1685 | rs11170466 (0.66) | 25371288 |
| rs1701704 | 12 | 56412487 | 56018703 | T | G | IKZF4 (12:56412487:T:G) | 0.3394 | 0.3525 | rs2292239 (0.88) | 19430480 |
| rs71459332 | 12 | 56418678 | 56024894 | C | T | IKZF4 (12:56418678:C:T) | 0.0708 | 0.6281 | | |
| rs78865465 | 12 | 56434146 | 56040362 | G | A | IKZF4 (12:56434146:G:A) | 0.0589 | 0.3509 | | |
| rs3809113 | 12 | 57848708 | 57454925 | C | G | CYP27B1 (12:57848708:C:G) | 0.6878 | 0.0168 | rs10877012 (0.07) | 21980299 |
| rs3184504 | 12 | 111884608 | 111446804 | T | C | SH2B3 (12:111884608:T:C) | 0.5329 | 0.4242 | rs653178 (1.00) | 25751624 |
| rs570074821 | 12 | 111926961 | 111489157 | A | G | SH2B3 (12:111926961:A:G) | 0.0019 | 0.2420 | | |
| rs238265 | 13 | 42930659 | 42356523 | T | G | AKAP11 (13:42930659:T:G) | 0.6952 | 0.1942 | | |
| rs9573641 | 13 | 76303932 | 75729796 | T | C | LMO7 (13:76303932:T:C) | 0.4486 | 0.1167 | rs539514 (0.70) | 21980299 |
| rs9517712 | 13 | 100079833 | 99427579 | T | C | GPR183 (13:100079833:T:C) | 0.7410 | 0.8330 | rs9585056 (0.98) | 25751624 |
| rs2104047 | 14 | 68754417 | 68287700 | T | C | RAD51B (14:68754417:T:C) | 0.7083 | 0.1810 | rs911263 (1.00) | 25751624 |
| rs17106304 | 14 | 69260511 | 68793794 | C | G | ZFP36L1 (14:69260511:C:G) | 0.6560 | 0.4150 | rs1465788 (0.75) | 19430480 |
| rs1350275 | 14 | 98489306 | 98022969 | T | G | 14q32 (14:98489306:T:G) | 0.6984 | 0.0420 | rs1456988 (0.97) | 25751624 |
| rs56994090 | 14 | 101306447 | 100840110 | T | C | DLK1 (14:101306447:T:C) | 0.4311 | 0.9988 | rs56994090 (1.00) | 25751624 |
| rs3783355 | 14 | 101308958 | 100842621 | G | A | DLK1 (14:101308958:G:A) | 0.2406 | 0.5634 | | |
| rs35134214 | 15 | 38865321 | 38573120 | CTG | C | RASGRP1 (15:38865321:CTG:C) | 0.1954 | 0.0728 | rs72727394 (0.89) | 25751624 |
| rs34593439 | 15 | 79234957 | 78942615 | G | A | CTSH (15:79234957:G:A) | 0.1082 | 0.3294 | rs34593439 (1.00) | 25751624 |
| rs12927355 | 16 | 11194771 | 11100914 | C | T | DEXI (16:11194771:C:T) | 0.3162 | 0.3532 | rs12927355 (1.00) | 25751624 |
| rs66718203 | 16 | 11428643 | 11334786 | C | G | DEXI (16:11428643:C:G) | 0.1740 | 0.0507 | rs193778 (0.07) | 25751624 |
| rs4238595 | 16 | 20343091 | 20331769 | T | C | GP2 (16:20343091:T:C) | 0.6805 | 0.4521 | | |
| rs34835 | 16 | 28499291 | 28487970 | A | G | IL27 (16:28499291:A:G) | 0.4299 | 0.3799 | | |
| rs231972 | 16 | 28538336 | 28527015 | A | C | IL27 (16:28538336:A:C) | 0.1180 | 0.2249 | rs151234 (0.83) | 25751624 |
| rs7187776 | 16 | 28857645 | 28846324 | A | G | IL27 (16:28857645:A:G) | 0.4090 | 0.1714 | | |
| rs55993634 | 16 | 75236763 | 75202865 | C | G | CTRB2 (16:75236763:C:G) | 0.0849 | 0.3716 | rs8056814 (0.87) | 25751624 |
| rs8046043 | 16 | 80284024 | 80250127 | G | C | 16q23 (16:80284024:G:C) | 0.3916 | 0.8014 | | |
| rs9891059 | 17 | 4300706 | 4397411 | T | G | UBE2G1 (17:4300706:T:G) | 0.1659 | 0.1207 | rs9906760 (0.30) | 25371288 |
| rs61759532 | 17 | 7240391 | 7337072 | C | T | NEURL4 (17:7240391:C:T) | 0.2346 | 0.3785 | | |
| rs62062613 | 17 | 7642035 | 7738717 | C | T | NEURL4 (17:7642035:C:T) | 0.1200 | 0.1339 | | |
| rs56750287 | 17 | 38062944 | 39906691 | A | C | GSDMB (17:38062944:A:C) | 0.1652 | 0.0774 | rs12453507 (0.84) | 25751624 |
| rs112401631 | 17 | 38764524 | 40608272 | T | A | CCR7 (17:38764524:T:A) | 0.0191 | 0.2938 | rs757411 (0.01) | 25751624 |
| rs35327136 | 17 | 43487574 | 45410208 | C | A | MAPT (17:43487574:C:A) | 0.1555 | 0.0142 | rs1052553 (0.69) | 25751624 |
| rs11651753 | 17 | 46029089 | 47951723 | C | T | PRR15L (17:46029089:C:T) | 0.4198 | 0.4572 | | |
| rs57209021 | 17 | 66164918 | 68168777 | C | T | 17q24 (17:66164918:C:T) | 0.2256 | 0.4710 | | |
| rs80262450 | 18 | 12818922 | 12818923 | G | A | PTPN2 (18:12818922:G:A) | 0.1214 | 0.5460 | rs1893217 (0.66) | 25751624 |
| rs2542162 | 18 | 12820900 | 12820901 | C | T | PTPN2 (18:12820900:C:T) | 0.4078 | 0.1309 | rs12971201 (1.00) | 25751624 |
| rs62097857 | 18 | 12857758 | 12857759 | G | A | PTPN2 (18:12857758:G:A) | 0.0388 | 0.5908 | | |
| rs1808094 | 18 | 67526026 | 69858790 | T | C | CD226 (18:67526026:T:C) | 0.5239 | 0.3313 | rs1615504 (1.00) | 25751624 |
| rs34536443 | 19 | 10463118 | 10352442 | G | C | TYK2 (19:10463118:G:C) | 0.0427 | 1.0000 | rs34536443 (1.00) | 25751624 |
| rs12720356 | 19 | 10469975 | 10359299 | A | C | TYK2 (19:10469975:A:C) | 0.0936 | 0.9653 | rs12720356 (1.00) | 25751624 |
| rs60652743 | 19 | 47205707 | 46702450 | A | G | PRKD2 (19:47205707:A:G) | 0.1657 | 0.4358 | rs402072 (0.98) | 25751624 |
| rs71353922 | 19 | 47211296 | 46708039 | G | T | PRKD2 (19:47211296:G:T) | 0.0853 | 0.1878 | | |
| rs601338 | 19 | 49206674 | 48703417 | G | A | FUT2 (19:49206674:G:A) | 0.4788 | 0.3078 | rs516246 (1.00) | 25751624 |
| rs6043405 | 20 | 1615544 | 1634898 | T | C | SIRPG (20:1615544:T:C) | 0.6608 | 0.2560 | rs6043409 (0.98) | 25751624 |
| rs2250268 | 20 | 1657741 | 1677095 | G | A | SIRPG (20:1657741:G:A) | 0.8301 | 0.2658 | | |
| rs12483582 | 21 | 43817769 | 42397660 | G | A | UBASH3A (21:43817769:G:A) | 0.4139 | 0.9714 | | |
| rs9981624 | 21 | 43825722 | 42405613 | G | C | UBASH3A (21:43825722:G:C) | 0.3291 | 0.6382 | rs11203202 (1.00) | 25751624 |
| rs1893592 | 21 | 43855067 | 42434957 | A | C | UBASH3A (21:43855067:A:C) | 0.3036 | 0.9627 | | |
| rs1055311 | 21 | 45709568 | 44289685 | C | T | AIRE (21:45709568:C:T) | 0.2968 | 0.0778 | | |
| rs74203920 | 21 | 45714294 | 44294411 | C | T | AIRE (21:45714294:C:T) | 0.0198 | 0.9901 | | |
| rs2074204 | 22 | 30403996 | 30008007 | C | T | HORMAD2 (22:30403996:C:T) | 0.2634 | 0.2376 | | |
| rs4820827 | 22 | 30477589 | 30081600 | T | C | HORMAD2 (22:30477589:T:C) | 0.6209 | 0.1346 | rs4820830 (1.00) | 25751624 |
| rs228963 | 22 | 37535948 | 37139908 | A | G | C1QTNF6 (22:37535948:A:G) | 0.4388 | 0.2645 | | |
| rs229527 | 22 | 37581485 | 37185445 | C | A | C1QTNF6 (22:37581485:C:A) | 0.4164 | 0.4042 | rs229533 (0.97) | 25751624 |

## Table 3.4. Fine mapped variants mapping in protein coding regions.

| Marker | Chrom. | Position | Signal name | PPA | Gene | Mutation type | Mutation | PolyPhen | SIFT |
|---|---|---|---|---|---|---|---|---|---|
| rs34536443 | 19 | 10463117 | TYK2 (19:10463118:G:C) | 1.0000 | TYK2 | missense | P1104A | probably damaging | deleterious |
| rs74203920 | 21 | 45714293 | AIRE (21:45714294:C:T) | 0.9901 | AIRE | missense | R471C | possibly damaging | deleterious |
| rs2476601 | 1 | 114377567 | PTPN22 (1:114377568:A:G) | 0.9820 | PTPN22 | missense | R620W | benign | tolerated |
| rs35667974 | 2 | 163124636 | IFIH1 (2:163124637:T:C) | 0.9784 | IFIH1 | missense | I923V | probably damaging | tolerated |
| rs12720356 | 19 | 10469974 | TYK2 (19:10469975:A:C) | 0.9653 | TYK2 | missense | I684S | possibly damaging | deleterious |
| rs74162075 | 2 | 163167418 | IFIH1 (2:163139399:A:C) | 0.4348 | IFIH1 | missense | N160D | probably damaging | deleterious |
| rs3184504 | 12 | 111884607 | SH2B3 (12:111884608:T:C) | 0.4242 | SH2B3 | missense | W262R | benign | tolerated |
| rs229527 | 22 | 37581484 | C1QTNF6 (22:37581485:C:A) | 0.4042 | C1QTNF6 | missense | G21V | benign | tolerated low confidence |
| rs601338 | 19 | 49206673 | FUT2 (19:49206674:G:A) | 0.3078 | FUT2 | stop gained | - | - | - |
| rs35947132 | 10 | 72360386 | PRF1 (10:72378489:C:G) | 0.2779 | PRF1 | missense | A91V | probably damaging | deleterious |
| rs6043409 | 20 | 1616205 | SIRPG (20:1615544:T:C) | 0.2095 | SIRPG | missense | V263A | benign | tolerated |
| rs681343 | 19 | 49206461 | FUT2 (19:49206674:G:A) | 0.1811 | FUT2 | stop gained | - | - | - |
| rs2289702 | 15 | 79237292 | CTSH (15:79234957:G:A) | 0.1751 | CTSH | missense | G11R | probably damaging | deleterious low confidence |
| rs3842753 | 11 | 2181059 | INS (11:2182224:A:T) | 0.1387 | INS | missense | H77P | benign | tolerated |
| rs1990760 | 2 | 163124050 | IFIH1 (2:163110536:A:G) | 0.0907 | IFIH1 | missense | A946T | benign | tolerated |
| rs2202683 | 1 | 212873073 | BATF3 (1:212852832:T:C) | 0.0782 | BATF3 | missense | V11I | benign | tolerated low confidence |
| rs34655914 | 1 | 38397340 | INPP5B (1:38347417:A:G) | 0.0650 | INPP5B | missense | S259L | benign | tolerated |
| rs35267671 | 1 | 38397368 | INPP5B (1:38347417:A:G) | 0.0547 | INPP5B | missense | G250C | probably damaging | deleterious |
| rs1050152 | 5 | 131676319 | IRF1 (5:131766763:TGTGATACCCCAA:T) | 0.0501 | SLC22A4 | missense | L503F | benign | tolerated |
| rs11557467 | 17 | 38028633 | GSDMB (17:38062944:A:C) | 0.0314 | ZPBP2 | missense | S173I | benign | tolerated |
| rs763361 | 18 | 67531641 | CD226 (18:67526026:T:C) | 0.0239 | CD226 | missense | S307G | benign | tolerated |
| rs10797300 | 16 | 28607195 | IL27 (16:28538336:A:C) | 0.0200 | SULT1A2 | missense | P19R | probably damaging | deleterious |
| rs1445898 | 5 | 35910528 | IL7R (5:35895725:A:T) | 0.0168 | CAPSL | missense | R85Q | benign | tolerated |
| rs2305479 | 17 | 38062216 | GSDMB (17:38062944:A:C) | 0.0147 | GSDMB | missense | G304R | probably damaging | deleterious |

## Table 3.5. Metadata for snATACseq experiments.

| Exp. ID | Donor ID | Sex | Age | BMI | Race | Tissue type | Tissue status | Tissue source | Technology | Duplicate reads (%) | Mito. reads (%) | Reads in peaks (%) | Final read count | # cells (post QC) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| nPOD-6282 | 6282 | Male | 14 | 41.5 | Caucasian | Whole pancreas | Frozen | JDRF nPOD | Combinatorial barcoding | 75.84% | 5.70% | 49.72% | 29,962,338 | 6,183 |
| nPOD-6251-1 | 6251 | Female | 33 | 29.5 | Caucasian | Whole pancreas | Frozen | JDRF nPOD | Combinatorial barcoding | 76.31% | 4.74% | 58.11% | 29,892,458 | 4,916 |
| nPOD-6251-2 | 6251 | Female | 33 | 29.5 | Caucasian | Whole pancreas | Frozen | JDRF nPOD | Combinatorial barcoding | 70.68% | 3.59% | 62.17% | 33,973,568 | 5,545 |
| nPOD-6004-1 | 6004 | Male | 33 | 30.9 | Caucasian | Whole pancreas | Frozen | JDRF nPOD | Combinatorial barcoding | 71.96% | 17.84% | 55.74% | 19,779,400 | 4,688 |
| nPOD-6004-2 | 6004 | Male | 33 | 30.9 | Caucasian | Whole pancreas | Frozen | JDRF nPOD | 10x Chromium Single Cell ATAC (v1.0) | 40.39% | 1.64% | 33.13% | 226,837,754 | 6,342 |
| nPOD-6007 | 6007 | Male | 9 | 20.1 | African American | Whole pancreas | Frozen | JDRF nPOD | Combinatorial barcoding | 71.58% | 11.59% | 51.33% | 35,464,636 | 6,435 |
| IIDP-AFC2208-1 | AFC2208 | Male | 32 | 32.3 | Caucasian | Purified pancreatic islets | Fresh | IIDP | Combinatorial barcoding | 77.33% | 1.00% | 67.99% | 52,164,796 | 5,582 |
| IIDP-AFC2208-2 | AFC2208 | Male | 32 | 32.3 | Caucasian | Purified pancreatic islets | Frozen | IIDP | Combinatorial barcoding | 13.31% | 1.61% | 42.31% | 36,216,124 | 4,186 |
| IIDP-AFEA331 | AFEA331 | Male | 45 | 29.3 | African American | Purified pancreatic islets | Fresh | IIDP | Combinatorial barcoding | 29.83% | 1.47% | 71.57% | 37,904,450 | 4,515 |
| IIDP-AFEP022 | AFEP022 | Male | 62 | 36.1 | African American | Purified pancreatic islets | Fresh | IIDP | Combinatorial barcoding | 45.92% | 8.87% | 73.11% | 57,423,960 | 6,031 |
| IIDP-AEHU156 | AEHU156 | Male | 37 | 29.5 | Caucasian | Purified pancreatic islets | Frozen | IIDP | Combinatorial barcoding | 47.57% | 6.66% | 44.83% | 72,522,452 | 10,530 |
| HemaCare-D205220 | D205220 | Male | 47 | - | Caucasian | Peripheral blood mononuclear cells | Frozen | HemaCare | 10x Chromium Single Cell ATAC (v1.0) | 28.44% | 0.77% | 44.85% | 214,376,844 | 8,449 |
| HemaCare-D105 | D105 | Male | 46 | - | Caucasian | Peripheral blood mononuclear cells | Frozen | HemaCare | 10x Chromium Single Cell ATAC (v1.0) | 34.23% | 0.29% | 46.07% | 204,447,412 | 6,308 |
| HemaCare-D147558 | D147558 | Male | 53 | - | Caucasian | Peripheral blood mononuclear cells | Frozen | HemaCare | 10x Chromium Single Cell ATAC (v1.0) | 38.43% | 0.33% | 44.82% | 149,151,488 | 6,136 |
| HemaCare-D270271 | D270271 | Male | 45 | - | Caucasian | Peripheral blood mononuclear cells | Frozen | HemaCare | 10x Chromium Single Cell ATAC (v1.0) | 53.51% | 0.38% | 52.68% | 105,523,078 | 4,660 |
| HemaCare-D182364 | D182364 | Male | 48 | - | Caucasian | Peripheral blood mononuclear cells | Frozen | HemaCare | 10x Chromium Single Cell ATAC (v1.0) | 40.98% | 0.16% | 36.49% | 144,472,996 | 3,942 |
| HemaCare-D273097 | D273097 | Female | 54 | - | Caucasian | Peripheral blood mononuclear cells | Frozen | HemaCare | 10x Chromium Single Cell ATAC (v1.0) | 30.18% | 0.32% | 38.55% | 154,634,340 | 5,323 |
| 10x-5k-1 | 10x_1 | Male | - | - | - | Peripheral blood mononuclear cells | - | - | 10x Chromium Single Cell ATAC (v1.0) | 52.34% | 0.24% | 54.32% | 189,560,042 | 3,577 |
| 10x-10k-1 | 10x_1 | Male | - | - | - | Peripheral blood mononuclear cells | - | - | 10x Chromium Single Cell ATAC (v1.0) | 43.07% | 0.16% | 42.46% | 381,665,308 | 6,346 |
| 10x-5k-2 | 10x_2 | Male | - | - | - | Peripheral blood mononuclear cells | - | - | 10x Chromium Single Cell ATAC (v1.0) | 42.23% | 0.16% | 51.73% | 118,206,092 | 3,374 |
| 10x-5k-3 | 10x_2 | Male | - | - | - | Peripheral blood mononuclear cells | - | - | 10x Chromium Single Cell ATAC (Next GEM v1.1) | 53.03% | 0.14% | 59.60% | 182,462,524 | 4,184 |
| 10x-10k-2 | 10x_2 | Male | - | - | - | Peripheral blood mononuclear cells | - | - | 10x Chromium Single Cell ATAC (v1.0) | 42.87% | 0.12% | 50.02% | 238,627,438 | 6,271 |
| 10x-10k-3 | 10x_2 | Male | - | - | - | Peripheral blood mononuclear cells | - | - | 10x Chromium Single Cell ATAC (Next GEM v1.1) | 46.76% | 0.16% | 52.03% | 380,318,842 | 8,034 |

## Table 3.6. Marker genes used for cell type identification.

| Cell type | Abbrev. | Marker genes | Identified modality | References (PMID) | Additional marker genes | Additional notes |
|---|---|---|---|---|---|---|
| hematopoietic stem cell | HSC | *CD34* | scRNA-seq | 24497003 | https://panglaodb.se/markers.html?cell_type=%27Hematopoietic%20stem%20cells%27 | *CD34* is listed as a marker gene for hematopoetic stem cells on PanglaoDB. |
| megakaryocyte | MKC | *GATA1, PPBP* | both | 16166640 | https://panglaodb.se/markers.html?cell_type=%27Megakaryocytes%27 | *GATA1* and *PPBP* are listed as a marker genes for megakaryocytes and platelets, respectively, on PanglaoDB. |
| plasmacytoid dendritic | pDC | *PTCRA* | both | 18854153 | https://panglaodb.se/markers.html?cell_type=%27Plasmacytoid%20dendritic%20cells%27 | *PTCRA* is listed as a marker gene for plasmacytoid dendritic on PanglaoDB. |
| pancreatic macrophage | panc. MΦ | *C1QB* | both | 29167450 | https://panglaodb.se/markers.html?cell_type=%27Macrophages%27 | *C1QB* encodes part of the complement protein C1q which is primarily expressed in macrophages. |
| classical monocyte | cMono | *FCAR, CD14* | both | 21937986 | https://panglaodb.se/markers.html?cell_type=%27Monocytes%27 | *FCAR* (CD89) encodes the Fc fragment of the IgA receptor which is a cell surface marker for myeloid lineage cells. |
| non-classical monocyte | ncMono | *HES4, FCGR3B* | both | 21653326 | https://panglaodb.se/markers.html?cell_type=%27Monocytes%27 | *HES4* has higher expression in non-classical monocytes than classical monocytes in the Human Protein Atlas. |
| conventional dendritic | cDC | *TLR3, CD1C* | both | 20561711 | https://panglaodb.se/markers.html?cell_type=%27Dendritic%20cells%27 | *CD1C* is a cell surface marker for sorting conventional dendritic cells. *CD1C* and *TLR3* are listed as marker genes for dendritic cells on PanglaoDB. |
| endothelial | endothel. | *CLDN5* | both | 22038628 | https://panglaodb.se/markers.html?cell_type=%27Endothelial%20cells%27 | *CLDN5* is involved in the regulation of endothelial cell motility. |
| quiescent stellate | qui. stell. | *RGS5* | both | 27667365 | https://panglaodb.se/markers.html?cell_type=%27Pancreatic%20stellate%20cells%27 | *RGS5* is listed as a marker gene for stellate cells on PanglaoDB. |
| activated stellate | act. stell. | *COL6A3* | both | | https://panglaodb.se/markers.html?cell_type=%27Pancreatic%20stellate%20cells%27 | *COL6A3* is listed as a marker gene for stellate cells on PanglaoDB. |
| schwann | schwann | *CDH19* | scRNA-seq | 28512649 | https://panglaodb.se/markers.html?cell_type=%27Peri-islet%20Schwann%20cells%27 | *CDH19* is a marker of schwann cell precursors. |
| ductal | ductal | *CFTR* | both | 27667365 | https://panglaodb.se/markers.html?cell_type=%27Ductal%20cells%27 | *CFTR* encodes a chloride channel and is the gene affected in cystic fibrosis. |
| acinar | acinar | *REG1A* | both | | https://panglaodb.se/markers.html?cell_type=%27Acinar%20cells%27 | *REG1A* is listed as a marker gene for acinar cells on PanglaoDB. |
| $GCG^{high}$ alpha | $GCG^{hi}$ alpha | | both (states in snAT | | | |
| $GCG^{low}$ alpha | $GCG^{lo}$ alpha | *GCG* | AC-seq) | 27667365 | https://panglaodb.se/markers.html?cell_type=%27Alpha%20cells%27 | *GCG* is the hallmark hormone produced by alpha cells. |
| $INS^{high}$ beta | $INS^{hi}$ beta | | both (states in snAT | | | |
| $INS^{low}$ beta | $INS^{lo}$ beta | *INS* | AC-seq) | 27667365 | https://panglaodb.se/markers.html?cell_type=%27Beta%20cells%27 | *INS* is the hallmark hormone produced by beta cells. |
| $SST^{high}$ delta | $SST^{hi}$ delta | | both (states in snAT | | | |
| $SST^{low}$ delta | $SST^{lo}$ delta | *SST* | AC-seq) | 27667365 | https://panglaodb.se/markers.html?cell_type=%27Delta%20cells%27 | *SST* is the hallmark hormone produced by delta cells. |
| gamma | gamma | *PPY* | both | 27667365 | https://panglaodb.se/markers.html?cell_type=%27Gamma%20(PP)%20cells%27 | *PPY* is the hallmark hormone produced by gamma (PP) cells. |
| naive CD4/CD8 T | naïve T | *CD3D, CD4, CD8B* | both (split into CD4 and CD8 in scRNA-seq) | | https://panglaodb.se/markers.html?cell_type=%27T%20cells%27 | Naïve T cells are typically characterized by lack of markers for other mature T cell types. |
| activated CD4 T | act. CD4 | *CD4* | both | | https://panglaodb.se/markers.html?cell_type=%27T%20helper%20cells%27 | *GATA1* is listed as a marker gene for megakaryocytes on PanglaoDB. |
| regulatory T | reg. CD4 | *CD4, FOXP3* | both | 25683611 | https://panglaodb.se/markers.html?cell_type=%27T%20regulatory%20cells%27 | *FOXP3* is a master regulator of regulatory T cell development and function. |
| memory CD8 T | mem. CD8 | *CD8B, KLRB1* | both | 26220166 | | *GATA1* is listed as a marker gene for megakaryocytes on PanglaoDB. |
| cytotoxic CD8 T | cyto. CD8 | *CD8B, IFNG* | both snAT | 11698428 | | *IFNG* is a cytokine secreted by cytotoxic T cells. |
| pancreatic CD8 T | panc. CD8 | *CD8B, CD69* | AC-seq | 30952804 | | *CD69* is a marker for tissue resident CD8 T cells. |
| cytotoxic NK | cyto. NK | *IFNG, GZMB* | both | 12355441 | https://panglaodb.se/markers.html?cell_type=%27NK%20cells%27 | *IFNG* and *GZMB* are listed as marker genes for NK cells on PanglaoDB. |
| adaptive NK | adap. NK | *XCL1* | both | 29429633 | https://panglaodb.se/markers.html?cell_type=%27NK%20cells%27 | *XCL1* is listed as a marker gene for NK cells on PanglaoDB. |
| naïve B | naïve B | *TCL1A* | both | 11607815 | https://panglaodb.se/markers.html?cell_type=%27B%20cells%20naive%27 | *TCL1A* is listed as a marker gene for naïve B cells on PanglaoDB. |
| memory B | mem. B | *CD80* | both | 24880458 | https://panglaodb.se/markers.html?cell_type=%27B%20cells%20memory%27 | *CD80* (aka B7-1) is a cell surface marker for memory B cells and is listed as a marker gene for memory B cells on PanglaoDB. |
| plasma B | plasma B | *TNFRSF17* | scRNA-seq | 14707116 | https://panglaodb.se/markers.html?cell_type=%27Plasma%20cells%27 | *TNFRSF17* (aka BCMA) is essential for plasma cell survival and is listed as a marker gene for plasma cells on PanglaoDB. |

