UNIVERSITY OF CALIFORNIA SAN DIEGO

Statistical strategies and resources for deciphering mechanisms of diabetes risk loci

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

in

Bioinformatics and Systems Biology

by

Anthony Aylward

Committee in charge:

Professor Kyle J. Gaulton, Chair
Professor Dorothy D. Sears, Co-Chair
Professor Graham McVicker
Professor Maike Sander
Professor Nicholas Webster
Professor Ronghui Xu

2021

The Dissertation of Anthony Aylward is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my parents, Gayle and Tom, and my brother Charlie.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I am truly thankful to my PhD advisor, Prof. Kyle Gaulton, for continuous support throughout my PhD, for pushing me to do my best, and for giving me space and autonomy when I needed it. Kyle put his faith in me as his first graduate student, and encouraged me to keep going even when I doubted myself. Second, I thank my committee members Prof. Dorothy D. Sears, Prof. Graham McVicker, Prof. Maike Sander, Prof. Nicholas Webster, and Prof. Ronghui Xu for their support and guidance during the development of this dissertation. It has been the most challenging project of my life to date, and they have all been there when I needed them. Third, I am grateful to my co-authors and fellow lab members for their invaluable contributions, especially during the formative years of our lab. Finally, I thank my family for their unerring love and support throughout my life.

Chapter 1, in full, is a reformatted reprint of material as it appears in Aylward A, Chiou J, Okino M-L, Kadakia N, Gaulton KJ. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. Hum Mol Genet. 2018. The dissertation author was a primary investigator and author of this paper.

Chapter 2, in full, is a reformatted reprint of material as it appears in appears in Aylward, A., Okino, M., Benaglio, P., Chiou, J., Beebe, E., Padilla, J.A., Diep, S., Gaulton, K.J. Glucocorticoid signaling in pancreatic islets modulates gene regulatory programs and genetic risk of type 2 diabetes. *PLOS Genet.* **17**, e1009531 (2021). The dissertation author was a primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Anthony Aylward, Mei-Lin Okino, Joshua Chiou, Jaspreet Kaur, Yleia Sanchez, Kyle J Gaulton.

Allelic imbalance mapping improves fine-mapping of diabetes risk loci. The dissertation author was a primary investigator and author of this paper.

VITA

2012 Bachelor of Science, Mathematics, University of California, Santa Barbara
2021 Doctor of Philosophy, Bioinformatics and Systems Biology, University of California San Diego

PUBLICATIONS

Anthony Aylward*, Joshua Chiou*, Mei-Lin Okino, Nikita Kadakia, Kyle J Gaulton. 2018. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Human Molecular Genetics*, ddy314, https://doi.org/10.1093/hmg/ddy314.

Anthony Aylward*, Mei-Lin Okino*, Paola Benaglio, Joshua Chiou, Elisha Beebe, Jose Andres Padilla, Sharlene Diep, Kyle J Gaulton. 2021. Glucocorticoid signaling in pancreatic islets modulates gene regulatory programs and genetic risk of type 2 diabetes. *PLOS Genetics.* 17, e1009531 (2021).

Ben Murrell, Steven Weaver, Martin D. Smith, Joel O Wertheim, Sasha Murell, **Anthony Aylward**, Kemal Eren, Tristan Pollner, Darren P. Martin, Davey M. Smith, Konrad Scheffler, Sergei L. Kosakovsky Pond. 2015. Gene-Wide identification of Episodic Selection. *Molecular Biology and Evolution*, 32 (5): 1365–1371.

William W. Greenwald, Joshua Chiou, Jian Yan, Yunjiang Qiu, Ning Dai, Allen Wang, Naoki Nariai, **Anthony Aylward**, Jee Yun Han, Nikita Kadakia, Laura Regue, Mei-Lin Okino, Frauke Drees, Dana Kramer, Nicholas Vinckier, Liliana Minichiello, David Gorkin, Joseph Avruch, Kelly A Frazer, Maike Sander, Bing Ren, Kyle J Gaulton. 2019. Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nature Communications* 10, 2078.

Paola Benaglio, Jacklyn Newsome, Jee Yun Han, Joshua Chiou, **Anthony Aylward**, Sierra Corban, Mei-Lin Okino, Jaspreet Kaur, David U Gorkin, Kyle J Gaulton. 2020. Mapping genetic effects on cell type-specific chromatin accessibility and annotating complex trait variants using single nucleus ATAC-seq. *bioRxiv*. https://doi.org/10.1101/2020.12.03.387894.

Ryan J Geusz, Allen Wang, Joshua Chiou, Joseph J Lancman, Nichole Wetton, Samy Kefalopoulou, Jinzhao Wang, Yunjaing Qiu, Jian Yan, **Anthony Aylward**, Bing Ren, P Duc Si Dong, Kyle J Gaulton, Maike Sander. 2021. *eLife*, 10:e59067

ABSTRACT OF THE DISSERTATION

Statistical strategies and resources for deciphering mechanisms of diabetes risk loci

By

Anthony Aylward

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California San Diego, 2021

Professor Kyle J. Gaulton, Chair

Professor Dorothy D. Sears, Co-Chair

The two most common forms of diabetes are type 1 diabetes (T1D) which is an autoimmune disorder and type 2 diabetes (T2D) which is a metabolic disorder. Large-scale genome-wide association studies (GWAS) have identified hundreds of loci associated with disease risk, but determining the molecular mechanisms of these loci is not trivial. One major

challenge in interpreting GWAS loci is that there is extensive linkage disequilibrium in the human genome where many variants will show association through linkage with the causal variant but not be causal themselves. A second challenge is that most GWAS loci map to non-coding regions of the genome and have no immediately obvious function or affected gene. Together these challenges motivate research to fine-map causal variants at diabetes risk loci and leverage epigenomic and functional genomic data to determine the mechanisms of fine-mapped variants.

In my work I developed strategies and created resources for fine-mapping diabetes risk signals identified in GWAS and determining their molecular mechanisms. In the first chapter, we identify genetic risk shared by T1D and T2D. We then fine-map causal variants at specific shared loci and perform molecular characterization of candidate causal variants at the shared risk loci *GLIS3* and *CTRB1/2* in pancreatic islets. In the second chapter, we use ATAC-seq and RNA-seq on dexamethasone-treated and untreated pancreatic islets to generate a map of glucocorticoid-responsive islet chromatin sites and gene expression, as well as genetic variants that interact with glucocorticoid signaling to affect islet regulation. We identify enrichment of T2D-associated variants in glucocorticoid-responsive islet chromatin and characterize a fine-mapped T2D risk variant with glucocorticoid-dependent effects on islet accessible chromatin and *SIX2/3* expression. Finally, in the third chapter we develop a novel framework for allelic imbalance mapping using ChIP-seq and ATAC-seq data. We quantify the allelic effects of variants on epigenomic sequencing data in islets and liver cells and demonstrate that these effects can help predict likely causal variants for expression QTLs and T2D risk loci. At the *HMG20A* locus, we identify a fine-mapped T2D risk variant with allelic imbalance in pancreatic islet accessible chromatin and validate allelic effects on pancreatic islets.

INTRODUCTION

Over the past decade, genome-wide association studies (GWAS) have powered an explosion of scientific discoveries revealing the genetic causes of complex disease. Over 50,000 genomic loci have been found to be associated with risk of complex human diseases such as heart disease, diabetes, and psychiatric disorders, and many other phenotypes[1–4]. While GWAS have been wildly successful in identifying risk loci, the molecular mechanisms through which these loci affect risk of disease remain almost completely unknown. The majority of risk loci map to non-coding DNA and likely affect disease risk by altering gene regulation[5–7]. Determining the regulatory function of these loci requires identifying the specific causal variants underlying disease risk, but this is complicated by extensive linkage disequilibrium in the human genome[8–11]. Fine-mapping can help narrow disease association signals to precise causal variants and can comprise statistical analyses of genetic data alone as well as techniques that leverage functional genomic data in these analyses[5,12–16].

Diabetes affects over 400 million individuals worldwide and is a major epidemic. The two major forms of diabetes are type 1 diabetes (T1D) and type 2 diabetes (T2D) which are both complex diseases, where GWAS of T1D have identified over 90 associated loci while T2D studies have identified over 200[17,18]. The great majority of these loci map to non-coding regions, and have no immediately obvious mechanisms. Fine-mapping plays an especially important role in GWAS studies of type 1 and Type 2 diabetes[19,20]. However, despite extensive research that has pinpointed causal variants at some loci[17,19,20,21(p1),22(p5)], the causal variants at most loci remain unknown. These efforts have benefited from integrating GWAS results with transcriptomic and epigenomic data representing disease-relevant tissues and cell types, such as expression QTL (eQTL) mapping, ChIP-seq, and ATAC-seq studies in pancreatic islets[20,23–26]. Hence, future

advances in understanding the genetic causes of diabetes will be facilitated by joint analyses of GWAS and other functional genomics datasets.

In this work we develop and apply strategies for fine-mapping and functional inference of variants at diabetes risk loci. In the first chapter, we leverage genetic data from multiple traits to fine-map diabetes loci. We define a broad relationship between genetic risk of T1D and T2D. We determine that regulatory sites active in pancreas and multiple other tissues enriched for variants associated with both traits. We then identify specific loci influencing risk of both T1D and T2D including *GLIS3* and *CTRB1/2*[27,28]. At these shared loci, we perform fine-mapping using GWAS data from both T1D and T2D to improve resolution of causal variants. By using trait-enriched regulatory annotations[29–32] we next identified candidate causal variants for these loci. Finally, we validated the molecular effects of these candidate variants using gene reporter assays in pancreatic islets.

In the second chapter, we determine the role of regulatory programs that respond to environmental stimuli in diabetes risk. Glucocorticoids are steroid hormones which regulate inflammatory, metabolic, and stress responses and are highly relevant to diabetes pathogenesis[33–37]. We therefore generated transcriptomic and epigenomic datasets in pancreatic islets cultured with the glucocorticoid dexamethasone as well as in untreated conditions. We map glucocorticoid-dependent changes in islet gene expression and chromatin accessibility, and identify transcriptional regulators of these changes and downstream effects on molecular pathways. Fasting glucose- and T2D-associated variants were enriched in glucocorticoid-responsive chromatin sites which were linked to glucocorticoid-responsive genes. Finally, we show that a likely causal T2D and glucose variant at the *SIX2/SIX3* locus has a glucocorticoid-dependent effect on regulatory activity in islets.

In the third chapter, we develop a framework for gaining additional insights from epigenomic sequencing data using allelic imbalance mapping. Allelic imbalance identifies heterozygous variants where the two alleles have differences in the readout of sequencing experiments such as RNA-seq, ChIP-seq, or ATAC-seq[38–41]. Traditional imbalance analysis requires independent genotyping data to determine which SNPs are heterozygous in the experimental sample. However, methods such as QuASAR have been developed to infer genotypes from epigenomic data[42]. We demonstrate that QuASAR enables accurate genotyping of the liver cancer cell line HepG2 and primary pancreatic islet cells from ChIP-seq and ATAC-seq data. We then develop a novel estimator of allelic effects from imbalance statistics which can be applied to even shallow sequence data. Application of this estimator to fine-mapping data for islet eQTL signals and T2D loci demonstrates that variants with evidence for imbalance in islet ATAC-seq are more likely to be causal. Furthermore, we prioritize candidate causal variants at T2D loci, including at the *HMG20A* locus where we validate allelic effects of a variant with evidence for allelic imbalance using gene reporter assays in pancreatic islet cells.

In summary, this work presents techniques and epigenomic mapping resources to improve fine-mapping and determine the molecular function of diabetes risk loci.

1. Chang M, He L, Cai L. An Overview of Genome-Wide Association Studies. *Methods Mol Biol*. 2018;1754:97-108. doi:10.1007/978-1-4939-7717-8_6

2. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20(8):467-484. doi:10.1038/s41576-019-0127-1

3. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F, Parkinson H. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120

4. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, Kubo M, Okada Y, Kamatani Y. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet*. 2018;50(3):390-400. doi:10.1038/s41588-018-0047-6

5. Cano-Gamez E, Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet*. 2020;11. doi:10.3389/fgene.2020.00424

6. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet*. 2015;24(R1):R102-110. doi:10.1093/hmg/ddv259

7. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLOS Genetics*. 2010;6(4):e1000888. doi:10.1371/journal.pgen.1000888

8. Jin H-J, Jung S, DebRoy AR, Davuluri RV. Identification and validation of regulatory SNPs that modulate transcription factor chromatin binding and gene expression in prostate cancer. *Oncotarget*. 2016;7(34):54616-54626. doi:10.18632/oncotarget.10520

9. Bonjoch L, Mur P, Arnau-Collell C, Vargas-Parra G, Shamloo B, Franch-Expósito S, Pineda M, Capellà G, Erman B, Castellví-Bel S. Approaches to functionally validate candidate genetic variants involved in colorectal cancer predisposition. *Molecular Aspects of Medicine*. 2019;69:27-40. doi:10.1016/j.mam.2019.03.004

10. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. *Nature*. 2001;411(6834):199-204. doi:10.1038/35075590

11. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet*. 2002;3(4):299-309. doi:10.1038/nrg777

12. Broekema RV, Bakker OB, Jonkers IH. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biology*. 10(1):190221. doi:10.1098/rsob.190221

13. Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJH, Shishkin AA, Hatan M, Carrasco-Alfonso MJ, Mayer D, Luckey CJ, Patsopoulos NA, Jager PLD, Kuchroo VK, Epstein CB, Daly MJ, Hafler DA, Bernstein BE. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337-343. doi:10.1038/nature13835

14. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 2018;19(8):491-504. doi:10.1038/s41576-018-0016-z

15. Pickrell JK. Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits. *Am J Hum Genet*. 2014;94(4):559-573. doi:10.1016/j.ajhg.2014.03.004

16. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying Causal Variants at Loci with Multiple Signals of Association. *Genetics*. 2014;198(2):497-508. doi:10.1534/genetics.114.167908

17. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Mägi R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss MH, Prins BP, Guo X, Bielak LF, Below JE, Bowden DW, Chambers JC, Kim YJ, Ng MCY, Petty LE, Sim X, Zhang W, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Ec Kardt K-U, Fischer K, Kardia SLR, Kronenberg F, Läll K, Liu C-T, Locke AE, Luan J, Ntalla I, Nylander V, Schönherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jørgensen ME, Jørgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stančáková A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50(11):1505-1513. doi:10.1038/s41588-018-0241-6

18. Chiou J, Geusz RJ, Okino M-L, Han JY, Miller M, Melton R, Beebe E, Benaglio P, Huang S, Korgaonkar K, Heller S, Kleger A, Preissl S, Gorkin DU, Sander M, Gaulton KJ. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature*. 2021;594(7863):398-402. doi:10.1038/s41586-021-03552-w

19. Onengut-Gumuscu S, Chen W-M, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Farber E, Bonnie JK, Szpak M, Schofield E, Achuthan P, Guo H, Fortune MD, Stevens H, Walker NM, Ward LD, Kundaje A, Kellis M, Daly MJ, Barrett JC, Cooper JD, Deloukas P, Todd JA, Wallace C, Concannon P, Rich SS. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet*. 2015;47(4):381-386. doi:10.1038/ng.3245

20. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, Mahajan A, Locke A, Rayner NW, Robertson N, Scott RA, Prokopenko I, Scott LJ, Green T, Sparso T, Thuillier D, Yengo L, Grallert H, Wahl S, Frånberg M, Strawbridge RJ, Kestler H, Chheda H, Eisele L, Gustafsson S, Steinthorsdottir V, Thorleifsson G, Qi L, Karssen LC, Leeuwen EM van, Willems SM, Li M, Chen H, Fuchsberger C, Kwan P, Ma C, Linderman M, Lu Y, Thomsen SK, Rundle JK, Beer NL, Bunt M van de, Chalisey A, Kang HM, Voight BF, Abecasis GR, Almgren P, Baldassarre D, Balkau B, Benediktsson R, Blüher M, Boeing H, Bonnycastle LL, Bottinger EP, Burtt NP, Carey J, Charpentier G, Chines PS, Cornelis MC, Couper DJ, Crenshaw AT, Dam RM van, Doney ASF, Dorkhan M, Edkins S, Eriksson JG, Esko T, Eury E, Fadista J, Flannick J, Fontanillas P, Fox C, Franks PW, Gertow K, Gieger C, Gigante B, Gottesman O, Grant GB, Grarup N, Groves CJ, Hassinen M, Have CT, Herder C, Holmen OL, Hreidarsson AB, Humphries SE, Hunter DJ, Jackson AU, Jonsson A, Jørgensen ME, Jørgensen T, Kao W-HL, Kerrison ND, Kinnunen L, Klopp N, Kong A, Kovacs P, Kraft P, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet*. 2015;47(12):1415-1425. doi:10.1038/ng.3437

21. Fogarty MP, Cannon ME, Vadlamudi S, Gaulton KJ, Mohlke KL. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet*. 2014;10(9):e1004633. doi:10.1371/journal.pgen.1004633

22. Roman TS, Cannon ME, Vadlamudi S, Buchkovich ML, Wolford BN, Welch RP, Morken MA, Kwon GJ, Varshney A, Kursawe R, Wu Y, Jackson AU, Program NI of HISC (NISC) CS, Erdos MR, Kuusisto J, Laakso M, Scott LJ, Boehnke M, Collins FS, Parker SCJ, Stitzel ML, Mohlke KL. A Type 2 Diabetes–Associated Functional Regulatory Variant in a Pancreatic Islet Enhancer at the ADCY5 Locus. *Diabetes*. 2017;66(9):2521-2530. doi:10.2337/db17-0464

23. Viñuela A, Varshney A, van de Bunt M, Prasad RB, Asplund O, Bennett A, Boehnke M, Brown AA, Erdos MR, Fadista J, Hansson O, Hatem G, Howald C, Iyengar AK, Johnson P, Krus U, MacDonald PE, Mahajan A, Manning Fox JE, Narisu N, Nylander V, Orchard P, Oskolkov N, Panousis NI, Payne A, Stitzel ML, Vadlamudi S, Welch R, Collins FS, Mohlke KL, Gloyn AL, Scott LJ, Dermitzakis ET, Groop L, Parker SCJ, McCarthy MI. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat Commun*. 2020;11(1):4912. doi:10.1038/s41467-020-18581-8

24. Cebola I. Pancreatic Islet Transcriptional Enhancers and Diabetes. *Curr Diab Rep*. 2019;19(12):145. doi:10.1007/s11892-019-1230-6

25. Chiou J, Zeng C, Cheng Z, Han JY, Schlichting M, Miller M, Mendez R, Huang S, Wang J, Sui Y, Deogaygay A, Okino M-L, Qiu Y, Sun Y, Kudtarkar P, Fang R, Preissl S, Sander M, Gorkin DU, Gaulton KJ. Single-cell chromatin accessibility identifies pancreatic islet cell type– and state-specific regulatory programs of diabetes risk. *Nature Genetics*. Published online April 1, 2021:1-12. doi:10.1038/s41588-021-00823-0

26. Gaulton KJ. Mechanisms of Type 2 Diabetes Risk Loci. *Curr Diab Rep*. 2017;17(9):72. doi:10.1007/s11892-017-0908-x

27. ‘t Hart LM, Fritsche A, Nijpels G, van Leeuwen N, Donnelly LA, Dekker JM, Alssema M, Fadista J, Carlotti F, Gjesing AP, Palmer CNA, van Haeften TW, Herzberg-Schäfer SA, Simonis-Bik AMC, Houwing-Duistermaat JJ, Helmer Q, Deelen J, Guigas B, Hansen T, Machicao F, Willemsen G, Heine RJ, Kramer MHH, Holst JJ, de Koning EJP, Häring H-U, Pedersen O, Groop L, de Geus EJC, Slagboom PE, Boomsma DI, Eekhoff EMW, Pearson ER, Diamant M. The CTRB1/2 Locus Affects Diabetes Susceptibility and Treatment via the Incretin Pathway. *Diabetes*. 2013;62(9):3275-3281. doi:10.2337/db13-0227

28. Dooley J, Tian L, Schonefeldt S, Delghingaro-Augusto V, Garcia-Perez JE, Pasciuto E, Di Marino D, Carr EJ, Oskolkov N, Lyssenko V, Franckaert D, Lagou V, Overbergh L, Vandenbussche J, Allemeersch J, Chabot-Roy G, Dahlstrom JE, Laybutt DR, Petrovsky N, Socha L, Gevaert K, Jetten AM, Lambrechts D, Linterman MA, Goodnow CC, Nolan CJ, Lesage S, Schlenner SM, Liston A. Genetic Predisposition for Beta Cell Fragility Underlies Type 1 and Type 2 Diabetes. *Nat Genet*. 2016;48(5):519-527. doi:10.1038/ng.3531

29. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, Pfenning AR, Wang X, Claussnitzer M, Yaping Liu, Coarfa C, Alan Harris R, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, David Hawkins R, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield

TK, Scott Hansen R, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh K-H, Feizi S, Karlic R, Kim A-R, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, Jager PLD, Farnham PJ, Fisher SJ, Haussler D, Jones SJM, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai L-H, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317-330. doi:10.1038/nature14248

30. Strawbridge RJ, Dupuis J, Prokopenko I, Barker A, Ahlqvist E, Rybin D, Petrie JR, Travers ME, Bouatia-Naji N, Dimas AS, Nica A, Wheeler E, Chen H, Voight BF, Taneera J, Kanoni S, Peden JF, Turrini F, Gustafsson S, Zabena C, Almgren P, Barker DJP, Barnes D, Dennison EM, Eriksson JG, Eriksson P, Eury E, Folkersen L, Fox CS, Frayling TM, Goel A, Gu HF, Horikoshi M, Isomaa B, Jackson AU, Jameson KA, Kajantie E, Kerr-Conte J, Kuulasmaa T, Kuusisto J, Loos RJF, Luan J, Makrilakis K, Manning AK, Martínez-Larrad MT, Narisu N, Nastase Mannila M, Ohrvik J, Osmond C, Pascoe L, Payne F, Sayer AA, Sennblad B, Silveira A, Stancáková A, Stirrups K, Swift AJ, Syvänen A-C, Tuomi T, van 't Hooft FM, Walker M, Weedon MN, Xie W, Zethelius B, DIAGRAM Consortium, GIANT Consortium, MuTHER Consortium, CARDIoGRAM Consortium, C4D Consortium, Ongen H, Mälarstig A, Hopewell JC, Saleheen D, Chambers J, Parish S, Danesh J, Kooner J, Ostenson C-G, Lind L, Cooper CC, Serrano-Ríos M, Ferrannini E, Forsen TJ, Clarke R, Franzosi MG, Seedorf U, Watkins H, Froguel P, Johnson P, Deloukas P, Collins FS, Laakso M, Dermitzakis ET, Boehnke M, McCarthy MI, Wareham NJ, Groop L, Pattou F, et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes*. 2011;60(10):2624-2634. doi:10.2337/db11-0415

31. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Mägi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JRB, Egan JM, Lajunen T, Grarup N, Sparsø T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proença C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Payne F, Roccasecca RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Böttcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen Y-DI, Chines P, Clarke R, Coin LJM, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day INM, de Geus EJC, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42(2):105-116. doi:10.1038/ng.520

32. Wheeler E, Leong A, Liu C-T, Hivert M-F, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, Chu AY, Zhang W, Wang X, Chen P, Maruthur NM, Porneala BC, Sharp SJ, Jia Y, Kabagambe EK, Chang L-C, Chen W-M, Elks CE, Evans DS, Fan Q, Giulianini F, Go MJ, Hottenga J-J, Hu Y, Jackson AU, Kanoni S, Kim YJ, Kleber ME, Ladenvall C, Lecoeur C, Lim S-H, Lu Y, Mahajan A, Marzi C, Nalls MA, Navarro P, Nolte IM, Rose LM, Rybin DV, Sanna S, Shi Y, Stram DO, Takeuchi F, Tan SP, van der Most PJ, Van Vliet-Ostaptchouk JV, Wong A, Yengo L, Zhao W, Goel A, Martinez Larrad MT, Radke D, Salo P, Tanaka T,

van Iperen EPA, Abecasis G, Afaq S, Alizadeh BZ, Bertoni AG, Bonnefond A, Böttcher Y, Bottinger EP, Campbell H, Carlson OD, Chen C-H, Cho YS, Garvey WT, Gieger C, Goodarzi MO, Grallert H, Hamsten A, Hartman CA, Herder C, Hsiung CA, 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, et al. 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*. 2017;14(9):e1002383. doi:10.1371/journal.pmed.1002383

33. Becker DE. Basic and Clinical Pharmacology of Glucocorticosteroids. *Anesth Prog*. 2013;60(1):25-32. doi:10.2344/0003-3006-60.1.25

34. Coutinho AE, Chapman KE. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. *Mol Cell Endocrinol*. 2011;335(1):2-13. doi:10.1016/j.mce.2010.04.005

35. Flammer JR, Rogatsky I. Minireview: Glucocorticoids in Autoimmunity: Unexpected Targets and Mechanisms. *Mol Endocrinol*. 2011;25(7):1075-1086. doi:10.1210/me.2011-0068

36. Bauerle KT, Harris C. Glucocorticoids and Diabetes. *Mo Med*. 2016;113(5):378-383.

37. Suh S, Park MK. Glucocorticoid-Induced Diabetes Mellitus: An Important but Overlooked Problem. *Endocrinol Metab (Seoul)*. 2017;32(2):180-189. doi:10.3803/EnM.2017.32.2.180

38. Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nature Reviews Genetics*. 2010;11(8):533-538. doi:10.1038/nrg2815

39. Hasin-Brumshtein Y, Hormozdiari F, Martin L, van Nas A, Eskin E, Lusis AJ, Drake TA. Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics*. 2014;15(1):471. doi:10.1186/1471-2164-15-471

40. Chen J, Rozowsky J, Galeev TR, Harmanci A, Kitchen R, Bedford J, Abyzov A, Kong Y, Regan L, Gerstein M. A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat Commun*. 2016;7(1):11101. doi:10.1038/ncomms11101

41. Zhang Q, Keles S. An empirical Bayes test for allelic-imbalance detection in ChIP-seq. *Biostatistics*. 2018;19(4):546-561. doi:10.1093/biostatistics/kxx060

42. Harvey CT, Moyerbrailean GA, Davis GO, Wen X, Luca F, Pique-Regi R. QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics*. 2015;31(8):1235-1242. doi:10.1093/bioinformatics/btu802

CHAPTER 1: Shared genetic risk contributes to type 1 and type 2 diabetes etiology

1.1 Abstract

The extent to which shared genetic risk contributes to T1D and T2D etiology is unknown. In this study, we generated T1D association data of 15k samples imputed into the HRC panel which we compared to published T2D association data imputed into 1000 Genomes. The effects of genetic variants on T1D and T2D risk at known loci and genome-wide were positively correlated. Increased risk of T1D and T2D was correlated with higher fasting insulin and glucose level and decreased birth weight, among other correlations. Variants with T1D and T2D association were further enriched in pancreatic, adipose, B cell, and endoderm regulatory elements. We fine-mapped causal variants at known loci and found evidence for co-localization at five signals, four of which had same direction of effect. Shared risk variants were associated with quantitative measures of islet function and early growth, and were expression QTLs in relevant tissues. We further identified a shared variant at *GLIS3* in islet accessible chromatin with allelic effects on enhancer activity. Our findings identify a shared genetic risk involving effects on islet function as well as insulin resistance, growth and development in the etiology of T1D and T2D, supporting a role for T2D-relevant processes in addition to autoimmunity in T1D risk.

1.2 Introduction

Diabetes affects over 400 million individuals worldwide and contributes to substantial morbidity and mortality[1]. Type 1 diabetes (T1D) is an autoimmune disease resulting in destruction of pancreatic beta cells, whereas type 2 diabetes (T2D) is a metabolic disease of insulin resistance and beta cell dysfunction[2]. Genetics plays a major role in both forms of diabetes, where

over 60 risk signals have been identified for T1D[3] and over 400 for T2D[4]. Roughly half of the genetic risk for T1D can be attributed to the HLA locus, and many known T1D risk loci affect immune function[2]. Conversely, the majority of known T2D risk loci appear to affect pancreatic islet and insulin resistance tissues such as adipocytes and skeletal muscle[5–9]. Outside of known loci there are many additional genetic factors influencing diabetes risk[7]. Pathophysiological links have been reported between T1D and T2D suggesting an underlying shared etiology[10,11], but the contribution of genetic variants to this shared etiology and the underlying molecular and physiological mechanisms are unknown.

Multiple genomic loci that affect risk of both T1D and T2D have been identified. One example is the *CTRB1* locus, where risk variants are correlated with chymotrypsin expression in the pancreas and pancreatic islets and GLP-1 mediated insulin secretion[12]. Another example is *GLIS3,* a gene that causes monogenic neonatal diabetes[13(p3)]. A linkage study in non-obese diabetic (NOD) mice identified an effect of the *GLIS3* locus on T1D progression, suggesting an underlying pancreatic beta cell phenotype[11]. This study further argued that beta cell 'fragility' involving the unfolded protein response leading to pronounced cell death underlies shared T1D and T2D risk[14]. However, the specific causal variants at shared risk loci, including whether the signals are the same or distinct, and the mechanisms of how they alter genomic and cellular functions to influence disease risk are unknown. Furthermore, shared loci appear to have both opposite (*CTRB1*) and same *(GLIS3)* direction of effect on T1D and T2D risk, and thus the broader relationship between their genetic effects on T1D and T2D is unclear.

Genome-wide association data of variant genotypes imputed into comprehensive reference panels enables understanding broad relationships to other traits and functional annotations[15–17]. In addition, these data enable fine-mapping of causal variants and mechanisms underlying diabetes risk at specific loci[7]. Previous fine-mapping studies of T1D and T2D loci

resolved sets of causal variants at many risk signals and the annotations enriched in these causal variant sets[18,19]. These studies revealed that the majority of risk signals for diabetes are non-coding and map in regulatory elements active in specific cell-types, such as lymphoid cells for T1D[7,18,19]. Projects such as ENCODE and the NIH Epigenome Roadmap have annotated regulatory elements in hundreds of human cells and tissues[20,21], while other studies have provided detailed regulatory maps of specific tissues[5,22]. Epigenomic annotations broadly enriched for disease signals can further be used to prioritize potential functions of causal variants overlapping these annotations for experimental validation[18].

Here, we studied genetic risk of T1D and T2D using comprehensive genome-wide association data for both traits. We identified positive correlations both genome-wide and at known loci between variant effects on T1D and T2D risk. Increased risk of T1D and T2D was further correlated with higher fasting insulin and glucose level and decreased birth weight, among other traits, and variants with T1D and T2D association were enriched in pancreatic islet, adipocyte, CD19+ B cell, and CD184+ endoderm regulatory elements. We identified evidence of co-localized risk signals for T1D and T2D at five loci, four of which had same direction of effect. Shared signals were associated with measures of beta cell function and early growth phenotypes, and were also quantitative trait loci for gene expression in relevant tissues. We fine-mapped casual variants at shared signals and identified a candidate variant at *GLIS3* in islet accessible chromatin with allelic effects on islet enhancer activity. Together our results provide evidence for a shared genetic risk underlying T1D and T2D etiology, and demonstrate a role for variants affecting both islet function and insulin resistance in addition to immune system activity in the genetic basis of T1D.

1.3 Results

1.3.1 Genetic variants have shared effects on T1D and T2D risk

We generated genome-wide association data for T1D using publicly-available genotype data of T1D case and control samples of European ancestry (**see Methods, Figure S1.1**). We imputed genotypes from each study into 39M variants in the Haplotype Reference Consortium (HRC) panel[23]. Imputed genotypes passing quality filters (r^2>.3) were tested for T1D association separately for different genotyping platforms using firth-biased regression including sex and the top 3 principal components as covariates. We then performed inverse variance weighted meta-analysis to combine results. We retained imputed variants tested in all samples with minor allele frequency (MAF) > .005, resulting in 8.5M variants. As expected, given comparable sample size to previous studies, variants with genome-wide significant association mapped to known loci (**Figure S1.1**).

We then determined the relationship between variant effects on T1D and T2D risk by comparing T1D association statistics with T2D association from the DIAGRAM consortium[24]. We first determined shared effects among variants at known risk loci for both traits excluding the MHC locus. There was an enrichment of nominal T1D association (P<.05) among 94 known T2D index variants relative to matched background variants (obs=19.1%, exp=7.8%, binomial P=3.2×10^{-4}) (**Figure 1.1A, Table S1.1)**. T2D index variants were also enriched for concordant direction of effect on T1D (57/94, binomial P=.037), including among those with nominal T1D association (T1D P<.05) (14/18, binomial P=.031) (**Figure 1.1B, Table S1.1**). We found significant directional concordance among the 14 variants with both nominal T1D association and same direction of effect on T2D using summary data from UK Biobank (UKBB) (12/14, binomial P=.013). Despite a net sharing in effects, several T2D loci had opposite effects on T1D risk including *CTRB1* and *TCF7L2* (**Figure 1.1B**). Conversely, there was less evidence for enrichment of nominal T2D

association (obs=12.2%, exp=7.3%, binomial P=.19) or concordant direction of effect (28/57, binomial P=1) among 57 known T1D variants (**Figure 1.1A, Table S1.2**).

We then determined the correlation between variant effects genome-wide on T1D and T2D risk. In these analyses, we used LD-score regression on the set of HapMap3 variants common to T1D and T2D association datasets (**see Methods**). We observed evidence for a positive correlation in the effects of variants genome-wide on T1D and T2D risk (R_g=.18) (**Figure 1.1C)**. We also identified positive correlation with T1D risk when using T2D association data imputed from different reference panels (GoT2D, HM2) (R_g=.18, R_g=.23) and from trans-ethnic cohorts (R_g=.22) (**Figure 1.1C**). To limit the potential effects of misdiagnosed diabetes on these results, we first generated association data using clinical definitions of T1D and T2D in the WTCCC and observed a positive correlation when using either WTCCC dataset (T2D-WTCCC R_g=.24; T1D-WTCCC R_g=.15) (**see Methods**). Second, we removed obese (BMI>30) samples from T1D cohorts and the positive correlation with T2D remained (R_g=.18) (**Figure 1.1C**). Finally, a positive correlation remained after removing variants in a 1MB window around 57 known T1D loci or 94 known T2D loci (no T1D R_g=.20, no T2D R_g=.16). These results demonstrate consistent evidence for correlated effects of variants genome-wide on T1D and T2D risk.

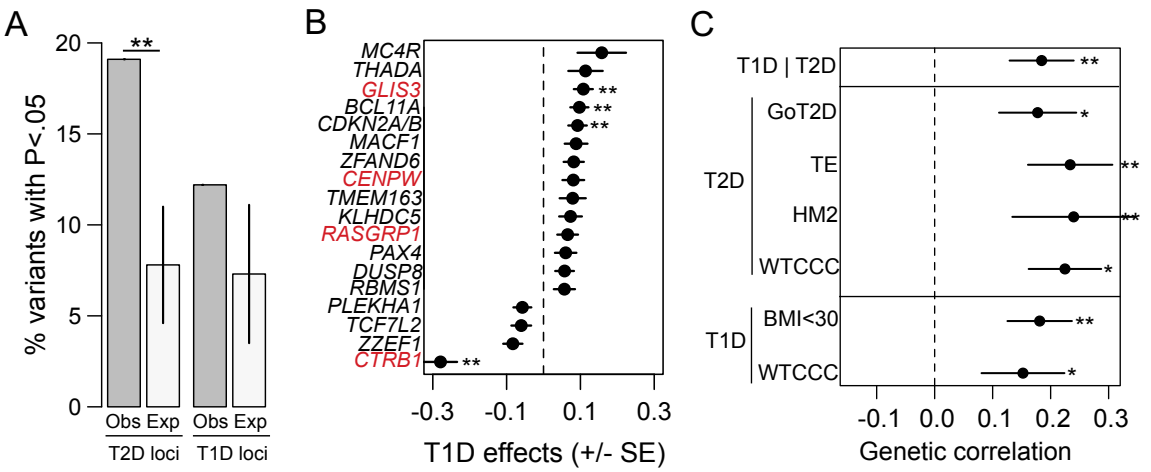


Figure 1.1. Shared effects of genetic variants on T1D and T2D risk. (A) Known T2D risk variants are significantly enriched for nominal T1D association (P<.05), whereas known T1D risk variants do not show evidence for enrichment of nominal T2D association. **P<.001 (B) Known T2D risk variants with nominal T1D association have concordant direction of effect on T1D risk (14/18, red=known T1D locus; **index variant T1D $P<5x10^{-4}$). Values are T1D effect size and SE. (C) Variants genome-wide have correlated effects on T1D and T2D risk using multiple datasets for each disease (WTCCC – Wellcome Trust Case Control Consortium, T2D TE – Mahajan et al, T2D HM2 – Morris et al 2012, T2D GoT2D – Fuchsberger et al 2016, T1D BMI<30 – T1D association data removing individuals with BMI>30). Values are genetic correlation estimates and SE. **P<.005, **P<.05

1.3.2 Mechanisms of shared variant effects on T1D and T2D risk

Given evidence for a positive correlation in variant effects on T1D and T2D, we sought to understand potential mechanisms underlying the shared effects. We first determined the correlation between T1D and T2D risk and relevant traits using LD score regression[25–28]. For T2D, there was a significant correlation between T2D risk and increased HbA1C level (R_g=.66, $P=2.7x10^{-18}$), fasting glucose level (R_g=.60, $P=7.3x10^{-14}$), fasting insulin level (R_g=.52, $P=8.5x10^{-12}$), HOMA-IR (R_g=.55, $P=7.2x10^{-9}$), and body-mass index (BMI) (R_g=.47, $P=1.1x10^{-38}$), and decreased birth weight (R_g=-.25, $P=3.0x10^{-8}$) (**Figure 1.2A**). There was also evidence for a correlation between T2D risk and increased proinsulin level (R_g=.21, P=.037) and male height at age 12 (12M; R_g=.10, P=.17) although the latter estimate was not significant. For T1D, we

observed a correlation between T1D risk and increased fasting proinsulin (R_g=.23, P=.017) and fasting insulin level (R_g=.15, P=.049) (**Figure 1.2A**). We also observed evidence for a correlation between T1D risk and decreased birth weight (R_g=-.10, P=.053), and increased male height at age 12 (12M; R_g=.18, P=.061) and fasting glucose level (R_g=.068, P=.29) although these estimates were not significant. We did not observe correlation between T1D and BMI (R_g=0.010, P=.74) or childhood obesity (R_g=-0.047, P=.47), the latter previously identified as an instrumental variable for T1D risk[29]. While there were significant correlations between T1D and other autoimmune traits such as inflammatory bowel disease (R_g=-.18, P=1.9x10^{-3}) and rheumatoid arthritis (R_g=.47, P=3.2x10^{-7}), we observed no such correlations for T2D (**Figure S1.2**).

We determined the extent to which traits correlated with both T1D and T2D risk might be driven through variants with shared effects on T1D and T2D. From genome-wide association data for T1D and T2D, we extracted variants with the same direction of effect and tested these variants for correlation to each trait using LD score regression. For both T1D and T2D, we observed stronger correlations with increased fasting glucose level (T1D shared R_g=.43, T2D shared R_g=.65), increased fasting insulin level (T1D shared R_g=.55, T2D shared R_g=.68), and decreased birth weight (T1D shared R_g=.25, T2D shared R_g=.29) among variants with same direction of effect (**Figure 1.2B**). We observed less evidence for pronounced correlation between shared effect T1D and T2D variants and fasting proinsulin level (T1D shared R_g=.33, T2D shared R_g=.28), and male height at age 12 (T1D shared R_g=.26, T2D shared R_g=.16) (**Figure 1.2B**).

We next determined functional annotations enriched for T1D and T2D associated variants. We used annotations of active enhancer and promoter elements in 98 cell types from the Epigenome roadmap project[21] and annotations of protein-coding gene exons and UTRs from GENCODE[30]. We tested for enrichment of each annotation for T1D and T2D risk using stratified LD score regression[16]. There was evidence for positive enrichment genome-wide of both T1D and

T2D association for variants in pancreatic islet (T1D Z=1.02, T2D Z=2.67), adipose nuclei (T1D Z=.09, T2D Z=1.52), CD19+ B cell (T1D Z=3.12, T2D Z=.31), CD184+ endoderm (T1D Z=.62, T2D Z=1.25), and pancreas (T1D Z=.41, T2D Z=.62) regulatory elements **(Figure 1.2C**). We also observed enrichments specific to each trait, most notably T1D association for immune regulatory elements such as T cell (Z=4.67) and fetal thymus (Z=1.83) (**Table S4**).

Given enrichment of multiple cell-types for both T1D and T2D association, we next tested to what extent these effects were driven through variants with same direction of effect on T1D and T2D. We obtained LD-pruned variants nominally associated (P<.05) with both T1D and T2D and with same direction of effect and tested for enrichment of overlap with each annotation compared to random sets of matched variants (**see Methods**). We observed significant enrichment of overlap with CD184+ endoderm (Fisher's P=.017), adipose nuclei (P=.018) and pancreatic islet (P=.040) regulatory sites (**Figure 1.2D**). We next repeated these analyses instead using variants with opposite direction of effect on T1D and T2D. We observed significant overlap of opposite effect variants with CD184+ endoderm (P=.031) and pancreatic islet regulatory elements (P=.020), suggesting that these cell-types are enriched in variants with both shared and opposite effects on T1D and T2D.

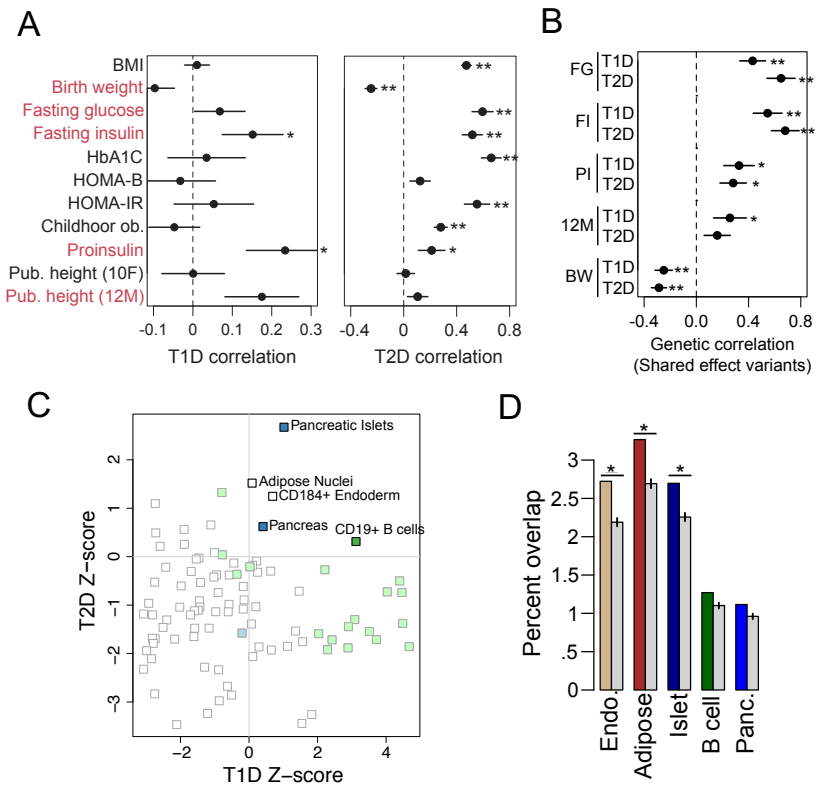


Figure 1.2. Mechanisms of variant effects on T1D and T2D risk. (A) Increased T1D risk (left) is correlated with increased fasting insulin level and proinsulin level (*P<.05), in addition to increased male height at age 12 (12M) and fasting glucose level, and decreased birth weight; Increased T2D risk (right) is correlated with increased HbA1C, fasting glucose, fasting insulin, HOMA-IR, BMI and childhood obesity, and decreased birth weight (**P<1x10^{-4}). Values are genetic correlation estimates and SE. (C) Variants with same direction of effect on T1D and T2D risk have stronger correlation with increased fasting insulin, glucose and proinsulin level, and decreased birth weight. (**P<1x10^{-4}, *P<.05). Values are genetic correlation estimates and SE. (D) Variants with T1D and T2D association are enriched for pancreatic islet, adipose, CD19+ B cell, and CD184+ endoderm regulatory sites. (blue = pancreatic, green = immune). (E) Variants with both nominal association (P<.05) and shared direction of effect on T1D and T2D risk are significantly enriched in endoderm, islet and adipose regulatory sites. (*P<.05). Values are percent overlap and CI.

1.3.3 Fine-mapping and functional annotation of known T1D risk loci

We next used association data to fine-map specific causal variants influencing T1D and T2D. For T2D we compiled fine-mapping data of 94 signals from previous studies (**see Methods**). As fine-mapping data for all known T1D loci have not been previously reported, we used T1D association statistics to fine-map 57 T1D risk signals excluding the MHC region. At each locus,

we considered the index variant for the locus and all variants in at least low LD (r^2>.1). We then used a Bayesian approach to calculate the posterior causal probability (PPA) for each variant, and 'credible sets' of variants explaining 99% of the total PPA (**see Methods, Figure 1.3A, Table S5)**. T1D credible sets contained a median of 66 variants, and 15 loci had 25 or fewer credible set variants. We compared fine-mapping for 34 loci common to our data and Immunochip fine-mapping[19], and found a strong correlation between T1D association for credible set variants (Pearson r=.93). Credible set sizes at these 34 loci were larger in our data than for Immunochip (median=37, Immunochip median=31), likely reflecting increased variant density. We also identified high probability variants not covered in Immunochip credible sets for example at *CTSH* (rs12592898, PPA=.19).

Given fine-mapping of known T1D and T2D signals, we next determined genomic annotations of candidate causal variants at these signals. For each signal, we calculated the cumulative PPA of variants overlapping T1D/T2D enriched annotations including pancreas, adipose, endoderm and immune cell regulatory elements as well as protein-coding exons. We then grouped signals based on the resulting cumulative PPA values for each annotation (**see Methods**). For T1D, signals mapped into distinct groups of immune cell regulatory elements (31 signals), pancreas regulatory elements (6 signals), and coding exons (4 signals) as well as 15 un-annotated signals (**Figure 1.3B**). For T2D, signals also mapped into distinct groups including pancreas regulatory elements (21 signals), adipose regulatory elements (15 signals), and coding exons (4 signals) (**Figure S1.3**). T1D pancreas signals were associated with T2D risk (median -log10(P)=1.37), whereas other T1D groups did not show evidence for T2D association (**Figure 1.3C**). Among T1D signals in the pancreas group were those with known T2D association such as *GLIS3* and *CTRB1*, as well as others with nominal T2D association such as *ERBB3.*

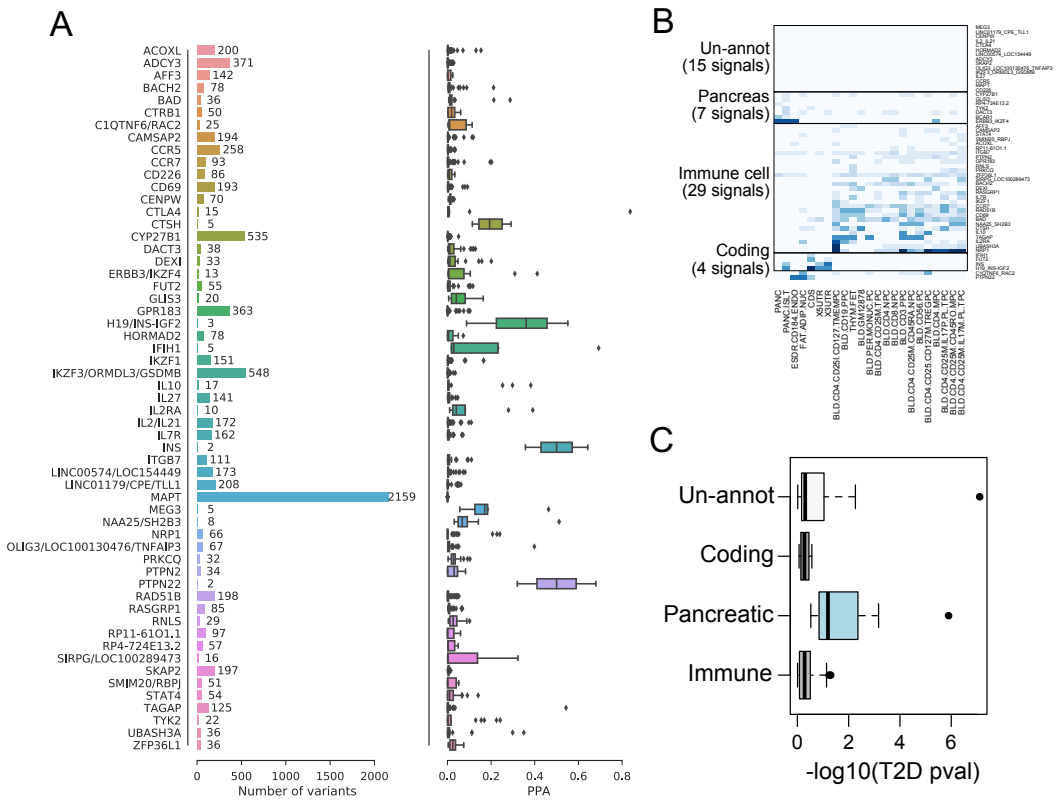


Figure 1.3. Fine-mapping and functional annotation of known T1D loci. (A) Fine-mapping of causal variant sets at 57 known T1D risk signals. (left) number of 99% credible set variants at each locus and (right) causal probabilities (PPA) of credible set variants at each locus. (B) Cumulative PPA values of 57 T1D signals in cell-type regulatory site and coding annotations. T1D signals mapped into four primary groups including immune cell regulatory sites (31 signals), pancreas regulatory sites (6 signals), and coding exons (4 signals). (C) T1D signals within different groups had distinct patterns of association with T2D, where T1D pancreas signals collectively had the strongest evidence for T2D association.

1.3.4 Shared T1D and T2D risk variants at *GLIS3* affect regulatory activity in islets

Several loci have been reported to influence risk of both T1D and T2D, but whether risk signals have shared or distinct causal variants is unknown. We cataloged 146 loci with known association to either form of diabetes and tested for shared causal variants using Bayesian co-localization (**see Methods, Table S6**). There was co-localization of risk signals (P_{shared}>.50) at three known T1D and T2D loci *CENPW* (P_{shared}=.88), *CTRB1* (P_{shared}=.88), and *GLIS3* (P_{shared}=.62) as well as evidence for putative co-localization of signals at two known T2D loci *BCL11A*

(P_{shared}=.73) and *THADA* (P_{shared}=.68) (**Figure 1.4A)**. All shared risk signals except for *CTRB1* had the same direction of effect on T1D and T2D risk. At *RASGRP1*, which has reported association to both T1D and T2D, we found no evidence for either state ($P_{distinct}$=.03, P_{shared}=.02) (**Table S5**). At several loci including *MTMR3* and *ZMIZ1*, there was evidence for two distinct T1D and T2D signals ($P_{distinct}$>.5) (**Figure 1.4A)**. We fine-mapped causal variants at co-localized signals by combining T1D and T2D evidence (**see Methods**). There was a reduction in credible set size at shared signals, including fewer than 10 variants at *GLIS3* (9 vars) and *CTRB1* (8 vars) (**Figure 1.4B, Figure S4, Table S7**). We further confirmed evidence (CLPP>.01) for shared causal variants at the *GLIS3* and *CTRB1* signals using eCAVIAR (**see Methods, Figure S4, Table S7**). Conversely, we did not find evidence for a shared variant at *CENPW* with eCAVIAR; furthermore, the variant with strongest T2D association (rs11759026) had nominal T1D association ($P=6.3x10^{-3}$), suggesting a potentially more complex causal variant structure at this locus.

We sought to understand mechanisms of how signals influence both T1D and T2D risk. We first examined quantitative trait associations at co-localized signals[25,31–33]. At *GLIS3*, risk alleles were associated with increased fasting glucose level (rs10758593 Z=4.51), and decreased HOMA-B (Z=-4.54) and birth weight (Z=-2.27) (**Figure 1.4C**). At *CTRB1*, risk alleles for T2D were nominally associated with higher fasting glucose (rs8056814 Z=2.27) and decreased birth weight (Z=-3.78). At *CENPW*, risk alleles were also nominally associated with higher fasting glucose (rs4565329 Z=2.32) and decreased birth weight (Z=2.97), as well as increased male pubertal height (12M, Z=3.14), height (Z=13), and earlier age of menarche (Z=-8.9). Among putative shared signals, variants at *THADA* were associated with increased fasting glucose level (Z=3.65) and decreased HOMA-B (Z=-4.23). We next examined expression QTL association and co-localization at shared signals in adipose, liver, blood, pancreas and pancreatic islets[34–36] (**see Methods**). Variants at *CENPW* were associated and co-localized with expression level of

CENPW in subcutaneous adipose tissue (rs4565329 $P=8.6x10^{-7}$) and islets ($P=1.9x10^{-3}$). Variants at *CTRB1* were associated and co-localized with *BCAR1* expression in whole blood ($P=2.5x10^{-4}$) and islets ($P=1.9x10^{-3}$); variants at this locus were also associated with *CFDP1* expression in subcutaneous and visceral adipose ($P=7.7x10^{-8}$, $P=7.8x10^{-8}$) and whole blood ($P=9.5x10^{-5}$) but these signals were all statistically distinct from the T1D/T2D signal ($P_{distinct}>.5$).