## Table 3.7. Transcription factors enriched for each cell type.

| **Cell type** | **Motif subfamily** | **Enriched TF motifs** (chromVAR z-score>2) | **Expressed subfamily TFs** (TPM>1) |
|---|---|---|---|
| acinar | ASC | Ascl2 | ASCL2 |
| acinar | ER-like | ESRRB, Esrra, Esrrg | ESRRA |
| acinar | FOXA | Foxa2, FOXA1 | FOXA2, FOXA3 |
| acinar | FOXC | FOXC2, FOXC1 | FOXC1 |
| acinar | HLH only | ID4 | ID2, ID3, ID4 |
| acinar | MESP | MSC | MESP1, MSC |
| acinar | NF-1 | NFIC, NFIA, NFIX | NFIC |
| acinar | NR2F | NR2F2 | NR2F2, NR2F6 |
| acinar | TWIST | FIGLA | PTF1A |
| cytotoxic CD8 T | TBrain-related | EOMES, TBX21, TBR1 | EOMES, TBX21 |
| cytotoxic CD8 T | ETS-like | ETS1 | ETS1, ETS2, ETV2, GABPA, FLI1, ETV3, ERF |
| cytotoxic CD8 T | Runt-related | RUNX1, RUNX3, RUNX2 | RUNX3 |
| activated stellate | ASC | Ascl2 | ASCL2 |
| activated stellate | ATF3-like | JDP2 | ATF3 |
| activated stellate | ATF4-related | ATF4 | ATF4, ATF5 |
| activated stellate | CEBP | CEBPA, CEBPB, CEBPE, CEBPG, CEBPD | CEBPB, CEBPG, CEBPD, DDIT3 |
| activated stellate | E2F | E2F6 | E2F4, E2F6 |
| activated stellate | EGR | EGR3 | EGR1, EGR2, EGR3, EGR4 |
| activated stellate | ETS-like | EWSR1-FLI1 | ETS2, ETV2, FEV, ERF |
| activated stellate | Fos | FOS, FOSL2, FOSL1 | FOS, FOSB, FOSL1, FOSL2 |
| activated stellate | TWIST | TWIST1 | HAND2 |
| activated stellate | HSF | HSF2, HSF1, HSF4 | HSF1, HSF2, HSF4 |
| activated stellate | Jun | JUN(var.2), JUND, JUNB, JUN | JUN |
| activated stellate | Large MAF | Mafb | MAF, MAFB |
| activated stellate | Small MAF | MAFF, MAFK | MAFF, MAFG, MAFK |
| activated stellate | MESP | Tcf21 | MESP1, MSC, TCF21 |
| activated stellate | More than 3 adjacent zinc fingers | MZF1 | MZF1 |
| activated stellate | NFAT-related | NFATC2, NFAT5, NFATC1, NFATC3 | NFATC4 |
| activated stellate | NFE2 | NFE2, BACH2, Nfe2l2 | NFE2L1, NFE2L2 |
| activated stellate | NF-1 | NFIC, NFIA | NFIC |
| activated stellate | NR3C | NR3C2, NR3C1, Ar | NR3C1 |
| activated stellate | NR4A | NR4A1 | NR4A1, NR4A2 |
| activated stellate | NFkappaB p65-like | RELB | RELA, RELB |
| activated stellate | STAT | STAT3, STAT1, Stat4, Stat6 | STAT1, STAT2, STAT3, STAT6 |
| activated stellate | TEF1-related | TEAD3, TEAD4, TEAD1, TEAD2 | TEAD2, TEAD3, TEAD4 |
| activated stellate | AP4 | TFAP4 | TFAP4 |
| adaptive NK | ELK-like | ELK4 | ELK1, ELK3, ELK4 |
| adaptive NK | TBrain-related | TBR1, TBX21, EOMES | EOMES, TBX21 |
| adaptive NK | ETS-like | ETS1, ERG, FLI1, ERF, Gabpa | ETS1, ETS2, ETV2, GABPA, FLI1, ETV3, ERF |
| adaptive NK | IRF | IRF9 | IRF1, IRF2, IRF3, IRF5, IRF7, IRF8, IRF9 |
| adaptive NK | Runt-related | RUNX1, RUNX3, RUNX2 | RUNX3 |
| adaptive NK | TBX6-related | MGA | TBX6 |
| classical monocyte | ATF4-related | ATF4 | ATF4, ATF5 |
| classical monocyte | CEBP | CEBPA, CEBPE, CEBPB, CEBPG, CEBPD | CEBPA, CEBPB, CEBPG, CEBPD, CEBPE, DDIT3 |
| classical monocyte | ETV6-like | ETV6 | ETV6, ETV7 |
| classical monocyte | Fos | FOSL2, FOSL1, FOS | FOS, FOSB, FOSL1, FOSL2 |
| classical monocyte | Jun | JUNB, JUN(var.2), JUND | JUN, NFE2, ATF2 |
| classical monocyte | NFE2 | Nfe2l2, BACH2 | NFE2, NFE2L1, NFE2L2, NFE2L3, BACH1 |
| classical monocyte | SPI-like | SPI1, SPIC, SPIB | SPI1 |
| conventional dendritic | ETV6-like | ETV6 | ETV6 |
| conventional dendritic | IRF | IRF1 | IRF1, IRF2, IRF3, IRF4, IRF5, IRF7, IRF8, IRF9 |
| conventional dendritic | SPI-like | SPI1, SPIC, SPIB | SPI1, SPIB |
| cytotoxic NK | TBrain-related | EOMES, TBX21, TBR1 | EOMES, TBX21 |
| cytotoxic NK | ETS-like | ETS1 | ETS1, ETS2, ETV2, GABPA, FLI1, ETV3, ERF |
| cytotoxic NK | Runt-related | RUNX1, RUNX3, RUNX2 | RUNX3 |
| ductal | ER-like | ESRRB, Esrra, Esrrg | ESRRA |
| ductal | FOXA | Foxa2, FOXA1 | FOXA2, FOXA3 |
| ductal | FOXC | FOXC2, FOXC1 | FOXC1 |
| ductal | GATA double | GATA2, Gata4, GATA6, Gata1 | GATA6 |
| ductal | HNF1-like | HNF1B, HNF1A | HNF1B |
| ductal | NFAT-related | NFATC2 | NFATC4 |
| ductal | NF-1 | NFIC, NFIA, NFIX | NFIC |
| ductal | NR5A | Nr5a2 | NR5A2 |
| ductal | ONECUT | ONECUT1, ONECUT2, ONECUT3 | ONECUT2 |
| ductal | Group E | SOX10 | SOX8, SOX9 |
| ductal | TEF1-related | TEAD3, TEAD4, TEAD2, TEAD1 | TEAD2, TEAD3 |
| endothelial | Fos | FOSL2, FOS, FOSL1 | FOS, FOSB, FOSL1, FOSL2 |
| endothelial | Jun | JUN(var.2), JUND, JUNB | JUN, ATF2 |
| endothelial | NFAT-related | NFATC2 | NFATC4 |
| endothelial | Group G | SOX15 | SOX15 |
| endothelial | Group B | Sox2, Sox3 | SOX2 |
| endothelial | Group E | SOX9 | SOX8, SOX9 |
| endothelial | STAT | Stat4, STAT1, STAT3 | STAT2, STAT3, STAT5A, STAT5B, STAT6 |
| gamma | ARID3 | Arid3a, Arid3b | ARID3A |
| gamma | ARX | Arx | ARX |
| gamma | ASC | Ascl2, ASCL1 | ASCL2 |
| gamma | E2F | E2F4, E2F6 | E2F4 |
| gamma | EGR | EGR1, EGR2, EGR3 | EGR1, EGR3, EGR4 |
| gamma | ETS-like | EWSR1-FLI1 | ETV2, FEV, ERF |
| gamma | Fos | FOS, FOSL2, FOSL1 | FOS, FOSB |
| gamma | FOXA | Foxa2, FOXA1 | FOXA2, FOXA3 |
| gamma | FOXJ | Foxj3, Foxj2 | FOXJ1 |
| gamma | HINFP-like | HINFP | HINFP |
| gamma | HOX2 | HOXB2, HOXA2 | HOXB2 |
| gamma | ISL | ISL2 | ISL1 |
| gamma | Jun | JUN(var.2), JUND, JUNB | JUN |
| gamma | Large MAF | NRL | MAF, MAFB |
| gamma | Small MAF | MAFF, MAFK | MAFG |
| gamma | MEIS | MEIS1, MEIS2, MEIS3 | MEIS2, MEIS3 |
| gamma | MESP | MSC, Tcf21 | MESP1 |
| gamma | MNX | MNX1 | MNX1 |
| gamma | More than 3 adjacent zinc fingers | MZF1(var.2) | MZF1 |
| gamma | Neurogenin-ATO | NEUROD1, NEUROG2 | NEUROD1, ATOH7 |
| gamma | NFE2 | NFE2, BACH2 | NFE2L1 |
| gamma | NF-1 | NFIC | NFIC |
| gamma | NK2.2 | NKX2-8 | NKX2-2 |
| gamma | NK6 | NKX6-2, NKX6-1 | NKX6-1 |
| gamma | RFX | Rfx1, RFX5, RFX3, RFX2, RFX4 | RFX6 |
| gamma | Group C | SOX4, Sox11 | SOX4, SOX12 |
| gamma | Group A | SRY | SRY |
| gamma | STAT | STAT3, STAT1, Stat4 | STAT3 |
| gamma | AP4 | TFAP4 | TFAP4 |
| gamma | DP1 | TFDP1 | TFDP1 |
| gamma | ZNF24-like | ZNF24 | ZNF24, GLI4, ZKSCAN1 |
| GCGhi alpha | ARID3 | Arid3a, Arid3b | ARID3A |
| GCGhi alpha | ARID5 | Arid5a | ARID5A |

## Table 3.7. Transcription factors enriched for each cell type, continued.

| Cell type | Motif subfamily | Enriched TF motifs (chromVAR z-score>2) | Expressed subfamily TFs (TPM>1) |
|---|---|---|---|
| GCGhi alpha | ASC | Ascl2, ASCL1 | ASCL1, ASCL2 |
| GCGhi alpha | Fos | FOS, FOSL1, FOSL2 | FOS, FOSB |
| GCGhi alpha | FOXA | Foxa2, FOXA1 | FOXA2, FOXA3 |
| GCGhi alpha | FOXJ | Foxj3, Foxj2 | FOXJ1 |
| GCGhi alpha | FOXP | FOXP3, FOXP1, FOXP2 | FOXP4 |
| GCGhi alpha | FOXQ | Foxq1 | FOXQ1 |
| GCGhi alpha | GATA double | Gata4, Gata1, GATA6 | GATA6 |
| GCGhi alpha | HNF1-like | HNF1B, HNF1A | HNF1A |
| GCGhi alpha | HOX2 | HOXB2 | HOXB2 |
| GCGhi alpha | ISL | ISL2 | ISL1 |
| GCGhi alpha | Jun | JUND, JUN(var.2), JUNB | JUN |
| GCGhi alpha | Large MAF | NRL, Mafb | MAF, MAFB |
| GCGhi alpha | Small MAF | MAFF, MAFK, MAFG | MAFG, MAFK |
| GCGhi alpha | MESP | MSC, Tcf21 | MESP1 |
| GCGhi alpha | MNX | MNX1 | MNX1 |
| GCGhi alpha | Neurogenin-ATO | NEUROD1, NEUROG2 | NEUROD1, ATOH7 |
| GCGhi alpha | NFAT-related | NFATC2, NFAT5, NFATC3, NFATC1 | NFATC4 |
| GCGhi alpha | NFE2 | BACH2, NFE2 | NFE2L1, NFE2L2 |
| GCGhi alpha | NK6 | NKX6-2 | NKX6-1 |
| GCGhi alpha | RFX | Rfx1, RFX3, RFX5, RFX2, RFX4 | RFX5, RFX6 |
| GCGhi alpha | AP4 | TFAP4 | TFAP4 |
| GCGlo alpha | ARID3 | Arid3a, Arid3b | ARID3A |
| GCGlo alpha | E2F | E2F4 | E2F4, E2F6 |
| GCGlo alpha | Fos | FOS, FOSL2, FOSL1 | FOS, FOSB |
| GCGlo alpha | FOXA | Foxa2, FOXA1 | FOXA2, FOXA3 |
| GCGlo alpha | FOXJ | Foxj3, Foxj2 | FOXJ1 |
| GCGlo alpha | FOXP | FOXP3, FOXP1, FOXP2 | FOXP4 |
| GCGlo alpha | HINFP-like | HINFP | HINFP |
| GCGlo alpha | HNF1-like | HNF1B, HNF1A | HNF1A |
| GCGlo alpha | HOX2 | HOXB2 | HOXB2 |
| GCGlo alpha | Jun | JUND, JUNB, JUN(var.2) | JUN |
| GCGlo alpha | Large MAF | NRL, Mafb | MAF, MAFB |
| GCGlo alpha | Small MAF | MAFF, MAFK, MAFG | MAFG, MAFK |
| GCGlo alpha | MNX | MNX1 | MNX1 |
| GCGlo alpha | NFAT-related | NFATC2, NFAT5 | NFATC4 |
| GCGlo alpha | NFE2 | NFE2, BACH2, Nfe2l2 | NFE2L1, NFE2L2 |
| GCGlo alpha | NK6 | NKX6-2 | NKX6-1 |
| GCGlo alpha | RFX | Rfx1 | RFX5, RFX6 |
| GCGlo alpha | ZNF24-like | ZNF24 | ZNF24, GLI4, ZNF232, ZKSCAN1 |
| INShi beta | ARID5 | Arid5a | ARID5A |
| INShi beta | ASC | Ascl2, ASCL1 | ASCL1, ASCL2 |
| INShi beta | Fos | FOS, FOSL2, FOSL1 | FOS |
| INShi beta | FOXA | Foxa2, FOXA1 | FOXA2, FOXA3 |
| INShi beta | FOXJ | Foxj3, Foxj2 | FOXJ1 |
| INShi beta | FOXP | FOXP3, FOXP2, FOXP1 | FOXP4 |
| INShi beta | HOX2 | HOXB2, HOXA2 | HOXB2 |
| INShi beta | ISL | ISL2 | ISL1 |
| INShi beta | Jun | JUND, JUN(var.2), JUNB | JUN |
| INShi beta | Large MAF | NRL, Mafb | MAFA, MAFB |
| INShi beta | Small MAF | MAFF, MAFK, MAFG | MAFG |
| INShi beta | MESP | MSC, Tcf21 | MESP1 |
| INShi beta | MNX | MNX1 | MNX1 |
| INShi beta | Neurogenin-ATO | NEUROD1, NEUROG2 | NEUROD1, ATOH7, OLIG1, OLIG3 |
| INShi beta | NFE2 | BACH2, NFE2, Nfe2l2 | NFE2L1, NFE2L2 |
| INShi beta | NK6 | NKX6-2, NKX6-1 | NKX6-1, NKX6-2 |
| INShi beta | PAX4-like | PAX4 | PAX6 |
| INShi beta | PDX | PDX1 | PDX1 |
| INShi beta | RFX | Rfx1, RFX5, RFX3, RFX2, RFX4 | RFX5, RFX6 |
| INShi beta | AP4 | TFAP4 | TFAP4 |
| INSlo beta | ARID5 | Arid5a | ARID5A |
| INSlo beta | ASC | Ascl2, ASCL1 | ASCL1, ASCL2 |
| INSlo beta | E2F | E2F4 | E2F4, E2F6 |
| INSlo beta | Fos | FOS, FOSL2, FOSL1 | FOS |
| INSlo beta | FOXA | Foxa2, FOXA1 | FOXA2, FOXA3 |
| INSlo beta | FOXJ | Foxj3, Foxj2 | FOXJ1 |
| INSlo beta | FOXP | FOXP3, FOXP1, FOXP2 | FOXP4 |
| INSlo beta | HOX2 | HOXB2, HOXA2 | HOXB2 |
| INSlo beta | ISL | ISL2 | ISL1 |
| INSlo beta | Jun | JUND, JUNB, JUN(var.2) | JUN |
| INSlo beta | Large MAF | NRL, Mafb | MAFA, MAFB |
| INSlo beta | Small MAF | MAFF, MAFK, MAFG | MAFG |
| INSlo beta | MESP | MSC, Tcf21 | MESP1 |
| INSlo beta | MNX | MNX1 | MNX1 |
| INSlo beta | Neurogenin-ATO | NEUROD1 | NEUROD1, ATOH7, OLIG1, OLIG3 |
| INSlo beta | NFE2 | NFE2, BACH2, Nfe2l2 | NFE2L1, NFE2L2 |
| INSlo beta | NK6 | NKX6-2, NKX6-1 | NKX6-1, NKX6-2 |
| INSlo beta | PAX4-like | PAX4 | PAX6 |
| INSlo beta | PDX | PDX1 | PDX1 |
| INSlo beta | RFX | Rfx1, RFX3, RFX5, RFX2, RFX4 | RFX5, RFX6 |
| INSlo beta | AP4 | TFAP4 | TFAP4 |
| INSlo beta | ZNF24-like | ZNF24 | ZNF24, GLI4, ZKSCAN1 |
| megakaryocyte | CEBP | CEBPA | CEBPA, CEBPB, CEBPG, CEBPD, CEBPE, DDIT3 |
| megakaryocyte | GATA double | GATA6, GATA2, Gata4, GATA3, GATA5, Gata1 | GATA1, GATA2, GATA3 |
| memory B | HLH only | ID4 | ID2, ID3 |
| memory B | IRF | IRF1, IRF9, IRF4, IRF8, IRF3, IRF2 | IRF1, IRF2, IRF3, IRF4, IRF5, IRF7, IRF8, IRF9 |
| memory B | NFkappaB p50-like | NFKB2, NFKB1 | NFKB1, NFKB2 |
| memory B | POU2 | POU2F2, Pou2f3, POU2F1 | POU2F2 |
| memory B | POU3 | POU3F4, POU3F2, POU3F1, POU3F3 | POU3F1 |
| memory B | POU5 | POU5F1B, POU5F1 | POU5F2 |
| memory B | Snail-like | SNAI2 | SNAI1, SNAI3 |
| memory B | SPI-like | SPI1, SPIC | SPI1, SPIB |
| memory B | E2A | TCF3, TCF4 | TCF3, TCF4 |
| memory B | ZEB | ZEB1 | ZEB2 |
| naive B | HLH only | ID4 | ID2, ID3 |
| naive B | IRF | IRF1 | IRF1, IRF2, IRF3, IRF4, IRF5, IRF7, IRF8, IRF9 |
| naive B | NFkappaB p50-like | NFKB2, NFKB1 | NFKB1, NFKB2 |
| naive B | POU2 | POU2F2, Pou2f3, POU2F1 | POU2F2 |
| naive B | POU3 | POU3F4 | POU3F1 |
| naive B | POU5 | POU5F1B, POU5F1 | POU5F2 |
| naive B | Snail-like | SNAI2 | SNAI1, SNAI3 |
| naive B | SPI-like | SPI1, SPIC | SPI1, SPIB |
| naive B | E2A | TCF3, TCF4 | TCF3, TCF4 |
| naive T | CTCF-like | CTCF | CTCF |
| memory CD8 T | CTCF-like | CTCF | CTCF |
| memory CD8 T | ELK-like | ELK4 | ELK1, ELK3, ELK4 |
| memory CD8 T | TBrain-related | TBX21, TBR1, EOMES | EOMES, TBX21 |
| memory CD8 T | ETS-like | ETS1, ERG, FLI1 | ETS1, ETS2, ETV2, GABPA, FLI1, ETV3, ERF |