Multiple shared T1D and T2D signals likely affect beta cell function, and thus we finally annotated variants in islet regulatory sites at these signals. We used accessible chromatin sites merged from ATAC-seq assays in six islet samples[34,37] (**Table S8**), chromatin states created from islet histone modification ChIP-seq data[5,38], islet transcription factor (TF) ChIP-seq sites[5], and TF footprints identified from islet ATAC-seq using CENTIPEDE[34] (**see Methods**). At *GLIS3*, rs4237150 (PPA=.20), rs10116772 (PPA=.15) and rs10814915 (PPA=.007) mapped in islet accessible chromatin and active enhancer elements; rs4237150 further mapped in islet ChIP-seq sites for multiple TFs (**Figure 1.4D, Table S7**). At *CTRB1*, rs8056814 (PPA=.91) also mapped in islet accessible chromatin and an active enhancer (**Figure S5, Table S7**). We tested these variants, and another *GLIS3* variant rs6476839, for effects on islet regulatory activity. We cloned sequence for variant alleles into reporter vectors in forward and reverse orientations, and transfected constructs into the islet cell line MIN6. As rs10116772 and rs10814915 were within 3bp, we cloned these variants in the same construct. At *GLIS3*, there were significant allelic effects on enhancer activity for rs4237150 (Two-sided t-test Fwd $P=1.2x10^{-4;}$ Rev P=.024); we observed weaker evidence for allelic effects in one orientation only for rs10116772+rs10814915 and rs6476839 (**Figure 1.4E).** We further identified evidence for allelic imbalance in islet ChIP-seq reads from samples estimated to be heterozygous for rs4237150, and this variant was in a TF binding footprint for IRX and NR3C1 (**Table S7**). At *CTRB1*, we observed significant allelic effects on repressor activity for rs8056814 (Fwd $P=.017^{;}$ Rev $P=6.7x10^{-4}$; **Figure S5**).

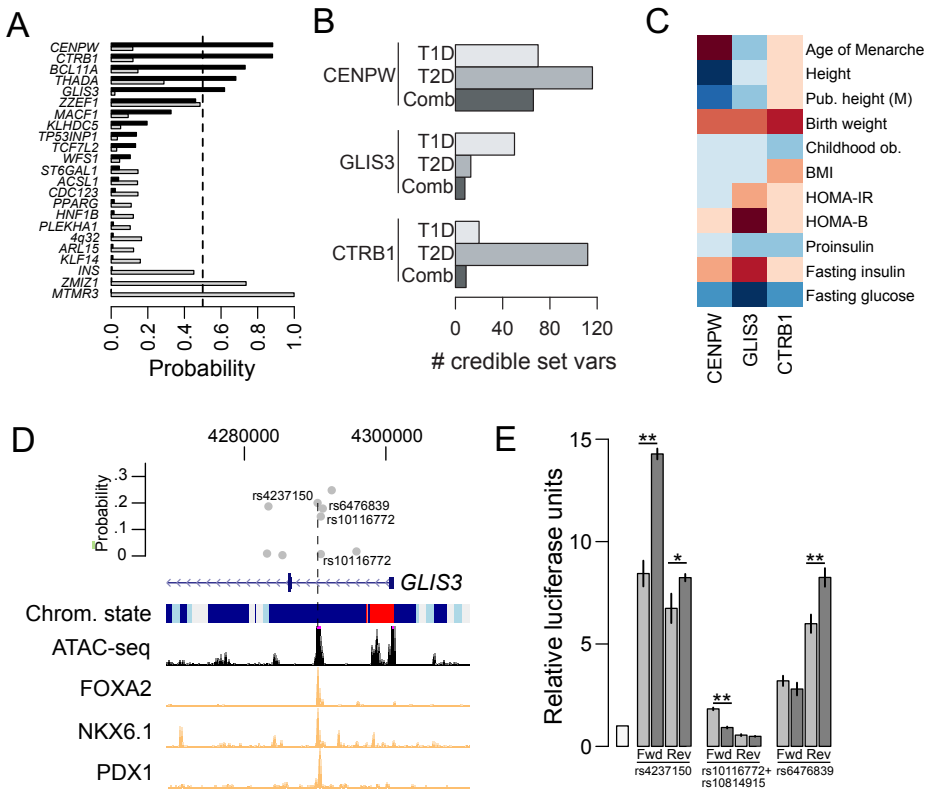


Figure 1.4. Shared T1D and T2D risk variants affect islet regulatory activity. (A) Five loci have evidence for a shared signal (P_{shared}>.50) influencing both T1D and T2D risk, and two have evidence for distinct signals ($P_{distinct}$>.50) (dark grey = P_{shared}, grey = $P_{distinct}$) (C) Number of 99% credible set variants at shared T1D and T2D risk loci. After combining T1D and T2D evidence the *GLIS3* and *CTRB1* signals have <10 variants. (C) Quantitative trait association at shared T1D and T2D signals. Values represent signed z-scores for the risk allele of the most likely causal variant (blue = positive, red = negative). For *CTRB1* z-scores are signed to the T2D risk allele. (D) Shared risk variants rs4237150, rs10116772, and rs10814915 at *GLIS3* are in islet active enhancer and accessible chromatin, and rs4237150 is also in islet TF ChIP-seq (states: dark blue = active enhancer, light blue = weak enhancer, red = active promoter) (E) Variants at *GLIS3* have allelic effects on enhancer activity in islet cells. Values are mean and SD. (N=3; *P<.05, ** P<.01).

1.4 Discussion

Comparison of variant effects on T1D and T2D genome-wide, across known loci, and at individual loci provide evidence for a shared genetic risk underlying the two major forms of diabetes. This shared risk involves variants with effects on islet function and insulin secretion in addition to insulin resistance and development, which are well established contributors to T2D but have not been broadly implicated in genetic risk of T1D. We also found strong enrichment of

T1D association among known T2D risk loci. While previous studies of T1D genetic risk have largely focused on the immune system given that many large effect T1D variants impact immune activity, our findings support an additional role for T2D-relevant processes in the genetic basis of T1D.

A recent study determined that a subset of patients with later-onset T1D are misdiagnosed with T2D[39]. This is unlikely to explain a positive correlation between T1D and T2D given that we observed no enrichment of T2D association or concordance in effect direction among known T1D variants, even among large effect T1D variants, and the correlation remained when using clinically defined T2D in the WTCCC with no T1D relatives, negative anti-GAD, and >1 year from diagnosis to insulin treatment. We also observed no evidence for correlations between T2D and other autoimmune traits, or enrichment of T2D association among T1D-relevant immune cell regulatory elements. Misdiagnosis of T2D as T1D is also an unlikely explanation of the positive correlation as it remains when using clinically defined T1D in the WTCCC with onset <17, insulin treatment from diagnosis for >6 months, and no monogenic diabetes, or when removing obese individuals from T1D cohorts. Furthermore, we found little evidence for directional consistency among largest effect T2D variants.

Reports have argued that islet dysfunction underlies shared etiology of T1D and T2D[11]. Our findings support a role for shared variants at *GLIS3* in islet function, where risk alleles were associated with increased fasting glucose level and decreased beta cell function. In addition, shared risk variants at *GLIS3* had allelic effects on islet enhancer activity and rs4237150 disrupted binding of the glucocorticoid receptor, which is involved in diabetes-relevant inflammatory response[40]. The mechanism of how these regulatory variants influence diabetes risk through *GLIS3* and/or other genes in islets remains to be uncovered. Putative shared risk signals at *THADA* were associated with increased glucose level and decreased beta cell function, in line

with a previous report[41], and variants at *BCL11A* have been reported to affect beta cell function[41]. Candidate genes at these loci are involved in apoptotic and stress-related processes[42(p3),43] and therefore altered activity could contribute to a fragile beta cell phenotype. Genome-wide, T1D and T2D associated variants were enriched in islet regulatory elements and correlated with increased fasting glucose level. Given the role of islet stress response in shared risk, studies mapping the islet epigenome and gene expression in diabetogenic stress conditions will help uncover additional relevant islet regulatory programs.

Shared variants at the *CTRB1* locus have opposite effects on T2D and T1D risk and have allelic effects on islet regulatory activity. A previous report identified correlation between risk variants and *CTRB1/2* expression in pancreas and pancreatic islets, albeit in a limited sample size[12(p1)]. In our study we identified evidence for association between these risk variants and expression of *BCAR1* in islets as well as whole blood, suggesting that the risk effects at this locus might involve multiple genes and cell types. Other loci have evidence for opposite effects on T1D and T2D such as *TCF7L2,* where T2D risk variants affect islet regulatory activity[6], *ZZEF1,* and a recently identified association at *HLA-DRB5*[44]. Heterogeneity in effect direction at specific loci has been observed in other contexts, for example, between T2D and cardiovascular disease and T2D and birth weight[27,45]. We further observed enrichment of nominally associated variants with opposite effects on T1D and T2D in islet regulatory elements, suggesting the potential of a broader role for aspects of pancreatic and islet function in opposed risk of T1D and T2D.

Another shared mechanism of T1D and T2D pathogenesis is through obesity and insulin resistance. The 'accelerator' hypothesis posits that weight gain and insulin resistance exacerbate beta cell stress and T1D progression in a manner similar to T2D pathogenesis[10]. Insulin resistance is also linked to chronic inflammation which is also involved in T1D pathophysiology. We identified support for this hypothesis through a correlation between increased fasting insulin level and T1D

and T2D risk. We also identified enrichment of T1D and T2D variants for adipose and B cell regulatory elements, cell types both involved in insulin resistance. We did not find significant correlation between T1D risk and BMI, or association with large effect obesity loci such as *FTO*. A recent study identified a causal relationship between childhood obesity and T1D risk, supporting a role for adolescent growth in T1D pathogenesis[29], though we did not observe a genome-wide correlation. There was, however, a positive correlation with male pubertal phenotypes, in line with increased prevalence of T1D in males in early adulthood[46], and risk variants at the *CENPW* locus were associated with pubertal phenotypes, height, and age of menarche[32,33]. Risk variants at this locus were also associated and co-localized with expression level of *CENPW* in subcutaneous adipose tissue, which has been implicated in the genetic basis of insulin resistance[47]. This supports a role for insulin resistance and growth in the shared etiology of T1D and T2D.

We also observed evidence for correlations with other traits, such as between increased T1D and T2D risk and decreased birth weight and increased proinsulin level. Previous studies have reported a correlation between low birth weight and increased T2D risk[27,48], although the potential link between birth weight and T1D risk is unclear[49]. Furthermore, variants in endoderm regulatory sites were enriched for T1D and T2D association, suggesting potential shared effects on developmental regulatory processes. Proinsulin is an autoantibody in T1D and higher proinsulin level could contribute to increased risk of developing T1D[50]. Conversely, impaired insulin processing is observed in beta cell dysfunction and thus could also represent a consequence of disease progression[51]. Additional studies will be needed to determine causal relationships between proinsulin level or birth weight and diabetes risk and the direction of these relationships.

In summary, a shared genetic risk contributes to the etiology of T1D and T2D which highlights the involvement of T2D-relevant processes and genetic factors in risk of T1D. These processes likely contribute to increased beta cell stress during T1D progression through both intrinsic islet functions as well as insulin resistance. Future studies will help determine the cellular mechanisms of these effects and provide novel avenues for disease management in particular for T1D.

1.5 Methods

T1D sample collection

For the type 1 diabetes GWAS, we compiled publicly available genotype-level data for case and control samples from the T1DGC (dbGAP: phs000180.v3.p2), GoKIND/GAIN (dbGAP: phs000018.v2.p1), DCCT-EDIC (dbGAP: phs000086.v3.p1), WTCCC1[52], and WTCCC2, which were either genotyped on Affymetrix or Illumina platforms (**Table S1**). Because the GoKIND/GAIN dataset contained family trios, we extracted only the proband samples. From the WTCCC1 samples, we used the T1D cohort as cases and the 1958 Birth Cohort (58BC), UK National Blood Service (NBS), and bipolar disorder (BP) cohorts as controls. Unlike a previous study for T1D[53], we did not include type 2 diabetes or hypertension from WTCCC1 as controls. From the WTCCC2 samples, we used control cohorts from the UK National Blood Service.

T1D quality control and imputation

We used the recommended individual and variant exclusion lists where available for 58BC, NBS, WTCCC1 T1D and BP. We used phenotype files for GoKIND/GAIN and DCCT-EDIC to exclude samples that were not reported of Caucasian ancestry. We used PLINK[54] (https://www.cog-genomics.org/plink2) to perform PCA with 1000 Genomes Project (1KGP)

samples to identify and remove outliers that did not overlap European 1KGP samples on PC1 and PC2. We used PLINK to calculate identity-by-descent (IBD) values between individuals. Pairs of individuals with at least second-degree relationships (IBD>.2) were pruned in a manner such that only one related individual was retained. For the NBS samples that overlapped between Affymetrix and Illumina platforms, we prioritized the samples genotyped on the Illumina platform. For each cohort, we filtered out variants with less than 95% call rate, less than 1% minor allele frequency (MAF), and extreme Hardy-Weinberg equilibrium values ($P<1x10^{-5}$). We also removed individuals with more than 5% missing genotypes. We then combined cohorts that were genotyped on similar platforms. After filtering steps, the total number of individuals available was 15,043, including 8,967 cases and 6,076 controls (**Table S3**). We imputed 347,083 (Affymetrix) and 500,096 (Illumina) autosomal variants separately into the HRC panel r1.1 using the Michigan Imputation Server[55], resulting in data for 39,117,105 variants. We excluded variants after imputation that had an imputation quality (R^2) less than 0.3, leaving 23,385,104 (Affymetrix) and 25,294,976 (Illumina) well-imputed variants.

T1D genome-wide association and meta-analysis

For Affymetrix and Illumina combined cohorts, we used PLINK to LD prune genotyped variants to create a set of independent variants. We then used PLINK to perform principal component analysis (PCA) and extracted the top 3 principal components (PCs). We used the firth bias-corrected logistic likelihood ratio test in EPACTS (https://genome.sph.umich.edu/wiki/EPACTS) to test variants for association to T1D separately for each genotyping platform. We used sex and the top 3 PCs as covariates, set a lower MAF threshold of 0.005, and used genotype dosages for association testing. For tri-allelic SNPs and cases where multiple variants mapped to the same genomic coordinates, we kept the variant with the highest MAF. We then used inverse-variance meta-analysis as implemented in METAL[56] on association results for 8,720,060 (Affymetrix) and 8,778,018 (Illumina) variants, keeping variants

that were tested on both platforms. We further removed genotyped variants that had an empirical R^2 (ER^2<.8) for either cohort and all variants in at least moderate LD (r^2>.5) with these variants. The ER2 is the Pearson correlation coefficient between the genotype of a variant and an imputed genotype derived by masking the genotype in a leave-one-out procedure during imputation. A total of 8,491,085 variants remained for downstream analyses.

To address the potential for misdiagnosed T2D cases in the T1D GWAS, we used phenotype data to remove 278 T1D cases with body-mass index (BMI)>30 from the DCCT and GoKIND/GAIN cohorts. We then re-ran the GWAS meta-analyses using the above methods.

WTCCC T2D genome-wide association

For the WTCCC T2D GWAS, we collected genotype data for a case cohort of T2D, and control cohorts from NBS and 58BC from WTCCC1[52]. We used sample exclusion lists to remove duplicated, related, or samples of non-Caucasian ancestry and variant exclusion lists to remove poorly genotyped variants. Prior to imputation, we also filtered out variants with less than 95% call rate, less than 1% MAF, and extreme Hardy-Weinberg equilibrium values ($P<1x10^{-5}$). We then imputed 412,388 genotyped variants from 1,924 T2D case samples and 2,939 control samples together into the HRC panel r1.1 using the Michigan imputation server. We excluded variants with low imputation quality (R^2<.3) and retained 22,520,888 well-imputed variants. We further filtered out potential artifacts by excluding genotyped variants with ER^2<.8 and all variants in at least moderate LD (r^2>.5) with these variants. We used sex and the top 3 PCs from LD-pruned genotypes as covariates, set a lower MAF threshold of 0.005, and used genotype dosages for association testing with the firth bias-corrected logistic regression.

Genetic enrichment analyses

We tested for enrichment of nominal association and concordance in effects among known T1D and T2D risk loci.

For T2D loci, we collected published credible sets of 49 signals on the Metabochip[18], 41 additional signals in GoT2D,[7] and 17 additional signals in DIAGRAM 1000G[24]. We removed all secondary association signals to retain only the primary signal at each locus. For the 94 resulting primary association signals, we then obtained the variant with the highest posterior probability. Where the most likely causal variant was not present in T1D association data, we used the next most likely causal variant. For each variant, we obtained the p-value for T1D association and direction of T1D effect for the T2D risk allele. We tested for enrichment of variants with nominal association (P<.05) by comparing to the expected percentage obtained from sets of matched variants from SNPsnap[57] using a binomial test.

We then determined concordance in T1D effect direction on T2D variants by calculating the number of variants with same effect direction and applying a binomial test. We further determined the concordance in effect direction in T1D association data in the UK Biobank (ICD10 code E10 from https://sites.google.com/broadinstitute.org/ukbbgwasresults/home) using a binomial test.

For T1D loci, we obtained the variant with the highest posterior probability in fine-mapping of 57 loci described the sections below. Where the top variant was not present in T2D association data we used the next most probable variant. For each variant, we obtained the p-value for T2D association and direction of T2D effect for the T1D risk allele. We tested for enrichment of nominal association (P<.05) by comparing to the expected percentage obtained from sets of matched variants from SNPsnap[57] using a binomial test.

We then determined concordance in T2D effect direction on T1D variants by calculating the number of variants with same effect direction and applying a binomial test.

Genetic correlation analyses

We tested for genetic correlation between T1D and T2D, related glycemic and anthropometric traits, and autoimmune diseases using LD score regression[15,58].

We collected quantitative trait data for fasting insulin level, fasting glucose level, HOMA-B, HOMA-IR, HbA1C, and proinsulin level from the MAGIC consortium[28,31,59], body-mass index (BMI) from the GIANT consortium[60], and pubertal height (12M, 10F), birth weight and childhood obesity from the EGG consortium[27,61]. For the UK Biobank, we obtained summary statistic data of 337k samples using T1D and T2D phenotypes defined from ICD10 codes E10 (T1D) and E11 (T2D) available at sites.google.com/broadinstitute.org/ukbbgwasresults/home. For T2D we obtained data from the GoT2D, HapMap2, and trans-ethnic GWAS studies from the DIAGRAM consortium website. We obtained summary statistic data for autoimmune traits including systemic lupus erythematosus[62], primary biliary cirrhosis[63], Crohn's disease[64], ulcerative colitis[64], inflammatory bowel disease[64], celiac disease[65], primary sclerosing cholangitis[66], autoimmune vitiligo[67], and rheumatoid arthritis[68] from Immunobase and the NHGRI-EBI GWAS catalog.

For each trait, we formatted summary statistics for variants in HapMap3 in order to retain well-imputed variants and to correctly orient variant alleles. We then ran LD score regression on the resulting formatted files using default LD scores.

We repeated the LD score regression analysis for T1D and T2D after removing variants in 1MB windows around 57 known T1D loci and removing variants in 1MB windows around 94 known T2D loci.

Genomic enrichment analyses

We considered active enhancer and promoter site annotations for 98 cell types from the Epigenome Roadmap project[21], along with annotations for coding exons from GENCODE[30]. We used stratified LD-score regression[16] to identify annotations that were enriched for signal in T1D and T2D association data. Stratified LD-score regression is a multiple regression, where the chi-squared statistics for a trait are regressed on LD-scores computed using variants from each of a set of functional annotations, and the estimated parameters quantify the relative contribution of each annotation to the total heritability.

For the five cell-types with positive enrichment for both T1D and T2D association (pancreatic islets, pancreas, adipose, CD19+ B cells, and CD184+ endoderm), we tested whether these annotations were enriched in variants with shared or opposite effects on T1D and T2D. We identified variants with P<.05 for both T1D and T2D association and in 1000 Genomes phase 3 data. For each of these variants i, we computed $z_{i,T1D} = \beta_{i,T1D} / SE_{i,T1D}$ and $z_{i,T2D} = \beta_{i,T2D} / SE_{i,T2D}$. We sorted them by the value of $| z_{i,T1D} + z_{i,T2D} |$ for LD-pruning purposes. After sorting, we pruned these variants using the SNPclip tool of LDlink[54] using EUR populations, a R^2>0.1 and MAF>0.01, resulting in 3856 and 2254 independent shared and opposite variants, respectively. We then generated sets of randomized, matched SNPs using SNPsnap[55]. We tested shared and opposite variants for enriched overlap compared to the average overlap across matched variant sets using a one-sided Fisher exact test.

Fine-mapping of causal variant sets

We used effect and standard error estimates to calculate a Bayes Factor[69] for each variant. We obtained 58 known loci for T1D from Immunobase and excluded the MHC locus (**Table S2**). We extracted the previously reported index variants and used PLINK to calculate r^2 values between 57 index variants and all common variants (MAF>.5) within a 5 MB window as done in a previous study[7]. We defined credible sets of variants for each locus as variants with r^2>.1 with the index variant. For each locus, we calculated the posterior probability of association (PPA) for each variant by dividing the Bayes Factor for each variant by the sum of Bayes Factors for the entire locus. We then calculated the 99% credible set by taking the set of variants for each locus that added up to 99% PPA. We compared our T1D credible sets to previously published Immunochip credible sets[19] by extracting 34 common loci between both studies. From the Immunochip study, we extracted only the primary signals. To directly compare p-values, we filtered for variants covered by both studies with non-missing p-values and calculated the Pearson correlation. To identify high probability variants not in Immunochip credible sets, we extracted variants from the 34 loci that were not in the Immunochip primary signal credible set and sorted by PPA.

Genomic annotations at fine-mapped signals

We considered active regulatory site annotations for cell-types enriched for T1D/T2D association along with annotations for coding exons and UTR regions from GENCODE[30]. For T1D we used fine-mapping data from 57 signals as described above. For T2D, we compiled publicly available fine-mapping credible sets based on publications up to 2017 for 94 primary signals from Metabochip, GoT2D and DIAGRAM 1000G studies. At each signal, we calculated a cumulative posterior causal probability (PPA) for each annotation as the sum of PPA values for variants overlapping that annotation. We then assigned T1D/T2D signals to groups based on the highest cumulative PPA value across annotations, considering signals with a cumulative PPA value less than .1 for all annotations as 'un-annotated'. For each T1D group we then calculated

the median association of signals in the group with T2D, and for each T2D group we calculated the median association with T1D.

Risk signal co-localization

We used a Bayesian co-localization method to determine loci at which T1D and T2D association data showed evidence of a causal variant shared by both traits[70]. At a given locus, the method takes as inputs Bayes Factors of association from two datasets and a specification of the prior probability that each is causal for one or both traits. From these a posterior probability (PP) is computed for each of five hypotheses:

H0: The locus contains no variant causal for either trait

H1: The locus contains a variant causal for trait 1 but none causal for trait 2

H2: The locus contains a variant causal for trait 2 but none for trait 1

H3 ($P_{distinct}$): The locus contains two distinct causal variants for trait 1 and trait 2

H4 (P_{shared}): The locus contains a variant causal for both trait 1 and trait 2

We used the default prior assumption that all variants at a locus are equally likely to be causal. This model has two important limitations: It assumes each locus has at most one causal variant, and the distinction between H3 and H4 may be confounded by cases of high LD. We considered the prior probability that a variant is associated with T1D or T2D as $1x10^{-4}$ and the prior probability that a variant is associated with both traits as $1x10^{-5}$.

We collected 94 T2D loci and 57 T1D loci, of which five have overlapping coordinates (*CENPW, GLIS3, RASGRP1, CTRB1, MTMR3*), for a total of 146 loci (**Table S3**). At each locus, we obtained a reported index variant and then extracted all variants in a 500kb window. For each variant, we calculated a Bayes Factor for T1D and T2D separately using the approach of

Wakefield[69]. We then applied the co-localization test to compare T1D and T2D Bayes Factors, and considered loci with H4 > .50 as shared. For loci with evidence for a shared risk variant, we then fine-mapped variants causal for the shared signal. For each locus, we multiplied T1D and T2D Bayes Factors at each variant, and then calculated the posterior causal probability (PPA) as the Bayes Factor divided by the sum of all variant Bayes Factors across the locus. We further calculated a cumulative PPA (cPPA) as the sum of PPA values for variants overlapping an annotation at a given locus.

To validate loci with evidence for a shared causal variant we further applied eCAVIAR, a co-localization method capable of modeling multiple causal variants[71]. For each locus, we chose a window of 100 variants on either side of the variant with the strongest combined T1D and T2D evidence. We provided Z-scores of T1D and T2D association together with pairwise LD statistics of European samples in 1000 Genomes Project v3 data for all variants within the window to eCAVIAR using default settings. For each variant in the window, eCAVIAR computed a co-localization posterior probability (CLPP), the probability that the variant is causal for the local signal in both traits. We considered loci to be co-localized using this approach with at least one variant with CLPP > 0.01 as recommended in the original study.

For quantitative trait association at shared risk variants, we obtained the most likely causal variant from combined T1D and T2D evidence. We extracted summary statistics for each trait and calculated a signed Z-score for the risk allele using effect size and standard error estimates.

Expression QTL association

We obtained gene expression QTL data for subcutaneous adipose, visceral adipose, liver, whole blood and liver from GTEx version 7[36]. For pancreatic islets, we meta-analyzed summary statistic results from two published studies[34,35] using METAL[56] and retained only variants test in

both studies and with minor allele frequency > .01. Using the most probable variant at each shared T1D/T2D signal we obtained the eQTL p-value for each protein-coding gene in a 1MB window around the variant. We considered genes with eQTL P<.005 to be associated with shared T1D/T2D variants. We then used Bayesian co-localization to determine whether the associated gene eQTL signals had evidence for being co-localized with the T1D/T2D signals, and considered eQTL signals shared where the shared probability (P_{shared}) was greater than the distinct probability ($P_{distinct}$).