## Table 3.7. Transcription factors enriched for each cell type, continued.

| **Cell type** | **Motif subfamily** | **Enriched TF motifs** (chromVAR z-score>2) | **Expressed subfamily TFs** (TPM>1) |
|---|---|---|---|
| memory CD8 T | ROR | RORC, RORB, RORA(var.2) | RORA, RORC |
| memory CD8 T | Runt-related | RUNX1, RUNX3, RUNX2 | RUNX3 |
| nonclassical monocyte | CEBP | CEBPA, CEBPG, CEBPE, CEBPB | CEBPA, CEBPB, CEBPG, CEBPD, DDIT3 |
| nonclassical monocyte | ETV6-like | ETV6 | ETV7 |
| nonclassical monocyte | SPI-like | SPI1, SPIC, SPIB | SPI1 |
| pancreatic macrophage | SPI-like | SPI1, SPIC | SPI1 |
| plasmacytoid dendritic | ASC | ASCL1, Ascl2 | ASCL2 |
| plasmacytoid dendritic | ETV6-like | ETV6 | ETV6 |
| plasmacytoid dendritic | HLH only | ID4 | ID2, ID3 |
| plasmacytoid dendritic | IRF | IRF1, IRF3, IRF9, IRF8, IRF4, IRF2, IRF5, IRF7 | IRF1, IRF2, IRF3, IRF4, IRF5, IRF7, IRF8, IRF9 |
| plasmacytoid dendritic | MESP | Tcf21, MSC | MSC |
| plasmacytoid dendritic | More than 3 adjacent zinc fingers | ZNF263, MZF1 | MZF1 |
| plasmacytoid dendritic | Factors with multiple dispersed zinc fingers | RREB1 | REST |
| plasmacytoid dendritic | Runt-related | RUNX1, RUNX3, RUNX2 | RUNX2, RUNX3 |
| plasmacytoid dendritic | Snail-like | SNAI2, SCRT1 | SNAI1, SNAI3 |
| plasmacytoid dendritic | SPI-like | SPI1, SPIC, SPIB | SPI1, SPIB |
| plasmacytoid dendritic | TAL-HEN | NHLH1 | TAL2, LYL1 |
| plasmacytoid dendritic | E2A | Tcf12, TCF4, TCF3 | TCF3, TCF4 |
| plasmacytoid dendritic | AP4 | TFAP4 | TFAP4 |
| plasmacytoid dendritic | ZEB | ZEB1 | ZEB2 |
| quiescent stellate | ATF3-like | JDP2 | ATF3 |
| quiescent stellate | CEBP | CEBPA | CEBPB, CEBPG, CEBPD, DDIT3 |
| quiescent stellate | Fos | FOS, FOSL2, FOSL1 | FOS, FOSB, FOSL1, FOSL2 |
| quiescent stellate | Jun | JUN(var.2), JUND, JUNB | JUN, ATF2 |
| quiescent stellate | NFAT-related | NFATC2, NFAT5, NFATC1, NFATC3 | NFATC4 |
| quiescent stellate | NFE2 | NFE2, BACH2, Nfe2l2 | NFE2L1, NFE2L2 |
| quiescent stellate | M | RBPJ | RBPJ |
| quiescent stellate | STAT | STAT3, STAT1, Stat4 | STAT2, STAT3, STAT6 |
| quiescent stellate | TEF1-related | TEAD3, TEAD4, TEAD2, TEAD1 | TEAD2, TEAD3 |
| SSThi delta | ASC | Ascl2, ASCL1 | ASCL1, ASCL2 |
| SSThi delta | Fos | FOS, FOSL2, FOSL1 | FOS, FOSB |
| SSThi delta | FOXA | Foxa2, FOXA1 | FOXA2, FOXA3 |
| SSThi delta | FOXJ | Foxj3, Foxj2 | FOXJ1 |
| SSThi delta | HOX2 | HOXB2, HOXA2 | HOXB2 |
| SSThi delta | ISL | ISL2 | ISL1 |
| SSThi delta | Jun | JUND, JUN(var.2), JUNB | JUN |
| SSThi delta | Small MAF | MAFF, MAFK | MAFG |
| SSThi delta | MESP | MSC, Tcf21 | MESP1 |
| SSThi delta | Neurogenin-ATO | NEUROD1 | NEUROD1 |
| SSThi delta | NFE2 | NFE2, BACH2 | NFE2L1 |
| SSThi delta | NK6 | NKX6-2, NKX6-1 | NKX6-1, NKX6-3 |
| SSThi delta | PAX4-like | PAX4 | PAX4, PAX6 |
| SSThi delta | PDX | PDX1 | PDX1 |
| SSThi delta | RFX | Rfx1, RFX3, RFX5, RFX2, RFX4 | RFX5, RFX6 |
| SSThi delta | AP4 | TFAP4 | TFAP4 |
| SSTlo delta | ASC | Ascl2, ASCL1 | ASCL1, ASCL2 |
| SSTlo delta | E2F | E2F4, E2F6 | E2F4, E2F6 |
| SSTlo delta | EGR | EGR3, EGR2, EGR1 | EGR1, EGR3, EGR4 |
| SSTlo delta | Fos | FOSL2, FOS, FOSL1 | FOS, FOSB |
| SSTlo delta | FOXA | Foxa2, FOXA1 | FOXA2, FOXA3 |
| SSTlo delta | FOXJ | Foxj3, Foxj2 | FOXJ1 |
| SSTlo delta | HOX2 | HOXB2, HOXA2 | HOXB2 |
| SSTlo delta | ISL | ISL2 | ISL1 |
| SSTlo delta | Jun | JUND, JUNB, JUN(var.2), JUND(var.2) | JUN |
| SSTlo delta | Large MAF | Mafb, NRL | MAFB |
| SSTlo delta | Small MAF | MAFF, MAFK, MAFG | MAFG |
| SSTlo delta | MESP | MSC, Tcf21 | MESP1 |
| SSTlo delta | Neurogenin-ATO | NEUROD1, NEUROG2 | NEUROD1 |
| SSTlo delta | NFE2 | NFE2, BACH2, Nfe2l2 | NFE2L1 |
| SSTlo delta | NK6 | NKX6-2, NKX6-1 | NKX6-1, NKX6-3 |
| SSTlo delta | PAX4-like | PAX4 | PAX4, PAX6 |
| SSTlo delta | PDX | PDX1 | PDX1 |
| SSTlo delta | RFX | Rfx1, RFX3, RFX5, RFX2, RFX4 | RFX5, RFX6 |
| SSTlo delta | Group C | SOX4, Sox11 | SOX4, SOX12 |
| SSTlo delta | Group E | SOX10 | SOX9 |
| SSTlo delta | AP4 | TFAP4 | TFAP4 |
| SSTlo delta | DP1 | TFDP1 | TFDP1 |
| SSTlo delta | ZNF24-like | ZNF24 | ZNF24, GLI4, ZKSCAN1 |

## Table 3.8. Annotation of pancreatic exocrine signals with accessible chromatin of stimulated immune and islet cells.

| Marker | Chrom | Position | Signal name | PPA | Exocrine | Stimulated immune | Stimulated islets |
|---|---|---|---|---|---|---|---|
| rs10489869 | 1 | 192515778 | RGS1 (1:192515849:A:G) | 0.101 | acinar | CD8pos T-S, Central memory CD8pos T-S, Effector memory CD8pos T-S, Gamma delta T-S, Mature NK-S, Memory Teffs-S, Memory Tregs-S, Naive Teffs-S, Regulatory T-S, Th1 precursors-S, Th2 precursors-S | |
| rs10801128 | 1 | 192515849 | RGS1 (1:192515849:A:G) | 0.112 | acinar | CD8pos T-S, Central memory CD8pos T-S, Effector CD4pos T-S, Effector memory CD8pos T-S, Gamma delta T-S, Memory Teffs-S, Memory Tregs-S, Naive Teffs-S, Regulatory T-S, Th17 precursors-S, Th1 precursors-S, Th2 precursors-S | |
| rs1016431 | 7 | 51020336 | COBL (7:51018492:GT:G) | 0.035 | acinar, ductal | | islets-S |
| rs1016432 | 7 | 51020410 | COBL (7:51018492:GT:G) | 0.039 | acinar, ductal | | islets-S |
| rs917072 | 7 | 51020808 | COBL (7:51018492:GT:G) | 0.036 | acinar | | |
| rs7795896 | 7 | 117086613 | CFTR (7:117086613:C:T) | 0.629 | ductal | | |
| rs7068821 | 10 | 90051035 | RNLS (10:90051035:G:T) | 0.448 | acinar | | |
| rs60888743 | 10 | 90051317 | RNLS (10:90051035:G:T) | 0.390 | acinar | | |
| rs4238595 | 16 | 20343091 | GP2 (16:20343091:T:C) | 0.452 | acinar, ductal | | |
| rs8062123 | 16 | 20343953 | GP2 (16:20343091:T:C) | 0.269 | acinar | | |
| rs8060932 | 16 | 20344077 | GP2 (16:20343091:T:C) | 0.244 | acinar | | |
| rs72802342 | 16 | 75234872 | CTRB2 (16:75236763:C:G) | 0.259 | acinar, ductal | | |
| rs8056814 | 16 | 75252327 | CTRB2 (16:75236763:C:G) | 0.301 | acinar | Bulk B-S, CD8pos T-S, Central memory CD8pos T-S, Effector CD4pos T-S, Effector memory CD8pos T-S, Follicular T Helper-S, Mem B-S, Memory Teffs-S, Memory Tregs-S, Naive B-S, Naive CD8 T-S, Naive Teffs-S, Regulatory T-S, Th17 precursors-S, Th1 precursors-S, Th2 precursors-S | |

## Table 3.9. Sequences for primers, oligonucleotides, and guide RNAs.

Luciferase reporter assay primers

| Experiment | Direction | Primer sequence |
|---|---|---|
| | Forward | TAGCGGTACCTAATGGGAAATCATGCCAACC |
| Cloning 180 bp | Reverse | AATAGAGCTCATGTGTGTGTGCTGGGATGT |
| | Forward | TAAGCAGGTACCTGGGTAGAAATTGGAAACACAA |
| Cloning 594 bp | Reverse | TGCTTAGAGCTCTTCCAACTCTTCATTCCTGGTT |
| | Forward | AGATGTAACTtATTAACATTAGAAAAAATAC |
| rs7795896 SDM | Reverse | AATTGTTCCTCTTTTTTTTTCAAATAG |

Electrophoretic mobility shift assay oligonucleotides

| Variant | Allele | Oligo sequence |
|---|---|---|
| | C | (5′ biotin)CAATTAGATGTAACTCATTAACATTAGAAAAA |
| rs7795896 | T | (5′ biotin)CAATTAGATGTAACTTATTAACATTAGAAAAA |

CRISPR guide RNAs (sgRNAs)

| sgRNA name | sgRNA sequence | Targeted region (hg19) | On-target score | Specificity score |
|---|---|---|---|---|
| rs7795896 enhancer guide 1 | GTAGTTGGCTTCCTCAGTAAG | chr7:117086829-117086849 | 63.3 | 61.5 |
| rs7795896 enhancer guide 2 | GAACAGTATGATTTACGTAA | chr7:117086506-117086525 | 57.1 | 77.4 |
| rs7795896 enhancer guide 3 | GAATGTACTATAAAATGTCT | chr7:117086190-117086209 | 59.5 | 41.6 |
| *CFTR* promoter | GCGCCCGAGAGACCATGCAG | chr7:117120135-117120154 | 71.9 | 43.1 |
| Non-targeting control | GTGACGTGCACCGCGGTGTG | N/A | N/A | N/A |

qPCR primers

| Gene | Direction | Primer sequence |
|---|---|---|
| *CFTR* | Forward | AGGAGGAACGCTCTATCG |
| *CFTR* | Reverse | GCAGACGCCTGTAACAAC |
| *CAPZA2* | Forward | CCGAATGGAGTCTGCACTGT |
| *CAPZA2* | Reverse | ACCTCCAACGACCATTCCAAA |
| *ST7* | Forward | TCAATCCTCATGTGCCAAAA |
| *ST7* | Reverse | TTCCCACGTACAATGCAAAA |
| *CTTNBP2* | Forward | AAGAGCGTGGCAAGAACAAG |
| *CTTNBP2* | Reverse | TGGGCCTCCTCTATGACTTTG |
| *NAA38* | Forward | ACATCAACCGAACTGTTGCC |
| *NAA38* | Reverse | CCAAATCAAGCGCAGAATCTGT |
| *ASZ1* | Forward | GAAAGCAATGACCATCGGAGA |
| *ASZ1* | Reverse | GGGGAGTCCATCCATACTGAAA |
| *WNT2* | Forward | CCGAGGTCAACTCTTCATGGT |
| *WNT2* | Reverse | CCTGGCACATTATCGCACAT |
| *TBP* | Forward | TGTGCACAGGAGCCAAGAGT |
| *TBP* | Reverse | ATTTTCTTGCTGCCAGTCTGG |
| Puromycin resistance | Forward | GGGTCACCGAGCTGCAAGAA |
| Puromycin resistance | Reverse | GCCTTCCATCTGTTGCTGCG |
| dCas9 | Forward | GAGTTGACGCCAAAGCAATC |
| dCas9 | Reverse | TACCAAACAGGCCGTTCTTC |

## 3.9 Data and Code Availability

Full summary statistics for the T1D GWAS have been deposited into the NHGRI-EBI GWAS catalog with study accession number GCST90012879. Sequencing data for snATAC-seq have been deposited into the NCBI Gene Expression Omnibus (GEO) with accession number GSE163160. Processed data files for snATAC-seq will be available through the Diabetes Epigenome Atlas (https://www.diabetesepigenome.org/).

Code used for processing snATAC-seq datasets and clustering cells is available at https://github.com/kjgaulton/pipelines/tree/master/T1D_snATAC_pipeline.

## 3.10 Acknowledgements

This work was supported by NIH grants DK112155, DK120429 and DK122607 to K.J.G and M.S., and T32 GM008666 to R.G. We thank Samantha Kuan in the Ren Lab at the LICR for assistance with sequencing.

nPOD: This research was performed with the support of the Network for Pancreatic Organ donors with Diabetes (nPOD; RRID:SCR_014641), a collaborative type 1 diabetes research project sponsored by JDRF (nPOD: 5-SRA-2018-557-Q-R) and The Leona M. & Harry B. Helmsley Charitable Trust (Grant #2018PG-T1D053). The content and views expressed are the responsibility of the authors and do not necessarily reflect the official view of nPOD. Organ Procurement Organizations (OPO) partnering with nPOD to provide research resources are listed at http://www.jdrfnpod.org/for-partners/npod-partners/.

DCCT/EDIC: The Diabetes Control and Complications Trial (DCCT) and its follow-up the Epidemiology of Diabetes Interventions and Complications (EDIC) study were conducted by the DCCT/EDIC Research Group and supported by National Institute of Health grants and contracts

and by the General Clinical Research Center Program, NCRR. The data from the DCCT/EDIC study were supplied by the NIDDK Central Repositories.

GENIE: The Genetics of Nephropathy, an International Effort (GENIE) study was conducted by the GENIE Investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The data from the GENIE study reported here were supplied by the GENIE investigators from the Broad Institute of MIT and Harvard, Queens University Belfast and the University of Dublin.

GoKinD: The Genetics of Kidneys in Diabetes (GoKinD) Study was conducted by the GoKinD Investigators and supported by the Juvenile Diabetes Research Foundation, the CDC, and the Special Statutory Funding Program for Type 1 Diabetes Research administered by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The data [and samples] from the GoKinD study were supplied by the NIDDK Central Repositories. This manuscript was not prepared in collaboration with Investigators of the GoKinD study and does not necessarily reflect the opinions or views of the GoKinD study, the NIDDK Central Repositories, or the NIDDK.

T1DGC: This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium (T1DGC), a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and the Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. The UK case series collection was additionally funded by the JDRF and Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Centre, at the Cambridge Institute for Medical Research, UK (CIMR), which is in receipt of a Wellcome Trust Strategic Award (079895). The data from the T1DGC study were supplied by dbGAP. This manuscript was not prepared in collaboration with

Investigators of the T1DGC study and does not necessarily reflect he opinions or views of the T1DGC study or the study sponsors.

T1DGC (ASP/UK GRID): This research was performed under the auspices of the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF).

WTCCC: This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113.

UK Biobank: Data from the UK Biobank was accessed under application 24058.

FinnGen: We want to acknowledge the participants and investigators of the FinnGen study.

CSGNM: We thank the participants of the Trinity Student Study. This study was supported by the Intramural Research Programs of the National Institutes of Health, the National Human Genome Research Institute, and the Eunice Kennedy Shriver National Institute of Child Health and Development.

NIMH Schizophrenia Controls: Funding support for the Genome-Wide Association of Schizophrenia Study was provided by the National Institute of Mental Health (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289 U01 MH46318, U01 MH79469, and U01 MH79470) and the genotyping of samples was provided

through the Genetic Association Information Network (GAIN). The datasets used for the analyses described in this manuscript were obtained from the database of Genotypes and Phenotypes (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000021.v3.p2. Samples and associated phenotype data for the Genome-Wide Association of Schizophrenia Study were provided by the Molecular Genetics of Schizophrenia Collaboration (PI: Pablo V. Gejman, Evanston Northwestern Healthcare (ENH) and Northwestern University, Evanston, IL, USA).

Neurodevelopmental Genomics: Support for the collection of the data for Philadelphia Neurodevelopment Cohort (PNC) was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. Subjects were recruited and genotyped through the Center for Applied Genomics (CAG) at The Children's Hospital in Philadelphia (CHOP). Phenotypic data collection occurred at the CAG/CHOP and at the Brain Behavior Laboratory, University of Pennsylvania.

eMERGE Network: Group Health Cooperative/University of Washington – Funding support for Alzheimer's Disease Patient Registry (ADPR) and Adult Changes in Thought (ACT) study was provided by a U01 from the National Institute on Aging (Eric B. Larson, PI, U01AG006781). A gift from the 3M Corporation was used to expand the ACT cohort. DNA aliquots sufficient for GWAS from ADPR Probable AD cases, who had been enrolled in Genetic Differences in Alzheimer's Cases and Controls (Walter Kukull, PI, R01 AG007584) and obtained under that grant, were made available to eMERGE without charge. Funding support for genotyping, which was performed at Johns Hopkins University, was provided by the NIH (U01HG004438). Genome-wide association analyses were supported through a Cooperative Agreement from the National Human Genome Research Institute, U01HG004610 (Eric B. Larson, PI). Mayo Clinic – Samples and associated genotype and phenotype data used in this study were provided by the Mayo Clinic. Funding support for the Mayo Clinic was provided through a

cooperative agreement with the National Human Genome Research Institute (NHGRI), Grant #: UOIHG004599; and by grant HL75794 from the National Heart Lung and Blood Institute (NHLBI). Funding support for genotyping, which was performed at The Broad Institute, was provided by the NIH (U01HG004424). Marshfield Clinic Research Foundation – Funding support for the Personalized Medicine Research Project (PMRP) was provided through a cooperative agreement (U01HG004608) with the National Human Genome Research Institute (NHGRI), with additional funding from the National Institute for General Medical Sciences (NIGMS) The samples used for PMRP analyses were obtained with funding from Marshfield Clinic, Health Resources Service Administration Office of Rural Health Policy grant number D1A RH00025, and Wisconsin Department of Commerce Technology Development Fund contract number TDF FYO10718. Funding support for genotyping, which was performed at Johns Hopkins University, was provided by the NIH (U01HG004438). Northwestern University – Samples and data used in this study were provided by the NUgene Project (www.nugene.org). Funding support for the NUgene Project was provided by the Northwestern University's Center for Genetic Medicine, Northwestern University, and Northwestern Memorial Hospital. Assistance with phenotype harmonization was provided by the eMERGE Coordinating Center (Grant number U01HG04603). This study was funded through the NIH, NHGRI eMERGE Network (U01HG004609). Funding support for genotyping, which was performed at The Broad Institute, was provided by the NIH (U01HG004424).

Vanderbilt University - Funding support for the Vanderbilt Genome-Electronic Records (VGER) project was provided through a cooperative agreement (U01HG004603) with the National Human Genome Research Institute (NHGRI) with additional funding from the National Institute of General Medical Sciences (NIGMS). The dataset and samples used for the VGER analyses were obtained from Vanderbilt University Medical Center's BioVU, which is supported by institutional funding and by the Vanderbilt CTSA grant UL1RR024975 from NCRR/NIH. Funding support for genotyping, which was performed at The Broad Institute, was provided by the NIH

(U01HG004424). Assistance with phenotype harmonization and genotype data cleaning was provided by the eMERGE Administrative Coordinating Center (U01HG004603) and the National Center for Biotechnology Information (NCBI). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000360.v3.p1.

This manuscript was not prepared in collaboration with investigators of these studies and does not necessarily reflect the opinions or views of the DCCT/EDIC, GENIE, GoKinD, T1DGC, WTCCC, studies or study groups, the NIDDK Central Repositories, the NIH, or the study sponsors.

Chapter 3, in full, has been accepted for publication of the material in “Cell type-specific genetic mechanisms of type 1 diabetes risk“. Joshua Chiou, Ryan J Geusz, Mei-Lin Okino, Jee Yun Han, Michael Miller, Rebecca Melton, Elisha Beebe, Paola Benaglio, Serina Huang, Katha Korgaonkar, Sandra Heller, Alexander Kleger, Sebastian Preissl, David U Gorkin, Maike Sander, Kyle J Gaulton. The dissertation author was the primary investigator and author of this paper.

## 3.11 Author Contributions

K.J.G and J.C. designed the study and wrote the manuscript. J.C. performed genetic association and single cell genomics analyses. R.G., M.O. and S.Huang performed molecular experiments of enhancer and variant function. R.M. and E.B. performed analyses of single cell gene expression. J.Y.H and M.M. generated single cell accessible chromatin data. P.B. and K.K. contributed to data analysis. D.U.G and S.P. supervised the generation of single cell accessible chromatin and contributed to data interpretation and analyses. M.S. supervised experiments related to enhancer function and contributed to data interpretation. S.Heller and A.K. contributed to interpretation of experimental data.

## 3.12 References


1. Katsarou, A., Gudbjörnsdottir, S., Rawshani, A., Dabelea, D., Bonifacio, E., Anderson, B. J., Jacobsen, L. M., Schatz, D. A. & Lernmark, Å. Type 1 diabetes mellitus. *Nat. Rev. Dis. Primer* **3**, 17016 (2017).

2. Onengut-Gumuscu, S., Chen, W.-M., Burren, O., Cooper, N. J., Quinlan, A. R., Mychaleckyj, J. C., Farber, E., Bonnie, J. K., Szpak, M., Schofield, E., Achuthan, P., Guo, H., Fortune, M. D., Stevens, H., Walker, N. M., Ward, L. D., Kundaje, A., Kellis, M., Daly, M. J., Barrett, J. C., Cooper, J. D., Deloukas, P., Type 1 Diabetes Genetics Consortium, Todd, J. A., Wallace, C., Concannon, P. & Rich, S. S. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).