Islet ATAC-seq and chromatin states

We utilized ATAC-seq data generated from four primary pancreatic islet samples as described in a separate study[72]. For each sample, we trimmed adaptor sequences from the reads with trim_galore (https://github.com/FelixKrueger/TrimGalore). The resulting sequences were aligned to sex-specific hg19 reference genomes using bwa mem[73]. We filtered reads to retain those in proper pairs and with mapping quality score greater than 30. We then removed duplicate and non-autosomal reads. We called sites individually for each sample with MACS2[74] at a q-value threshold of .05 with the following options "—no-model", "—shift -100", "—extsize 200". We removed sites that overlapped genomic regions blacklisted by the ENCODE consortium[20]. We merged sites from these 4 samples and two previously generated in islets[34] with bedtools[75] to obtain a comprehensive set of ATAC-seq peaks in human islets.

We used islet chromatin states described separately[34]. In brief, we used previously published data[5,38] from ChIP-seq assays generated in islets and for which there was matching input sequence from the same sample. For each assay and input, we aligned reads to the human genome hg19 using bwa samse and bwa aln[73] with a flag to trim reads at a quality threshold of less than 15. We converted the alignments to bam format and sorted the bam files. We then removed duplicate reads, and further filtered reads that had a mapping quality score below 30.

Sequence data from the same assay in the same sample were then pooled. We defined chromatin states from ChIP-seq data using ChromHMM[76] with a 9 state model. We assigned the resulting states names based on the resulting patterns.

ATAC-seq footprint analysis

To identify haplotype-aware motifs within ATAC-seq footprints overlapping accessible chromatin sites, we searched accessible chromatin sites from four ATAC-seq samples for instances of motifs from JASPAR, SELEX, ENCODE and *de novo* motifs identified in our data[77]. We used vcf2diploid[78] (https://github.com/abyzovlab/vcf2diploid) to create individual-specific diploid genomes by mapping our phased, imputed genotypes onto hg19 using only SNPs and ignoring indels. Then, we used fimo[79] to scan the personalized genomes for our compiled database of motifs, limiting the sequences scanned to those derived from islet accessible chromatin. For fimo, we used the default parameters for p-value threshold ($1x10^{-4}$) and a background GC content of 40.9% based on hg19.

CENTIPEDE[80] was used to discover footprint sites for each motif, using the discovered motif instances within ATAC-seq peaks. For each motif, we used the make_cut_matrix utility from atactk (https://github.com/ParkerLab/atactk) to calculate a cut-site matrix that contained counts of the number of Tn5 integrations within a window defined by ±100 bp from each motif occurrence for both forward and reverse strands. This cut-site matrix was provided as input to CENTIPEDE along with regions for each motif occurrence to model the posterior probability that a given motif occurrence was bound by a TF. We defined footprints for a given motif as regions that had a posterior probability ≥ 0.99. We combined footprints from our samples with a previously published set of footprints in pancreatic islets[34].

We further identified variants predicted to disrupt each footprint[24]. We calculated the entropy score for a variant position in a footprint using the position frequency matrix for each motif. For each base at a given position bp and the frequency of the base at that position f, we calculated the entropy as:

$$Entropy = \sum_{bp} f(bp) \times \log_2 f(bp).$$

A footprint was considered disrupted if a variant fell in a conserved position in the motif (Entropy<1.0).

Luciferase reporter assays

To test for allelic differences in enhancer activity at rs4237150, rs10116772 and rs8056814, we cloned sequences containing the alt or ref allele in forward and reverse orientation upstream of the minimal promoter of firefly luciferase vector pGL4.23 (Promega) using KpnI and SacI restriction sites.

Primer sequences were:

rs4237150

Fwd: TTACGCGGTACCACACACTTCTGTAAATCAGGTCAG,
TCATAGGAGCTCGAAGCAGTTTGTTTGCTGGC

Rev: TTACGCGAGCTCACACACTTCTGTAAATCAGGTCAG,
TCATAGGGTACCGAAGCAGTTTGTTTGCTGGC

rs6476839

Fwd: GTCGGTACCTCGCAATTCAATCAAGGACA,
GCTGAGCTCCAGGCACATGTTTGCACTTT

Rev: GTCGAGTCGTCGCAATTCAATCAAGGACA,
GCTGGTACCCAGGCACATGTTTGCACTTT

rs10116772+rs10814915

Fwd: GTCGGTACCTTCATTAATGCCGCCTTTTC,

GCTGAGCTCTGAATTGCGAAATGTGCTTC

Rev: GTCGAGTCGTTCATTAATGCCGCCTTTTC,

GCTGGTACCTGAATTGCGAAATGTGCTTC

rs8056814

Fwd: TAAGCAGGTACCTGGGTGACAGAGTGAGACTCC,

TGCTTAGAGCTCGGTGTTTCCGCCTAACACTG

Rev: TAAGCAGAGCTCTGGGTGACAGAGTGAGACTCC,

TGCTTAGGTACCGGTGTTTCCGCCTAACACTG

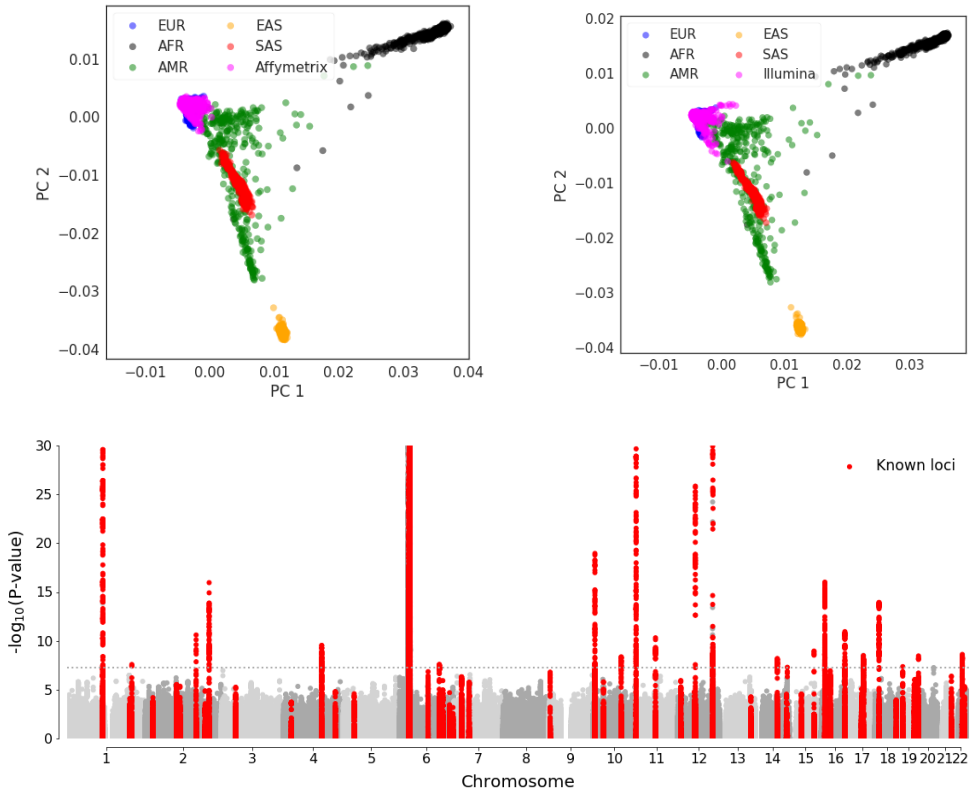
MIN6 beta cells were seeded into 6 (or 12)-well trays at 1 million cells per well. At 80% confluency, cells were co-transfected with 400ng of the experimental firefly luciferase vector pGL4.23 containing the alt or ref allele in either orientation or an empty vector and 50ng of the vector pRL-SV40 (Promega) using the Lipofectamine 3000 reagent. All transfections were done in triplicate. Cells were lysed 48 hours after transfection and assayed for Firefly and Renilla luciferase activities using the Dual-Luciferase Reporter system (Promega). Firefly activity was normalized to Renilla activity and compared to the empty vector and normalized results were expressed as fold change compared to empty vector control per allele. A two-sided t-test was used to compare the luciferase activity between the two alleles in each orientation.

Allelic imbalance analysis

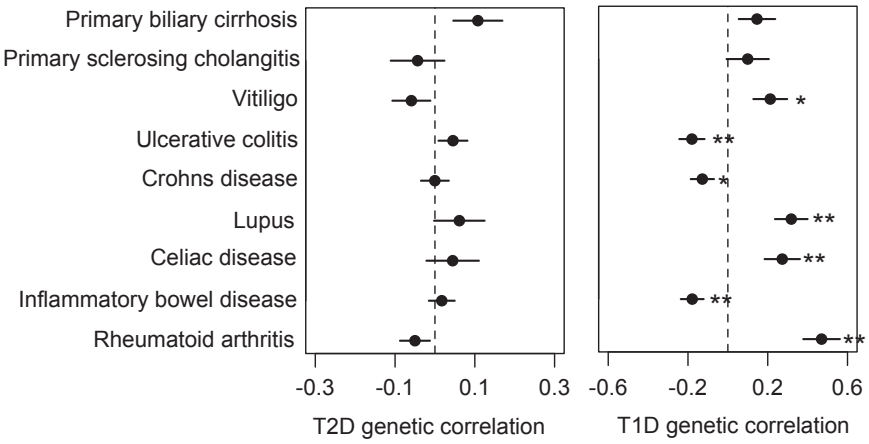
We collected ChIP-seq data from assays in primary islet cells from multiple sources[5,38,81–84]. We aligned sequence data using bwa samse[73], filtered out mitochondrial reads, and removed duplicates using Picard software. For each sample we applied QuASAR[85] to obtain estimated genotypes. A total of 6 samples were determined to be heterozygous at rs4237150 with probability

of being homozygous < 10^{-4}. For these samples we also inferred heterozygosity at rs10116772, due to high linkage and by imputation into 1000 Genomes v3 variants via the Michigan Imputation Server[55]. Across these 6 samples, a total of 8 datasets had more than 5 reads overlapping rs4237150 – FOXA2 (1), H3K27ac (3), PDX1 (2), NKX6-1 (2). We applied WASP[86] to each dataset to correct for reference mapping bias. We then pooled read counts for risk and protective alleles at rs4237150 and rs10116772 and applied a two-sided binomial test for allelic imbalance.

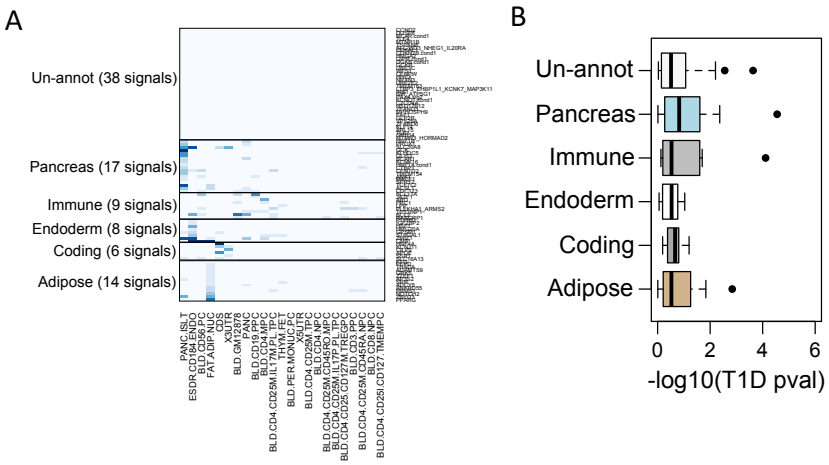
1.6 Supplementary Figures



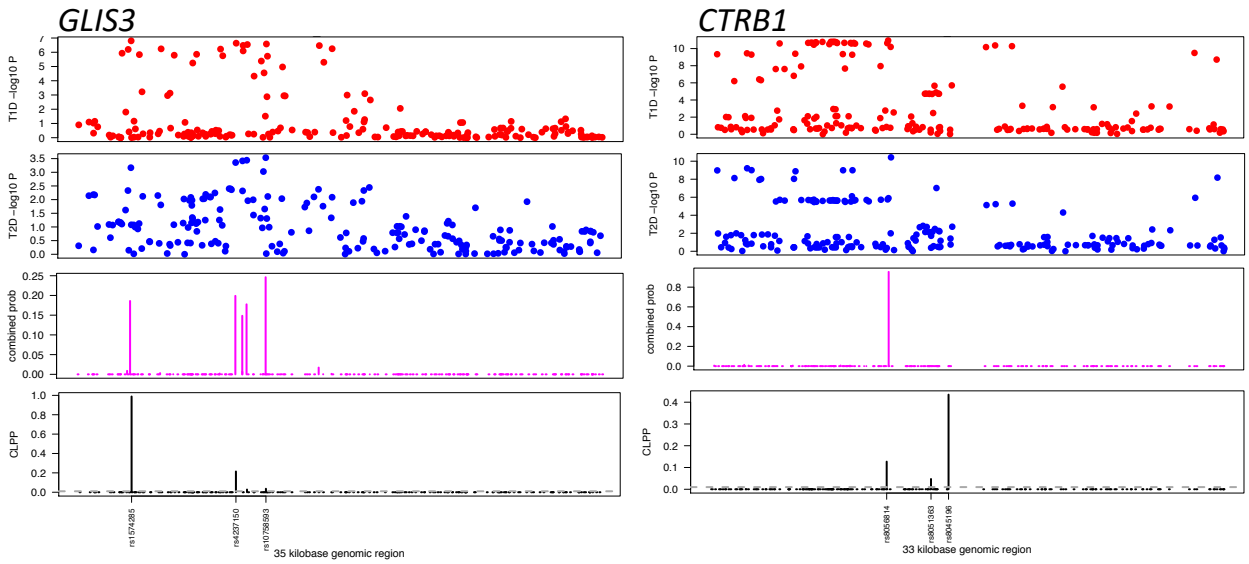
Supplementary Figure S1.1. Genome-wide association study of T1D case and control samples. (A) Principal component plots showing the ancestry of samples genotyped on Affymetrix and Illumina arrays as compared to the super populations of the 1000 Genomes Project after QC measures. EUR = European, AFR = African, AMR = Americas, EAS = East Asian, and SAS = South Asian. (B) Manhattan plot plotting chromosomal positions (hg19) and the negative log10- P-values, with known T1D loci highlighted in red.



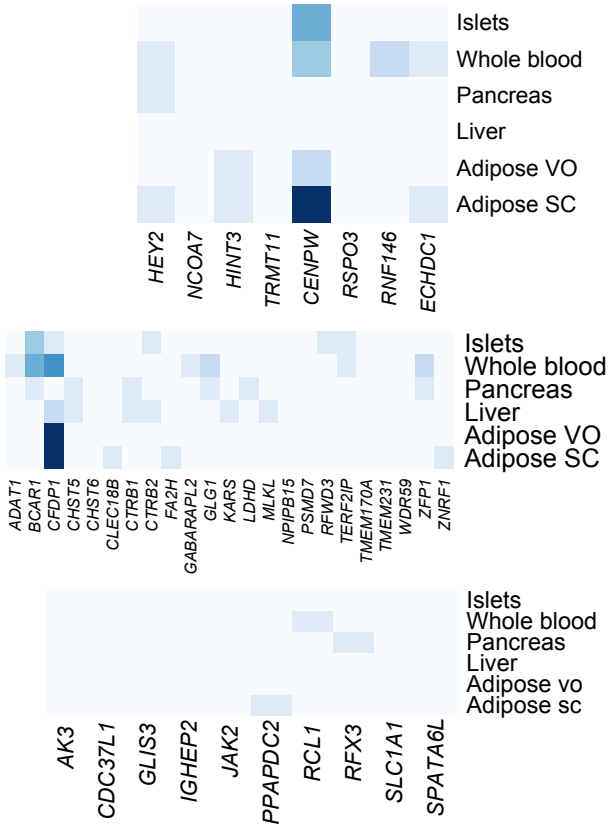
Supplementary Figure S1.2. T1D and T2D genetic correlation with autoimmune traits. Genetic correlations between T2D (left) and T1D (right) and autoimmune traits using LD score regression. T1D has significant correlations with multiple autoimmune traits including rheumatoid arthritis, inflammatory bowel disease, and celiac disease whereas T2D has no significant correlation with any trait. **P<.005, *P<.05.



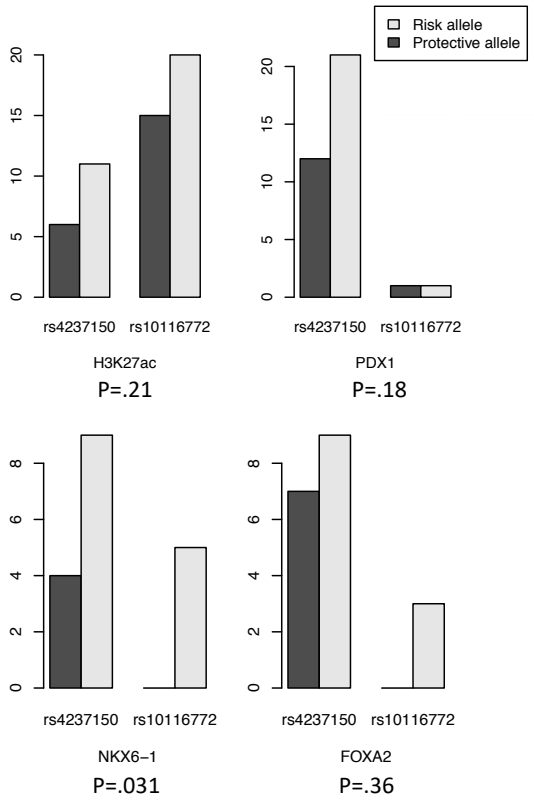
Supplementary Figure S1.3. Genomic annotations and T1D association and fine-mapped T2D loci. (A) Cumulative PPA values of 94 primary T2D signals in cell-type regulatory site and coding annotations. T2D signals mapped into six groups including pancreatic regulatory sites (21 signals), adipose regulatory sites (15 signals), and coding exons (4 signals) in addition to un-annotated signals. (B) T2D signals within different groups had distinct patterns of association with T1D, where T2D pancreas signals had the strongest T1D association.



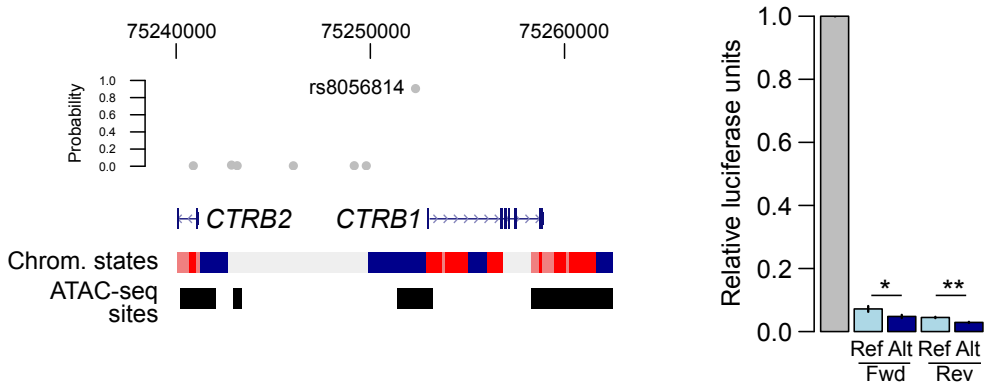
Supplementary Figure S1.4. Shared T1D and T2D signals at the *GLIS3* and *CTRB1* loci. (top) P-values of variant associations with T1D (red) and T2D (blue) at the *GLIS3* and *CTRB1* loci. Causal probability of variants at the shared *GLIS3* and *CTRB1* signals by (middle) combining T1D and T2D evidence in Bayesian fine-mapping, and (bottom) modeling shared causal variants using eCAVIAR. Variants at each signal have high causal probabilities in both analyses.



Supplementary Figure S1.5. Expression QTL association at shared T1D and T2D signals. Expression QTL (eQTL) association data for the most likely causal variant at shared T1D/T2D signals *CENPW, CTRB1,* and *GLIS3* for subcutaneous adipose (adipose sc), visceral adipose (adipose vo), liver, whole blood, pancreas and pancreatic islets. Values are the -log10(P) for eQTL association.



Supplementary Figure S1.6. Allelic imbalance in islet regulatory activity at *GLIS3.* Read counts in samples heterozygote for rs4237150 and rs10116772 in pancreatic islet FOXA1, NKX6.1, PDX1 and H3K27ac ChIP-seq assays (risk allele counts = light grey, protective allele = dark grey). The risk allele had increased read counts in all assays. P-values for binomial tests are listed below each assay.



Supplementary Figure S1.7. Shared variant at *CTRB1* affects islet regulatory activity. (A) Plot of candidate causal variants at the shared *CTRB1* signal. Variant rs8056814 has a high probability (PPA=.90) of being causal for T1D and T2D risk, and maps in an islet accessible chromatin site and an islet active enhancer upstream of *CTRB1.* (B) Luciferase reporter assay of sequence surrounding rs8056814 alleles in the islet cell line MIN6. All constructs had reduced activity compared to the empty vector. The T2D risk allele of rs8056814 has increased activity compared to the T2D protective allele. Values are fold change and SD. (N=3; *P<.05, **P<.001).

1.7 Data availability

Supplementary datasets are available from the online version of this article: https://doi.org/10.1093/hmg/ddy314

1.8 Acknowledgements

The guarantor of this study is KJG. This work in this manuscript supported in part by NIH/NIDDK award DK112155 and ADA award 1-17-JDF-027 to KJG. 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 from the GoKinD study were supplied by the

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

Chapter 1, in full, is a reformatted reprint of material as it appears in Aylward A, Chiou J, Okino M-L, Kadakia N, Gaulton KJ. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. Hum Mol Genet. 2018. The dissertation author was a primary investigator and author of this paper.

1.9 Author Information

Anthony Aylward[#,1], Joshua Chiou[#,2], Mei-Lin Okino[3], Nikita Kadakia[3], Kyle J Gaulton[*,3]

1. Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
2. Biomedical Sciences Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
3. Department of Pediatrics, University of California San Diego, La Jolla CA

# These authors contributed equally to this work

* Corresponding author:

Kyle Gaulton

1059 Cellular and Molecular Medicine East

UC San Diego

San Diego CA 92091

kgaulton@ucsd.edu

K.J.G. designed the study**;** K.J.G, A.J.A. and J.C**.** wrote the manuscript and performed genetic and genomic analyses; M.O. and N.K. performed experiments and contributed to analyses.

1.10 References

1. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*. 2006;3(11):e442. doi:10.1371/journal.pmed.0030442

2. Atkinson MA. The Pathogenesis and Natural History of Type 1 Diabetes. *Cold Spring Harb Perspect Med*. 2012;2(11). doi:10.1101/cshperspect.a007641

3. Pociot F. Type 1 diabetes genome-wide association studies: not to be lost in translation. *Clin Transl Immunology*. 2017;6(12):e162. doi:10.1038/cti.2017.51

4. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Mägi R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss M, Prins BP, Guo X, Bielak LF, Consortium D, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Eckardt K-U, Fischer K, Kardia SL, Kronenberg F, Läll K, Liu C-T, Locke AE, Luan J 'an, Ntalla I, Nylander V, Schönherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jørgensen ME, Jørgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stančáková A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, Grallert H, Metspalu A, Dehghan A, Köttgen A, Abecasis G, Meigs JB, Rotter JI, Marchini J, et al. Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *bioRxiv*. Published online January 9, 2018:245506. doi:10.1101/245506

5. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Morán I, Gómez-Marín C, van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T, Cebola I, García-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Gómez-Skarmeta JL, Müller F, McCarthy MI, Ferrer J. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet*. 2014;46(2):136-143. doi:10.1038/ng.2870

6. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, Berney T, Montanya E, Mohlke KL, Lieb JD, Ferrer J. A map of open chromatin in human pancreatic islets. *Nat Genet*. 2010;42(3):255-259. doi:10.1038/ng.530

7. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, Rivas MA, Perry JRB, Sim X, Blackwell TW, Robertson NR, Rayner NW, Cingolani P, Locke AE, Fernandez Tajes J, Highland HM, Dupuis J, Chines PS, Lindgren CM, Hartl C, Jackson AU, Chen H, Huyghe JR, van de Bunt M, Pearson RD, Kumar A, Müller-Nurasyid M, Grarup N, Stringham HM, Gamazon ER, Lee J, Chen Y, Scott RA, Below JE, Chen P, Huang J, Go MJ, Stitzel ML, Pasko D, Parker SCJ, Varga TV, Green T, Beer NL, Day-Williams AG, Ferreira T, Fingerlin T, Horikoshi M, Hu C, Huh I, Ikram MK, Kim B-J, Kim Y, Kim YJ, Kwon M-S, Lee J, Lee S, Lin K-H, Maxwell TJ, Nagai Y, Wang X, Welch RP, Yoon J, Zhang W, Barzilai N, Voight BF, Han B-G, Jenkinson CP, Kuulasmaa T, Kuusisto J, Manning A, Ng MCY, Palmer ND, Balkau B, Stancáková A, Abboud HE, Boeing H, Giedraitis V, Prabhakaran D, Gottesman O, Scott J, Carey J, Kwan P, Grant G, Smith JD, Neale BM, Purcell S, Butterworth AS, Howson JMM, Lee HM, Lu Y, Kwak S-H, Zhao W, Danesh J, Lam VKL, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536(7614):41-47. doi:10.1038/nature18642

8. Claussnitzer M, Dankel SN, 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 MJ, Mejhert N, Dahlman I, Hansson O, Hauck SM, 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*. 2014;156(1-2):343-358. doi:10.1016/j.cell.2013.10.058

9. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puviindran V, Abdennur NA, Liu J, Svensson P-A, Hsu Y-H, Drucker DJ, Mellgren G, Hui C-C, Hauner H, Kellis M. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *New England Journal of Medicine*. 2015;373(10):895-907. doi:10.1056/NEJMoa1502214

10. Wilkin TJ. The accelerator hypothesis: weight gain as the missing link between Type I and Type II diabetes. *Diabetologia*. 2001;44(7):914-922. doi:10.1007/s001250100548

11. Dooley J, Tian L, Schonefeldt S, Delghingaro-Augusto V, Garcia-Perez JE, Pasciuto E, Di Marino D, Carr EJ, Oskolkov N, Lyssenko V, Franckaert D, Lagou V, Overbergh L, Vandenbussche J, Allemeersch J, Chabot-Roy G, Dahlstrom JE, Laybutt DR, Petrovsky N, Socha L, Gevaert K, Jetten AM, Lambrechts D, Linterman MA, Goodnow CC, Nolan CJ, Lesage S, Schlenner SM, Liston A. Genetic predisposition for beta cell fragility underlies type 1 and type 2 diabetes. *Nat Genet*. 2016;48(5):519-527. doi:10.1038/ng.3531