3. Barrett, J. C., Clayton, D. G., Concannon, P., Akolkar, B., Cooper, J. D., Erlich, H. A., Julier, C., Morahan, G., Nerup, J., Nierras, C., Plagnol, V., Pociot, F., Schuilenburg, H., Smyth, D. J., Stevens, H., Todd, J. A., Walker, N. M., Rich, S. S., & Type 1 Diabetes Genetics Consortium. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).

4. Bradfield, J. P., Qu, H.-Q., Wang, K., Zhang, H., Sleiman, P. M., Kim, C. E., Mentch, F. D., Qiu, H., Glessner, J. T., Thomas, K. A., Frackelton, E. C., Chiavacci, R. M., Imielinski, M., Monos, D. S., Pandey, R., Bakay, M., Grant, S. F. A., Polychronakos, C. & Hakonarson, H. A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci. *PLOS Genet.* **7**, e1002293 (2011).

5. Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., Taliun, S. A. G., Corvelo, A., Gogarten, S. M., Kang, H. M., Pitsillides, A. N., LeFaive, J., Lee, S., Tian, X., Browning, B. L., Das, S., Emde, A.-K., Clarke, W. E., Loesch, D. P., Shetty, A. C., Blackwell, T. W., Wong, Q., Aguet, F., Albert, C., Alonso, A., Ardlie, K. G., Aslibekyan, S., Auer, P. L., Barnard, J., Barr, R. G., Becker, L. C., Beer, R. L., Benjamin, E. J., Bielak, L. F., Blangero, J., Boehnke, M., Bowden, D. W., Brody, J. A., Burchard, E. G., Cade, B. E., Casella, J. F., Chalazan, B., Chen, Y.-D. I., Cho, M. H., Choi, S. H., Chung, M. K., Clish, C. B., Correa, A., Curran, J. E., Custer, B., Darbar, D., Daya, M., Andrade, M. de, DeMeo, D. L., Dutcher, S. K., Ellinor, P. T., Emery, L. S., Fatkin, D., Forer, L., Fornage, M., Franceschini, N., Fuchsberger, C., Fullerton, S. M., Germer, S., Gladwin, M. T., Gottlieb, D. J., Guo, X., Hall, M. E., He, J., Heard-Costa, N. L., Heckbert, S. R., Irvin, M. R., Johnsen, J. M., Johnson, A. D., Kardia, S. L. R., Kelly, T., Kelly, S., Kenny, E. E., Kiel, D. P., Klemmer, R., Konkle, B. A., Kooperberg, C., Köttgen, A., Lange, L. A., Lasky-Su, J., Levy, D., Lin, X., Lin, K.-H., Liu, C., Loos, R. J. F., Garman, L., Gerszten, R., Lubitz, S. A., Lunetta, K. L., Mak, A. C. Y., Manichaikul, A., Manning, A. K., Mathias, R. A., McManus, D. D., McGarvey, S. T., Meigs, J. B., Meyers, D. A., Mikulla, J. L., Minear, M. A., Mitchell, B., Mohanty, S., Montasser, M. E., Montgomery, C., Morrison, A. C., Murabito, J. M., Natale, A., Natarajan, P., Nelson, S. C., North, K. E., O'Connell, J. R., Palmer, N. D., Pankratz, N., Peloso, G. M., Peyser, P. A., Post, W. S., Psaty, B. M., Rao, D. C., Redline, S., Reiner, A. P., Roden, D., Rotter, J. I., Ruczinski, I., Sarnowski, C., Schoenherr, S., Seo, J.-S., Seshadri, S., Sheehan, V. A., Shoemaker, M. B., Smith, A. V., Smith, N. L., Smith, J. A., Sotoodehnia, N., Stilp, A. M., Tang, W., Taylor, K. D., Telen, M., Thornton, T. A., Tracy, R. P., Berg, D. J. V. D., Vasan, R. S., Viaud-Martinez, K. A., Vrieze, S., Weeks, D. E., Weir, B. S., Weiss, S. T., Weng, L.-C., Willer, C. J., Zhang, Y., Zhao, X., Arnett, D. K., Ashley-Koch, A. E., Barnes, K. C., Boerwinkle, E., Gabriel, S., Gibbs, R., Rice, K. M., Rich, S. S., Silverman, E., Qasba, P., Gan, W., Trans-Omics for Precision Medicine (TOPMed) Program,

Topm. P. G. W. G., Papanicolaou, G. J., Nickerson, D. A., Browning, S. R., Zody, M. C., Zöllner, S., Wilson, J. G., Cupples, L. A., Laurie, C. C., Jaquish, C. E., Hernandez, R. D., O'Connor, T. D. & Abecasis, G. R. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv* 563866 (2019) doi:10.1101/563866.

6. Aylward, A., Chiou, J., Okino, M.-L., Kadakia, N. & Gaulton, K. J. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Hum. Mol. Genet.* (2018) doi:10.1093/hmg/ddy314.

7. Raeder, H., Johansson, S., Holm, P. I., Haldorsen, I. S., Mas, E., Sbarra, V., Nermoen, I., Eide, S. A., Grevle, L., Bjørkhaug, L., Sagen, J. V., Aksnes, L., Søvik, O., Lombardo, D., Molven, A. & Njølstad, P. R. Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nat. Genet.* **38**, 54–62 (2006).

8. Ramos, P. S., Criswell, L. A., Moser, K. L., Comeau, M. E., Williams, A. H., Pajewski, N. M., Chung, S. A., Graham, R. R., Zidovetzki, R., Kelly, J. A., Kaufman, K. M., Jacob, C. O., Vyse, T. J., Tsao, B. P., Kimberly, R. P., Gaffney, P. M., Alarcón-Riquelme, M. E., Harley, J. B., Langefeld, C. D., & International Consortium on the Genetics of Systemic Erythematosus. A comprehensive analysis of shared loci between systemic lupus erythematosus (SLE) and sixteen autoimmune diseases reveals limited genetic overlap. *PLoS Genet.* **7**, e1002406 (2011).

9. Chiou, J., Zeng, C., Cheng, Z., Han, J. Y., Schlichting, M., Huang, S., Wang, J., Sui, Y., Deogaygay, A., Okino, M.-L., Qiu, Y., Sun, Y., Kudtarkar, P., Fang, R., Preissl, S., Sander, M., Gorkin, D. & Gaulton, K. J. Single cell chromatin accessibility reveals pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. *bioRxiv* 693671 (2019) doi:10.1101/693671.

10. Gibson-Corley, K. N., Meyerholz, D. K. & Engelhardt, J. F. Pancreatic Pathophysiology in Cystic Fibrosis. *J. Pathol.* **238**, 311–320 (2016).

11. Sharer, N., Schwarz, M., Malone, G., Howarth, A., Painter, J., Super, M. & Braganza, J. Mutations of the cystic fibrosis gene in patients with chronic pancreatitis. *N. Engl. J. Med.* **339**, 645–652 (1998).

12. Diaferia, G. R., Balestrieri, C., Prosperini, E., Nicoli, P., Spaggiari, P., Zerbi, A. & Natoli, G. Dissection of transcriptional and cis-regulatory control of differentiation in human pancreatic cancer. *EMBO J.* **35**, 595–617 (2016).

13. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lead analysts:, Laboratory, Data Analysis &Coordinating Center (LDACC):, NIH program management:, Biospecimen collection:, Pathology:, eQTL manuscript working group:, Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

14. McWilliams, R. R., Petersen, G. M., Rabe, K. G., Holtegaard, L. M., Lynch, P. J., Bishop, M. D. & Highsmith, W. E. Cystic fibrosis transmembrane conductance regulator (CFTR) gene mutations and risk for pancreatic adenocarcinoma. *Cancer* **116**, 203–209 (2010).

15. Noone, P. G., Zhou, Z., Silverman, L. M., Jowell, P. S., Knowles, M. R. & Cohn, J. A. Cystic fibrosis gene mutations and pancreatitis risk: relation to epithelial ion transport and trypsin inhibitor gene mutations. *Gastroenterology* **121**, 1310–1319 (2001).

16. Virostko, J., Williams, J., Hilmes, M., Bowman, C., Wright, J. J., Du, L., Kang, H., Russell, W. E., Powers, A. C. & Moore, D. J. Pancreas Volume Declines During the First Year After Diagnosis of Type 1 Diabetes and Exhibits Altered Diffusion at Disease Onset. *Diabetes Care* **42**, 248–257 (2019).

17. Campbell-Thompson, M., Rodriguez-Calvo, T. & Battaglia, M. Abnormalities of the Exocrine Pancreas in Type 1 Diabetes. *Curr. Diab. Rep.* **15**, 79 (2015).

18. Camunas-Soler, J., Dai, X.-Q., Hang, Y., Bautista, A., Lyon, J., Suzuki, K., Kim, S. K., Quake, S. R. & MacDonald, P. E. Patch-Seq Links Single-Cell Transcriptomes to Human Islet Dysfunction in Diabetes. *Cell Metab.* **31**, 1017-1031.e4 (2020).

19. Hart, N. J., Aramandla, R., Poffenberger, G., Fayolle, C., Thames, A. H., Bautista, A., Spigelman, A. F., Babon, J. A. B., DeNicola, M. E., Dadi, P. K., Bush, W. S., Balamurugan, A. N., Brissova, M., Dai, C., Prasad, N., Bottino, R., Jacobson, D. A., Drumm, M. L., Kent, S. C., MacDonald, P. E. & Powers, A. C. Cystic fibrosis-related diabetes is caused by islet loss and inflammation. *JCI Insight* **3**, (2018).

20. Valle, A., Giamporcaro, G. M., Scavini, M., Stabilini, A., Grogan, P., Bianconi, E., Sebastiani, G., Masini, M., Maugeri, N., Porretti, L., Bonfanti, R., Meschi, F., De Pellegrin, M., Lesma, A., Rossini, S., Piemonti, L., Marchetti, P., Dotta, F., Bosi, E. & Battaglia, M. Reduction of circulating neutrophils precedes and accompanies type 1 diabetes. *Diabetes* **62**, 2072–2077 (2013).

21. Navis, A. & Bagnat, M. Loss of cftr function leads to pancreatic destruction in larval zebrafish. *Dev. Biol.* **399**, 237–248 (2015).

22. Lin, Y., Nakatochi, M., Hosono, Y., Ito, H., Kamatani, Y., Inoko, A., Sakamoto, H., Kinoshita, F., Kobayashi, Y., Ishii, H., Ozaka, M., Sasaki, T., Matsuyama, M., Sasahira, N., Morimoto, M., Kobayashi, S., Fukushima, T., Ueno, M., Ohkawa, S., Egawa, N., Kuruma, S., Mori, M., Nakao, H., Adachi, Y., Okuda, M., Osaki, T., Kamiya, S., Wang, C., Hara, K., Shimizu, Y., Miyamoto, T., Hayashi, Y., Ebi, H., Kohmoto, T., Imoto, I., Kasugai, Y., Murakami, Y., Akiyama, M., Ishigaki, K., Matsuda, K., Hirata, M., Shimada, K., Okusaka, T., Kawaguchi, T., Takahashi, M., Watanabe, Y., Kuriki, K., Kadota, A., Okada, R., Mikami, H., Takezaki, T., Suzuki, S., Yamaji, T., Iwasaki, M., Sawada, N., Goto, A., Kinoshita, K., Fuse, N., Katsuoka, F., Shimizu, A., Nishizuka, S. S., Tanno, K., Suzuki, K., Okada, Y., Horikoshi, M., Yamauchi, T., Kadowaki, T., Yu, H., Zhong, J., Amundadottir, L. T., Doki, Y., Ishii, H., Eguchi, H., Bogumil, D., Haiman, C. A., Le Marchand, L., Mori, M., Risch, H., Setiawan, V. W., Tsugane, S., Wakai, K., Yoshida, T., Matsuda, F., Kubo, M., Kikuchi, S. & Matsuo, K. Genome-wide association meta-analysis identifies GP2 gene risk variants for pancreatic cancer. *Nat. Commun.* **11**, 3175 (2020).

23. Wolpin, B. M., Rizzato, C., Kraft, P., Kooperberg, C., Petersen, G. M., Wang, Z., Arslan, A. A., Beane-Freeman, L., Bracci, P. M., Buring, J., Canzian, F., Duell, E. J., Gallinger, S., Giles, G. G., Goodman, G. E., Goodman, P. J., Jacobs, E. J., Kamineni, A., Klein, A. P., Kolonel, L. N., Kulke, M. H., Li, D., Malats, N., Olson, S. H., Risch, H. A., Sesso, H. D., Visvanathan, K., White, E., Zheng, W., Abnet, C. C., Albanes, D., Andreotti, G., Austin, M. A., Barfield, R., Basso, D., Berndt, S. I., Boutron-Ruault, M.-C., Brotzman, M., Büchler, M. W., Bueno-de-Mesquita, H. B., Bugert, P., Burdette, L., Campa, D., Caporaso, N. E., Capurso, G., Chung, C., Cotterchio, M., Costello, E., Elena, J., Funel, N., Gaziano, J. M., Giese, N. A., Giovannucci, E. L., Goggins, M., Gorman, M. J., Gross, M., Haiman, C. A., Hassan, M., Helzlsouer, K. J., Henderson, B. E., Holly, E. A., Hu, N., Hunter, D. J., Innocenti, F., Jenab, M., Kaaks, R., Key, T. J., Khaw, K.-T., Klein, E. A., Kogevinas, M., Krogh, V., Kupcinskas, J., Kurtz, R. C., LaCroix, A., Landi, M. T., Landi, S., Le Marchand, L., Mambrini, A., Mannisto, S., Milne, R. L., Nakamura, Y., Oberg, A. L., Owzar, K., Patel, A. V., Peeters, P. H. M., Peters, U., Pezzilli, R., Piepoli, A., Porta, M., Real, F. X., Riboli, E., Rothman, N., Scarpa, A., Shu, X.-O., Silverman, D. T., Soucek, P., Sund, M., Talar-Wojnarowska, R., Taylor, P. R., Theodoropoulos, G. E., Thornquist, M., Tjønneland, A., Tobias, G. S., Trichopoulos, D., Vodicka, P., Wactawski-Wende, J., Wentzensen, N., Wu, C., Yu, H., Yu, K., Zeleniuch-Jacquotte, A., Hoover, R., Hartge, P., Fuchs, C., Chanock, S. J., Stolzenberg-Solomon, R. S. & Amundadottir, L. T. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat. Genet.* **46**, 994–1000 (2014).

24. Johansson, B. B., Fjeld, K., El Jellas, K., Gravdal, A., Dalva, M., Tjora, E., Ræder, H., Kulkarni, R. N., Johansson, S., Njølstad, P. R. & Molven, A. The role of the carboxyl ester lipase (CEL) gene in pancreatic disease. *Pancreatol. Off. J. Int. Assoc. Pancreatol. IAP Al* **18**, 12–19 (2018).

25. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. & Sham, P. C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

26. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., van Duijn, C. M., Gillies, C. E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J., Boomsma, D. I., Branham, K., Breen, G., Brummet, C., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F. S., Corbin, L., Davey Smith, G., Dedoussis, G., Dorr, M., Farmaki, A.-E., Ferrucci, L., Forer, L., Fraser, R. M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O. L., Hveem, K., Kretzler, M., Lee, J., McGue, M., Meitinger, T., Melzer, D., Min, J., Mohlke, K. L., Vincent, J., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, B., Sala, C., Salomaa, V., Schlessinger, D., Schoenheer, S., Slagboom, P. E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van den Berg, L., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M. G., Wilson, J. F., Frayling, T., de Bakker, P., Swertz, M. A., McCarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., Soranzo, N., Walter, K., Swaroop, A., Cucca, F., Anderson, C., Boehnke, M., McCarthy, M. I., Durbin, R., Abecasis, G. & Marchini, J. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

27. 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A. & Abecasis, G. R. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

28. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G. R. & Fuchsberger, C. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

29. Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., Wolford, B. N., LeFaive, J., VandeHaar, P., Gagliano, S. A., Gifford, A., Bastarache, L. A., Wei, W.-Q., Denny, J. C., Lin, M., Hveem, K., Kang, H. M., Abecasis, G. R., Willer, C. J. & Lee, S. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

30. Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M. J., Price, A. L. & Neale, B. M. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

31. Evangelou, M., Smyth, D. J., Fortune, M. D., Burren, O. S., Walker, N. M., Guo, H., Onengut-Gumuscu, S., Chen, W.-M., Concannon, P., Rich, S. S., Todd, J. A. & Wallace, C. A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. *Genet. Epidemiol.* **38**, 661–670 (2014).

32. Benner, C., Spencer, C. C. A., Havulinna, A. S., Salomaa, V., Ripatti, S. & Pirinen, M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinforma. Oxf. Engl.* **32**, 1493–1501 (2016).

33. Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L. & Neale, B. M. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

34. Ji, S.-G., Juran, B. D., Mucha, S., Folseraas, T., Jostins, L., Melum, E., Kumasaka, N., Atkinson, E. J., Schlicht, E. M., Liu, J. Z., Shah, T., Gutierrez-Achury, J., Boberg, K. M., Bergquist, A., Vermeire, S., Eksteen, B., Durie, P. R., Farkkila, M., Müller, T., Schramm, C., Sterneck, M., Weismüller, T. J., Gotthardt, D. N., Ellinghaus, D., Braun, F., Teufel, A., Laudes, M., Lieb, W., Jacobs, G., Beuers, U., Weersma, R. K., Wijmenga, C., Marschall, H.-U., Milkiewicz, P., Pares, A., Kontula, K., Chazouillères, O., Invernizzi, P., Goode, E., Spiess, K., Moore, C., Sambrook, J., Ouwehand, W. H., Roberts, D. J., Danesh, J., Floreani, A., Gulamhusein, A. F., Eaton, J. E., Schreiber, S., Coltescu, C., Bowlus, C. L., Luketic, V. A., Odin, J. A., Chopra, K. B., Kowdley, K. V., Chalasani, N., Manns, M. P., Srivastava, B., Mells, G., Sandford, R. N., Alexander, G., Gaffney, D. J., Chapman, R. W., Hirschfield, G. M., de Andrade, M., UK-PSC Consortium, International IBD Genetics Consortium, International PSC Study Group, Rushbrook, S. M., Franke, A., Karlsen, T. H., Lazaridis, K. N. & Anderson, C. A. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat. Genet.* **49**, 269–273 (2017).

35. Bentham, J., Morris, D. L., Graham, D. S. C., Pinder, C. L., Tombleson, P., Behrens, T. W., Martín, J., Fairfax, B. P., Knight, J. C., Chen, L., Replogle, J., Syvänen, A.-C., Rönnblom, L., Graham, R. R., Wither, J. E., Rioux, J. D., Alarcón-Riquelme, M. E. & Vyse, T. J. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**, 1457–1464 (2015).

36. Cordell, H. J., Han, Y., Mells, G. F., Li, Y., Hirschfield, G. M., Greene, C. S., Xie, G., Juran, B. D., Zhu, D., Qian, D. C., Floyd, J. A. B., Morley, K. I., Prati, D., Lleo, A., Cusi, D., Canadian-US PBC Consortium, Italian PBC Genetics Study Group, UK-PBC Consortium, Gershwin, M. E., Anderson, C. A., Lazaridis, K. N., Invernizzi, P., Seldin, M. F., Sandford, R. N., Amos, C. I. & Siminovitch, K. A. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat. Commun.* **6**, 8019 (2015).

37. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., Graham, R. R., Manoharan, A., Ortmann, W., Bhangale, T., Denny, J. C., Carroll, R. J., Eyler, A. E., Greenberg, J. D., Kremer, J. M., Pappas, D. A., Jiang, L., Yin, J., Ye, L., Su, D.-F., Yang, J., Xie, G., Keystone, E., Westra, H.-J., Esko, T., Metspalu, A., Zhou, X., Gupta, N., Mirel, D., Stahl, E. A., Diogo, D., Cui, J., Liao, K., Guo, M. H., Myouzen, K., Kawaguchi, T., Coenen, M. J. H., van Riel, P. L. C. M., van de Laar, M. A. F. J., Guchelaar, H.-J., Huizinga, T. W. J., Dieudé, P., Mariette, X., Bridges, S. L., Zhernakova, A., Toes, R. E. M., Tak, P. P., Miceli-Richard, C., Bang, S.-Y., Lee, H.-S., Martin, J., Gonzalez-Gay, M. A., Rodriguez-Rodriguez, L., Rantapää-Dahlqvist, S., Ärlestig, L., Choi, H. K., Kamatani, Y., Galan, P., Lathrop, M., Eyre, S., Bowes, J., Barton, A., de Vries, N., Moreland, L. W., Criswell, L. A., Karlson, E. W., Taniguchi, A., Yamada, R., Kubo, M., Liu, J. S., Bae, S.-C., Worthington, J., Padyukov, L., Klareskog, L., Gregersen, P. K., Raychaudhuri, S., Stranger, B. E., De Jager, P. L., Franke, L., Visscher, P. M., Brown, M. A., Yamanaka, H., Mimori, T., Takahashi, A., Xu, H., Behrens, T. W., Siminovitch, K. A., Momohara, S., Matsuda, F., Yamamoto, K. & Plenge, R. M. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).

38. de Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., Jostins, L., Rice, D. L., Gutierrez-Achury, J., Ji, S.-G., Heap, G., Nimmo, E. R., Edwards, C., Henderson, P., Mowat, C., Sanderson, J., Satsangi, J., Simmons, A., Wilson, D. C., Tremelling, M., Hart, A., Mathew, C. G., Newman, W. G., Parkes, M., Lees, C. W., Uhlig, H., Hawkey, C., Prescott, N. J., Ahmad, T., Mansfield, J. C., Anderson, C. A. & Barrett, J. C. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).

39. Dubois, P. C. A., Trynka, G., Franke, L., Hunt, K. A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G. A. R., Adány, R., Aromaa, A., Bardella, M. T., van den Berg, L. H., Bockett, N. A., de la Concha, E. G., Dema, B., Fehrmann, R. S. N., Fernández-Arquero, M., Fiatal, S., Grandone, E., Green, P. M., Groen, H. J. M., Gwilliam, R., Houwen, R. H. J., Hunt, S. E., Kaukinen, K., Kelleher, D., Korponay-Szabo, I., Kurppa, K., MacMathuna, P., Mäki, M., Mazzilli, M. C., McCann, O. T., Mearin, M. L., Mein, C. A., Mirza, M. M., Mistry, V., Mora, B., Morley, K. I., Mulder, C. J., Murray, J. A., Núñez, C., Oosterom, E., Ophoff, R. A., Polanco, I., Peltonen, L., Platteel, M., Rybak, A., Salomaa, V., Schweizer, J. J., Sperandeo, M. P., Tack, G. J., Turner, G., Veldink, J. H., Verbeek, W. H. M., Weersma, R. K., Wolters, V. M., Urcelay, E., Cukrowska, B., Greco, L., Neuhausen, S. L., McManus, R., Barisani, D., Deloukas, P., Barrett, J. C., Saavalainen, P., Wijmenga, C. & van Heel, D. A. Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).