12. 't Hart LM, Fritsche A, Nijpels G, van Leeuwen N, Donnelly LA, Dekker JM, Alssema M, Fadista J, Carlotti F, Gjesing AP, Palmer CNA, van Haeften TW, Herzberg-Schäfer SA, Simonis-Bik AMC, Houwing-Duistermaat JJ, Helmer Q, Deelen J, Guigas B, Hansen T, Machicao F, Willemsen G, Heine RJ, Kramer MHH, Holst JJ, de Koning EJP, Häring H-U, Pedersen O, Groop L, de Geus EJC, Slagboom PE, Boomsma DI, Eekhoff EMW, Pearson ER, Diamant M. The CTRB1/2 locus affects diabetes susceptibility and treatment via the incretin pathway. *Diabetes*. 2013;62(9):3275-3281. doi:10.2337/db13-0227

13. Senée V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec J-C, Charon C, Nicolino M, Boileau P, Cavener DR, Bougnères P, Taha D, Julier C. Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet*. 2006;38(6):682-687. doi:10.1038/ng1802

14. Liston A, Todd JA, Lagou V. Beta-Cell Fragility As a Common Underlying Risk Factor in Type 1 and Type 2 Diabetes. *Trends Mol Med*. 2017;23(2):181-194. doi:10.1016/j.molmed.2016.12.005

15. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211

16. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, Anttila V, Xu H, Zang C, Farh K, Ripke S, Day FR, ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, RACI Consortium, Purcell S, Stahl E, Lindstrom S, Perry JRB, Okada Y, Raychaudhuri S, Daly MJ, Patterson N, Neale BM, Price AL. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235. doi:10.1038/ng.3404

17. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. *Nature Reviews Genetics*. 2017;18(2):117-127. doi:10.1038/nrg.2016.142

18. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, Mahajan A, Locke A, Rayner NW, Robertson N, Scott RA, Prokopenko I, Scott LJ, Green T, Sparso T, Thuillier D, Yengo L, Grallert H, Wahl S, Frånberg M, Strawbridge RJ, Kestler H, Chheda H, Eisele L,

Gustafsson S, Steinthorsdottir V, Thorleifsson G, Qi L, Karssen LC, van Leeuwen EM, Willems SM, Li M, Chen H, Fuchsberger C, Kwan P, Ma C, Linderman M, Lu Y, Thomsen SK, Rundle JK, Beer NL, van de Bunt M, Chalisey A, Kang HM, Voight BF, Abecasis GR, Almgren P, Baldassarre D, Balkau B, Benediktsson R, Blüher M, Boeing H, Bonnycastle LL, Bottinger EP, Burtt NP, Carey J, Charpentier G, Chines PS, Cornelis MC, Couper DJ, Crenshaw AT, van Dam RM, Doney ASF, Dorkhan M, Edkins S, Eriksson JG, Esko T, Eury E, Fadista J, Flannick J, Fontanillas P, Fox C, Franks PW, Gertow K, Gieger C, Gigante B, Gottesman O, Grant GB, Grarup N, Groves CJ, Hassinen M, Have CT, Herder C, Holmen OL, Hreidarsson AB, Humphries SE, Hunter DJ, Jackson AU, Jonsson A, Jørgensen ME, Jørgensen T, Kao W-HL, Kerrison ND, Kinnunen L, Klopp N, Kong A, Kovacs P, Kraft P, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet*. 2015;47(12):1415-1425. doi:10.1038/ng.3437

19. Onengut-Gumuscu S, Chen W-M, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Farber E, Bonnie JK, Szpak M, Schofield E, Achuthan P, Guo H, Fortune MD, Stevens H, Walker NM, Ward LD, Kundaje A, Kellis M, Daly MJ, Barrett JC, Cooper JD, Deloukas P, Type 1 Diabetes Genetics Consortium, Todd JA, Wallace C, Concannon P, Rich SS. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet*. 2015;47(4):381-386. doi:10.1038/ng.3245

20. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247

21. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh K-H, Feizi S, Karlic R, Kim A-R, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJM, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai L-H, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317-330. doi:10.1038/nature14248

22. Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, Rosen ED. Comparative epigenomic analysis of murine and human adipogenesis. *Cell*. 2010;143(1):156-169. doi:10.1016/j.cell.2010.09.006

23. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett J, Boomsma DI, Branham K, Breen G, Brummet C, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin L, Davey Smith G, Dedoussis G, Dorr M, Farmaki A-E, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee J, McGue M, Meitinger T, Melzer D, Min J, Mohlke KL, 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 PE, 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 MG, Wilson JF, Frayling T, de Bakker P, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

24. Scott RA, Scott LJ, Mägi R, Marullo L, Gaulton KJ, Kaakinen M, Pervjakova N, Pers TH, Johnson AD, Eicher JD, Jackson AU, Ferreira T, Lee Y, Ma C, Steinthorsdottir V, Thorleifsson G, Qi L, Van Zuydam NR, Mahajan A, Chen H, Almgren P, Voight BF, Grallert H, Müller-Nurasyid M, Ried JS, Rayner WN, Robertson N, Karssen LC, van Leeuwen EM, Willems SM, Fuchsberger C, Kwan P, Teslovich TM, Chanda P, Li M, Lu Y, Dina C, Thuillier D, Yengo L, Jiang L, Sparso T, Kestler HA, Chheda H, Eisele L, Gustafsson S, Frånberg M, Strawbridge RJ, Benediktsson R, Hreidarsson AB, Kong A, Sigurðsson G, Kerrison ND, Luan J 'an, 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 DJ, Pankow JS, Grarup N, Have CT, Jørgensen ME, Jørgensen T, Linneberg A, Cornelis MC, van Dam RM, Hunter DJ, Kraft P, Sun Q, Edkins S, Owen KR, Perry JR, Wood AR, Zeggini E, Tajes-Fernandes J, Abecasis GR, Bonnycastle LL, Chines PS, Stringham HM, Koistinen HA, Kinnunen L, Sennblad B, Mühleisen TW, Nöthen MM, Pechlivanis S, Baldassarre D, Gertow K, Humphries SE, Tremoli E, Klopp N, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes*. Published online May 31, 2017. doi:10.2337/db16-1253

25. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu C-T, Bielak LF, Prokopenko I, Amin N, Barnes D, Cadby G, Hottenga J-J, Ingelsson E, Jackson AU, Johnson T, Kanoni S, Ladenvall C, Lagou V, Lahti J, Lecoeur C, Liu Y, Martinez-Larrad MT, Montasser ME, Navarro P, Perry JRB, Rasmussen-Torvik LJ, Salo P, Sattar N, Shungin D, Strawbridge RJ, Tanaka T, van Duijn CM, An P, de Andrade M, Andrews JS, Aspelund T, Atalay M, Aulchenko Y, Balkau B, Bandinelli S, Beckmann JS, Beilby JP, Bellis C, Bergman RN, Blangero J, Boban M, Boehnke M, Boerwinkle E, Bonnycastle LL, Boomsma DI, Borecki IB, Böttcher Y, Bouchard C, Brunner E, Budimir D, Campbell H, Carlson O, Chines PS, Clarke R, Collins FS, Corbatón-Anchuelo A, Couper D, de Faire U, Dedoussis GV, Deloukas P, Dimitriou M, Egan JM, Eiriksdottir G, Erdos MR, Eriksson JG, Eury E, Ferrucci L, Ford I, Forouhi NG, Fox CS, Franzosi MG, Franks PW, Frayling TM, Froguel P, Galan P, de Geus E, Gigante B, Glazer NL, Goel A, Groop L, Gudnason V, Hallmans G, Hamsten A, Hansson O, Harris TB, Hayward C, Heath S, Hercberg S, Hicks AA, Hingorani A, Hofman A, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet*. 2012;44(6):659-669. doi:10.1038/ng.2274

26. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J 'an, Mägi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Zhao JH, Zhao W, Chen J, Fehrmann R, Hedman ÅK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Leach IM, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stančáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Ärnlöv J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Blüher M, Böhringer S, Bonnycastle LL,

Böttcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Chen Y-DI, Clarke R, Daw EW, de Craen AJM, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206. doi:10.1038/nature14177

27. Horikoshi M, Beaumont RN, Day FR, Warrington NM, Kooijman MN, Fernandez-Tajes J, Feenstra B, van Zuydam NR, Gaulton KJ, Grarup N, Bradfield JP, Strachan DP, Li-Gao R, Ahluwalia TS, Kreiner E, Rueedi R, Lyytikäinen L-P, Cousminer DL, Wu Y, Thiering E, Wang CA, Have CT, Hottenga J-J, Vilor-Tejedor N, Joshi PK, Boh ETH, Ntalla I, Pitkänen N, Mahajan A, van Leeuwen EM, Joro R, Lagou V, Nodzenski M, Diver LA, Zondervan KT, Bustamante M, Marques-Vidal P, Mercader JM, Bennett AJ, Rahmioglu N, Nyholt DR, Ma RCW, Tam CHT, Tam WH, CHARGE Consortium Hematology Working Group, Ganesh SK, van Rooij FJA, Jones SE, Loh P-R, Ruth KS, Tuke MA, Tyrrell J, Wood AR, Yaghootkar H, Scholtens DM, Paternoster L, Prokopenko I, Kovacs P, Atalay M, Willems SM, Panoutsopoulou K, Wang X, Carstensen L, Geller F, Schraut KE, Murcia M, van Beijsterveldt CEM, Willemsen G, Appel EVR, Fonvig CE, Trier C, Tiesler CMT, Standl M, Kutalik Z, Bonàs-Guarch S, Hougaard DM, Sánchez F, Torrents D, Waage J, Hollegaard MV, de Haan HG, Rosendaal FR, Medina-Gomez C, Ring SM, Hemani G, McMahon G, Robertson NR, Groves CJ, Langenberg C, Luan J, Scott RA, Zhao JH, Mentch FD, MacKenzie SM, Reynolds RM, Early Growth Genetics (EGG) Consortium, Lowe WL, Tönjes A, et al. Genome-wide associations for birth weight and correlations with adult disease. *Nature*. 2016;538(7624):248-252. doi:10.1038/nature19806

28. Strawbridge RJ, Dupuis J, Prokopenko I, Barker A, Ahlqvist E, Rybin D, Petrie JR, Travers ME, Bouatia-Naji N, Dimas AS, Nica A, Wheeler E, Chen H, Voight BF, Taneera J, Kanoni S, Peden JF, Turrini F, Gustafsson S, Zabena C, Almgren P, Barker DJP, Barnes D, Dennison EM, Eriksson JG, Eriksson P, Eury E, Folkersen L, Fox CS, Frayling TM, Goel A, Gu HF, Horikoshi M, Isomaa B, Jackson AU, Jameson KA, Kajantie E, Kerr-Conte J, Kuulasmaa T, Kuusisto J, Loos RJF, Luan J 'an, Makrilakis K, Manning AK, Martínez-Larrad MT, Narisu N, Nastase Mannila M, Ohrvik J, Osmond C, Pascoe L, Payne F, Sayer AA, Sennblad B, Silveira A, Stancáková A, Stirrups K, Swift AJ, Syvänen A-C, Tuomi T, van 't Hooft FM, Walker M, Weedon MN, Xie W, Zethelius B, DIAGRAM Consortium, GIANT Consortium, MuTHER Consortium, CARDIoGRAM Consortium, C4D Consortium, Ongen H, Mälarstig A, Hopewell JC, Saleheen D, Chambers J, Parish S, Danesh J, Kooner J, Ostenson C-G, Lind L, Cooper CC, Serrano-Ríos M, Ferrannini E, Forsen TJ, Clarke R, Franzosi MG, Seedorf U, Watkins H, Froguel P, Johnson P, Deloukas P, Collins FS, Laakso M, Dermitzakis ET, Boehnke M, McCarthy MI, Wareham NJ, Groop L, Pattou F, et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes*. 2011;60(10):2624-2634. doi:10.2337/db11-0415

29. Censin JC, Nowak C, Cooper N, Bergsten P, Todd JA, Fall T. Childhood adiposity and risk of type 1 diabetes: A Mendelian randomization study. *PLOS Medicine*. 2017;14(8):e1002362. doi:10.1371/journal.pmed.1002362

30. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, 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 JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012;22(9):1760-1774. doi:10.1101/gr.135350.111

31. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Mägi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JRB, Egan JM, Lajunen T, Grarup N, Sparsø T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proença C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J 'an, Elliott A, McCarroll SA, Payne F, Roccasecca RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Böttcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen Y-DI, Chines P, Clarke R, Coin LJM, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day INM, Geus EJC de, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42(2):105-116. doi:10.1038/ng.520

32. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J 'an, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Mägi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Mateo Leach I, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Bolton JL, Böttcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. 2014;46(11):1173-1186. doi:10.1038/ng.3097

33. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, Ruth KS, Whalen S, Sarkar AK, Albrecht E, Altmaier E, Amini M, Barbieri CM, Boutin T, Campbell A, Demerath E, Giri A, He C, Hottenga JJ, Karlsson R, Kolcic I, Loh P-R, Lunetta KL, Mangino M, Marco B, McMahon G, Medland SE, Nolte IM, Noordam R, Nutile T, Paternoster L, Perjakova N, Porcu E, Rose LM, Schraut KE, Segrè AV, Smith AV, Stolk L, Teumer A, Andrulis IL, Bandinelli S, Beckmann MW, Benitez J, Bergmann S, Bochud M, Boerwinkle E, Bojesen SE, Bolla MK, Brand JS, Brauch H, Brenner H, Broer L, Brüning T, Buring JE, Campbell H, Catamo E, Chanock S, Chenevix-Trench G, Corre T, Couch FJ, Cousminer DL, Cox A, Crisponi L, Czene K, Davey Smith G, de Geus EJCN, de Mutsert R, De Vivo I, Dennis J, Devilee P, dos-Santos-Silva I, Dunning AM, Eriksson JG, Fasching PA, Fernández-Rhodes L, Ferrucci L, Flesch-Janys D, Franke L, Gabrielson M, Gandin I, Giles GG, Grallert H, Gudbjartsson DF, Guénel P, Hall P, Hallberg E, Hamann U, Harris TB, Hartman CA, Heiss G, Hooning MJ, Hopper JL, Hu F, Hunter DJ, Ikram MA, Im HK, Järvelin M-R, Joshi PK, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet*. 2017;49(6):834-841. doi:10.1038/ng.3841

34. Varshney A, Scott LJ, Welch RP, Erdos MR, Chines PS, Narisu N, Albanus RD, Orchard P, Wolford BN, Kursawe R, Vadlamudi S, Cannon ME, Didion JP, Hensley J, Kirilusha A, NISC

Comparative Sequencing Program, Bonnycastle LL, Taylor DL, Watanabe R, Mohlke KL, Boehnke M, Collins FS, Parker SCJ, Stitzel ML. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci USA*. 2017;114(9):2301-2306. doi:10.1073/pnas.1621192114

35. van de Bunt M, Manning Fox JE, Dai X, Barrett A, Grey C, Li L, Bennett AJ, Johnson PR, Rajotte RV, Gaulton KJ, Dermitzakis ET, MacDonald PE, McCarthy MI, Gloyn AL. Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. *PLoS Genet*. 2015;11(12):e1005694. doi:10.1371/journal.pgen.1005694

36. 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 CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213. doi:10.1038/nature24277

37. Greenwald WW, Chiou J, Yan J, Qiu Y, Dai N, Wang A, Nariai N, Aylward A, Han JY, Kadakia N, Barrufet L, Okino M-L, Drees F, Vinckier N, Minichiello L, Gorkin D, Avruch J, Frazer K, Sander M, Ren B, Gaulton K. Pancreatic islet chromatin accessibility and conformation defines distal enhancer networks of type 2 diabetes risk. *bioRxiv*. Published online April 20, 2018:299388. doi:10.1101/299388

38. Parker SCJ, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, van Bueren KL, Chines PS, Narisu N, NISC Comparative Sequencing Program, Black BL, Visel A, Pennacchio LA, Collins FS, 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 USA*. 2013;110(44):17921-17926. doi:10.1073/pnas.1317023110

39. Thomas NJ, Jones SE, Weedon MN, Shields BM, Oram RA, Hattersley AT. Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank. *Lancet Diabetes Endocrinol*. Published online November 30, 2017. doi:10.1016/S2213-8587(17)30362-5

40. Barnes PJ. Anti-inflammatory actions of glucocorticoids: molecular mechanisms. *Clin Sci*. 1998;94(6):557-572.

41. Simonis-Bik AM, Nijpels G, van Haeften TW, Houwing-Duistermaat JJ, Boomsma DI, Reiling E, van Hove EC, Diamant M, Kramer MHH, Heine RJ, Maassen JA, Slagboom PE, Willemsen G, Dekker JM, Eekhoff EM, de Geus EJ, 't Hart LM. Gene variants in the novel type 2 diabetes loci CDC123/CAMK1D, THADA, ADAMTS9, BCL11A, and MTNR1B affect

different aspects of pancreatic beta-cell function. *Diabetes*. 2010;59(1):293-301. doi:10.2337/db09-1048

42. Nogueira TC, Paula FM, Villate O, Colli ML, Moura RF, Cunha DA, Marselli L, Marchetti P, Cnop M, Julier C, Eizirik DL. GLIS3, a susceptibility gene for type 1 and type 2 diabetes, modulates pancreatic beta cell apoptosis via regulation of a splice variant of the BH3-only protein Bim. *PLoS Genet*. 2013;9(5):e1003532. doi:10.1371/journal.pgen.1003532

43. Dombroski BA, Nayak RR, Ewens KG, Ankener W, Cheung VG, Spielman RS. Gene expression and genetic variation in response to endoplasmic reticulum stress in human cells. *Am J Hum Genet*. 2010;86(5):719-729. doi:10.1016/j.ajhg.2010.03.017

44. Zhao W, Rasheed A, Tikkanen E, Lee J-J, Butterworth AS, Howson JMM, Assimes TL, Chowdhury R, Orho-Melander M, Damrauer S, Small A, Asma S, Imamura M, Yamauch T, Chambers JC, Chen P, Sapkota BR, Shah N, Jabeen S, Surendran P, Lu Y, Zhang W, Imran A, Abbas S, Majeed F, Trindade K, Qamar N, Mallick NH, Yaqoob Z, Saghir T, Rizvi SNH, Memon A, Rasheed SZ, Memon F-U-R, Mehmood K, Ahmed N, Qureshi IH, Tanveer-Us-Salam null, Iqbal W, Malik U, Mehra N, Kuo JZ, Sheu WH-H, Guo X, Hsiung CA, Juang J-MJ, Taylor KD, Hung Y-J, Lee W-J, Quertermous T, Lee I-T, Hsu C-C, Bottinger EP, Ralhan S, Teo YY, Wang T-D, Alam DS, Di Angelantonio E, Epstein S, Nielsen SF, Nordestgaard BG, Tybjaerg-Hansen A, Young R, CHD Exome+ Consortium, Benn M, Frikke-Schmidt R, Kamstrup PR, EPIC-CVD Consortium, EPIC-Interact Consortium, Michigan Biobank, Jukema JW, Sattar N, Smit R, Chung R-H, Liang K-W, Anand S, Sanghera DK, Ripatti S, Loos RJF, Kooner JS, Tai ES, Rotter JI, Chen Y-DI, Frossard P, Maeda S, Kadowaki T, Reilly M, Pare G, Melander O, Salomaa V, Rader DJ, Danesh J, Voight BF, Saleheen D. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat Genet*. 2017;49(10):1450-1457. doi:10.1038/ng.3943

45. Zhao W, Rasheed A, Tikkanen E, Lee J-J, Butterworth AS, Howson JMM, Assimes TL, Chowdhury R, Orho-Melander M, Damrauer S, Small A, Asma S, Imamura M, Yamauch T, Chambers JC, Chen P, Sapkota BR, Shah N, Jabeen S, Surendran P, Lu Y, Zhang W, Imran A, Abbas S, Majeed F, Trindade K, Qamar N, Mallick NH, Yaqoob Z, Saghir T, Rizvi SNH, Memon A, Rasheed SZ, Memon F-R, Mehmood K, Ahmed N, Qureshi IH, Tanveer-us-Salam, Iqbal W, Malik U, Mehra N, Kuo JZ, Sheu WH-H, Guo X, Hsiung CA, Juang J-MJ, Taylor KD, Hung Y-J, Lee W-J, Quertermous T, Lee I-T, Hsu C-C, Bottinger EP, Ralhan S, Teo YY, Wang T-D, Alam DS, Di Angelantonio E, Epstein S, Nielsen SF, Nordestgaard BG, Tybjaerg-Hansen A, Young R, CHD Exome+ Consortium, Benn M, Frikke-Schmidt R, Kamstrup PR, Epic-Cvd Consortium, EPIC-Interact Consortium, Michigan Biobank, Jukema JW, Sattar N, Smit R, Chung R-H, Liang K-W, Anand S, Sanghera DK, Ripatti S, Loos RJF, Kooner JS, Tai ES, Rotter JI, Chen Y-DI, Frossard P, Maeda S, Kadowaki T, Reilly M, Pare G, Melander O, Salomaa V, Rader DJ, Danesh J, Voight BF, Saleheen D. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat Genet*. 2017;advance online publication. doi:10.1038/ng.3943

46. Gale EA, Gillespie KM. Diabetes and gender. *Diabetologia*. 2001;44(1):3-15. doi:10.1007/s001250051573

47. Lotta LA, Gulati P, Day FR, Payne F, Ongen H, van de Bunt M, Gaulton KJ, Eicher JD, Sharp SJ, Luan J 'an, De Lucia Rolfe E, Stewart ID, Wheeler E, Willems SM, Adams C, Yaghootkar H, EPIC-InterAct Consortium, Cambridge FPLD1 Consortium, Forouhi NG, Khaw K-T, Johnson AD, Semple RK, Frayling T, Perry JRB, Dermitzakis E, McCarthy MI, Barroso I,

Wareham NJ, Savage DB, Langenberg C, O'Rahilly S, Scott RA. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet*. 2017;49(1):17-26. doi:10.1038/ng.3714

48. Barker DJP. The developmental origins of chronic adult disease. *Acta Paediatr Suppl*. 2004;93(446):26-33.

49. Stene LC, Magnus P, Lie RT, Søvik O, Joner G, Group TNCDS. Birth weight and childhood onset type 1 diabetes: population based cohort study. *BMJ*. 2001;322(7291):889-892. doi:10.1136/bmj.322.7291.889

50. Böhmer K, Keilacker H, Kuglin B, Hübinger A, Bertrams J, Gries FA, Kolb H. Proinsulin autoantibodies are more closely associated with type 1 (insulin-dependent) diabetes mellitus than insulin autoantibodies. *Diabetologia*. 1991;34(11):830-834.

51. Røder ME, Porte D, Schwartz RS, Kahn SE. Disproportionately elevated proinsulin levels reflect the degree of impaired B cell secretory capacity in patients with noninsulin-dependent diabetes mellitus. *J Clin Endocrinol Metab*. 1998;83(2):604-608. doi:10.1210/jcem.83.2.4544

52. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-678. doi:10.1038/nature05911

53. Bradfield JP, Qu H-Q, Wang K, Zhang H, Sleiman PM, Kim CE, Mentch FD, Qiu H, Glessner JT, Thomas KA, Frackelton EC, Chiavacci RM, Imielinski M, Monos DS, Pandey R, Bakay M, Grant SFA, Polychronakos C, Hakonarson H. A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci. *PLOS Genetics*. 2011;7(9):e1002293. doi:10.1371/journal.pgen.1002293

54. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795

55. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh P-R, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287. doi:10.1038/ng.3656

56. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340

57. Pers TH, Timshel P, Hirschhorn JN. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics*. 2015;31(3):418-420. doi:10.1093/bioinformatics/btu655

58. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan L, Perry JRB, Patterson N, Robinson

EB, Daly MJ, Price AL, Neale BM. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406

59. Wheeler E, Leong A, Liu C-T, Hivert M-F, Strawbridge RJ, Podmore C, Li M, Yao J, Sim X, Hong J, Chu AY, Zhang W, Wang X, Chen P, Maruthur NM, Porneala BC, Sharp SJ, Jia Y, Kabagambe EK, Chang L-C, Chen W-M, Elks CE, Evans DS, Fan Q, Giulianini F, Go MJ, Hottenga J-J, Hu Y, Jackson AU, Kanoni S, Kim YJ, Kleber ME, Ladenvall C, Lecoeur C, Lim S-H, Lu Y, Mahajan A, Marzi C, Nalls MA, Navarro P, Nolte IM, Rose LM, Rybin DV, Sanna S, Shi Y, Stram DO, Takeuchi F, Tan SP, van der Most PJ, Van Vliet-Ostaptchouk JV, Wong A, Yengo L, Zhao W, Goel A, Martinez Larrad MT, Radke D, Salo P, Tanaka T, van Iperen EPA, Abecasis G, Afaq S, Alizadeh BZ, Bertoni AG, Bonnefond A, Böttcher Y, Bottinger EP, Campbell H, Carlson OD, Chen C-H, Cho YS, Garvey WT, Gieger C, Goodarzi MO, Grallert H, Hamsten A, Hartman CA, Herder C, Hsiung CA, 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 'an, Lyssenko V, Meitinger T, Miki T, Miljkovic I, Moon S, et al. 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*. 2017;14(9):e1002383. doi:10.1371/journal.pmed.1002383

60. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, Strawbridge RJ, Pers TH, Fischer K, Justice AE, Workalemahu T, Wu JMW, Buchkovich ML, Heard-Costa NL, Roman TS, Drong AW, Song C, Gustafsson S, Day FR, Esko T, Fall T, Kutalik Z, Luan J, Randall JC, Scherag A, Vedantam S, Wood AR, Chen J, Fehrmann R, Karjalainen J, Kahali B, Liu C-T, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bragg-Gresham JL, Buyske S, Demirkan A, Ehret GB, Feitosa MF, Goel A, Jackson AU, Johnson T, Kleber ME, Kristiansson K, Mangino M, Leach IM, Medina-Gomez C, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Stančáková A, Sung YJ, Tanaka T, Teumer A, Van Vliet-Ostaptchouk JV, Yengo L, Zhang W, Albrecht E, Ärnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Böhringer S, Bonnet F, Böttcher Y, Bruinenberg M, Carba DB, Caspersen IH, Clarke R, Daw EW, Deelen J, Deelman E, Delgado G, Doney AS, Eklund N, Erdos MR, Estrada K, Eury E, Friedrich N, Garcia ME, Giedraitis V, Gigante B, Go AS, Golay A, Grallert H, Grammer TB, Gräßler J, Grewal J, Groves CJ, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518(7538):187-196. doi:10.1038/nature14132

61. Felix JF, Bradfield JP, Monnereau C, van der Valk RJP, 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 MM, Cousminer DL, Marsh JA, Lehtimäki T, Curtin JA, Vioque J, Ahluwalia TS, Myhre R, Price TS, Vilor-Tejedor N, Yengo L, Grarup N, Ntalla I, Ang W, Atalay M, Bisgaard H, Blakemore AI, 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 CMA, Hirschhorn JN, Hofman A, Holm J-C, Horikoshi M, Hottenga JJ, Huang J, Kadarmideen HN, Kähönen M, Kiess W, Lakka H-M, Lakka TA, Lewin AM, Liang L, Lyytikäinen L-P, Ma B, Magnus P, McCormack SE, McMahon G, Mentch FD, Middeldorp CM, Murray CS, Pahkala K, Pers TH, Pfäffle R, Postma DS, Power C, Simpson A, Sengpiel V, Tiesler CMT, Torrent M, Uitterlinden AG, van Meurs JB, Vinding R, Waage J, Wardle J, Zeggini E, Zemel BS, Dedoussis GV, Pedersen O, Froguel P, Sunyer J, Plomin R, Jacobsson B, Hansen T, Gonzalez JR, Custovic A, Raitakari OT, Pennell CE, Widén E, Boomsma DI, Koppelman GH, Sebert S, Järvelin M-R, Hyppönen E, et al. Genome-wide

association analysis identifies three new susceptibility loci for childhood body mass index. *Hum Mol Genet*. 2016;25(2):389-403. doi:10.1093/hmg/ddv472

62. Bentham J, Morris DL, Graham DSC, Pinder CL, Tombleson P, Behrens TW, Martín J, Fairfax BP, Knight JC, Chen L, Replogle J, Syvänen A-C, Rönnblom L, Graham RR, Wither JE, Rioux JD, Alarcón-Riquelme ME, Vyse TJ. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet*. 2015;47(12):1457-1464. doi:10.1038/ng.3434

63. Cordell HJ, Han Y, Mells GF, Li Y, Hirschfield GM, Greene CS, Xie G, Juran BD, Zhu D, Qian DC, Floyd JAB, Morley KI, Prati D, Lleo A, Cusi D, Canadian-US PBC Consortium, Italian PBC Genetics Study Group, UK-PBC Consortium, Gershwin ME, Anderson CA, Lazaridis KN, Invernizzi P, Seldin MF, Sandford RN, Amos CI, Siminovitch KA. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat Commun*. 2015;6:8019. doi:10.1038/ncomms9019

64. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, Jostins L, Rice DL, Gutierrez-Achury J, Ji S-G, Heap G, Nimmo ER, Edwards C, Henderson P, Mowat C, Sanderson J, Satsangi J, Simmons A, Wilson DC, Tremelling M, Hart A, Mathew CG, Newman WG, Parkes M, Lees CW, Uhlig H, Hawkey C, Prescott NJ, Ahmad T, Mansfield JC, Anderson CA, Barrett JC. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*. 2017;49(2):256-261. doi:10.1038/ng.3760

65. Dubois PCA, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zhernakova A, Heap GAR, Adány R, Aromaa A, Bardella MT, van den Berg LH, Bockett NA, de la Concha EG, Dema B, Fehrmann RSN, Fernández-Arquero M, Fiatal S, Grandone E, Green PM, Groen HJM, Gwilliam R, Houwen RHJ, Hunt SE, Kaukinen K, Kelleher D, Korponay-Szabo I, Kurppa K, MacMathuna P, Mäki M, Mazzilli MC, McCann OT, Mearin ML, Mein CA, Mirza MM, Mistry V, Mora B, Morley KI, Mulder CJ, Murray JA, Núñez C, Oosterom E, Ophoff RA, Polanco I, Peltonen L, Platteel M, Rybak A, Salomaa V, Schweizer JJ, Sperandeo MP, Tack GJ, Turner G, Veldink JH, Verbeek WHM, Weersma RK, Wolters VM, Urcelay E, Cukrowska B, Greco L, Neuhausen SL, McManus R, Barisani D, Deloukas P, Barrett JC, Saavalainen P, Wijmenga C, van Heel DA. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet*. 2010;42(4):295-302. doi:10.1038/ng.543

66. Ji S-G, Juran BD, Mucha S, Folseraas T, Jostins L, Melum E, Kumasaka N, Atkinson EJ, Schlicht EM, Liu JZ, Shah T, Gutierrez-Achury J, Boberg KM, Bergquist A, Vermeire S, Eksteen B, Durie PR, Farkkila M, Müller T, Schramm C, Sterneck M, Weismüller TJ, Gotthardt DN, Ellinghaus D, Braun F, Teufel A, Laudes M, Lieb W, Jacobs G, Beuers U, Weersma RK, Wijmenga C, Marschall H-U, Milkiewicz P, Pares A, Kontula K, Chazouillères O, Invernizzi P, Goode E, Spiess K, Moore C, Sambrook J, Ouwehand WH, Roberts DJ, Danesh J, Floreani A, Gulamhusein AF, Eaton JE, Schreiber S, Coltescu C, Bowlus CL, Luketic VA, Odin JA, Chopra KB, Kowdley KV, Chalasani N, Manns MP, Srivastava B, Mells G, Sandford RN, Alexander G, Gaffney DJ, Chapman RW, Hirschfield GM, de Andrade M, UK-PSC Consortium, International IBD Genetics Consortium, International PSC Study Group, Rushbrook SM, Franke A, Karlsen TH, Lazaridis KN, Anderson CA. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat Genet*. 2017;49(2):269-273. doi:10.1038/ng.3745

67. Jin Y, Andersen G, Yorgov D, Ferrara TM, Ben S, Brownson KM, Holland PJ, Birlea SA, Siebert J, Hartmann A, Lienert A, van Geel N, Lambert J, Luiten RM, Wolkerstorfer A, Wietze van der Veen JP, Bennett DC, Taïeb A, Ezzedine K, Kemp EH, Gawkrodger DJ, Weetman AP, Kõks S, Prans E, Kingo K, Karelson M, Wallace MR, McCormack WT, Overbeck A, Moretti S, Colucci R, Picardo M, Silverberg NB, Olsson M, Valle Y, Korobko I, Böhm M, Lim HW, Hamzavi I, Zhou L, Mi Q-S, Fain PR, Santorico SA, Spritz RA. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat Genet*. 2016;48(11):1418-1424. doi:10.1038/ng.3680

68. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, Graham RR, Manoharan A, Ortmann W, Bhangale T, Denny JC, Carroll RJ, Eyler AE, Greenberg JD, Kremer JM, Pappas DA, 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 EA, Diogo D, Cui J, Liao K, Guo MH, Myouzen K, Kawaguchi T, Coenen MJH, van Riel PLCM, van de Laar MAFJ, Guchelaar H-J, Huizinga TWJ, Dieudé P, Mariette X, Bridges SL, Zhernakova A, Toes REM, Tak PP, Miceli-Richard C, Bang S-Y, Lee H-S, Martin J, Gonzalez-Gay MA, Rodriguez-Rodriguez L, Rantapää-Dahlqvist S, Ärlestig L, Choi HK, Kamatani Y, Galan P, Lathrop M, Eyre S, Bowes J, Barton A, de Vries N, Moreland LW, Criswell LA, Karlson EW, Taniguchi A, Yamada R, Kubo M, Liu JS, Bae S-C, Worthington J, Padyukov L, Klareskog L, Gregersen PK, Raychaudhuri S, Stranger BE, De Jager PL, Franke L, Visscher PM, Brown MA, Yamanaka H, Mimori T, Takahashi A, Xu H, Behrens TW, Siminovitch KA, Momohara S, Matsuda F, Yamamoto K, Plenge RM. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014;506(7488):376-381. doi:10.1038/nature12873

69. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol*. 2009;33(1):79-86. doi:10.1002/gepi.20359

70. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383. doi:10.1371/journal.pgen.1004383

71. Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet*. 2016;99(6):1245-1260. doi:10.1016/j.ajhg.2016.10.003

72. Greenwald W, Chiou J, Yan J, Dai N, Qiu Y, Wang A, Aylward A, Han JY, Okino M-L, Drees F, Vinckier N, Gorkin D, Avruch J, Sander M, Frazer K, Ren B, Gaulton K. Pancreatic islet chromatin accessibility and conformation defines distal enhancer networks of type 2 diabetes risk. Published online Manuscript in preparation.

73. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595. doi:10.1093/bioinformatics/btp698

74. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137. doi:10.1186/gb-2008-9-9-r137

75. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033

76. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*. 2012;9(3):215-216. doi:10.1038/nmeth.1906

77. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37(Web Server issue):W202-208. doi:10.1093/nar/gkp335

78. Rozowsky J, Abyzov A, Wang J, Alves P, Raha D, Harmanci A, Leng J, Bjornson R, Kong Y, Kitabayashi N, Bhardwaj N, Rubin M, Snyder M, Gerstein M. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol*. 2011;7:522. doi:10.1038/msb.2011.54

79. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27(7):1017-1018. doi:10.1093/bioinformatics/btr064

80. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res*. 2011;21(3):447-455. doi:10.1101/gr.112623.110

81. Bhandare R, Schug J, Le Lay J, Fox A, Smirnova O, Liu C, Naji A, Kaestner KH. Genome-wide analysis of histone modifications in human pancreatic islets. *Genome Res*. 2010;20(4):428-433. doi:10.1101/gr.102038.109

82. Khoo C, Yang J, Weinrott SA, Kaestner KH, Naji A, Schug J, Stoffers DA. Research resource: the pdx1 cistrome of pancreatic islets. *Mol Endocrinol*. 2012;26(3):521-533. doi:10.1210/me.2011-1231

83. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, Welch R, Parker SCJ, Boyle AP, Scott LJ, NISC Comparative Sequencing Program, Margulies EH, Boehnke M, Furey TS, Crawford GE, Collins FS. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab*. 2010;12(5):443-455. doi:10.1016/j.cmet.2010.09.012

84. Wang A, Yue F, Li Y, Xie R, Harper T, Patel NA, Muth K, Palmer J, Qiu Y, Wang J, Lam DK, Raum JC, Stoffers DA, Ren B, Sander M. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell*. 2015;16(4):386-399. doi:10.1016/j.stem.2015.02.013

85. Harvey CT, Moyerbrailean GA, Davis GO, Wen X, Luca F, Pique-Regi R. QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics*. 2015;31(8):1235-1242. doi:10.1093/bioinformatics/btu802

86. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods*. 2015;12(11):1061-1063. doi:10.1038/nmeth.3582

CHAPTER 2: Glucocorticoid signaling in pancreatic islets modulates gene regulatory programs and genetic risk of type 2 diabetes

2.1 Abstract

Glucocorticoids are key regulators of glucose homeostasis and pancreatic islet function, but the gene regulatory programs driving responses to glucocorticoid signaling in islets and the contribution of these programs to diabetes risk are unknown. In this study we used ATAC- seq and RNA-seq to map chromatin accessibility and gene expression from eleven primary human islet samples cultured *in vitro* with the glucocorticoid dexamethasone at multiple doses and durations. We identified thousands of accessible chromatin sites and genes with significant changes in activity in response to glucocorticoids. Chromatin sites up-regulated in glucocorticoid signaling were prominently enriched for glucocorticoid receptor binding sites and up-regulated genes were enriched for ion transport and lipid metabolism, whereas down-regulated chromatin sites and genes were enriched for inflammatory, stress response and proliferative processes. Genetic variants associated with glucose levels and T2D risk were enriched in glucocorticoid-responsive chromatin sites, including fine-mapped variants at 51 known signals. Among fine-mapped variants in glucocorticoid-responsive chromatin, a likely casual variant at the 2p21 locus had glucocorticoid-dependent allelic effects on beta cell enhancer activity and affected *SIX2* and *SIX3* expression. Our results provide a comprehensive map of islet regulatory programs in response to glucocorticoids through which we uncover a role for islet glucocorticoid signaling in mediating genetic risk of T2D.

2.2 Introduction

Glucocorticoids are steroid hormones produced by the adrenal cortex which broadly regulate inflammatory, metabolic and stress responses and are widely used in the treatment of immune disorders [1–3]. The metabolic consequences of glucocorticoid action are directly relevant to diabetes pathogenesis, as chronic glucocorticoid exposure causes hyperglycemia and steroid-induced diabetes and endogenous excess of glucocorticoids causes Cushing's syndrome in which diabetes is a common co-morbidity [4,5]. Glucocorticoids contribute to the development of diabetes both through insulin resistance and obesity via effects on adipose, liver and muscle, as well as through pancreatic islet dysfunction [4]. In islets, glucocorticoid signaling has been shown to modulate numerous processes such as insulin secretion, ion channel activity, cAMP signaling, proliferation and development [6–11].

The effects of glucocorticoids on cellular function are largely mediated through regulation of transcriptional activity. Glucocorticoids diffuse through the cell membrane into cytoplasm and bind the glucocorticoid receptor (GR), which is then translocated into the nucleus where it binds DNA and modulates the transcriptional program [12–15]. Gene activity can be affected by GR via direct genomic binding and regulation as well as indirectly through physical interaction with other transcriptional regulators [13–17]. Previous studies have profiled glucocorticoid signaling by mapping genomic locations of GR binding and other epigenomic features such as histone modifications and chromatin accessibility in response to endogenous glucocorticoids such as cortisol or analogs such as dexamethasone [13,14,18,19]. Studies have also shown that the genomic function of GR is largely mediated via binding to regions of accessible chromatin [20,21].

Genetic studies have identified hundreds of genomic loci that contribute to diabetes risk and which primarily map to non-coding sequence and affect gene regulation [22–25]. Risk variants

for type 2 diabetes (T2D) are enriched for pancreatic islet regulatory sites [22–24,26,27], while type 1 diabetes (T1D) risk variants are enriched for immune cell as well as islet regulatory sites. The specific mechanisms of most risk variants in islets are unknown, however, which is critical for understanding the genes and pathways involved in disease pathogenesis and for the development of novel therapeutic strategies. Previous studies of islet chromatin have focused predominantly on normal, non-disease states [27–33], although recent evidence has shown that diabetes risk variants can interact with environmental stimuli to affect islet chromatin and gene regulatory programs [34].

The effects of glucocorticoid and other steroid hormone signaling on islet regulatory programs and how these signals interact with diabetes risk variants, however, are largely unknown. In this study we profiled islet accessible chromatin and gene expression in primary human pancreatic islets exposed *in vitro* to the glucocorticoid dexamethasone. Glucocorticoid signaling had widespread effects on islet accessible chromatin and gene expression levels. Up-regulated chromatin sites were strongly enriched for glucocorticoid receptor binding and up-regulated genes were enriched for processes related to ion channel activity and steroid and lipid metabolism. Conversely, down-regulated sites and genes were involved in inflammation, stress response and proliferation. Genetic variants affecting T2D risk and glucose levels were significantly enriched in glucocorticoid-responsive chromatin sites, including a likely causal variant at the *SIX2/3* locus which had glucocorticoid-dependent effects on beta cell enhancer activity and affected *SIX2* and *SIX3* expression. Together our results provide a comprehensive map of islet gene regulatory programs in response to glucocorticoids which will facilitate a greater mechanistic understanding of glucocorticoid signaling and its role in islet function and diabetes risk.

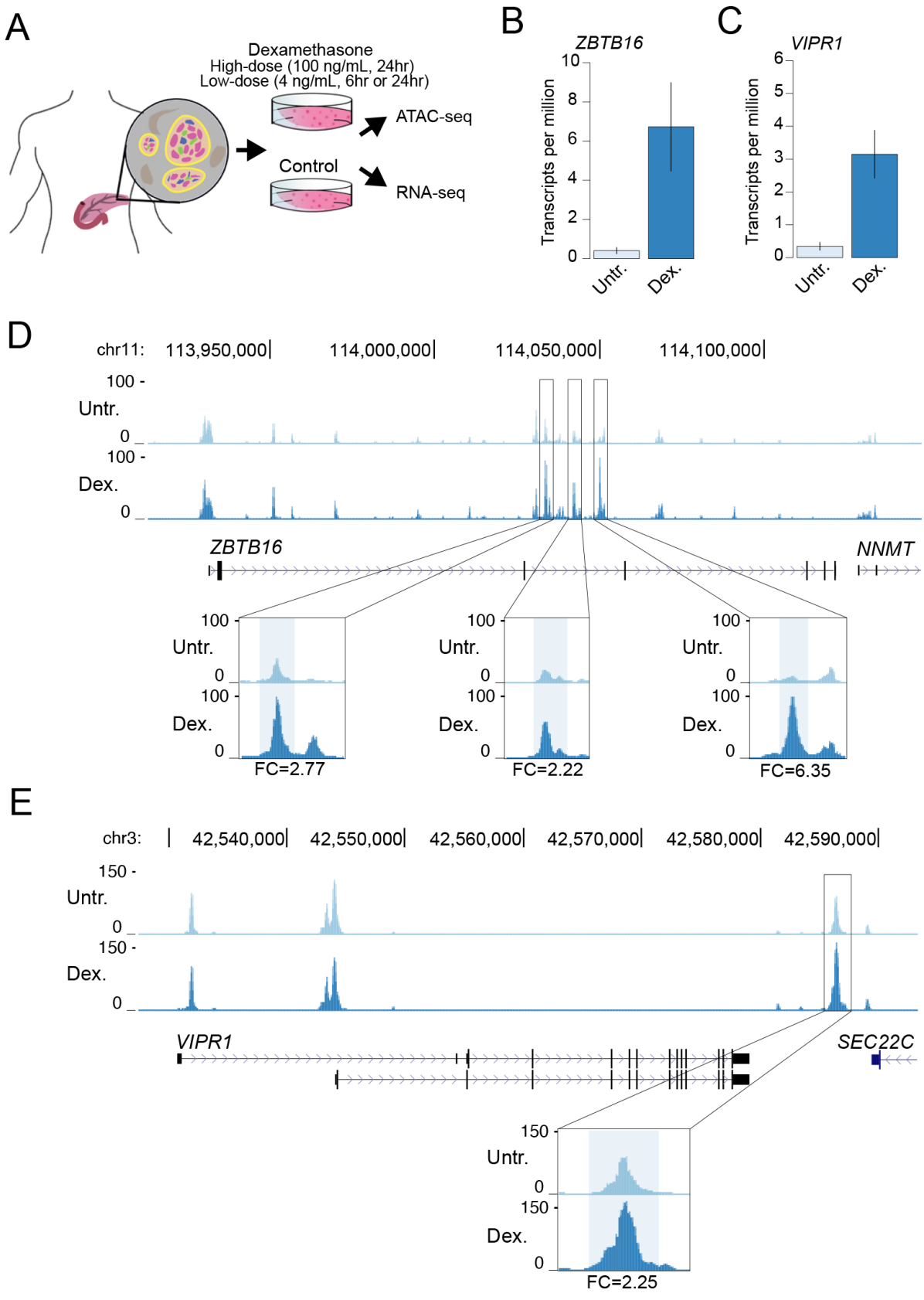
2.3 Results

2.3.1 Map of gene regulation in pancreatic islets in response to glucocorticoid signaling

In order to determine the effects of glucocorticoid signaling on pancreatic islet regulation, we cultured primary islet cells *in vitro* with dexamethasone at several different doses (100 ng/mL for 24hr, 4 ng/mL for 6hr and 24hr) as well as in untreated conditions and measured accessible chromatin and gene expression levels in both treated and untreated cells. An overview of the study design is provided in Figure 2.1A.

We assayed gene expression in dexamethasone-treated and untreated islets from 6 total samples using RNA-seq (S1 Table; see Methods). Across replicate samples we observed changes in expression levels of genes both known to be induced by dexamethasone such as *ZBTB16* [35–37] and *VIPR1* [38] as well as those suppressed by dexamethasone such as *IL11* [39] in both the high-dose (100 ng/mL) and low-dose (4 ng/mL) treatments (**Figures 2.1B, 2.1C, S2.1A, S2.1B and S2.1C**). We next assayed accessible chromatin in dexamethasone-treated and untreated islets from 9 total samples using ATAC-seq (**S1 Table; see Methods**). Across replicate samples we observed reproducible changes in islet accessible chromatin signal concordant with changes in gene expression. For example, accessible chromatin signal was notably induced at several sites proximal to the *ZBTB16* and *VIPR1* genes in dexamethasone-treated compared to untreated islets in both high- and low-dose treatments (**Figures 2.1D, 2.1E, S2.2, S2.3 and S2.4**). Similarly, accessible chromatin signal was reduced at a site proximal to the *IL11* promoter in glucocorticoid-treated compared to untreated islets (**Figure S2.5**).

Figure 2.1. A map of gene regulation in pancreatic islets in response to glucocorticoid signaling. (A) Overview of study design. Primary pancreatic islet samples were split and separately cultured in normal conditions and including the glucocorticoid dexamethasone at either a high-dose (100ng/mL for 24hr) or low-dose (4 ng/mL for 6hr or 24hr) treatment, and then profiled for gene expression and accessible chromatin using RNA-seq and ATAC-seq assays. Genes with known induction in glucocorticoid signaling (B) *ZBTB16* and (C) *VIPR1* had increased expression in glucocorticoid-treated islets compared to untreated islets. Values represent mean and standard error. (D) At the *ZBTB16* locus several accessible chromatin sites intronic to *ZBTB16* had increased accessibility in glucocorticoid treated (Dex.) compared to untreated (Untr.) islets. (E) At the *VIPR1* locus an accessible chromatin site downstream of *VIPR1* had increased accessibility in glucocorticoid treated (Dex.) compared to untreated (Untr.) islets. Values in D and E represent RPKM normalized ATAC-seq read counts. Fold-change (FC) in accessible chromatin signal in glucocorticoid treatment compared to untreated indicated at highlighted sites. All results shown are for the high-dose treatment.


A
Dexamethasone
High-dose (100 ng/mL, 24hr)
Low-dose (4 ng/mL, 6hr or 24hr)
ATAC-seq
Control
RNA-seq
B
ZBTB16
Transcripts per million
Untr.
Dex.
C
VIPR1
Transcripts per million
Untr.
Dex.
D
chr11: 113,950,000 114,000,000 114,050,000 114,100,000
Untr.
Dex.
ZBTB16
NNMT
FC=2.77
FC=2.22
FC=6.35
E
chr3: 42,540,000 42,550,000 42,560,000 42,570,000 42,580,000 42,590,000
Untr.
Dex.
VIPR1
SEC22C
FC=2.25

2.3.2 Islet accessible chromatin sites with differential activity in response to glucocorticoid signaling

To understand the effects of glucocorticoid signaling on accessible chromatin in islets at a genome-wide level, we first performed principal components analysis (PCA) using normalized read counts in chromatin sites for each treated and untreated islet ATAC-seq sample (**see Methods**). We observed reproducible differences in accessible chromatin profiles in dexamethasone-treated compared to untreated islets across replicate samples, where the effects of low-dose treatment (4 ng/mL, n=3) were intermediate to high-dose treatment (100 ng/mL, n=6) relative to untreated samples (n=9) (**Figure 2.2A**).

We then identified specific islet accessible chromatin sites with significant differential activity in glucocorticoid treatment compared to untreated control cells. We first defined a canonical set of 127,228 islet accessible chromatin sites genome-wide by comparing replicate samples using IDR (**see Methods, S2 Table**). Among these canonical sites, there were 2,688 sites with significant evidence (FDR<.10) for differential activity in glucocorticoid signaling at high-dose treatment (**Figure 2.2B and S3 Table**). Among these 2,688 glucocorticoid-responsive sites, 1,992 had up-regulated activity and 695 had down-regulated activity in glucocorticoid treated compared to untreated cells (**Figure 2.2B and S3 Table**). The majority of sites (95%) with differential activity were already accessible in untreated islets, suggesting that sites induced by glucocorticoid signaling are typically not activated *de novo.* Furthermore, a majority of differentially accessible sites (2,453, 91%) were not proximal to promoter regions, suggesting they act via distal regulation of gene activity. At low-dose treatment, 373 sites had differential activity (FDR<.10) in glucocorticoid signaling, where the majority (350) were up-regulated (**S3 Table**). Among sites with differential activity in either treatment, the effects in high- and low-dose were highly concordant (Spearman r=.72, $P<2.2x10^{-16}$) (**Figure S2.6A and S2.6B**).

We next characterized transcriptional regulators underlying changes in glucocorticoid-responsive islet chromatin. First, we identified TF motifs enriched in genomic sequence underneath sites up-regulated and down-regulated in glucocorticoid-treated islets (**see Methods**). The most enriched sequence motifs in up-regulated sites for both high- and low-dose treatment were for glucocorticoid and other steroid hormone response elements (high-dose: GRE $P=1x10^{-340}$, ARE $P=1x10^{-302}$, PGR $P=1x10^{-280}$; low-dose: GRE $P=1x10^{-73}$, ARE $P=1x10^{-66}$, PGR $P=1x10^{-62}$), in addition to lesser enrichment for TFs relevant to islet function (FOXA1: high-dose $P=1x10^{-5}$, low-dose $P=1x10^{-3}$) (**Figure 2.2C and S4 Table)**. Conversely, down-regulated sites in high-dose treatment were most enriched for sequence motifs for STAT TFs (STAT3 $P=1x10^{-9}$, STAT1 $P=1x10^{-8}$) followed by TFs involved in islet function (NKX6.1 $P=1x10^{-7}$, FOXA1 $P=1x10^{-6}$) (**Figure 2.2C and S4 Table**). Next, we determined enrichment of glucocorticoid-responsive chromatin sites for ChIP-seq TF-binding sites previously identified by the ENCODE project. We observed strongest enrichment of up-regulated accessible chromatin sites in both high- and low-dose treatment for glucocorticoid receptor (NR3C1) binding sites (high-dose ratio=3.7, $P=1.7x10^{-294}$, low-dose ratio=5.4, $P=2.1x10^{-129}$), and less pronounced enrichment for binding sites of FOXA1 (high-dose ratio=1.7, $P=1.6x10^{-55}$; low-dose ratio=2.3, $P=2.3x10^{-30}$) and other TFs (**Figure 2.2D and S4 Table**). Down-regulated sites were most enriched for STAT binding (STAT3 ratio=2.1, $P=7.6x10^{-41}$) as well as enhancer binding TFs such as FOS/JUN (FOS ratio=1.5, $P=2.3x10^{-17}$; JUN ratio=1.7, $P=1.3x10^{-14}$) and P300 (ratio=1.4, $P=2.9x10^{-11}$) (**Figure 2.2D and S4 Table**).