40. Jin, Y., Andersen, G., Yorgov, D., Ferrara, T. M., Ben, S., Brownson, K. M., Holland, P. J., Birlea, S. A., Siebert, J., Hartmann, A., Lienert, A., van Geel, N., Lambert, J., Luiten, R. M., Wolkerstorfer, A., Wietze van der Veen, J. P., Bennett, D. C., Taïeb, A., Ezzedine, K., Kemp, E. H., Gawkrodger, D. J., Weetman, A. P., Kõks, S., Prans, E., Kingo, K., Karelson, M., Wallace, M. R., McCormack, W. T., Overbeck, A., Moretti, S., Colucci, R., Picardo, M., Silverberg, N. B., Olsson, M., Valle, Y., Korobko, I., Böhm, M., Lim, H. W., Hamzavi, I., Zhou, L., Mi, Q.-S., Fain, P. R., Santorico, S. A. & Spritz, R. A. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat. Genet.* **48**, 1418–1424 (2016).

41. Paternoster, L., Standl, M., Waage, J., Baurecht, H., Hotze, M., Strachan, D. P., Curtin, J. A., Bønnelykke, K., Tian, C., Takahashi, A., Esparza-Gordillo, J., Alves, A. C., Thyssen, J. P., den Dekker, H. T., Ferreira, M. A., Altmaier, E., Sleiman, P. M., Xiao, F. L., Gonzalez, J. R., Marenholz, I., Kalb, B., Yanes, M. P., Xu, C.-J., Carstensen, L., Groen-Blokhuis, M. M., Venturini, C., Pennell, C. E., Barton, S. J., Levin, A. M., Curjuric, I., Bustamante, M., Kreiner-Møller, E., Lockett, G. A., Bacelis, J., Bunyavanich, S., Myers, R. A., Matanovic, A., Kumar, A., Tung, J. Y., Hirota, T., Kubo, M., McArdle, W. L., Henderson, A. J., Kemp, J. P., Zheng, J., Smith, G. D., Rüschendorf, F., Bauerfeind, A., Lee-Kirsch, M. A., Arnold, A., Homuth, G., Schmidt, C. O., Mangold, E., Cichon, S., Keil, T., Rodríguez, E., Peters, A., Franke, A., Lieb, W., Novak, N., Fölster-Holst, R., Horikoshi, M., Pekkanen, J., Sebert, S., Husemoen, L. L., Grarup, N., de Jongste, J. C., Rivadeneira, F., Hofman, A., Jaddoe, V. W., Pasmans, S. G., Elbert, N. J., Uitterlinden, A. G., Marks, G. B., Thompson, P. J., Matheson, M. C., Robertson, C. F., Australian Asthma Genetics Consortium (AAGC), Ried, J. S., Li, J., Zuo, X. B., Zheng, X. D., Yin, X. Y., Sun, L. D., McAleer, M. A., O'Regan, G. M., Fahy, C. M., Campbell, L. E., Macek, M., Kurek, M., Hu, D., Eng, C., Postma, D. S., Feenstra, B., Geller, F., Hottenga, J. J., Middeldorp, C. M., Hysi, P., Bataille, V., Spector, T., Tiesler, C. M., Thiering, E., Pahukasahasram, B., Yang, J. J., Imboden, M., Huntsman, S., Vilor-Tejedor, N., Relton, C. L., Myhre, R., Nystad, W., Custovic, A., Weiss, S. T., Meyers, D. A., Söderhäll, C., Melén, E., Ober, C., Raby, B. A., Simpson, A., Jacobsson, B., Holloway, J. W., Bisgaard, H., Sunyer, J., Hensch, N. M. P., Williams, L. K., Godfrey, K. M., Wang, C. A., Boomsma, D. I., Melbye, M., Koppelman, G. H., Jarvis, D., McLean, W. I., Irvine, A. D., Zhang, X. J., Hakonarson, H., Gieger, C., Burchard, E. G., Martin, N. G., Duijts, L., Linneberg, A., Jarvelin, M.-R., Noethen, M. M., Lau, S., Hübner, N., Lee, Y.-A., Tamari, M., Hinds, D. A., Glass, D., Brown, S. J., Heinrich, J., Evans, D. M. & Weidinger, S. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat. Genet.* **47**, 1449–1456 (2015).

42. López-Isac, E., Acosta-Herrera, M., Kerick, M., Assassi, S., Satpathy, A. T., Granja, J., Mumbach, M. R., Beretta, L., Simeón, C. P., Carreira, P., Ortego-Centeno, N., Castellvi, I., Bossini-Castillo, L., Carmona, F. D., Orozco, G., Hunzelmann, N., Distler, J. H. W., Franke, A., Lunardi, C., Moroncini, G., Gabrielli, A., de Vries-Bouwstra, J., Wijmenga, C., Koeleman, B. P. C., Nordin, A., Padyukov, L., Hoffmann-Vold, A.-M., Lie, B., European Scleroderma Group†, Proudman, S., Stevens, W., Nikpour, M., Australian Scleroderma Interest Group (ASIG), Vyse, T., Herrick, A. L., Worthington, J., Denton, C. P., Allanore, Y., Brown, M. A., Radstake, T. R. D. J., Fonseca, C., Chang, H. Y., Mayes, M. D. & Martin, J. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat. Commun.* **10**, 4955 (2019).

43. Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., Sealock, J., Karlsson, I. K., Hägg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I. S., Andersen, F., Bergh, S., Bettella, F., Bjornsson, S., Brækhus,

A., Bråthen, G., de Leeuw, C., Desikan, R. S., Djurovic, S., Dumitrescu, L., Fladby, T., Hohman, T. J., Jonsson, P. V., Kiddle, S. J., Rongve, A., Saltvedt, I., Sando, S. B., Selbæk, G., Shoai, M., Skene, N. G., Snaedal, J., Stordal, E., Ulstein, I. D., Wang, Y., White, L. R., Hardy, J., Hjerling-Leffler, J., Sullivan, P. F., van der Flier, W. M., Dobson, R., Davis, L. K., Stefansson, H., Stefansson, K., Pedersen, N. L., Ripke, S., Andreassen, O. A. & Posthuma, D. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).

44. Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M. H., Prins, B. P., Guo, X., Bielak, L. F., Below, J. E., Bowden, D. W., Chambers, J. C., Kim, Y. J., Ng, M. C. Y., Petty, L. E., Sim, X., Zhang, W., Bennett, A. J., Bork-Jensen, J., Brummett, C. M., Canouil, M., Ec Kardt, K.-U., Fischer, K., Kardia, S. L. R., Kronenberg, F., Läll, K., Liu, C.-T., Locke, A. E., Luan, J., Ntalla, I., Nylander, V., Schönherr, S., Schurmann, C., Yengo, L., Bottinger, E. P., Brandslund, I., Christensen, C., Dedoussis, G., Florez, J. C., Ford, I., Franco, O. H., Frayling, T. M., Giedraitis, V., Hackinger, S., Hattersley, A. T., Herder, C., Ikram, M. A., Ingelsson, M., Jørgensen, M. E., Jørgensen, T., Kriebel, J., Kuusisto, J., Ligthart, S., Lindgren, C. M., Linneberg, A., Lyssenko, V., Mamakou, V., Meitinger, T., Mohlke, K. L., Morris, A. D., Nadkarni, G., Pankow, J. S., Peters, A., Sattar, N., Stančáková, A., Strauch, K., Taylor, K. D., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Witte, D. R., Dupuis, J., Peyser, P. A., Zeggini, E., Loos, R. J. F., Froguel, P., Ingelsson, E., Lind, L., Groop, L., Laakso, M., Collins, F. S., Jukema, J. W., Palmer, C. N. A., Grallert, H., Metspalu, A., Dehghan, A., Köttgen, A., Abecasis, G. R., Meigs, J. B., Rotter, J. I., Marchini, J., Pedersen, O., Hansen, T., Langenberg, C., Wareham, N. J., Stefansson, K., Gloyn, A. L., Morris, A. P., Boehnke, M. & McCarthy, M. I. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).

45. Nelson, C. P., Goel, A., Butterworth, A. S., Kanoni, S., Webb, T. R., Marouli, E., Zeng, L., Ntalla, I., Lai, F. Y., Hopewell, J. C., Giannakopoulou, O., Jiang, T., Hamby, S. E., Di Angelantonio, E., Assimes, T. L., Bottinger, E. P., Chambers, J. C., Clarke, R., Palmer, C. N. A., Cubbon, R. M., Ellinor, P., Ermel, R., Evangelou, E., Franks, P. W., Grace, C., Gu, D., Hingorani, A. D., Howson, J. M. M., Ingelsson, E., Kastrati, A., Kessler, T., Kyriakou, T., Lehtimäki, T., Lu, X., Lu, Y., März, W., McPherson, R., Metspalu, A., Pujades-Rodriguez, M., Ruusalepp, A., Schadt, E. E., Schmidt, A. F., Sweeting, M. J., Zalloua, P. A., AlGhalayini, K., Keavney, B. D., Kooner, J. S., Loos, R. J. F., Patel, R. S., Rutter, M. K., Tomaszewski, M., Tzoulaki, I., Zeggini, E., Erdmann, J., Dedoussis, G., Björkegren, J. L. M., EPIC-CVD Consortium, CARDIoGRAMplusC4D, UK Biobank CardioMetabolic Consortium CHD working group, Schunkert, H., Farrall, M., Danesh, J., Samani, N. J., Watkins, H. & Deloukas, P. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).

46. Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J. R. I., Gaspar, H. A., de Leeuw, C. A., Steinberg, S., Pavlides, J. M. W., Trzaskowski, M., Byrne, E. M., Pers, T. H., Holmans, P. A., Richards, A. L., Abbott, L., Agerbo, E., Akil, H., Albani, D., Alliey-Rodriguez, N., Als, T. D., Anjorin, A., Antilla, V., Awasthi, S., Badner, J. A., Bækvad-Hansen, M., Barchas, J. D., Bass, N., Bauer, M., Belliveau, R., Bergen, S. E., Pedersen, C. B., Bøen, E., Boks, M. P., Boocock, J., Budde, M., Bunney, W., Burmeister, M., Bybjerg-Grauholm, J., Byerley, W., Casas, M., Cerrato, F., Cervantes, P., Chambert, K., Charney, A. W., Chen, D., Churchhouse, C., Clarke, T.-K., Coryell, W., Craig, D. W., Cruceanu, C., Curtis, D., Czerski, P. M., Dale, A. M., de Jong, S., Degenhardt, F., Del-

Favero, J., DePaulo, J. R., Djurovic, S., Dobbyn, A. L., Dumont, A., Elvsåshagen, T., Escott-Price, V., Fan, C. C., Fischer, S. B., Flickinger, M., Foroud, T. M., Forty, L., Frank, J., Fraser, C., Freimer, N. B., Frisén, L., Gade, K., Gage, D., Garnham, J., Giambartolomei, C., Pedersen, M. G., Goldstein, J., Gordon, S. D., Gordon-Smith, K., Green, E. K., Green, M. J., Greenwood, T. A., Grove, J., Guan, W., Guzman-Parra, J., Hamshere, M. L., Hautzinger, M., Heilbronner, U., Herms, S., Hipolito, M., Hoffmann, P., Holland, D., Huckins, L., Jamain, S., Johnson, J. S., Juréus, A., Kandaswamy, R., Karlsson, R., Kennedy, J. L., Kittel-Schneider, S., Knowles, J. A., Kogevinas, M., Koller, A. C., Kupka, R., Lavebratt, C., Lawrence, J., Lawson, W. B., Leber, M., Lee, P. H., Levy, S. E., Li, J. Z., Liu, C., Lucae, S., Maaser, A., MacIntyre, D. J., Mahon, P. B., Maier, W., Martinsson, L., McCarroll, S., McGuffin, P., McInnis, M. G., McKay, J. D., Medeiros, H., Medland, S. E., Meng, F., Milani, L., Montgomery, G. W., Morris, D. W., Mühleisen, T. W., Mullins, N., Nguyen, H., Nievergelt, C. M., Adolfsson, A. N., Nwulia, E. A., O'Donovan, C., Loohuis, L. M. O., Ori, A. P. S., Oruc, L., Ösby, U., Perlis, R. H., Perry, A., Pfennig, A., Potash, J. B., Purcell, S. M., Regeer, E. J., Reif, A., Reinbold, C. S., Rice, J. P., Rivas, F., Rivera, M., Roussos, P., Ruderfer, D. M., Ryu, E., Sánchez-Mora, C., Schatzberg, A. F., Scheftner, W. A., Schork, N. J., Shannon Weickert, C., Shehktman, T., Shilling, P. D., Sigurdsson, E., Slaney, C., Smeland, O. B., Sobell, J. L., Søholm Hansen, C., Spijker, A. T., St Clair, D., Steffens, M., Strauss, J. S., Streit, F., Strohmaier, J., Szelinger, S., Thompson, R. C., Thorgeirsson, T. E., Treutlein, J., Vedder, H., Wang, W., Watson, S. J., Weickert, T. W., Witt, S. H., Xi, S., Xu, W., Young, A. H., Zandi, P., Zhang, P., Zöllner, S., eQTLGen Consortium, BIOS Consortium, Adolfsson, R., Agartz, I., Alda, M., Backlund, L., Baune, B. T., Bellivier, F., Berrettini, W. H., Biernacka, J. M., Blackwood, D. H. R., Boehnke, M., Børglum, A. D., Corvin, A., Craddock, N., Daly, M. J., Dannlowski, U., Esko, T., Etain, B., Frye, M., Fullerton, J. M., Gershon, E. S., Gill, M., Goes, F., Grigoroiu-Serbanescu, M., Hauser, J., Hougaard, D. M., Hultman, C. M., Jones, I., Jones, L. A., Kahn, R. S., Kirov, G., Landén, M., Leboyer, M., Lewis, C. M., Li, Q. S., Lissowska, J., Martin, N. G., Mayoral, F., McElroy, S. L., McIntosh, A. M., McMahon, F. J., Melle, I., Metspalu, A., Mitchell, P. B., Morken, G., Mors, O., Mortensen, P. B., Müller-Myhsok, B., Myers, R. M., Neale, B. M., Nimgaonkar, V., Nordentoft, M., Nöthen, M. M., O'Donovan, M. C., Oedegaard, K. J., Owen, M. J., Paciga, S. A., Pato, C., Pato, M. T., Posthuma, D., Ramos-Quiroga, J. A., Ribasés, M., Rietschel, M., Rouleau, G. A., Schalling, M., Schofield, P. R., Schulze, T. G., Serretti, A., Smoller, J. W., Stefansson, H., Stefansson, K., Stordal, E., Sullivan, P. F., Turecki, G., Vaaler, A. E., Vieta, E., Vincent, J. B., Werge, T., Nurnberger, J. I., Wray, N. R., Di Florio, A., Edenberg, H. J., Cichon, S., Ophoff, R. A., Scott, L. J., Andreassen, O. A., Kelsoe, J., Sklar, P., & Bipolar Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).

47. Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J. H., Clarke, T.-K., Coleman, J. I. R., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G. E., Crowley, C. A., Dashti, H. S., Davies, G., Deary, I. J., Degenhardt, F., Derks, E. M., Direk, N., Dolan, C. V., Dunn, E. C., Eley, T. C., Eriksson, N., Escott-Price, V., Kiadeh, F. H. F., Finucane, H. K., Forstner, A. J., Frank, J., Gaspar, H. A., Gill, M., Giusti-Rodríguez, P., Goes, F. S., Gordon, S. D., Grove, J., Hall, L. S., Hannon, E., Hansen, C. S., Hansen, T. F., Herms, S., Hickie, I. B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.-J., Hougaard, D. M., Hu, M., Hyde, C. L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles, J. A., Kohane, I. S., Kraft, J., Kretzschmar, W. W., Krogh, J., Kutalik, Z., Lane, J. M., Li, Y., Li, Y., Lind, P. A., Liu, X., Lu, L., MacIntyre, D. J., MacKinnon, D. F., Maier, R. M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S. E., Mehta, D., Middeldorp, C. M., Mihailov, E., Milaneschi, Y.,

Milani, L., Mill, J., Mondimore, F. M., Montgomery, G. W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M. G., Nyholt, D. R., O'Reilly, P. F., Oskarsson, H., Owen, M. J., Painter, J. N., Pedersen, C. B., Pedersen, M. G., Peterson, R. E., Pettersson, E., Peyrot, W. J., Pistis, G., Posthuma, D., Purcell, S. M., Quiroz, J. A., Qvist, P., Rice, J. P., Riley, B. P., Rivera, M., Saeed Mirza, S., Saxena, R., Schoevers, R., Schulte, E. C., Shen, L., Shi, J., Shyn, S. I., Sigurdsson, E., Sinnamon, G. B. C., Smit, J. H., Smith, D. J., Stefansson, H., Steinberg, S., Stockmeier, C. A., Streit, F., Strohmaier, J., Tansey, K. E., Teismann, H., Teumer, A., Thompson, W., Thomson, P. A., Thorgeirsson, T. E., Tian, C., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A. G., Umbricht, D., Van der Auwera, S., van Hemert, A. M., Viktorin, A., Visscher, P. M., Wang, Y., Webb, B. T., Weinsheimer, S. M., Wellmann, J., Willemsen, G., Witt, S. H., Wu, Y., Xi, H. S., Yang, J., Zhang, F., eQTLGen, 23andMe, Arolt, V., Baune, B. T., Berger, K., Boomsma, D. I., Cichon, S., Dannlowski, U., de Geus, E. C. J., DePaulo, J. R., Domenici, E., Domschke, K., Esko, T., Grabe, H. J., Hamilton, S. P., Hayward, C., Heath, A. C., Hinds, D. A., Kendler, K. S., Kloiber, S., Lewis, G., Li, Q. S., Lucae, S., Madden, P. F. A., Magnusson, P. K., Martin, N. G., McIntosh, A. M., Metspalu, A., Mors, O., Mortensen, P. B., Müller-Myhsok, B., Nordentoft, M., Nöthen, M. M., O'Donovan, M. C., Paciga, S. A., Pedersen, N. L., Penninx, B. W. J. H., Perlis, R. H., Porteous, D. J., Potash, J. B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T. G., Smoller, J. W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M. M., Werge, T., Winslow, A. R., Lewis, C. M., Levinson, D. F., Breen, G., Børglum, A. D., Sullivan, P. F., & Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

48. Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., Churchhouse, C., Dellenvall, K., Demontis, D., De Rubeis, S., Devlin, B., Djurovic, S., Dumont, A. L., Goldstein, J. I., Hansen, C. S., Hauberg, M. E., Hollegaard, M. V., Hope, S., Howrigan, D. P., Huang, H., Hultman, C. M., Klei, L., Maller, J., Martin, J., Martin, A. R., Moran, J. L., Nyegaard, M., Nærland, T., Palmer, D. S., Palotie, A., Pedersen, C. B., Pedersen, M. G., Poterba, T., Poulsen, J. B., St Pourcain, B., Qvist, P., Rehnström, K., Reichenberg, A., Reichert, J., Robinson, E. B., Roeder, K., Roussos, P., Saemundsen, E., Sandin, S., Satterstrom, F. K., Smith, G. D., Stefansson, H., Steinberg, S., Stevens, C. R., Sullivan, P. F., Turley, P., Walters, G. B., Xu, X., Stefansson, K., Geschwind, D. H., Nordentoft, M., Hougaard, D. M., Werge, T., Mors, O., Mortensen, P. B., Neale, B. M., Daly, M. J. & Børglum, A. D. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

49. Watson, H. J., Yilmaz, Z., Thornton, L. M., Hübel, C., Coleman, J. R. I., Gaspar, H. A., Bryois, J., Hinney, A., Leppä, V. M., Mattheisen, M., Medland, S. E., Ripke, S., Yao, S., Giusti-Rodríguez, P., Anorexia Nervosa Genetics Initiative, Hanscombe, K. B., Purves, K. L., Eating Disorders Working Group of the Psychiatric Genomics Consortium, Adan, R. A. H., Alfredsson, L., Ando, T., Andreassen, O. A., Baker, J. H., Berrettini, W. H., Boehm, I., Boni, C., Perica, V. B., Buehren, K., Burghardt, R., Cassina, M., Cichon, S., Clementi, M., Cone, R. D., Courtet, P., Crow, S., Crowley, J. J., Danner, U. N., Davis, O. S. P., de Zwaan, M., Dedoussis, G., Degortes, D., DeSocio, J. E., Dick, D. M., Dikeos, D., Dina, C., Dmitrzak-Weglarz, M., Docampo, E., Duncan, L. E., Egberts, K., Ehrlich, S., Escaramís, G., Esko, T., Estivill, X., Farmer, A., Favaro, A., Fernández-Aranda, F., Fichter, M. M., Fischer, K., Föcker, M., Foretova, L., Forstner, A. J., Forzan, M., Franklin, C. S., Gallinger, S., Giegling, I., Giuranna, J., Gonidakis, F., Gorwood, P., Mayora, M. G., Guillaume, S., Guo, Y., Hakonarson, H., Hatzikotoulas, K., Hauser, J., Hebebrand, J., Helder, S. G., Herms, S., Herpertz-Dahlmann, B., Herzog, W., Huckins, L. M.,

Hudson, J. I., Imgart, H., Inoko, H., Janout, V., Jiménez-Murcia, S., Julià, A., Kalsi, G., Kaminská, D., Kaprio, J., Karhunen, L., Karwautz, A., Kas, M. J. H., Kennedy, J. L., Keski-Rahkonen, A., Kiezebrink, K., Kim, Y.-R., Klareskog, L., Klump, K. L., Knudsen, G. P. S., La Via, M. C., Le Hellard, S., Levitan, R. D., Li, D., Lilenfeld, L., Lin, B. D., Lissowska, J., Luykx, J., Magistretti, P. J., Maj, M., Mannik, K., Marsal, S., Marshall, C. R., Mattingsdal, M., McDevitt, S., McGuffin, P., Metspalu, A., Meulenbelt, I., Micali, N., Mitchell, K., Monteleone, A. M., Monteleone, P., Munn-Chernoff, M. A., Nacmias, B., Navratilova, M., Ntalla, I., O'Toole, J. K., Ophoff, R. A., Padyukov, L., Palotie, A., Pantel, J., Papezova, H., Pinto, D., Rabionet, R., Raevuori, A., Ramoz, N., Reichborn-Kjennerud, T., Ricca, V., Ripatti, S., Ritschel, F., Roberts, M., Rotondo, A., Rujescu, D., Rybakowski, F., Santonastaso, P., Scherag, A., Scherer, S. W., Schmidt, U., Schork, N. J., Schosser, A., Seitz, J., Slachtova, L., Slagboom, P. E., Slof-Op 't Landt, M. C. T., Slopien, A., Sorbi, S., Świątkowska, B., Szatkiewicz, J. P., Tachmazidou, I., Tenconi, E., Tortorella, A., Tozzi, F., Treasure, J., Tsitsika, A., Tyszkiewicz-Nwafor, M., Tziouvas, K., van Elburg, A. A., van Furth, E. F., Wagner, G., Walton, E., Widen, E., Zeggini, E., Zerwas, S., Zipfel, S., Bergen, A. W., Boden, J. M., Brandt, H., Crawford, S., Halmi, K. A., Horwood, L. J., Johnson, C., Kaplan, A. S., Kaye, W. H., Mitchell, J. E., Olsen, C. M., Pearson, J. F., Pedersen, N. L., Strober, M., Werge, T., Whiteman, D. C., Woodside, D. B., Stuber, G. D., Gordon, S., Grove, J., Henders, A. K., Juréus, A., Kirk, K. M., Larsen, J. T., Parker, R., Petersen, L., Jordan, J., Kennedy, M., Montgomery, G. W., Wade, T. D., Birgegård, A., Lichtenstein, P., Norring, C., Landén, M., Martin, N. G., Mortensen, P. B., Sullivan, P. F., Breen, G. & Bulik, C. M. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).

50. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

51. Wuttke, M., Li, Y., Li, M., Sieber, K. B., Feitosa, M. F., Gorski, M., Tin, A., Wang, L., Chu, A. Y., Hoppmann, A., Kirsten, H., Giri, A., Chai, J.-F., Sveinbjornsson, G., Tayo, B. O., Nutile, T., Fuchsberger, C., Marten, J., Cocca, M., Ghasemi, S., Xu, Y., Horn, K., Noce, D., van der Most, P. J., Sedaghat, S., Yu, Z., Akiyama, M., Afaq, S., Ahluwalia, T. S., Almgren, P., Amin, N., Ärnlöv, J., Bakker, S. J. L., Bansal, N., Baptista, D., Bergmann, S., Biggs, M. L., Biino, G., Boehnke, M., Boerwinkle, E., Boissel, M., Bottinger, E. P., Boutin, T. S., Brenner, H., Brumat, M., Burkhardt, R., Butterworth, A. S., Campana, E., Campbell, A., Campbell, H., Canouil, M., Carroll, R. J., Catamo, E., Chambers, J. C., Chee, M.-L., Chee, M.-L., Chen, X., Cheng, C.-Y., Cheng, Y., Christensen, K., Cifkova, R., Ciullo, M., Concas, M. P., Cook, J. P., Coresh, J., Corre, T., Sala, C. F., Cusi, D., Danesh, J., Daw, E. W., de Borst, M. H., De Grandi, A., de Mutsert, R., de Vries, A. P. J., Degenhardt, F., Delgado, G., Demirkan, A., Di Angelantonio, E., Dittrich, K., Divers, J., Dorajoo, R., Eckardt, K.-U., Ehret, G., Elliott, P., Endlich, K., Evans, M. K., Felix, J. F., Foo, V. H. X., Franco, O. H., Franke, A., Freedman, B. I., Freitag-Wolf, S., Friedlander, Y., Froguel, P., Gansevoort, R. T., Gao, H., Gasparini, P., Gaziano, J. M., Giedraitis, V., Gieger, C., Girotto, G., Giulianini, F., Gögele, M., Gordon, S. D., Gudbjartsson, D. F., Gudnason, V., Haller, T., Hamet, P., Harris, T. B., Hartman, C. A., Hayward, C., Hellwege, J. N., Heng, C.-K., Hicks, A. A., Hofer, E., Huang, W., Hutri-Kähönen, N., Hwang, S.-J., Ikram, M. A., Indridason, O. S., Ingelsson, E., Ising, M., Jaddoe, V. W. V., Jakobsdottir, J., Jonas, J. B., Joshi, P. K., Josyula, N. S., Jung, B., Kähönen, M., Kamatani, Y., Kammerer, C. M., Kanai, M., Kastarinen, M., Kerr, S. M., Khor, C.-C., Kiess, W., Kleber, M. E., Koenig, W., Kooner, J. S., Körner, A., Kovacs, P., Kraja, A. T., Krajcoviechova, A., Kramer, H., Krämer, B. K., Kronenberg, F., Kubo, M., Kühnel, B., Kuokkanen, M., Kuusisto, J., La Bianca, M., Laakso, M., Lange, L. A., Langefeld, C. D., Lee, J. J.-M., Lehne, B., Lehtimäki, T., Lieb, W., Lifelines Cohort Study, Lim, S.-C., Lind, L., Lindgren, C. M., Liu, J., Liu, J., Loeffler, M., Loos, R. J. F., Lucae, S., Lukas, M. A., Lyytikäinen, L.-P., Mägi, R., Magnusson, P. K. E., Mahajan, A., Martin,

N. G., Martins, J., März, W., Mascalzoni, D., Matsuda, K., Meisinger, C., Meitinger, T., Melander, O., Metspalu, A., Mikaelsdottir, E. K., Milaneschi, Y., Miliku, K., Mishra, P. P., V. A. Million Veteran Program, Mohlke, K. L., Mononen, N., Montgomery, G. W., Mook-Kanamori, D. O., Mychaleckyj, J. C., Nadkarni, G. N., Nalls, M. A., Nauck, M., Nikus, K., Ning, B., Nolte, I. M., Noordam, R., O'Connell, J., O'Donoghue, M. L., Olafsson, I., Oldehinkel, A. J., Orho-Melander, M., Ouwehand, W. H., Padmanabhan, S., Palmer, N. D., Palsson, R., Penninx, B. W. J. H., Perls, T., Perola, M., Pirastu, M., Pirastu, N., Pistis, G., Podgornaia, A. I., Polasek, O., Ponte, B., Porteous, D. J., Poulain, T., Pramstaller, P. P., Preuss, M. H., Prins, B. P., Province, M. A., Rabelink, T. J., Raffield, L. M., Raitakari, O. T., Reilly, D. F., Rettig, R., Rheinberger, M., Rice, K. M., Ridker, P. M., Rivadeneira, F., Rizzi, F., Roberts, D. J., Robino, A., Rossing, P., Rudan, I., Rueedi, R., Ruggiero, D., Ryan, K. A., Saba, Y., Sabanayagam, C., Salomaa, V., Salvi, E., Saum, K.-U., Schmidt, H., Schmidt, R., Schöttker, B., Schulz, C.-A., Schupf, N., Shaffer, C. M., Shi, Y., Smith, A. V., Smith, B. H., Soranzo, N., Spracklen, C. N., Strauch, K., Stringham, H. M., Stumvoll, M., Svensson, P. O., Szymczak, S., Tai, E.-S., Tajuddin, S. M., Tan, N. Y. Q., Taylor, K. D., Teren, A., Tham, Y.-C., Thiery, J., Thio, C. H. L., Thomsen, H., Thorleifsson, G., Toniolo, D., Tönjes, A., Tremblay, J., Tzoulaki, I., Uitterlinden, A. G., Vaccargiu, S., van Dam, R. M., van der Harst, P., van Duijn, C. M., Velez Edward, D. R., Verweij, N., Vogelezang, S., Völker, U., Vollenweider, P., Waeber, G., Waldenberger, M., Wallentin, L., Wang, Y. X., Wang, C., Waterworth, D. M., Bin Wei, W., White, H., Whitfield, J. B., Wild, S. H., Wilson, J. F., Wojczynski, M. K., Wong, C., Wong, T.-Y., Xu, L., Yang, Q., Yasuda, M., Yerges-Armstrong, L. M., Zhang, W., Zonderman, A. B., Rotter, J. I., Bochud, M., Psaty, B. M., Vitart, V., Wilson, J. G., Dehghan, A., Parsa, A., Chasman, D. I., Ho, K., Morris, A. P., Devuyst, O., Akilesh, S., Pendergrass, S. A., Sim, X., Böger, C. A., Okada, Y., Edwards, T. L., Snieder, H., Stefansson, K., Hung, A. M., Heid, I. M., Scholz, M., Teumer, A., Köttgen, A. & Pattaro, C. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972 (2019).

52. Nielsen, J. B., Thorolfsdottir, R. B., Fritsche, L. G., Zhou, W., Skov, M. W., Graham, S. E., Herron, T. J., McCarthy, S., Schmidt, E. M., Sveinbjornsson, G., Surakka, I., Mathis, M. R., Yamazaki, M., Crawford, R. D., Gabrielsen, M. E., Skogholt, A. H., Holmen, O. L., Lin, M., Wolford, B. N., Dey, R., Dalen, H., Sulem, P., Chung, J. H., Backman, J. D., Arnar, D. O., Thorsteinsdottir, U., Baras, A., O'Dushlaine, C., Holst, A. G., Wen, X., Hornsby, W., Dewey, F. E., Boehnke, M., Kheterpal, S., Mukherjee, B., Lee, S., Kang, H. M., Holm, H., Kitzman, J., Shavit, J. A., Jalife, J., Brummett, C. M., Teslovich, T. M., Carey, D. J., Gudbjartsson, D. F., Stefansson, K., Abecasis, G. R., Hveem, K. & Willer, C. J. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).

53. Tachmazidou, I., Hatzikotoulas, K., Southam, L., Esparza-Gordillo, J., Haberland, V., Zheng, J., Johnson, T., Koprulu, M., Zengini, E., Steinberg, J., Wilkinson, J. M., Bhatnagar, S., Hoffman, J. D., Buchan, N., Süveges, D., arcOGEN Consortium, Yerges-Armstrong, L., Smith, G. D., Gaunt, T. R., Scott, R. A., McCarthy, L. C. & Zeggini, E. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat. Genet.* **51**, 230–236 (2019).

54. Wheeler, E., Leong, A., Liu, C.-T., Hivert, M.-F., Strawbridge, R. J., Podmore, C., Li, M., Yao, J., Sim, X., Hong, J., Chu, A. Y., Zhang, W., Wang, X., Chen, P., Maruthur, N. M., Porneala, B. C., Sharp, S. J., Jia, Y., Kabagambe, E. K., Chang, L.-C., Chen, W.-M., Elks, C. E., Evans, D. S., Fan, Q., Giulianini, F., Go, M. J., Hottenga, J.-J., Hu, Y., Jackson, A. U., Kanoni, S., Kim, Y. J., Kleber, M. E., Ladenvall, C., Lecoeur, C., Lim, S.-H., Lu, Y., Mahajan, A., Marzi, C., Nalls, M. A., Navarro, P., Nolte, I. M., Rose, L. M., Rybin, D. V., Sanna, S., Shi, Y., Stram, D. O., Takeuchi, F., Tan, S. P., van der Most, P. J., Van Vliet-Ostaptchouk, J. V.,

Wong, A., Yengo, L., Zhao, W., Goel, A., Martinez Larrad, M. T., Radke, D., Salo, P., Tanaka, T., van Iperen, E. P. A., Abecasis, G., Afaq, S., Alizadeh, B. Z., Bertoni, A. G., Bonnefond, A., Böttcher, Y., Bottinger, E. P., Campbell, H., Carlson, O. D., Chen, C.-H., Cho, Y. S., Garvey, W. T., Gieger, C., Goodarzi, M. O., Grallert, H., Hamsten, A., Hartman, C. A., Herder, C., Hsiung, C. A., Huang, J., Igase, M., Isono, M., Katsuya, T., Khor, C.-C., Kiess, W., Kohara, K., Kovacs, P., Lee, J., Lee, W.-J., Lehne, B., Li, H., Liu, J., Lobbens, S., Luan, J., Lyssenko, V., Meitinger, T., Miki, T., Miljkovic, I., Moon, S., Mulas, A., Müller, G., Müller-Nurasyid, M., Nagaraja, R., Nauck, M., Pankow, J. S., Polasek, O., Prokopenko, I., Ramos, P. S., Rasmussen-Torvik, L., Rathmann, W., Rich, S. S., Robertson, N. R., Roden, M., Roussel, R., Rudan, I., Scott, R. A., Scott, W. R., Sennblad, B., Siscovick, D. S., Strauch, K., Sun, L., Swertz, M., Tajuddin, S. M., Taylor, K. D., Teo, Y.-Y., Tham, Y. C., Tönjes, A., Wareham, N. J., Willemsen, G., Wilsgaard, T., Hingorani, A. D., EPIC-CVD Consortium, EPIC-InterAct Consortium, Lifelines Cohort Study, Egan, J., Ferrucci, L., Hovingh, G. K., Jula, A., Kivimaki, M., Kumari, M., Njølstad, I., Palmer, C. N. A., Serrano Ríos, M., Stumvoll, M., Watkins, H., Aung, T., Blüher, M., Boehnke, M., Boomsma, D. I., Bornstein, S. R., Chambers, J. C., Chasman, D. I., Chen, Y.-D. I., Chen, Y.-T., Cheng, C.-Y., Cucca, F., de Geus, E. J. C., Deloukas, P., Evans, M. K., Fornage, M., Friedlander, Y., Froguel, P., Groop, L., Gross, M. D., Harris, T. B., Hayward, C., Heng, C.-K., Ingelsson, E., Kato, N., Kim, B.-J., Koh, W.-P., Kooner, J. S., Körner, A., Kuh, D., Kuusisto, J., Laakso, M., Lin, X., Liu, Y., Loos, R. J. F., Magnusson, P. K. E., März, W., McCarthy, M. I., Oldehinkel, A. J., Ong, K. K., Pedersen, N. L., Pereira, M. A., Peters, A., Ridker, P. M., Sabanayagam, C., Sale, M., Saleheen, D., Saltevo, J., Schwarz, P. E., Sheu, W. H. H., Snieder, H., Spector, T. D., Tabara, Y., Tuomilehto, J., van Dam, R. M., Wilson, J. G., Wilson, J. F., Wolffenbuttel, B. H. R., Wong, T. Y., Wu, J.-Y., Yuan, J.-M., Zonderman, A. B., Soranzo, N., Guo, X., Roberts, D. J., Florez, J. C., Sladek, R., Dupuis, J., Morris, A. P., Tai, E.-S., Selvin, E., Rotter, J. I., Langenberg, C., Barroso, I. & Meigs, J. B. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med.* **14**, e1002383 (2017).

55. Horikoshi, M., Beaumont, R. N., Day, F. R., Warrington, N. M., Kooijman, M. N., Fernandez-Tajes, J., Feenstra, B., van Zuydam, N. R., Gaulton, K. J., Grarup, N., Bradfield, J. P., Strachan, D. P., Li-Gao, R., Ahluwalia, T. S., Kreiner, E., Rueedi, R., Lyytikäinen, L.-P., Cousminer, D. L., Wu, Y., Thiering, E., Wang, C. A., Have, C. T., Hottenga, J.-J., Vilor-Tejedor, N., Joshi, P. K., Boh, E. T. H., Ntalla, I., Pitkänen, N., Mahajan, A., van Leeuwen, E. M., Joro, R., Lagou, V., Nodzenski, M., Diver, L. A., Zondervan, K. T., Bustamante, M., Marques-Vidal, P., Mercader, J. M., Bennett, A. J., Rahmioglu, N., Nyholt, D. R., Ma, R. C. W., Tam, C. H. T., Tam, W. H., CHARGE Consortium Hematology Working Group, Ganesh, S. K., van Rooij, F. J. A., Jones, S. E., Loh, P.-R., Ruth, K. S., Tuke, M. A., Tyrrell, J., Wood, A. R., Yaghootkar, H., Scholtens, D. M., Paternoster, L., Prokopenko, I., Kovacs, P., Atalay, M., Willems, S. M., Panoutsopoulou, K., Wang, X., Carstensen, L., Geller, F., Schraut, K. E., Murcia, M., van Beijsterveldt, C. E. M., Willemsen, G., Appel, E. V. R., Fonvig, C. E., Trier, C., Tiesler, C. M. T., Standl, M., Kutalik, Z., Bonàs-Guarch, S., Hougaard, D. M., Sánchez, F., Torrents, D., Waage, J., Hollegaard, M. V., de Haan, H. G., Rosendaal, F. R., Medina-Gomez, C., Ring, S. M., Hemani, G., McMahon, G., Robertson, N. R., Groves, C. J., Langenberg, C., Luan, J., Scott, R. A., Zhao, J. H., Mentch, F. D., MacKenzie, S. M., Reynolds, R. M., Early Growth Genetics (EGG) Consortium, Lowe, W. L., Tönjes, A., Stumvoll, M., Lindi, V., Lakka, T. A., van Duijn, C. M., Kiess, W., Körner, A., Sørensen, T. I. A., Niinikoski, H., Pahkala, K., Raitakari, O. T., Zeggini, E., Dedoussis, G. V., Teo, Y.-Y., Saw, S.-M., Melbye, M., Campbell, H., Wilson, J. F., Vrijheid, M., de Geus, E. J. C. N., Boomsma, D. I., Kadarmideen, H. N., Holm, J.-C., Hansen, T., Sebert, S., Hattersley, A. T., Beilin, L. J., Newnham, J. P., Pennell, C. E., Heinrich, J., Adair, L. S., Borja, J. B., Mohlke, K. L., Eriksson, J. G., Widén, E., Kähönen, M., Viikari, J. S.,

Lehtimäki, T., Vollenweider, P., Bønnelykke, K., Bisgaard, H., Mook-Kanamori, D. O., Hofman, A., Rivadeneira, F., Uitterlinden, A. G., Pisinger, C., Pedersen, O., Power, C., Hyppönen, E., Wareham, N. J., Hakonarson, H., Davies, E., Walker, B. R., Jaddoe, V. W. V., Järvelin, M.-R., Grant, S. F. A., Vaag, A. A., Lawlor, D. A., Frayling, T. M., Smith, G. D., Morris, A. P., Ong, K. K., Felix, J. F., Timpson, N. J., Perry, J. R. B., Evans, D. M., McCarthy, M. I. & Freathy, R. M. Genome-wide associations for birth weight and correlations with adult disease. *Nature* **538**, 248–252 (2016).

56. Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., Visscher, P. M., & GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).

57. Jiang, X., O'Reilly, P. F., Aschard, H., Hsu, Y.-H., Richards, J. B., Dupuis, J., Ingelsson, E., Karasik, D., Pilz, S., Berry, D., Kestenbaum, B., Zheng, J., Luan, J., Sofianopoulou, E., Streeten, E. A., Albanes, D., Lutsey, P. L., Yao, L., Tang, W., Econs, M. J., Wallaschofski, H., Völzke, H., Zhou, A., Power, C., McCarthy, M. I., Michos, E. D., Boerwinkle, E., Weinstein, S. J., Freedman, N. D., Huang, W.-Y., Van Schoor, N. M., van der Velde, N., Groot, L. C. P. G. M. de, Enneman, A., Cupples, L. A., Booth, S. L., Vasan, R. S., Liu, C.-T., Zhou, Y., Ripatti, S., Ohlsson, C., Vandenput, L., Lorentzon, M., Eriksson, J. G., Shea, M. K., Houston, D. K., Kritchevsky, S. B., Liu, Y., Lohman, K. K., Ferrucci, L., Peacock, M., Gieger, C., Beekman, M., Slagboom, E., Deelen, J., Heemst, D. van, Kleber, M. E., März, W., de Boer, I. H., Wood, A. C., Rotter, J. I., Rich, S. S., Robinson-Cohen, C., den Heijer, M., Jarvelin, M.-R., Cavadino, A., Joshi, P. K., Wilson, J. F., Hayward, C., Lind, L., Michaëlsson, K., Trompet, S., Zillikens, M. C., Uitterlinden, A. G., Rivadeneira, F., Broer, L., Zgaga, L., Campbell, H., Theodoratou, E., Farrington, S. M., Timofeeva, M., Dunlop, M. G., Valdes, A. M., Tikkanen, E., Lehtimäki, T., Lyytikäinen, L.-P., Kähönen, M., Raitakari, O. T., Mikkilä, V., Ikram, M. A., Sattar, N., Jukema, J. W., Wareham, N. J., Langenberg, C., Forouhi, N. G., Gundersen, T. E., Khaw, K.-T., Butterworth, A. S., Danesh, J., Spector, T., Wang, T. J., Hyppönen, E., Kraft, P. & Kiel, D. P. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nat. Commun.* **9**, 260 (2018).

58. Manning, A. K., Hivert, M.-F., Scott, R. A., Grimsby, J. L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.-T., Bielak, L. F., Prokopenko, I., Amin, N., Barnes, D., Cadby, G., Hottenga, J.-J., Ingelsson, E., Jackson, A. U., Johnson, T., Kanoni, S., Ladenvall, C., Lagou, V., Lahti, J., Lecoeur, C., Liu, Y., Martinez-Larrad, M. T., Montasser, M. E., Navarro, P., Perry, J. R. B., Rasmussen-Torvik, L. J., Salo, P., Sattar, N., Shungin, D., Strawbridge, R. J., Tanaka, T., van Duijn, C. M., An, P., de Andrade, M., Andrews, J. S., Aspelund, T., Atalay, M., Aulchenko, Y., Balkau, B., Bandinelli, S., Beckmann, J. S., Beilby, J. P., Bellis, C., Bergman, R. N., Blangero, J., Boban, M., Boehnke, M., Boerwinkle, E., Bonnycastle, L. L., Boomsma, D. I., Borecki, I. B., Böttcher, Y., Bouchard, C., Brunner, E., Budimir, D., Campbell, H., Carlson, O., Chines, P. S., Clarke, R., Collins, F. S., Corbatón-Anchuelo, A., Couper, D., de Faire, U., Dedoussis, G. V., Deloukas, P., Dimitriou, M., Egan, J. M., Eiriksdottir, G., Erdos, M. R., Eriksson, J. G., Eury, E., Ferrucci, L., Ford, I., Forouhi, N. G., Fox, C. S., Franzosi, M. G., Franks, P. W., Frayling, T. M., Froguel, P., Galan, P., de Geus, E., Gigante, B., Glazer, N. L., Goel, A., Groop, L., Gudnason, V., Hallmans, G., Hamsten, A., Hansson, O., Harris, T. B., Hayward, C., Heath, S., Hercberg, S., Hicks, A. A., Hingorani, A., Hofman, A., Hui, J., Hung, J., Jarvelin, M.-R., Jhun, M. A., Johnson, P. C. D., Jukema, J. W., Jula, A., Kao, W. H., Kaprio, J., Kardia, S. L. R., Keinanen-Kiukaanniemi, S., Kivimaki, M., Kolcic, I., Kovacs, P., Kumari, M., Kuusisto, J., Kyvik, K. O., Laakso, M., Lakka, T., Lannfelt, L., Lathrop, G. M., Launer, L. J., Leander, K., Li, G., Lind, L., Lindstrom, J., Lobbens, S., Loos, R. J. F., Luan, J., Lyssenko, V., Mägi, R., Magnusson, P. K.

E., Marmot, M., Meneton, P., Mohlke, K. L., Mooser, V., Morken, M. A., Miljkovic, I., Narisu, N., O'Connell, J., Ong, K. K., Oostra, B. A., Palmer, L. J., Palotie, A., Pankow, J. S., Peden, J. F., Pedersen, N. L., Pehlic, M., Peltonen, L., Penninx, B., Pericic, M., Perola, M., Perusse, L., Peyser, P. A., Polasek, O., Pramstaller, P. P., Province, M. A., Räikkönen, K., Rauramaa, R., Rehnberg, E., Rice, K., Rotter, J. I., Rudan, I., Ruokonen, A., Saaristo, T., Sabater-Lleal, M., Salomaa, V., Savage, D. B., Saxena, R., Schwarz, P., Seedorf, U., Sennblad, B., Serrano-Rios, M., Shuldiner, A. R., Sijbrands, E. J. G., Siscovick, D. S., Smit, J. H., Small, K. S., Smith, N. L., Smith, A. V., Stančáková, A., Stirrups, K., Stumvoll, M., Sun, Y. V., Swift, A. J., Tönjes, A., Tuomilehto, J., Trompet, S., Uitterlinden, A. G., Uusitupa, M., Vikström, M., Vitart, V., Vohl, M.-C., Voight, B. F., Vollenweider, P., Waeber, G., Waterworth, D. M., Watkins, H., Wheeler, E., Widen, E., Wild, S. H., Willems, S. M., Willemsen, G., Wilson, J. F., Witteman, J. C. M., Wright, A. F., Yaghootkar, H., Zelenika, D., Zemunik, T., Zgaga, L., Consortium, Dia. G. R. A. M. (DIAGRAM), Consortium, T. M. T. H. E. R. (MUTHER), Wareham, N. J., McCarthy, M. I., Barroso, I., Watanabe, R. M., Florez, J. C., Dupuis, J., Meigs, J. B. & Langenberg, C. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).