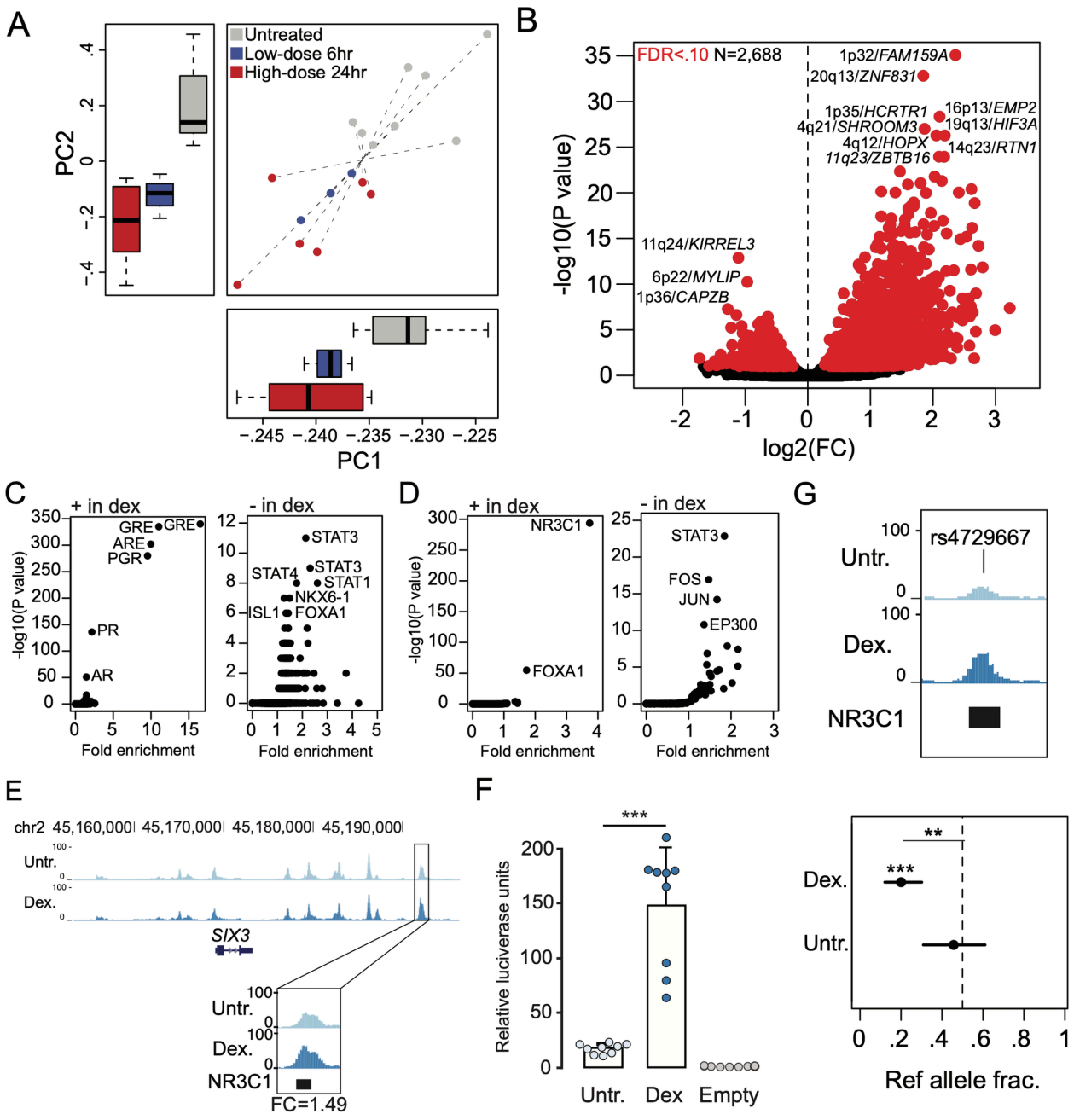
Accessible chromatin sites with significant up-regulation in glucocorticoid signaling compared to untreated islets included a site that mapped to the *SIX2/SIX3* locus (**Figure 2.2E and S3 Table**), which also harbors genetic variants associated with fasting glucose level and risk of T2D. The glucocorticoid-responsive site at this locus also directly overlapped a NR3C1 ChIP-seq site identified by the ENCODE project (**Figure 2.2E**). We tested the glucocorticoid-induced

site at this locus (high-dose fold-change=1.49; P=1.0x10^{-5}; low-dose fold-change=1.51; P=4.4x10^{-4}) for enhancer activity in luciferase gene reporter assays in dexamethasone-treated and untreated MIN6 mouse insulinoma cells. We observed a significant increase in enhancer activity in dexamethasone-treated cells relative to untreated cells (T-test P=1.65x10^{-6}) (**Figure 2.2F**), confirming that this site is highly induced in response to glucocorticoid signaling.

Environmental stimuli can interact with genetic variation to affect chromatin accessibility and gene regulation. We therefore determined the effects of genetic variants on islet accessible chromatin in both glucocorticoid-treated and untreated conditions using allelic imbalance mapping. We performed microarray genotyping of seven islet samples and imputed genotypes into 39M variants (**see Methods**). For variants overlapping islet chromatin sites we obtained read counts in samples heterozygote for that variant, corrected for mapping bias using WASP and modeled the resulting counts for imbalance using a beta-binomial test. We then identified variants with evidence (FDR<.10) for allelic imbalance in accessible chromatin from either glucocorticoid-treated or untreated islets (**S5 Table**). Among imbalanced variants, we further identified those with significant differences in allelic effects (FDR<.10) between glucocorticoid-treated and untreated islets (**S5 Table, see Methods**). For example, variant rs4729667 at 7q22 mapped in a glucocorticoid-responsive site bound by GR and had significantly stronger imbalance in glucocorticoid-treated islets (GC ref frac.=.20, untr. ref frac.=.46; P=3.9x10^{-3}) (**Figure 2.2G, S5 Table**). Conversely, variant rs2291583 at 10p12 in a glucocorticoid-responsive site had significantly stronger imbalance in untreated islets (GC ref frac=.39, untr. ref frac.=.28; P=9.6x10^{-4}) (**S5 Table**).

These results demonstrate that glucocorticoid signaling broadly affects accessible chromatin in islets including sites both up-regulated through glucocorticoid receptor activity and down-regulated through the activity of STAT and other TFs.

Figure 2.2. Glucocorticoid signaling affects chromatin accessibility in pancreatic islets. (A) Principal components plot showing ATAC-seq signal for high-dose (red) and low-dose (blue) glucocorticoid-treated islets and untreated (grey) islets from 9 total donors. Dashed lines connect assays from the same sample, and box plots on each axis represent the distribution of principal components of samples for each condition. (B) Volcano plot of sites with differential chromatin accessibility in glucocorticoid treated compared to untreated islets. Sites with significant differential activity (FDR<.10) are highlighted in red. The sites with the most significant changes are labelled with the locus and the nearest gene. (C) Transcription factor (TF) sequence motifs enriched in differential chromatin sites with increased activity (+ in dex) and decreased activity (- in dex) in glucocorticoid-treated islets. (D) Enrichment of ChIP-seq sites from ENCODE for 160 TFs in differential chromatin sites with increased activity (+ in dex) and decreased activity (- in dex) in glucocorticoid-treated islets. (E) A chromatin site at the *SIX2/3* locus had increased activity in glucocorticoid-treated islets and overlapped a ChIP-seq site for the glucocorticoid receptor (GR/NR3C1) (top). Fold-change (FC) in accessible chromatin signal in glucocorticoid treatment compared to untreated indicated at the highlighted site for high-dose treatment. (F) The differential site at *SIX2/3* had glucocorticoid-dependent effects on enhancer activity in gene reporter assays in MIN6 cells (bottom). Values represent mean and standard deviation. (G) Variant rs4729667 mapped in a chromatin site with increased activity in glucocorticoid-treated islets and had stronger allelic imbalance in chromatin accessibility in glucocorticoid-treated compared to untreated islets. Values represent ref allele fraction and 95% confidence intervals. For panels B, C and D the values shown are from results using high-dose treatment. **P<.01, ***P<$1x10^{-4}$


A
Untreated
Low-dose 6hr
High-dose 24hr
PC2
PC1
B
FDR<.10 N=2,688
1p32/FAM159A
20q13/ZNF831
1p35/HCRTR1
4q21/SHROOM3
16p13/EMP2
19q13/HIF3A
4q12/HOPX
11q23/ZBTB16
14q23/RTN1
11q24/KIRREL3
6p22/MYLIP
1p36/CAPZB
-log10(P value)
log2(FC)
C
+ in dex
- in dex
GRE
ARE
PGR
PR
AR
STAT3
STAT4
STAT1
NKX6-1
ISL1
FOXA1
-log10(P value)
Fold enrichment
D
+ in dex
- in dex
NR3C1
FOXA1
STAT3
FOS
JUN
EP300
-log10(P value)
Fold enrichment
G
rs4729667
Untr.
Dex.
NR3C1
E
chr2 45,160,000 45,170,000 45,180,000 45,190,000
Untr.
Dex.
SIX3
Untr.
Dex.
NR3C1
FC=1.49
F
***
Relative luciverase units
Untr.
Dex
Empty
**
***
Dex.
Untr.
Ref allele frac.

2.3.3 Genes and pathways with differential regulation in islets in response to glucocorticoid signaling

We next sought to determine the effects of glucocorticoid treatment on gene expression levels. We first performed PCA using gene transcript counts from untreated and dexamethasone-treated islet samples at each treatment dose and duration obtained from RNA-seq assays (**see Methods**). There were again reproducible differences in expression levels across replicate samples, where the effects of low-dose treatment (4 ng/mL at 24hr, n=3; 4 ng/mL at 6hr, n=3) were intermediate to high-dose treatment (100 ng/mL at 24hr, n=6) relative to untreated samples (n=6) (**Figure 2.3A)**.

We identified specific genes with differential expression in response to glucocorticoids compared to untreated islet samples using DESeq2 (**see Methods**). There were 2,837 genes with significant evidence for differential expression (FDR<0.10) in glucocorticoid signaling at high-dose treatment (**S6 Table**). Among these genes, 1,348 (47%) were up-regulated and 1,489 (53%) were down-regulated in response to glucocorticoids compared to untreated islets (**Figure 2.3B)**. Genes with the most significant up-regulation included *EDN3* (log2(FC)=1.44, FDR=2.42×10^{-81}), *FAM115C* (log2(FC)=1.52, FDR=3.61x10-^{75}), *METTL7A* (log2(FC)=1.81, FDR=4.36×10^{-71}), *PRR15L* (log2(FC)=2.20, FDR=9.55×10^{-62}), and *CCND3* (log2(FC)=0.95, FDR=9.05×10^{-60}). Conversely, genes with most significant down-regulation included *PCSK1* (log2(FC)=-1.21, FDR=2.05×10^{-61}), *KLHL41* (log2(FC)=-1.31, FDR=8.84×10^{-59}), *DHRS2* (log2(FC)=-1.41, FDR=2.19×10^{-49}) and *CD36* (log2(FC)=-1.21, FDR=2.41×10^{-49}) (**Figure 2.3B**). At low-dose treatment 775 and 848 genes had differential expression (FDR<.10) at 6hr and 24hr, respectively (**S6 Table and Figure S2.7A and S2.7B**). Among genes differentially expressed in either treatment, the effects in high- and low-dose were highly concordant (24hr low-dose r=.91, P<2.2×10^{-16}; 6hr low-dose r=.86, P<2.2×10^{-16}) (**Figure S2.7C and S2.7D and S2.7E**).

We determined whether changes in gene expression in glucocorticoid signaling were driven through accessible chromatin, by testing for enrichment of glucocorticoid-responsive chromatin sites for proximity to differentially expressed genes. Glucocorticoid-responsive chromatin sites were significantly more likely to map within 100kb of a gene with glucocorticoid-responsive expression compared to other chromatin sites in islets (high-dose: OR=1.48, $P=9.9x10^{-20}$; low-dose: OR=4.91, $P=6.5x10^{-36}$)**.** We next performed these analyses separately for sites up- and down-regulated in glucocorticoid signaling. There was significant enrichment of sites with increased activity in glucocorticoid signaling within 100kb of genes with up-regulated expression specifically (up-reg OR=2.9, $P=3.8x10^{-82}$, down-reg OR=0.51, $P=2.1x10^{-19}$) (**Figure 2.3C**). Similarly, sites with decreased activity in glucocorticoid signaling were enriched within 100kb of genes with down-regulated expression (down-reg OR=2.0, $P=6.2x10^{-13}$, up-reg OR=0.48, $P=1.6x10^{-7}$) (**Figure 2.3C**). Furthermore, we also observed an enrichment of glucocorticoid-responsive chromatin sites for closer proximity to genes with glucocorticoid-responsive expression compared to background sites (Kolmogorov-Smirnov $P=3.4x10^{-11}$) (**Figure 2.3D**).

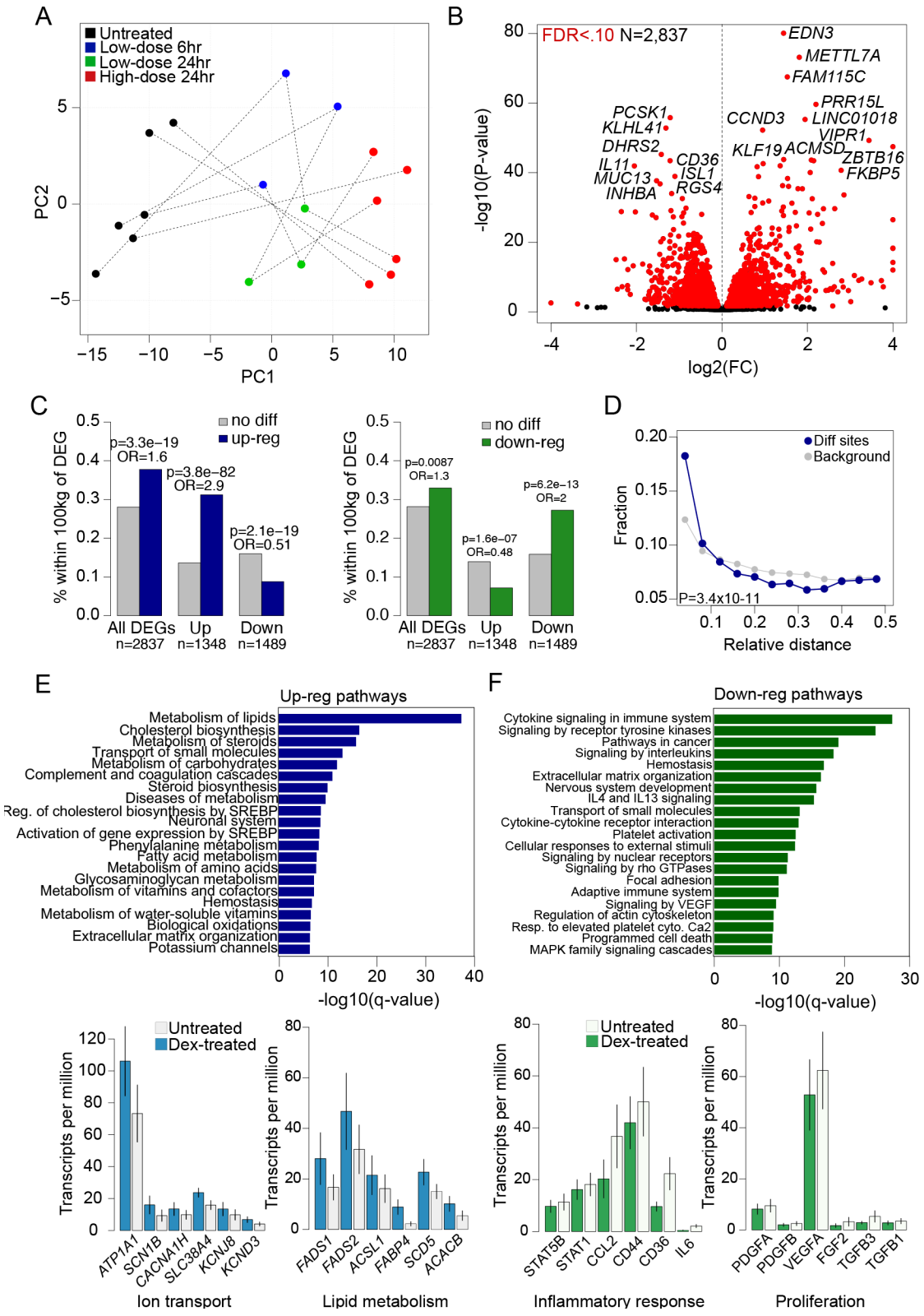
In order to understand the molecular pathways affected by glucocorticoid activity in islets, we tested genes up- and down-regulated in glucocorticoid signaling for gene set enrichment using pathway and gene ontology (GO) terms (**see Methods**). Up-regulated genes in high-dose treatment were enriched for gene sets related to steroid metabolism (steroid metabolic process $FDR=8.94x10^{-30}$), lipid metabolism (lipid biosynthetic process $FDR=1.93x10^{-32}$), potassium and other ion transport (potassium channels $FDR=5.71x10^{-7}$; regulation of ion transport $FDR=1.93x10^{-17}$), and extracellular matrix organization ($FDR=3.68x10^{-7}$) **(Figure 2.3E and S7 Table).** Similar gene sets were enriched among genes up-regulated in low-dose treatments (**S7 Table)**. Numerous genes that function in ion transport were up-regulated in glucocorticoid signaling; for

example *ATP1A1, SCN1B, SCNN1A, CACNA1H, CACNG4, SLC38A4, TRPV6* as well as potassium channel genes including *KCNJ2, KCNAB1, KCNF1, KCNJ8,* and *KCND3* (**Fig 2.3E and S6 Table**). Up-regulated genes also included numerous that function in lipid metabolism including *FADS1, FADS2, ACSL1, SCD5, FABP4, ACACB,* and *ANGPTL4* (**Figure 2.3E and S6 Table**).

Conversely, genes down-regulated in glucocorticoid signaling were enriched for inflammatory response (cytokine signaling in immune system FDR=$2.2x10^{-27}$, signaling by interleukins FDR=$9.50x10^{-19}$), extracellular matrix, cell adhesion and morphogenesis (extracellular matrix organization FDR=$1.53x10^{-17}$, regulation of cell adhesion FDR=$2.48x10^{-42}$, cellular component morphogenesis FDR=$2.45x10^{-37}$), and cell differentiation and proliferation terms (neg. regulation of cell differentiation FDR=$2.18x10^{-36}$) (**Figure 2.3F and S7 Table)**. Similar gene sets were enriched among genes down-regulated in low-dose treatments (**S7 Table)**. Down-regulated genes included those involved in the inflammatory response such as *IL6, STAT5B, STAT3, STAT4, SMAD3, CXCL12, CCL2, CD44, CD36, RELB, IRF1,* extracellular matrix formation such matrix metalloproteinase genes such as *MMP3, MMP7, MMP9* and matrix components such as *LAMA4* and *LAMC2*, islet function and pancreatic differentiation such as *ISL1, PAX6, NKX6-1, HES1* and *JAG1,* and proliferation and growth factors such as *PDGFA, PDGFB, FGF2, TGFB3* and *VEGFA* (**Figure 2.3F and S6 Table**).

These results demonstrate that glucocorticoid signaling in islets up-regulates genes involved in steroid and lipid metabolism and ion channel activity, and down-regulates key genes in islet function as well as genes involved in inflammation, proliferation and extracellular matrix formation.

Figure 2.3. Glucocorticoid signaling affects gene expression levels in pancreatic islets. (A) Principal components plot of gene expression from high-dose (red) and low-dose (green 24hr, blue 6hr) glucocorticoid-treated and untreated (black) islets from a total of 6 samples. Dashed lines connect assays from the same sample. (B) Volcano plot showing genes with differential expression in glucocorticoid-treated islets compared to untreated islets. Genes with significantly differential expression (FDR<.10) are highlighted in red, and genes with pronounced changed in expression are listed. (C) Percentage of accessible chromatin sites with up-regulated activity (left) and down-regulated activity (right) in glucocorticoid-treated islets within 100kb of differentially expressed genes (DEGs) compared to chromatin sites without differential activity. (D) Relative distance metric (from bedtools reldist) between accessible chromatin sites with differential activity (dex) and genes with differential expression compared to all chromatin sites (background). (E) Biological pathway terms enriched among genes with up-regulated expression in glucocorticoid-treated islets (top), and the expression level of selected genes annotated with ion transport and lipid metabolism terms in glucocorticoid-treated and untreated islets (bottom). Values represent mean expression and standard error. (F) Biological pathway terms enriched among genes with up-regulated expression in glucocorticoid-treated islets (top), and the expression level of selected genes annotated with inflammatory response and proliferation pathway terms in glucocorticoid-treated and untreated islets (bottom). Values represent mean expression and standard error. For panels B, C, D and E the values shown are from results using high-dose treatment.


A
Untreated
Low-dose 6hr
Low-dose 24hr
High-dose 24hr
PC1
PC2
B
FDR<.10 N=2,837
EDN3
METTL7A
FAM115C
PRR15L
LINC01018
VIPR1
ZBTB16
FKBP5
CCND3
KLF19
ACMSD
PCSK1
KLHL41
DHRS2
IL11
MUC13
INHBA
CD36
ISL1
RGS4
-log10(P-value)
log2(FC)
C
no diff
up-reg
p=3.3e−19
OR=1.6
p=3.8e−82
OR=2.9
p=2.1e−19
OR=0.51
% within 100kg of DEG
All DEGs
n=2837
Up
n=1348
Down
n=1489
no diff
down-reg
p=0.0087
OR=1.3
p=1.6e−07
OR=0.48
p=6.2e−13
OR=2
D
Diff sites
Background
Fraction
P=3.4x10-11
Relative distance
E
Up-reg pathways
Metabolism of lipids
Cholesterol biosynthesis
Metabolism of steroids
Transport of small molecules
Metabolism of carbohydrates
Complement and coagulation cascades
Steroid biosynthesis
Diseases of metabolism
Reg. of cholesterol biosynthesis by SREBP
Neuronal system
Activation of gene expression by SREBP
Phenylalanine metabolism
Fatty acid metabolism
Metabolism of amino acids
Glycosaminoglycan metabolism
Metabolism of vitamins and cofactors
Hemostasis
Metabolism of water-soluble vitamins
Biological oxidations
Extracellular matrix organization
Potassium channels
-log10(q-value)
F
Down-reg pathways
Cytokine signaling in immune system
Signaling by receptor tyrosine kinases
Pathways in cancer
Signaling by interleukins
Hemostasis
Extracellular matrix organization
Nervous system development
IL4 and IL13 signaling
Transport of small molecules
Cytokine-cytokine receptor interaction
Platelet activation
Cellular responses to external stimuli
Signaling by nuclear receptors
Signaling by rho GTPases
Focal adhesion
Adaptive immune system
Signaling by VEGF
Regulation of actin cytoskeleton
Resp. to elevated platelet cyto. Ca2
Programmed cell death
MAPK family signaling cascades
-log10(q-value)
Untreated
Dex-treated
Transcripts per million
ATP1A1
SCN1B
CACNA1H
SLC38A4
KCNU8
KCND3
Ion transport
FADS1
FADS2
ACSL1
FABP4
SCD5
ACACB
Lipid metabolism
STAT5B
STAT1
CCL2
CD44
CD36
IL6
Inflammatory response
PDGFA
PDGFB
VEGFA
FGF2
TGFB3
TGFB1
Proliferation

2.3.4 T2D and glucose associated variants map in glucocorticoid-responsive islet chromatin

Genetic variants associated with diabetes risk are enriched in pancreatic islet regulatory elements. As these studies have been performed primarily using non-diabetic donors in normal (untreated) conditions, however, the role of environmental stimuli in modulating diabetes-relevant genetic effects on islet chromatin is largely unknown. We therefore tested for enrichment of diabetes and fasting glycemia associated variants in glucocorticoid-responsive islet chromatin sites using fgwas [35] (**see Methods**). We observed enrichment of variants influencing T2D risk and blood sugar (glucose) levels in chromatin sites with differential activity in both high- and low-dose glucocorticoid treatment (T2D high-dose ln(enrich)=3.71, 95% CI=3.03,4.25; T2D low-dose ln(enrich)=4.23, 95% CI=2.66,5.20; blood sugar high-dose ln(enrich)=3.92, 95% CI=0.86,5.70; blood sugar low-dose ln(enrich)=6.20, 95% CI=3.92,8.42) (**Figure 2.4A**). Conversely, we observed no evidence for enrichment of T1D risk variants (high-dose ln(enrich)=-28.00, 95% CI=-48.00,3.39; low-dose ln(enrich)=-23.82, 95% CI=-43.8,5.29) (**Figure 2.4A**).

We next catalogued fine-mapped variants overlapping glucocorticoid-responsive islet chromatin using 99% credible sets of T2D and glucose level signals from DIAMANTE and Biobank Japan (BBJ) [22,36] (**see Methods**). We identified 126 fine-mapped variants at 51 signals that overlapped a glucocorticoid-responsive site (**S8 Table)**. We further identified 511 variants genome-wide in glucocorticoid-responsive sites with at least nominal evidence for T2D association (P<.005) in DIAMANTE or BBJ GWAS (**S8 Table**). We prioritized potential target genes of T2D- and glucose-associated variants in glucocorticoid-responsive chromatin by identifying genes proximal to these sites with differential expression. For example, T2D-associated variants at the 11q12 locus mapped in a site induced by glucocorticoids proximal to *SCD5* and *TMEM150C* which both had up-regulated expression (**Figure 2.4B and S3 and S8 Table**). Similarly, T2D-associated variants at the 4q31 locus mapped in a site down-regulated in

glucocorticoids proximal to *FBXW7* which had down-regulated expression (**Figure S2.7A and S3 and S8 Tables**). Outside of known T2D loci we observed additional examples such as at the 7p15 locus where rs1107376 (T2D $P=2.2x10^{-4}$) mapped in a glucocorticoid-induced site proximal to *NPY* which had glucocorticoid-stimulated expression (**Figure S2.7B and S3 and S8 Tables)**. At 71 T2D- or glucose-associated variants we further observed evidence for association with target gene expression (eQTL) in islets (**S8 Table**); for example, rs1107376 was an islet eQTL for *NPY* ($P=2.2x10^{-21}$).