59. Day, F. R., Ruth, K. S., Thompson, D. J., Lunetta, K. L., Pervjakova, N., Chasman, D. I., Stolk, L., Finucane, H. K., Sulem, P., Bulik-Sullivan, B., Esko, T., Johnson, A. D., Elks, C. E., Franceschini, N., He, C., Altmaier, E., Brody, J. A., Franke, L. L., Huffman, J. E., Keller, M. F., McArdle, P. F., Nutile, T., Porcu, E., Robino, A., Rose, L. M., Schick, U. M., Smith, J. A., Teumer, A., Traglia, M., Vuckovic, D., Yao, J., Zhao, W., Albrecht, E., Amin, N., Corre, T., Hottenga, J.-J., Mangino, M., Smith, A. V., Tanaka, T., Abecasis, G. R., Andrulis, I. L., Anton-Culver, H., Antoniou, A. C., Arndt, V., Arnold, A. M., Barbieri, C., Beckmann, M. W., Beeghly-Fadiel, A., Benitez, J., Bernstein, L., Bielinski, S. J., Blomqvist, C., Boerwinkle, E., Bogdanova, N. V., Bojesen, S. E., Bolla, M. K., Borresen-Dale, A.-L., Boutin, T. S., Brauch, H., Brenner, H., Brüning, T., Burwinkel, B., Campbell, A., Campbell, H., Chanock, S. J., Chapman, J. R., Chen, Y.-D. I., Chenevix-Trench, G., Couch, F. J., Coviello, A. D., Cox, A., Czene, K., Darabi, H., De Vivo, I., Demerath, E. W., Dennis, J., Devilee, P., Dörk, T., dos-Santos-Silva, I., Dunning, A. M., Eicher, J. D., Fasching, P. A., Faul, J. D., Figueroa, J., Flesch-Janys, D., Gandin, I., Garcia, M. E., García-Closas, M., Giles, G. G., Girotto, G. G., Goldberg, M. S., González-Neira, A., Goodarzi, M. O., Grove, M. L., Gudbjartsson, D. F., Guénel, P., Guo, X., Haiman, C. A., Hall, P., Hamann, U., Henderson, B. E., Hocking, L. J., Hofman, A., Homuth, G., Hooning, M. J., Hopper, J. L., Hu, F. B., Huang, J., Humphreys, K., Hunter, D. J., Jakubowska, A., Jones, S. E., Kabisch, M., Karasik, D., Knight, J. A., Kolcic, I., Kooperberg, C., Kosma, V.-M., Kriebel, J., Kristensen, V., Lambrechts, D., Langenberg, C., Li, J., Li, X., Lindström, S., Liu, Y., Luan, J., Lubinski, J., Mägi, R., Mannermaa, A., Manz, J., Margolin, S., Marten, J., Martin, N. G., Masciullo, C., Meindl, A., Michailidou, K., Mihailov, E., Milani, L., Milne, R. L., Müller-Nurasyid, M., Nalls, M., Neale, B. M., Nevanlinna, H., Neven, P., Newman, A. B., Nordestgaard, B. G., Olson, J. E., Padmanabhan, S., Peterlongo, P., Peters, U., Petersmann, A., Peto, J., Pharoah, P. D. P., Pirastu, N. N., Pirie, A., Pistis, G., Polasek, O., Porteous, D., Psaty, B. M., Pylkäs, K., Radice, P., Raffel, L. J., Rivadeneira, F., Rudan, I., Rudolph, A., Ruggiero, D., Sala, C. F., Sanna, S., Sawyer, E. J., Schlessinger, D., Schmidt, M. K., Schmidt, F., Schmutzler, R. K., Schoemaker, M. J., Scott, R. A., Seynaeve, C. M., Simard, J., Sorice, R., Southey, M. C., Stöckl, D., Strauch, K., Swerdlow, A., Taylor, K. D., Thorsteinsdottir, U., Toland, A. E., Tomlinson, I., Truong, T., Tryggvadottir, L., Turner, S. T., Vozzi, D., Wang, Q., Wellons, M., Willemsen, G., Wilson, J. F., Winqvist, R., Wolffenbuttel, B. B. H. R., Wright, A. F., Yannoukakos, D., Zemunik, T., Zheng, W., Zygmunt, M., Bergmann, S., Boomsma, D. I., Buring, J. E., Ferrucci, L., Montgomery, G. W., Gudnason, V., Spector, T. D., van Duijn, C. M., Alizadeh, B. Z., Ciullo, M., Crisponi, L., Easton, D. F., Gasparini, P. P., Gieger, C., Harris, T. B., Hayward, C., Kardia, S. L. R., Kraft, P., McKnight, B., Metspalu, A., Morrison, A. C., Reiner,

A. P., Ridker, P. M., Rotter, J. I., Toniolo, D., Uitterlinden, A. G., Ulivi, S., Völzke, H., Wareham, N. J., Weir, D. R., Yerges-Armstrong, L. M., The PRACTICAL Consortium, kConFab Investigators, Aocs Investigators, Generation Scotland, EPIC-InterAct Consortium, LifeLines Cohort Study, Price, A. L., Stefansson, K., Visser, J. A., Ong, K. K., Chang-Claude, J., Murabito, J. M., Perry, J. R. B. & Murray, A. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* **47**, 1294–1303 (2015).

60. Day, F. R., Thompson, D. J., Helgason, H., Chasman, D. I., Finucane, H., Sulem, P., Ruth, K. S., Whalen, S., Sarkar, A. K., Albrecht, E., Altmaier, E., Amini, M., Barbieri, C. M., Boutin, T., Campbell, A., Demerath, E., Giri, A., He, C., Hottenga, J. J., Karlsson, R., Kolcic, I., Loh, P.-R., Lunetta, K. L., Mangino, M., Marco, B., McMahon, G., Medland, S. E., Nolte, I. M., Noordam, R., Nutile, T., Paternoster, L., Perjakova, N., Porcu, E., Rose, L. M., Schraut, K. E., Segrè, A. V., Smith, A. V., Stolk, L., Teumer, A., Andrulis, I. L., Bandinelli, S., Beckmann, M. W., Benitez, J., Bergmann, S., Bochud, M., Boerwinkle, E., Bojesen, S. E., Bolla, M. K., Brand, J. S., Brauch, H., Brenner, H., Broer, L., Brüning, T., Buring, J. E., Campbell, H., Catamo, E., Chanock, S., Chenevix-Trench, G., Corre, T., Couch, F. J., Cousminer, D. L., Cox, A., Crisponi, L., Czene, K., Davey Smith, G., de Geus, E. J. C. N., de Mutsert, R., De Vivo, I., Dennis, J., Devilee, P., dos-Santos-Silva, I., Dunning, A. M., Eriksson, J. G., Fasching, P. A., Fernández-Rhodes, L., Ferrucci, L., Flesch-Janys, D., Franke, L., Gabrielson, M., Gandin, I., Giles, G. G., Grallert, H., Gudbjartsson, D. F., Guénel, P., Hall, P., Hallberg, E., Hamann, U., Harris, T. B., Hartman, C. A., Heiss, G., Hooning, M. J., Hopper, J. L., Hu, F., Hunter, D. J., Ikram, M. A., Im, H. K., Järvelin, M.-R., Joshi, P. K., Karasik, D., Kellis, M., Kutalik, Z., LaChance, G., Lambrechts, D., Langenberg, C., Launer, L. J., Laven, J. S. E., Lenarduzzi, S., Li, J., Lind, P. A., Lindstrom, S., Liu, Y., Luan, J., Mägi, R., Mannermaa, A., Mbarek, H., McCarthy, M. I., Meisinger, C., Meitinger, T., Menni, C., Metspalu, A., Michailidou, K., Milani, L., Milne, R. L., Montgomery, G. W., Mulligan, A. M., Nalls, M. A., Navarro, P., Nevanlinna, H., Nyholt, D. R., Oldehinkel, A. J., O'Mara, T. A., Padmanabhan, S., Palotie, A., Pedersen, N., Peters, A., Peto, J., Pharoah, P. D. P., Pouta, A., Radice, P., Rahman, I., Ring, S. M., Robino, A., Rosendaal, F. R., Rudan, I., Rueedi, R., Ruggiero, D., Sala, C. F., Schmidt, M. K., Scott, R. A., Shah, M., Sorice, R., Southey, M. C., Sovio, U., Stampfer, M., Steri, M., Strauch, K., Tanaka, T., Tikkanen, E., Timpson, N. J., Traglia, M., Truong, T., Tyrer, J. P., Uitterlinden, A. G., Edwards, D. R. V., Vitart, V., Völker, U., Vollenweider, P., Wang, Q., Widen, E., van Dijk, K. W., Willemsen, G., Winqvist, R., Wolffenbuttel, B. H. R., Zhao, J. H., Zoledziewska, M., Zygmunt, M., Alizadeh, B. Z., Boomsma, D. I., Ciullo, M., Cucca, F., Esko, T., Franceschini, N., Gieger, C., Gudnason, V., Hayward, C., Kraft, P., Lawlor, D. A., Magnusson, P. K. E., Martin, N. G., Mook-Kanamori, D. O., Nohr, E. A., Polasek, O., Porteous, D., Price, A. L., Ridker, P. M., Snieder, H., Spector, T. D., Stöckl, D., Toniolo, D., Ulivi, S., Visser, J. A., Völzke, H., Wareham, N. J., Wilson, J. F., The LifeLines Cohort Study, The InterAct Consortium, kConFab/AOCS Investigators, Endometrial Cancer Association Consortium, Ovarian Cancer Association Consortium, Practical Consortium, Spurdle, A. B., Thorsteindottir, U., Pollard, K. S., Easton, D. F., Tung, J. Y., Chang-Claude, J., Hinds, D., Murray, A., Murabito, J. M., Stefansson, K., Ong, K. K. & Perry, J. R. B. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **49**, 834–841 (2017).

61. Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., Young, H., Zabaneh, D., Hägg, S., Hansell, N. K., Karlsson, I. K., Linnarsson, S., Montgomery, G. W., Muñoz-Manchado, A. B., Quinlan, E. B., Schumann, G., Skene, N. G., Webb, B. T., White, T., Arking, D. E., Avramopoulos, D., Bilder, R. M., Bitsios, P., Burdick, K.

E., Cannon, T. D., Chiba-Falek, O., Christoforou, A., Cirulli, E. T., Congdon, E., Corvin, A., Davies, G., Deary, I. J., DeRosse, P., Dickinson, D., Djurovic, S., Donohoe, G., Conley, E. D., Eriksson, J. G., Espeseth, T., Freimer, N. A., Giakoumaki, S., Giegling, I., Gill, M., Glahn, D. C., Hariri, A. R., Hatzimanolis, A., Keller, M. C., Knowles, E., Koltai, D., Konte, B., Lahti, J., Le Hellard, S., Lencz, T., Liewald, D. C., London, E., Lundervold, A. J., Malhotra, A. K., Melle, I., Morris, D., Need, A. C., Ollier, W., Palotie, A., Payton, A., Pendleton, N., Poldrack, R. A., Räikkönen, K., Reinvang, I., Roussos, P., Rujescu, D., Sabb, F. W., Scult, M. A., Smeland, O. B., Smyrnis, N., Starr, J. M., Steen, V. M., Stefanis, N. C., Straub, R. E., Sundet, K., Tiemeier, H., Voineskos, A. N., Weinberger, D. R., Widen, E., Yu, J., Abecasis, G., Andreassen, O. A., Breen, G., Christiansen, L., Debrabant, B., Dick, D. M., Heinz, A., Hjerling-Leffler, J., Ikram, M. A., Kendler, K. S., Martin, N. G., Medland, S. E., Pedersen, N. L., Plomin, R., Polderman, T. J. C., Ripke, S., van der Sluis, S., Sullivan, P. F., Vrieze, S. I., Wright, M. J. & Posthuma, D. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).

62. Strawbridge, R. J., Dupuis, J., Prokopenko, I., Barker, A., Ahlqvist, E., Rybin, D., Petrie, J. R., Travers, M. E., Bouatia-Naji, N., Dimas, A. S., Nica, A., Wheeler, E., Chen, H., Voight, B. F., Taneera, J., Kanoni, S., Peden, J. F., Turrini, F., Gustafsson, S., Zabena, C., Almgren, P., Barker, D. J. P., Barnes, D., Dennison, E. M., Eriksson, J. G., Eriksson, P., Eury, E., Folkersen, L., Fox, C. S., Frayling, T. M., Goel, A., Gu, H. F., Horikoshi, M., Isomaa, B., Jackson, A. U., Jameson, K. A., Kajantie, E., Kerr-Conte, J., Kuulasmaa, T., Kuusisto, J., Loos, R. J. F., Luan, J., Makrilakis, K., Manning, A. K., Martínez-Larrad, M. T., Narisu, N., Nastase Mannila, M., Ohrvik, J., Osmond, C., Pascoe, L., Payne, F., Sayer, A. A., Sennblad, B., Silveira, A., Stancáková, A., Stirrups, K., Swift, A. J., Syvänen, A.-C., Tuomi, T., van 't Hooft, F. M., Walker, M., Weedon, M. N., Xie, W., Zethelius, B., DIAGRAM Consortium, GIANT Consortium, MuTHER Consortium, CARDIoGRAM Consortium, C4D Consortium, Ongen, H., Mälarstig, A., Hopewell, J. C., Saleheen, D., Chambers, J., Parish, S., Danesh, J., Kooner, J., Ostenson, C.-G., Lind, L., Cooper, C. C., Serrano-Ríos, M., Ferrannini, E., Forsen, T. J., Clarke, R., Franzosi, M. G., Seedorf, U., Watkins, H., Froguel, P., Johnson, P., Deloukas, P., Collins, F. S., Laakso, M., Dermitzakis, E. T., Boehnke, M., McCarthy, M. I., Wareham, N. J., Groop, L., Pattou, F., Gloyn, A. L., Dedoussis, G. V., Lyssenko, V., Meigs, J. B., Barroso, I., Watanabe, R. M., Ingelsson, E., Langenberg, C., Hamsten, A. & Florez, J. C. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624–2634 (2011).

63. Saxena, R., Hivert, M.-F., Langenberg, C., Tanaka, T., Pankow, J. S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A. U., Kao, W. H. L., Li, M., Glazer, N. L., Manning, A. K., Luan, J., Stringham, H. M., Prokopenko, I., Johnson, T., Grarup, N., Boesgaard, T. W., Lecoeur, C., Shrader, P., O'Connell, J., Ingelsson, E., Couper, D. J., Rice, K., Song, K., Andreasen, C. H., Dina, C., Köttgen, A., Le Bacquer, O., Pattou, F., Taneera, J., Steinthorsdottir, V., Rybin, D., Ardlie, K., Sampson, M., Qi, L., van Hoek, M., Weedon, M. N., Aulchenko, Y. S., Voight, B. F., Grallert, H., Balkau, B., Bergman, R. N., Bielinski, S. J., Bonnefond, A., Bonnycastle, L. L., Borch-Johnsen, K., Böttcher, Y., Brunner, E., Buchanan, T. A., Bumpstead, S. J., Cavalcanti-Proença, C., Charpentier, G., Chen, Y.-D. I., Chines, P. S., Collins, F. S., Cornelis, M., Crawford, G. J., Delplanque, J., Doney, A., Egan, J. M., Erdos, M. R., Firmann, M., Forouhi, N. G., Fox, C. S., Goodarzi, M. O., Graessler, J., Hingorani, A., Isomaa, B., Jørgensen, T., Kivimaki, M., Kovacs, P., Krohn, K., Kumari, M., Lauritzen, T., Lévy-Marchal, C., Mayor, V., McAteer, J. B., Meyre, D., Mitchell, B. D., Mohlke, K. L., Morken, M. A., Narisu, N., Palmer, C. N. A., Pakyz, R., Pascoe, L., Payne, F., Pearson, D., Rathmann, W., Sandbaek, A., Sayer, A. A., Scott, L. J., Sharp, S. J., Sijbrands, E., Singleton, A., Siscovick, D. S., Smith, N. L., Sparsø, T., Swift, A. J., Syddall, H., Thorleifsson, G., Tönjes, A., Tuomi, T.,

Tuomilehto, J., Valle, T. T., Waeber, G., Walley, A., Waterworth, D. M., Zeggini, E., Zhao, J. H., Consortium, G. & Investigators, the M. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).

64. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).

65. Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., Strawbridge, R. J., Pers, T. H., Fischer, K., Justice, A. E., Workalemahu, T., Wu, J. M. W., Buchkovich, M. L., Heard-Costa, N. L., Roman, T. S., Drong, A. W., Song, C., Gustafsson, S., Day, F. R., Esko, T., Fall, T., Kutalik, Z., Luan, J., Randall, J. C., Scherag, A., Vedantam, S., Wood, A. R., Chen, J., Fehrmann, R., Karjalainen, J., Kahali, B., Liu, C.-T., Schmidt, E. M., Absher, D., Amin, N., Anderson, D., Beekman, M., Bragg-Gresham, J. L., Buyske, S., Demirkan, A., Ehret, G. B., Feitosa, M. F., Goel, A., Jackson, A. U., Johnson, T., Kleber, M. E., Kristiansson, K., Mangino, M., Leach, I. M., Medina-Gomez, C., Palmer, C. D., Pasko, D., Pechlivanis, S., Peters, M. J., Prokopenko, I., Stančáková, A., Sung, Y. J., Tanaka, T., Teumer, A., Van Vliet-Ostaptchouk, J. V., Yengo, L., Zhang, W., Albrecht, E., Ärnlöv, J., Arscott, G. M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A. J., Berne, C., Blüher, M., Böhringer, S., Bonnet, F., Böttcher, Y., Bruinenberg, M., Carba, D. B., Caspersen, I. H., Clarke, R., Daw, E. W., Deelen, J., Deelman, E., Delgado, G., Doney, A. S., Eklund, N., Erdos, M. R., Estrada, K., Eury, E., Friedrich, N., Garcia, M. E., Giedraitis, V., Gigante, B., Go, A. S., Golay, A., Grallert, H., Grammer, T. B., Gräßler, J., Grewal, J., Groves, C. J., Haller, T., Hallmans, G., Hartman, C. A., Hassinen, M., Hayward, C., Heikkilä, K., Herzig, K.-H., Helmer, Q., Hillege, H. L., Holmen, O., Hunt, S. C., Isaacs, A., Ittermann, T., James, A. L., Johansson, I., Juliusdottir, T., Kalafati, I.-P., Kinnunen, L., Koenig, W., Kooner, I. K., Kratzer, W., Lamina, C., Leander, K., Lee, N. R., Lichtner, P., Lind, L., Lindström, J., Lobbens, S., Lorentzon, M., Mach, F., Magnusson, P. K., Mahajan, A., McArdle, W. L., Menni, C., Merger, S., Mihailov, E., Milani, L., Mills, R., Moayyeri, A., Monda, K. L., Mooijaart, S. P., Mühleisen, T. W., Mulas, A., Müller, G., Müller-Nurasyid, M., Nagaraja, R., Nalls, M. A., Narisu, N., Glorioso, N., Nolte, I. M., Olden, M., Rayner, N. W., Renstrom, F., Ried, J. S., Robertson, N. R., Rose, L. M., Sanna, S., Scharnagl, H., Scholtens, S., Sennblad, B., Seufferlein, T., Sitlani, C. M., Smith, A. V., Stirrups, K., Stringham, H. M., Sundström, J., Swertz, M. A., Swift, A. J., Syvänen, A.-C., Tayo, B. O., Thorand, B., Thorleifsson, G., Tomaschitz, A., Troffa, C., van Oort, F. V., Verweij, N., Vonk, J. M., Waite, L. L., Wennauer, R., Wilsgaard, T., Wojczynski, M. K., Wong, A., Zhang, Q., Zhao, J. H., Brennan, E. P., Choi, M., Eriksson, P., Folkersen, L., Franco-Cereceda, A., Gharavi, A. G., Hedman, Å. K., Hivert, M.-F., Huang, J., Kanoni, S., Karpe, F., Keildson, S., Kiryluk, K., Liang, L., Lifton, R. P., Ma, B., McKnight, A. J., McPherson, R., Metspalu, A., Min, J. L., Moffatt, M. F., Montgomery, G. W., Murabito, J. M., Nicholson, G., Nyholt, D. R., Olsson, C., Perry, J. R., Reinmaa, E., Salem, R. M., Sandholm, N., Schadt, E. E., Scott, R. A., Stolk, L., Vallejo, E. E., Westra, H.-J., Zondervan, K. T., Amouyel, P., Arveiler, D., Bakker, S. J., Beilby, J., Bergman, R. N., Blangero, J., Brown, M. J., Burnier, M., Campbell, H., Chakravarti, A., Chines, P. S., Claudi-Boehm, S., Collins, F. S., Crawford, D. C., Danesh, J., de Faire, U., de Geus, E. J., Dörr, M., Erbel, R., Eriksson, J. G., Farrall, M., Ferrannini, E., Ferrières, J., Forouhi, N. G., Forrester, T., Franco, O. H., Gansevoort, R. T., Gieger, C., Gudnason, V., Haiman, C. A., Harris, T. B., Hattersley, A. T., Heliövaara, M., Hicks, A. A., Hingorani, A. D., Hoffmann, W., Hofman, A., Homuth, G., Humphries, S. E., Hyppönen, E., Illig, T., Jarvelin, M.-R., Johansen, B., Jousilahti, P., Jula, A. M., Kaprio, J., Kee, F., Keinanen-Kiukaanniemi, S. M., Kooner, J. S., Kooperberg, C., Kovacs, P., Kraja, A. T., Kumari, M., Kuulasmaa, K., Kuusisto, J., Lakka, T. A., Langenberg, C., Le Marchand, L., Lehtimäki, T., Lyssenko, V., Männistö, S., Marette, A., Matise, T. C., McKenzie, C. A., McKnight, B., Musk, A. W., Möhlenkamp, S., Morris, A. D., Nelis, M., Ohlsson, C., Oldehinkel, A. J., Ong, K. K., Palmer, L. J., Penninx, B. W., Peters, A.,

Pramstaller, P. P., Raitakari, O. T., Rankinen, T., Rao, D., Rice, T. K., Ridker, P. M., Ritchie, M. D., Rudan, I., Salomaa, V., Samani, N. J., Saramies, J., Sarzynski, M. A., Schwarz, P. E., Shuldiner, A. R., Staessen, J. A., Steinthorsdottir, V., Stolk, R. P., Strauch, K., Tönjes, A., Tremblay, A., Tremoli, E., Vohl, M.-C., Völker, U., Vollenweider, P., Wilson, J. F., Witteman, J. C., Adair, L. S., Bochud, M., Boehm, B. O., Bornstein, S. R., Bouchard, C., Cauchi, S., Caulfield, M. J., Chambers, J. C., Chasman, D. I., Cooper, R. S., Dedoussis, G., Ferrucci, L., Froguel, P., Grabe, H.-J., Hamsten, A., Hui, J., Hveem, K., Jöckel, K.-H., Kivimaki, M., Kuh, D., Laakso, M., Liu, Y., März, W., Munroe, P. B., Njølstad, I., Oostra, B. A., Palmer, C. N., Pedersen, N. L., Perola, M., Pérusse, L., Peters, U., Power, C., Quertermous, T., Rauramaa, R., Rivadeneira, F., Saaristo, T. E., Saleheen, D., Sinisalo, J., Slagboom, P. E., Snieder, H., Spector, T. D., Stefansson, K., Stumvoll, M., Tuomilehto, J., Uitterlinden, A. G., Uusitupa, M., van der Harst, P., Veronesi, G., Walker, M., Wareham, N. J., Watkins, H., Wichmann, H.-E., Abecasis, G. R., Assimes, T. L., Berndt, S. I., Boehnke, M., Borecki, I. B., Deloukas, P., Franke, L., Frayling, T. M., Groop, L. C., Hunter, D. J., Kaplan, R. C., O'Connell, J. R., Qi, L., Schlessinger, D., Strachan, D. P., Thorsteinsdottir, U., van Duijn, C. M., Willer, C. J., Visscher, P. M., Yang, J., Hirschhorn, J. N., Zillikens, M. C., McCarthy, M. I., Speliotes, E. K., North, K. E., Fox, C. S., Barroso, I., Franks, P. W., Ingelsson, E., Heid, I. M., Loos, R. J., Cupples, L. A., Morris, A. P., Lindgren, C. M. & Mohlke, K. L. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).

66. Cousminer, D. L., Berry, D. J., Timpson, N. J., Ang, W., Thiering, E., Byrne, E. M., Taal, H. R., Huikari, V., Bradfield, J. P., Kerkhof, M., Groen-Blokhuis, M. M., Kreiner-Møller, E., Marinelli, M., Holst, C., Leinonen, J. T., Perry, J. R. B., Surakka, I., Pietiläinen, O., Kettunen, J., Anttila, V., Kaakinen, M., Sovio, U., Pouta, A., Das, S., Lagou, V., Power, C., Prokopenko, I., Evans, D. M., Kemp, J. P., St Pourcain, B., Ring, S., Palotie, A., Kajantie, E., Osmond, C., Lehtimäki, T., Viikari, J. S., Kähönen, M., Warrington, N. M., Lye, S. J., Palmer, L. J., Tiesler, C. M. T., Flexeder, C., Montgomery, G. W., Medland, S. E., Hofman, A., Hakonarson, H., Guxens, M., Bartels, M., Salomaa, V., ReproGen Consortium, Murabito, J. M., Kaprio, J., Sørensen, T. I. A., Ballester, F., Bisgaard, H., Boomsma, D. I., Koppelman, G. H., Grant, S. F. A., Jaddoe, V. W. V., Martin, N. G., Heinrich, J., Pennell, C. E., Raitakari, O. T., Eriksson, J. G., Smith, G. D., Hyppönen, E., Järvelin, M.-R., McCarthy, M. I., Ripatti, S., Widén, E., & Early Growth Genetics (EGG) Consortium. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Hum. Mol. Genet.* **22**, 2735–2747 (2013).

67. Taal, H. R., Pourcain, B. S., Thiering, E., Das, S., Mook-Kanamori, D. O., Warrington, N. M., Kaakinen, M., Kreiner-Møller, E., Bradfield, J. P., Freathy, R. M., Geller, F., Guxens, M., Cousminer, D. L., Kerkhof, M., Timpson, N. J., Ikram, M. A., Beilin, L. J., Bønnelykke, K., Buxton, J. L., Charoen, P., Chawes, B. L. K., Eriksson, J., Evans, D. M., Hofman, A., Kemp, J. P., Kim, C. E., Klopp, N., Lahti, J., Lye, S. J., McMahon, G., Mentch, F. D., Müller, M., O'Reilly, P. F., Prokopenko, I., Rivadeneira, F., Steegers, E. A. P., Sunyer, J., Tiesler, C., Yaghootkar, H., Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium, Breteler, M. M. B., Debette, S., Fornage, M., Gudnason, V., Launer, L. J., van der Lugt, A., Mosley, T. H., Seshadri, S., Smith, A. V., Vernooij, M. W., Early Genetics & Lifecourse Epidemiology (EAGLE) consortium, Blakemore, A. I., Chiavacci, R. M., Feenstra, B., Fernandez-Benet, J., Grant, S. F. A., Hartikainen, A.-L., van der Heijden, A. J., Iñiguez, C., Lathrop, M., McArdle, W. L., Mølgaard, A., Newnham, J. P., Palmer, L. J., Palotie, A., Pouta, A., Ring, S. M., Sovio, U., Standl, M., Uitterlinden, A. G., Wichmann, H.-E., Vissing, N. H., DeCarli, C., van Duijn, C. M., McCarthy, M. I., Koppelman, G. H., Estivill, X., Hattersley, A. T., Melbye, M., Bisgaard, H., Pennell, C. E., Widen, E., Hakonarson, H., Smith, G. D., Heinrich, J., Jarvelin, M.-R., Early Growth Genetics (EGG) Consortium & Jaddoe, V. W. V. Common

variants at 12q15 and 12q24 are associated with infant head circumference. *Nat. Genet.* **44**, 532–538 (2012).

68. Teumer, A., Chaker, L., Groeneweg, S., Li, Y., Di Munno, C., Barbieri, C., Schultheiss, U. T., Traglia, M., Ahluwalia, T. S., Akiyama, M., Appel, E. V. R., Arking, D. E., Arnold, A., Astrup, A., Beekman, M., Beilby, J. P., Bekaert, S., Boerwinkle, E., Brown, S. J., De Buyzere, M., Campbell, P. J., Ceresini, G., Cerqueira, C., Cucca, F., Deary, I. J., Deelen, J., Eckardt, K.-U., Ekici, A. B., Eriksson, J. G., Ferrrucci, L., Fiers, T., Fiorillo, E., Ford, I., Fox, C. S., Fuchsberger, C., Galesloot, T. E., Gieger, C., Gögele, M., De Grandi, A., Grarup, N., Greiser, K. H., Haljas, K., Hansen, T., Harris, S. E., van Heemst, D., den Heijer, M., Hicks, A. A., den Hollander, W., Homuth, G., Hui, J., Ikram, M. A., Ittermann, T., Jensen, R. A., Jing, J., Jukema, J. W., Kajantie, E., Kamatani, Y., Kasbohm, E., Kaufman, J.-M., Kiemeney, L. A., Kloppenburg, M., Kronenberg, F., Kubo, M., Lahti, J., Lapauw, B., Li, S., Liewald, D. C. M., Lim, E. M., Linneberg, A., Marina, M., Mascalzoni, D., Matsuda, K., Medenwald, D., Meisinger, C., Meulenbelt, I., De Meyer, T., Meyer zu Schwabedissen, H. E., Mikolajczyk, R., Moed, M., Netea-Maier, R. T., Nolte, I. M., Okada, Y., Pala, M., Pattaro, C., Pedersen, O., Petersmann, A., Porcu, E., Postmus, I., Pramstaller, P. P., Psaty, B. M., Ramos, Y. F. M., Rawal, R., Redmond, P., Richards, J. B., Rietzschel, E. R., Rivadeneira, F., Roef, G., Rotter, J. I., Sala, C. F., Schlessinger, D., Selvin, E., Slagboom, P. E., Soranzo, N., Sørensen, T. I. A., Spector, T. D., Starr, J. M., Stott, D. J., Taes, Y., Taliun, D., Tanaka, T., Thuesen, B., Tiller, D., Toniolo, D., Uitterlinden, A. G., Visser, W. E., Walsh, J. P., Wilson, S. G., Wolffenbuttel, B. H. R., Yang, Q., Zheng, H.-F., Cappola, A., Peeters, R. P., Naitza, S., Völzke, H., Sanna, S., Köttgen, A., Visser, T. J. & Medici, M. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. *Nat. Commun.* **9**, (2018).

69. Jansen, P. R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A. R., de Leeuw, C. A., Benjamins, J. S., Muñoz-Manchado, A. B., Nagel, M., Savage, J. E., Tiemeier, H., White, T., 23andMe Research Team, Tung, J. Y., Hinds, D. A., Vacic, V., Wang, X., Sullivan, P. F., van der Sluis, S., Polderman, T. J. C., Smit, A. B., Hjerling-Leffler, J., Van Someren, E. J. W. & Posthuma, D. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).

70. van der Valk, R. J. P., Kreiner-Møller, E., Kooijman, M. N., Guxens, M., Stergiakouli, E., Sääf, A., Bradfield, J. P., Geller, F., Hayes, M. G., Cousminer, D. L., Körner, A., Thiering, E., Curtin, J. A., Myhre, R., Huikari, V., Joro, R., Kerkhof, M., Warrington, N. M., Pitkänen, N., Ntalla, I., Horikoshi, M., Veijola, R., Freathy, R. M., Teo, Y.-Y., Barton, S. J., Evans, D. M., Kemp, J. P., St Pourcain, B., Ring, S. M., Davey Smith, G., Bergström, A., Kull, I., Hakonarson, H., Mentch, F. D., Bisgaard, H., Chawes, B., Stokholm, J., Waage, J., Eriksen, P., Sevelsted, A., Melbye, M., Early Genetics and Lifecourse Epidemiology (EAGLE) Consortium, van Duijn, C. M., Medina-Gomez, C., Hofman, A., de Jongste, J. C., Taal, H. R., Uitterlinden, A. G., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Armstrong, L. L., Eriksson, J., Palotie, A., Bustamante, M., Estivill, X., Gonzalez, J. R., Llop, S., Kiess, W., Mahajan, A., Flexeder, C., Tiesler, C. M. T., Murray, C. S., Simpson, A., Magnus, P., Sengpiel, V., Hartikainen, A.-L., Keinanen-Kiukaanniemi, S., Lewin, A., Da Silva Couto Alves, A., Blakemore, A. I., Buxton, J. L., Kaakinen, M., Rodriguez, A., Sebert, S., Vaarasmaki, M., Lakka, T., Lindi, V., Gehring, U., Postma, D. S., Ang, W., Newnham, J. P., Lyytikäinen, L.-P., Pahkala, K., Raitakari, O. T., Panoutsopoulou, K., Zeggini, E., Boomsma, D. I., Groen-Blokhuis, M., Ilonen, J., Franke, L., Hirschhorn, J. N., Pers, T. H., Liang, L., Huang, J., Hocher, B., Knip, M., Saw, S.-M., Holloway, J. W., Melén, E., Grant, S. F. A., Feenstra, B., Lowe, W. L., Widén, E., Sergeyev, E., Grallert, H., Custovic, A., Jacobsson, B., Jarvelin, M.-R., Atalay, M., Koppelman, G. H., Pennell, C. E., Niinikoski, H., Dedoussis, G. V., Mccarthy, M. I., Frayling, T. M., Sunyer,

J., Timpson, N. J., Rivadeneira, F., Bønnelykke, K., Jaddoe, V. W. V., & Early Growth Genetics (EGG) Consortium. A novel common variant in DCST2 is associated with length in early life and height in adulthood. *Hum. Mol. Genet.* **24**, 1155–1168 (2015).

71. Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M. L., Mora, S., Beckmann, J. S., Bragg-Gresham, J. L., Chang, H.-Y., Demirkan, A., Den Hertog, H. M., Do, R., Donnelly, L. A., Ehret, G. B., Esko, T., Feitosa, M. F., Ferreira, T., Fischer, K., Fontanillas, P., Fraser, R. M., Freitag, D. F., Gurdasani, D., Heikkilä, K., Hyppönen, E., Isaacs, A., Jackson, A. U., Johansson, Å., Johnson, T., Kaakinen, M., Kettunen, J., Kleber, M. E., Li, X., Luan, J., Lyytikäinen, L.-P., Magnusson, P. K. E., Mangino, M., Mihailov, E., Montasser, M. E., Müller-Nurasyid, M., Nolte, I. M., O'Connell, J. R., Palmer, C. D., Perola, M., Petersen, A.-K., Sanna, S., Saxena, R., Service, S. K., Shah, S., Shungin, D., Sidore, C., Song, C., Strawbridge, R. J., Surakka, I., Tanaka, T., Teslovich, T. M., Thorleifsson, G., Van den Herik, E. G., Voight, B. F., Volcik, K. A., Waite, L. L., Wong, A., Wu, Y., Zhang, W., Absher, D., Asiki, G., Barroso, I., Been, L. F., Bolton, J. L., Bonnycastle, L. L., Brambilla, P., Burnett, M. S., Cesana, G., Dimitriou, M., Doney, A. S. F., Döring, A., Elliott, P., Epstein, S. E., Ingi Eyjolfsson, G., Gigante, B., Goodarzi, M. O., Grallert, H., Gravito, M. L., Groves, C. J., Hallmans, G., Hartikainen, A.-L., Hayward, C., Hernandez, D., Hicks, A. A., Holm, H., Hung, Y.-J., Illig, T., Jones, M. R., Kaleebu, P., Kastelein, J. J. P., Khaw, K.-T., Kim, E., Klopp, N., Komulainen, P., Kumari, M., Langenberg, C., Lehtimäki, T., Lin, S.-Y., Lindström, J., Loos, R. J. F., Mach, F., McArdle, W. L., Meisinger, C., Mitchell, B. D., Müller, G., Nagaraja, R., Narisu, N., Nieminen, T. V. M., Nsubuga, R. N., Olafsson, I., Ong, K. K., Palotie, A., Papamarkou, T., Pomilla, C., Pouta, A., Rader, D. J., Reilly, M. P., Ridker, P. M., Rivadeneira, F., Rudan, I., Ruokonen, A., Samani, N., Scharnagl, H., Seeley, J., Silander, K., Stančáková, A., Stirrups, K., Swift, A. J., Tiret, L., Uitterlinden, A. G., van Pelt, L. J., Vedantam, S., Wainwright, N., Wijmenga, C., Wild, S. H., Willemsen, G., Wilsgaard, T., Wilson, J. F., Young, E. H., Zhao, J. H., Adair, L. S., Arveiler, D., Assimes, T. L., Bandinelli, S., Bennett, F., Bochud, M., Boehm, B. O., Boomsma, D. I., Borecki, I. B., Bornstein, S. R., Bovet, P., Burnier, M., Campbell, H., Chakravarti, A., Chambers, J. C., Chen, Y.-D. I., Collins, F. S., Cooper, R. S., Danesh, J., Dedoussis, G., de Faire, U., Feranil, A. B., Ferrières, J., Ferrucci, L., Freimer, N. B., Gieger, C., Groop, L. C., Gudnason, V., Gyllensten, U., Hamsten, A., Harris, T. B., Hingorani, A., Hirschhorn, J. N., Hofman, A., Hovingh, G. K., Hsiung, C. A., Humphries, S. E., Hunt, S. C., Hveem, K., Iribarren, C., Järvelin, M.-R., Jula, A., Kähönen, M., Kaprio, J., Kesäniemi, A., Kivimaki, M., Kooner, J. S., Koudstaal, P. J., Krauss, R. M., Kuh, D., Kuusisto, J., Kyvik, K. O., Laakso, M., Lakka, T. A., Lind, L., Lindgren, C. M., Martin, N. G., März, W., McCarthy, M. I., McKenzie, C. A., Meneton, P., Metspalu, A., Moilanen, L., Morris, A. D., Munroe, P. B., Njølstad, I., Pedersen, N. L., Power, C., Pramstaller, P. P., Price, J. F., Psaty, B. M., Quertermous, T., Rauramaa, R., Saleheen, D., Salomaa, V., Sanghera, D. K., Saramies, J., Schwarz, P. E. H., Sheu, W. H.-H., Shuldiner, A. R., Siegbahn, A., Spector, T. D., Stefansson, K., Strachan, D. P., Tayo, B. O., Tremoli, E., Tuomilehto, J., Uusitupa, M., van Duijn, C. M., Vollenweider, P., Wallentin, L., Wareham, N. J., Whitfield, J. B., Wolffenbuttel, B. H. R., Ordovas, J. M., Boerwinkle, E., Palmer, C. N. A., Thorsteinsdottir, U., Chasman, D. I., Rotter, J. I., Franks, P. W., Ripatti, S., Cupples, L. A., Sandhu, M. S., Rich, S. S., Boehnke, M., Deloukas, P., Kathiresan, S., Mohlke, K. L., Ingelsson, E., Abecasis, G. R., & Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).

72. Felix, J. F., Bradfield, J. P., Monnereau, C., van der Valk, R. J. P., Stergiakouli, E., Chesi, A., Gaillard, R., Feenstra, B., Thiering, E., Kreiner-Møller, E., Mahajan, A., Pitkänen, N., Joro, R., Cavadino, A., Huikari, V., Franks, S., Groen-Blokhuis, M. M., Cousminer, D. L., Marsh, J. A., Lehtimäki, T., Curtin, J. A., Vioque, J., Ahluwalia, T. S., Myhre, R., Price, T. S., Vilor-

Tejedor, N., Yengo, L., Grarup, N., Ntalla, I., Ang, W., Atalay, M., Bisgaard, H., Blakemore, A. I., Bonnefond, A., Carstensen, L., Bone Mineral Density in Childhood Study (BMDCS), Early Genetics and Lifecourse Epidemiology (EAGLE) consortium, Eriksson, J., Flexeder, C., Franke, L., Geller, F., Geserick, M., Hartikainen, A.-L., Haworth, C. M. A., Hirschhorn, J. N., Hofman, A., Holm, J.-C., Horikoshi, M., Hottenga, J. J., Huang, J., Kadarmideen, H. N., Kähönen, M., Kiess, W., Lakka, H.-M., Lakka, T. A., Lewin, A. M., Liang, L., Lyytikäinen, L.-P., Ma, B., Magnus, P., McCormack, S. E., McMahon, G., Mentch, F. D., Middeldorp, C. M., Murray, C. S., Pahkala, K., Pers, T. H., Pfäffle, R., Postma, D. S., Power, C., Simpson, A., Sengpiel, V., Tiesler, C. M. T., Torrent, M., Uitterlinden, A. G., van Meurs, J. B., Vinding, R., Waage, J., Wardle, J., Zeggini, E., Zemel, B. S., Dedoussis, G. V., Pedersen, O., Froguel, P., Sunyer, J., Plomin, R., Jacobsson, B., Hansen, T., Gonzalez, J. R., Custovic, A., Raitakari, O. T., Pennell, C. E., Widén, E., Boomsma, D. I., Koppelman, G. H., Sebert, S., Järvelin, M.-R., Hyppönen, E., McCarthy, M. I., Lindi, V., Harri, N., Körner, A., Bønnelykke, K., Heinrich, J., Melbye, M., Rivadeneira, F., Hakonarson, H., Ring, S. M., Smith, G. D., Sørensen, T. I. A., Timpson, N. J., Grant, S. F. A., Jaddoe, V. W. V., Early Growth Genetics (EGG) Consortium, & Bone Mineral Density in Childhood Study BMDCS. Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum. Mol. Genet.* **25**, 389–403 (2016).

73. Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D. U., Zhang, Y., Sos, B. C., Afzal, V., Dickel, D. E., Kuan, S., Visel, A., Pennacchio, L. A., Zhang, K. & Ren, B. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).

74. Cusanovich, D. A., Daza, R., Adey, A., Pliner, H., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. & Shendure, J. Multiplex Single Cell Profiling of Chromatin Accessibility by Combinatorial Cellular Indexing. *Science* **348**, 910–914 (2015).

75. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **26**, 589–595 (2010).

76. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).

77. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

78. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. & Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

79. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 1–12 (2019).

80. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database J. Biol. Databases Curation* **2019**, (2019).

81. Xin, Y., Gutierrez, G. D., Okamoto, H., Kim, J., Lee, A.-H., Adler, C., Ni, M., Yancopoulos, G. D., Murphy, A. J. & Gromada, J. Pseudotime Ordering of Single Human β-Cells Reveals

States of Insulin Production and Unfolded Protein Response. *Diabetes* db180365 (2018) doi:10.2337/db18-0365.

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

83. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

84. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 1–5 (2019).

85. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

86. Arda, H. E., Tsai, J., Rosli, Y. R., Giresi, P., Bottino, R., Greenleaf, W. J., Chang, H. Y. & Kim, S. K. A Chromatin Basis for Cell Lineage and Disease Risk in the Human Pancreas. *Cell Syst.* **7**, 310-322.e4 (2018).

87. Calderon, D., Nguyen, M. L. T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J. V., Knowles, D. A., Gao, Z., Blaeschke, F., Parent, A. V., Burt, T. D., Anderson, M. S., Criswell, L. A., Greenleaf, W. J., Marson, A. & Pritchard, J. K. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**, 1494–1505 (2019).

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

89. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

90. Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S. R., Tan, G., Baranasic, D., Arenillas, D. J., Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W. W., Parcy, F. & Mathelier, A. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).

91. Wingender, E., Schoeps, T., Haubrock, M. & Dönitz, J. TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res.* **43**, D97-102 (2015).

92. Pliner, H. A., Packer, J. S., McFaline-Figueroa, J. L., Cusanovich, D. A., Daza, R. M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., Adey, A. C., Steemers, F. J., Shendure, J. & Trapnell, C. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71**, 858-871.e8 (2018).

93. Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B. L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., Steward, C., Harte, R., Lin, M., Howald, C., Tanzer, A., Derrien, T., Chrast, J., Walters, N., Balasubramanian, S., Pei,

B., Tress, M., Rodriguez, J. M., Ezkurdia, I., van Baren, J., Brent, M., Haussler, D., Kellis, M., Valencia, A., Reymond, A., Gerstein, M., Guigó, R. & Hubbard, T. J. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).

94. Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, RACI Consortium, Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R. B., Okada, Y., Raychaudhuri, S., Daly, M. J., Patterson, N., Neale, B. M. & Price, A. L. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

95. Cusanovich, D. A., Hill, A. J., Aghamirzaie, D., Daza, R. M., Pliner, H. A., Berletch, J. B., Filippova, G. N., Huang, X., Christiansen, L., DeWitt, W. S., Lee, C., Regalado, S. G., Read, D. F., Steemers, F. J., Disteche, C. M., Trapnell, C. & Shendure, J. A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324.e18 (2018).

96. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet. Epidemiol.* **33**, 79–86 (2009).

97. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).

98. Namkung, W., Lee, J. A., Ahn, W., Han, W., Kwon, S. W., Ahn, D. S., Kim, K. H. & Lee, M. G. Ca2+ activates cystic fibrosis transmembrane conductance regulator- and Cl- -dependent HCO3 transport in pancreatic duct cells. *J. Biol. Chem.* **278**, 200–207 (2003).

99. Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J. & Root, D. E. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

100. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G. & Zhang, F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).

101. Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M. & Weissman, J. S. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, (2016).

102. Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C. & Plagnol, V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).

103. Hormozdiari, F., van de Bunt, M., Segrè, A. V., Li, X., Joo, J. W. J., Bilow, M., Sul, J. H., Sankararaman, S., Pasaniuc, B. & Eskin, E. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).

104. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 1–9 (2019).