At the 2p21 locus associated with glucose level, lead variant rs12712928 (BBJ beta=.068, $P=7.4x10^{-46}$) mapped in a chromatin site with increased activity in glucocorticoid signaling and was proximal to *SIX2* and *SIX3* which both had glucocorticoid-induced expression (**Figure 2.4C and 2.4D and S8 Table**). This variant had the highest posterior probability in fine-mapping data (PPA=.89), suggesting it is likely causal for glucose association at this locus. This variant also had evidence for T2D association in BBJ (beta=.048, $P=2.1x10^{-6}$) and DIAMANTE (beta=.022, P=.012), and was the lead variant at a T2D signal recently reported in East Asians ($P=1.8x10^{-14}$) [37]. We therefore tested whether rs12712928 affected enhancer activity using sequence around variant alleles in untreated and dexamethasone treated MIN6 cells (**see Methods**). The glucose increasing and T2D risk allele C had significantly reduced enhancer activity in both glucocorticoid-treated (T-test $P=2.5x10^{-6}$) and untreated cells (T-test $P=3.2x10^{-4}$) (**Figure 2.4E**). However, the allelic differences at this variant were more pronounced in glucocorticoid-treated cells (ref/alt ratio GC=6.85, 95% CI=3.4,10.2; untreated=1.78, 95% CI=1.23,2.32, permutation test $P=5.1x10^{-3}$) (**Figure 2.4F**). We also observed evidence that rs12712928 was an islet eQTL for *SIX3* and *SIX2* (*SIX3* $P=5.1x10^{-23}$, *SIX2* $P=8.2x10^{-10}$; **Figure 2.4G**), where the T2D risk allele was correlated with reduced expression of both genes. Glucose level and T2D association at this locus was strongly co-localized with the *SIX3* and *SIX2* eQTLs (BBJ T2D shared *SIX3* PP=89%, *SIX2* PP=98%; BBJ blood sugar shared *SIX3* PP=98%, *SIX2* PP=99%) (**Figure 2.4G**).

These results reveal that variants associated with T2D and glucose level are enriched in glucocorticoid-responsive chromatin sites in islets, including variants such as rs12712928 at the *SIX2/3* locus which interact with glucocorticoid signaling directly to affect islet regulation.

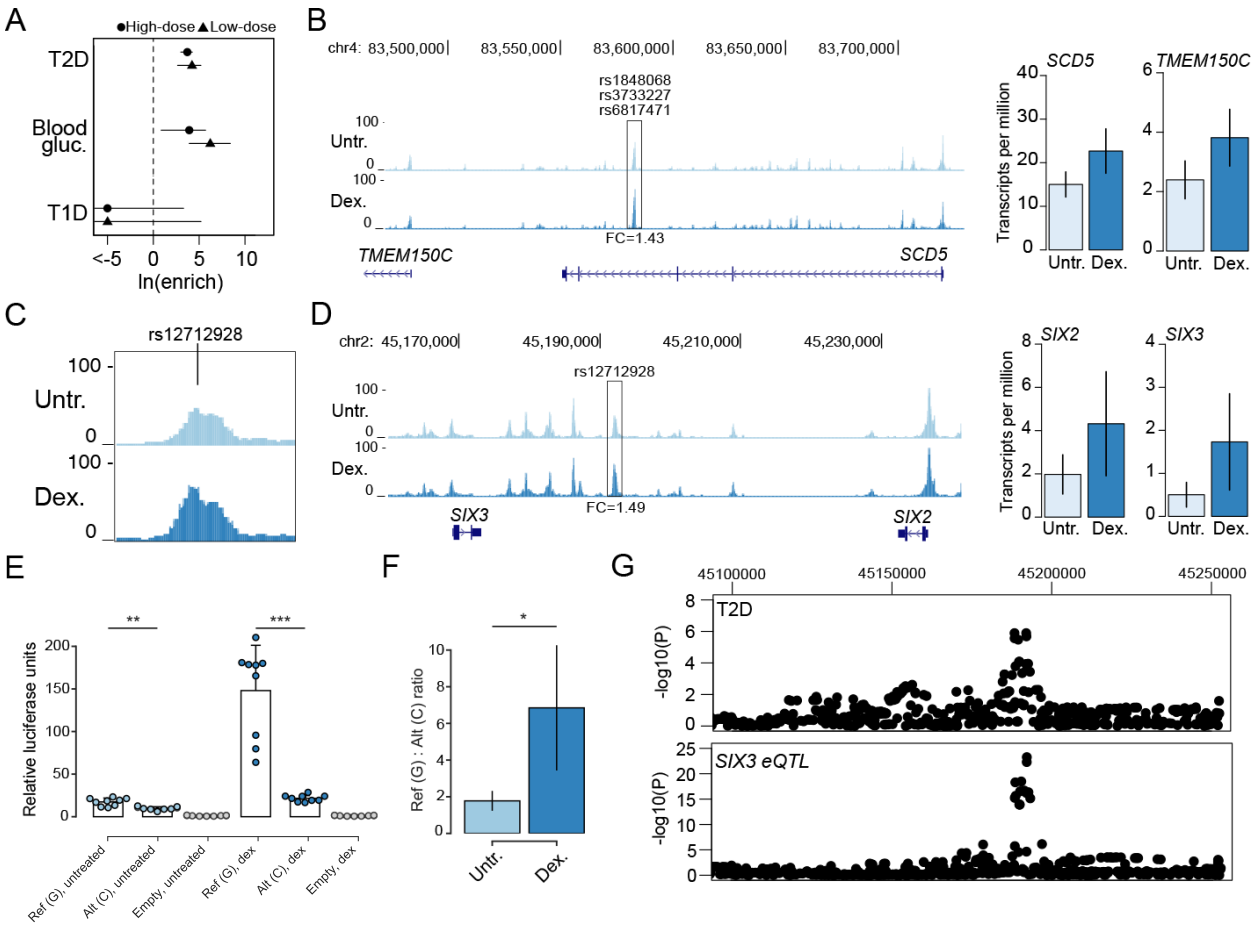


Figure 2.4. Type 2 diabetes and glucose associated variants affect glucocorticoid-responsive islet regulatory programs. (A) Enrichment of variants associated with type 1 diabetes (T1D), type 2 diabetes (T2D) and blood sugar (glucose) levels for differential chromatin sites in high-dose and low-dose glucocorticoid-treated islets. Values represent log enrichment estimates and 95% confidence intervals. (B) Multiple fine-mapped T2D variants at the *SCD5/TMEM150C* locus mapped in a glucocorticoid-responsive islet accessible chromatin site. Both the *SCD5* and *TMEM150C* genes had increased expression in glucocorticoid-treated islets. Genome browser tracks represent RPKM normalized ATAC-seq signal, and bar plots represent mean expression and standard error. (C, D) Variant rs12712928 with evidence for blood sugar and T2D association mapped in a glucocorticoid-responsive chromatin site at the *SIX2/3* locus. Both the *SIX2* and *SIX3* genes had increased expression in glucocorticoid-treated islets. Genome browser tracks represent RPKM normalized ATAC-seq signal, and bar plots represent mean expression and standard error. (E) Variant rs12712928 had significant allelic effects on enhancer activity in gene reporter assays in MIN6 cells. Values represent mean and standard deviation. (F) The allelic effects of rs12712928 were more pronounced in glucocorticoid-treated relative to untreated islets. Values represent fold-change and 95% CI. (G) The T2D association signal at *SIX2/3* was colocalized with an eQTL for *SIX3* expression in islets. For panels B, C and D the values shown are from results using high-dose treatment. For panels B and D, the fold-change (FC) in accessible chromatin signal in glucocorticoid treatment compared to untreated is indicated at highlighted sites. ***P<1x10^{-4}, **P<1x10^{-3}, *P<1x10^{-2}

2.4 Discussion

Our study demonstrates the relevance of islet chromatin dynamics in response to corticosteroid signaling to T2D pathogenesis, including T2D risk variants that interact with corticosteroid activity directly to affect islet chromatin. In a similar manner, variants mediating epigenomic responses of pancreatic islets to proinflammatory cytokines were recently shown to contribute to genetic risk of T1D [34]. Numerous environmental signals and external conditions modulate pancreatic islet function and contribute to the pathophysiology and genetic basis of diabetes, yet the epigenomic and transcriptional responses of islets to disease-relevant stimuli have not been extensively measured. Future studies of islet chromatin and gene regulation exposed to additional stimuli will therefore likely continue providing additional insight into diabetes risk.

Glucocorticoid signaling led to broad changes in accessible chromatin, which up-regulated the expression of proximal genes enriched for processes related to ion channels and transport, in particular potassium channels. Potassium ion concentrations modulate calcium influx and insulin secretion in beta cells [38], and in disruption of ion channel function leads to impaired glucose-induced insulin secretion and diabetes [39]. Glucocorticoids have been shown to suppress calcium influx while preserving insulin secretion via cAMP [7], and in line with this finding we observed evidence for increased activity of potassium channel and cAMP signaling genes and decreased activity of phosphodiesterase genes. Up-regulated genes were also strong enriched in lipid metabolism pathways, which has been shown to regulate insulin secretion and contribute to diabetes [40,41]. Several up-regulated genes *PER1* and *CRY2* are also components of the circadian clock, and previous studies have shown that endogenous glucocorticoid release is under control of circadian rhythms and therefore may contribute to downstream regulation of the clock [42].

Conversely, glucocorticoid signaling down-regulated inflammatory programs, in line with previous reports and the known function of glucocorticoids [2,17,43], as well as key genes involved in islet function such as *NKX6-1, PAX6, RFX6,* and *ISL1*. Our findings further suggest that down-regulation of gene activity in glucocorticoid signaling is mediated through the activity of STAT and other TFs at proximal accessible chromatin sites, either through reduced TF expression or inhibition by GR. We also observed enrichment of FOXA binding in sites both up- and down-regulated in glucocorticoid signaling, suggesting these TFs mark sites that are broadly responsive to signal-dependent TF activity in islets in line with their known function as pioneer factors.

Genetic variants near the homeobox TFs *SIX2* and *SIX3* influence glucose levels [44,45], and our results provide evidence that both of these TFs operate downstream of glucocorticoid signaling and that the variants interact with this signaling program directly to influence glucose levels and risk of T2D. A previous study identified association between this locus and glucose levels in Chinese samples and demonstrated allelic effects of the same variant on islet enhancer activity and binding of the TF GABP [45], further supporting the likely causality of this variant. *SIX2* and *SIX3* have been widely studied for their role in forebrain, kidney and other tissue development [46–51]. In islets, both *SIX2* and *SIX3* have been shown to increase expression in adult compared to juvenile islets, and induction of *SIX3* expression in EndoC-bH1 cells and juvenile islets enhanced islet function, insulin content and secretion and may contribute to the suppression of proliferative programs [52]. In line with this finding, the glucose-lowering and T2D protective allele of the likely causal variant increased islet enhancer activity and *SIX2/3* expression.

Our *in vitro* experimental model mimics the environment of pancreatic islets under hormone signaling, albeit for a small number of treatments and conditions. Given the similarity in binding motifs of many nuclear hormone receptors and the enrichment of glucocorticoid responsive sites for androgen and progesterone receptor motifs, the effects of GR on islet gene

regulation may overlap with other nuclear receptors by acting on shared chromatin sites [53(p1)]. Studies of other tissues have profiled glucocorticoid signaling across a broader range of experimental conditions and identified dose- and temporally-dependent effects on gene regulatory programs [1415], and in islets dose- and temporally-dependent effects of glucocorticoids may impact insulin secretion and other islet functions. Future studies profiling the genomic activity of nuclear receptors in islets across a greater breadth of experimental conditions will therefore help further shed light into the role of hormone signaling dynamics in islet gene regulation and diabetes pathogenesis.

2.5 Methods

Ethics statement

All studies were approved by the Institutional Review Board of the University of California San Diego.

Human islet samples

Human islet samples were obtained through the Integrated Islet Distribution Program (IIDP), University of Alberta and Prodo labs. Islet samples were further enriched using a dithizone stain. Islets were cultured for 24hr at approximately 10mL media/1k islets in 10cm dishes at 37C, 5% CO2 in CMRL 1066 media supplemented with 10% FBS, 1X pen-strep, 8mM glucose, 2mM L-glutamine, 1mM sodium pyruvate, 10mM HEPES, and 250ng/mL Amphotericin B. Treated islets had dexamethasone (Sigma) added in the culture media at either 100 ng/mL for 24hr, 4ng/mL for 24hr or 4 ng/mL for 6hr.

ATAC-seq assays

Islet samples were collected and centrifuged at 500xg for 3 minutes, then washed twice in HBSS, and resuspended in nuclei permeabilization buffer consisting of 5% BSA, 0.2% IGEPAL-CA630, 1mM DTT, and 1X complete EDTA-free protease inhibitor (Sigma) in 1X PBS. Islets were homogenized using a chilled glass dounce homogenizer and incubated on a tube rotator for 10 mins before being filtered through a 30uM filter (sysmex) and centrifuged at 500xg in a 4C microcentrifuge to pellet nuclei. Nuclei were resuspended in Tagmentation Buffer (Illumina) and counted using a Countess II Automated Cell Counter (Thermo). Approximately 50,000 nuclei were transferred to a 0.2mL PCR tube and volume was adjusted to 22.5uL with Tagmentation Buffer. 2.5uL TDE1 (Illumina) was added to each tagmentation reaction and mixed with gentle pipetting. Transposition reactions were incubated at 37C for 30 minutes. Tagmentation reactions were cleaned up using 2X reaction volume of Ampure XP beads (Beckman Coulter) and eluted in 20uL Buffer EB (Qiagen). 10uL tagmented DNA prepared as described above was used in a 25uL PCR reaction using NEBNext High-Fidelity Master Mix (New England Biolabs) and Nextera XT Dual-Indexed primers (Nextera). Final libraries were double size selected using Ampure XP beads and eluted in a final volume of 20uL Buffer EB. Libraries were analyzed using the Qubit HS DNA assay (Thermo) and Agilent 2200 Bioanalyzer (Agilent Biotechnologies). Sample libraries were sequenced on Illumina HiSeq 4000 using 100bp paired-end reads except for samples Isl10, Isl11 and Isl12 which were sequenced on Illumina NovaSeq 6000 using 100bp paired-end reads.

RNA-seq assays

RNA was isolated from treated and untreated islets using RNeasy Mini kit (Qiagen) and submitted to the UCSD Institute for Genomic Medicine to prepare and sequence ribodepleted RNA libraries. Sample libraries were sequenced on Illumina HiSeq4000 using 100bp paired-end reads except for samples Isl10, Isl11 and Isl12 which were sequenced on Illumina NovaSeq 6000 using 100bp paired-end reads.

ATAC-seq data processing

We trimmed reads using Trim Galore with options '–paired' and '–quality 10', then aligned them to the hg19 reference genome using BWA [54] mem with the '-M' flag. We then used samtools [55] to fix mate pairs, sort and index read alignments, used Picard (http://broadinstitute.github.io/picard/) to mark duplicate reads, and used samtools [55] to filer reads with flags '-q 30', '-f 3', '-F 3332'. We then calculated the percentage of mitochondrial reads and percentage of reads mapping to blacklisted regions and removed all mitochondrial reads. We calculated a TSS enrichment score for each ATAC-seq experiment using the Python package 'tssenrich'. To obtain read depth signal tracks, we used bamCoverage [56] to obtain bigWig files for each alignment with signal normalization using RPKM.

Identifying differential chromatin sites

We first used Irreproducible Discovery Rate (IDR) to define a set of canonical ATAC-seq sites for differential analysis. In brief, for each condition separately, we pooled reads across all assays and randomly split the pooled reads into two 'pseudo-replicates'. For the pooled and 'pseudo-replicate' data we called candidate peaks using MACS2 [57] with the parameters '—extsize 150 –keep-dup all –shift -75 –nomodel -p 0.01'. We applied IDR to the 'pseudo-replicate' candidate peak calls and obtained the number of peaks at an IDR threshold of .01. We then sorted and filtered the pooled candidate peak calls based on this number. Finally, we merged the resulting peaks across conditions, where if two peaks overlapped, we retained the more significant peak, and considered these canonical sites for downstream analyses.

The set of alignments for each assay were then supplied as inputs to the R function featureCounts from the Rsubread [58] package to generate a matrix of read counts within each canonical site. We applied the R function DESeqDataSetFromMatrix from the DESeq2 [59(p2)]

package to the read count matrix with default parameters then applied the DESeq function including donor as a variable to model paired samples. We considered sites differentially accessible with FDR<0.1, as computed by the Benjamini-Hochberg method.

We determined the percentage of differential sites with increased activity in glucocorticoids that overlapped a site active in untreated samples, as well as the percentage of differential sites proximal to a gene promoter defined as 5kb upstream of the transcription start site.

Principal components analysis

We first defined input sites by merging overlapping (1bp or more) peaks identified in at least two experiments across all ATAC-seq experiments. We then constructed a read count matrix using edgeR [60] and calculated normalization factors using the 'calcNormFactors' function. We applied the voom transformation [61] and used the 'removeBatchEffect' function from limma [62] to regress out batch effects and sample quality effects (using TSS enrichment as a proxy for sample quality). We then restricted the read count matrix to the 100,000 most variable peaks and performed PCA analysis using the core R function 'prcomp' with rank 2.

TF enrichment analysis

Differentially accessible chromatin sites were analyzed for sequence motif enrichment compared to a background of all chromatin sites tested for differential activity using HOMER [63] and a masked hg19 reference genome with the command `findMotifsGenome.pl <bed file> <masked hg19> <output dir> -bg <background bed file> -size 200 -p 8 -bits -preparse -preparsedDir tmp`. We used the TF sequence motif database provided with the HOMER software. For TF ChIP-seq enrichment, we obtained ChIP-seq binding sites for 160 TFs generated by the ENCODE project [64] and tested for enrichment of binding in differential accessible chromatin sites compared to a background of all remaining chromatin sites genome-wide without differential

activity. For each TF we calculated a 2x2 contingency table of overlap with differential sites and non-differential sites, determined significance using a Fisher test and calculated a fold-enrichment of overlap in differential compared to non-differential sites.

RNA-seq data processing and analysis

Paired-end RNA-Seq reads were aligned to the genome using STAR [65] (2.5.3a) with a splice junction database built from the Gencode v19 gene annotation [66]. Gene expression values were quantified using the RSEM package (1.3.1) and filtered for >0.1 TPM on average per sample. Raw expression counts from the remaining 20,480 genes were normalized using variance stabilizing transformation (vst) from DESeq2 [59(p2)] and corrected for sample batch effects using limma removeBatchEffect. Principal component analysis was performed in R using the prcomp function. To identify differentially expressed genes between treated and untreated samples we obtained raw expression counts from RSEM [67] for the 20,480 genes and applied DESeq2 [59(p2)] with default settings including donor as a cofactor to model paired samples. To identify enriched GO terms in up and down-regulated genes, we applied GSEA [68] using Gene Ontology terms and KEGG/REACTOME pathway terms. We excluded gene sets with large numbers of genes in enrichment tests.

Proximity of differential chromatin sites to differentially expressed genes

We calculated the percentage of differential accessible chromatin sites mapping within 100kb of (i) all differentially expressed genes, (ii) up-regulated genes and (iii) down-regulated genes compared to non-differentially accessible sites, and determined the significance and odds ratio using a Fisher exact test. We calculated a relative distance metric with bedtools [69] (reldist function) using either differential chromatin sites or a background of all islet accessible chromatin sites as the "a" argument and differentially expressed genes as the "b" argument. We compared

the distribution of relative distances from differential sites to the distribution from background sites using a Kolmogorov-Smirnov test.

Sample genotyping and imputation

Non-islet tissue was collected for seven samples during islet picking and used for genomic DNA extraction using the PureLink genomic DNA kit (Invitrogen). Genotyping was performed using Infinium Omni2.5-8 arrays (Illumina) at the UCSD Institute for Genomic Medicine. We called genotypes using GenomeStudio (v.2.0.4) with default settings. We then used PLINK [70] to filter out variants with 1) minor allele frequency (MAF) less than 0.01 in the Haplotype Reference Consortium (HRC) [71] panel r1.1 and 2) ambiguous A/T or G/C alleles with MAF greater than 0.4. For variants that passed these filters, we imputed genotypes into the HRC reference panel r1.1 using the Michigan Imputation Server with minimac4. Post imputation, we removed imputed genotypes with low imputation quality (R2<.3).

Allelic imbalance mapping

We identified heterozygous variant calls in each sample with read depth of at least 10 in both untreated and treated cells, and then used WASP [72] to correct for reference mapping bias. We retained variants in each sample where both alleles were identified at least 3 times across untreated and treated cells. We then merged read counts at heterozygous SNPs from all samples in untreated and treated cells separately. We fit a beta-binomial model to the observed allele counts using the method of NPBin [73]. The parameters of the beta-binomial model were a=40.78 and b=39.26 with over-dispersion of .012 for untreated samples and a=41.76 and b=40.10 with over-dispersion of .012 for glucocorticoid-treated samples. We called imbalanced variants from the merged counts using a beta-binomial test, and then calculated q-values from the resulting beta-binomial p-values. We considered variants significant at FDR<.10.

Heterogeneous allelic imbalance

For all variants with significant allelic imbalance in either glucocorticoid-treated or untreated conditions, we tested for heterogeneity in imbalance between conditions. We used Pearson's chi-squared test as implemented in the "prop.test" function of R. We calculated q-values from the resulting p-values and considered variants significant at FDR<.10.

Genetic association analysis

We tested glucocorticoid-responsive chromatin sites for enrichment of diabetes association using genome-wide association data for T1D [74], T2D from the DIAMANTE consortium [22], and blood sugar (glucose) from the Japan Biobank study [44]. For each study we retained variants with minor allele frequency (MAF)>.05 and tested for enrichment of high-dose and low-dose differential sites using fgwas [35] with a window size of 1Mb.

We then cataloged all variants in glucocorticoid-responsive chromatin sites in T2D and glucose fine-mapping data and with nominal association (P<.005) genome-wide. For DIAMANTE, we used fine-mapping results provided with the study. For the Japan Biobank, we fine-mapped signals ourselves using summary statistics. We calculated approximate Bayes factors (ABF) for each variant as described previously [75]. We compiled index variants for each locus and defined variants within a 5 Mb window and at least low linkage (r^2>0.1) in the East Asian subset of 1000 Genomes [76] with each index. For each variant, we calculated posterior probabilities of associations (PPA) by dividing the variant ABF by the sum of ABF for the locus. We defined 99% credible sets by sorting variants by descending PPA and retaining variants up to a cumulative probability of 99%. For each variant in glucocorticoid-responsive chromatin, we identified protein-coding genes in GENCODE v33 with differential expression and where the gene body mapped within 100kb of the variant.

Expression QTL analyses

We obtained islet expression QTL data from a published study [77]. We extracted variant associations at the *SIX2/SIX3* locus and tested for colocalization between T2D and blood sugar association in the Biobank Japan study and *SIX2* and *SIX3* eQTLs using a Bayesian approach [78]. We considered signals colocalized with shared PP greater than 80%.

Gene reporter assays

To test for allelic differences in enhancer activity at the *SIX2/3* locus, we cloned human DNA sequences (Coriell) containing the reference allele upstream of the minimal promoter in the luciferase reporter vector pGL4.23 (Promega) using the enzymes Sac I and Kpn I. A construct containing the alternate allele was then created using the NEB Q5 SDM kit (New England Biolabs). The primer sequences used were as follows:

Cloning FWD AGCTAGGTACCCCTCATCTGCCTTTCTGGAC

Cloning REV TAACTGAGCTCCAGTGGGTATTGCTGCTTCC

SDM FWD TGCATTGTTTcCTGTCCTGAAGACGAGC

SDM REV GGGGGTGCCTGCATCTGC

MIN6 cells were seeded at approximately 2.5E05 cells/cm^2 into a 48-well plate. The day after passaging into the 48-well plate, cells were co-transfected with 250ng of experimental firefly luciferase vector pGL4.23 containing the alt or ref allele in the forward direction or an empty pGL4.23 vector, and 15ng pRL-SV40 Renilla luciferase vector (Promega) using the Lipofectamine 3000 reagent. Cells were fed culture media and stimulated where applicable 24 hours post-transfection. For stimulation 100 ng/mL dexamethasone (Sigma) was added to the culture media.

Cells were lysed 48 hours post transfection and assayed using the Dual-Luciferase Reporter system (Promega). Firefly activity was normalized to Renilla activity and normalized results were expressed as fold change compared to the luciferase activity of the empty vector. The python package 'luciferase' was then used to remove batch effects. A two-sided t-test was used to compare the luciferase activity between the two alleles or between treatments. A permutation test was used to compare the allelic ratio of luciferase activity between the two treatments, based on 100,000 permutations of the allele labels.

2.6 Supplementary Figures

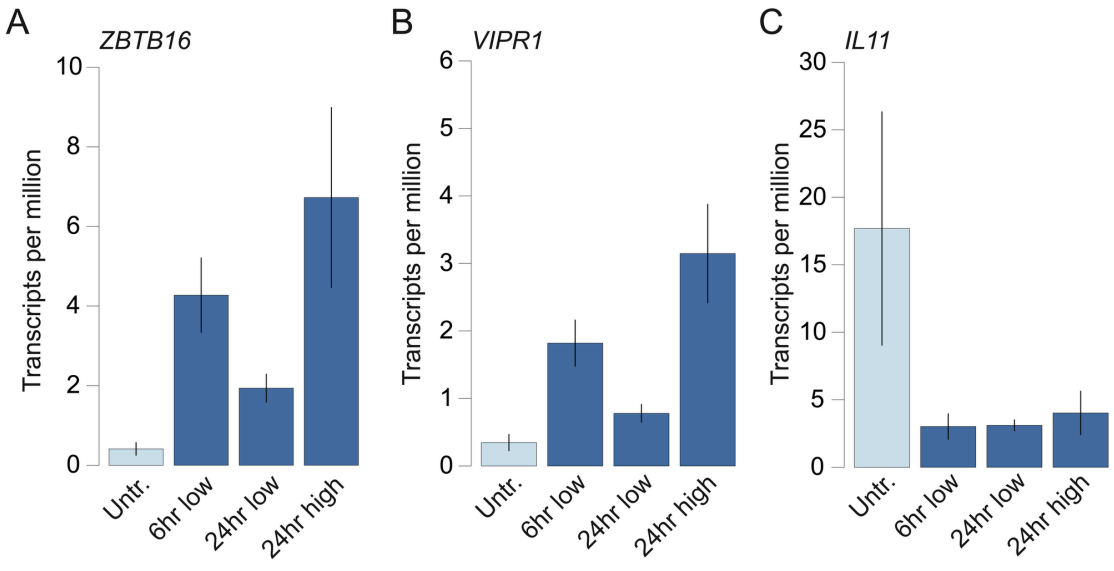


Figure S2.1. Gene expression in islets in response to different doses and durations of glucocorticoid treatment. Expression level of (A) *ZBTB16,* (B) *VIPR1* and (C) *IL11* in high-dose (100ng/mL for 24hr), low-dose (4ng/mL for 24hr or 6hr) glucocorticoid-treated or untreated islets. Values represent mean expression and standard error.

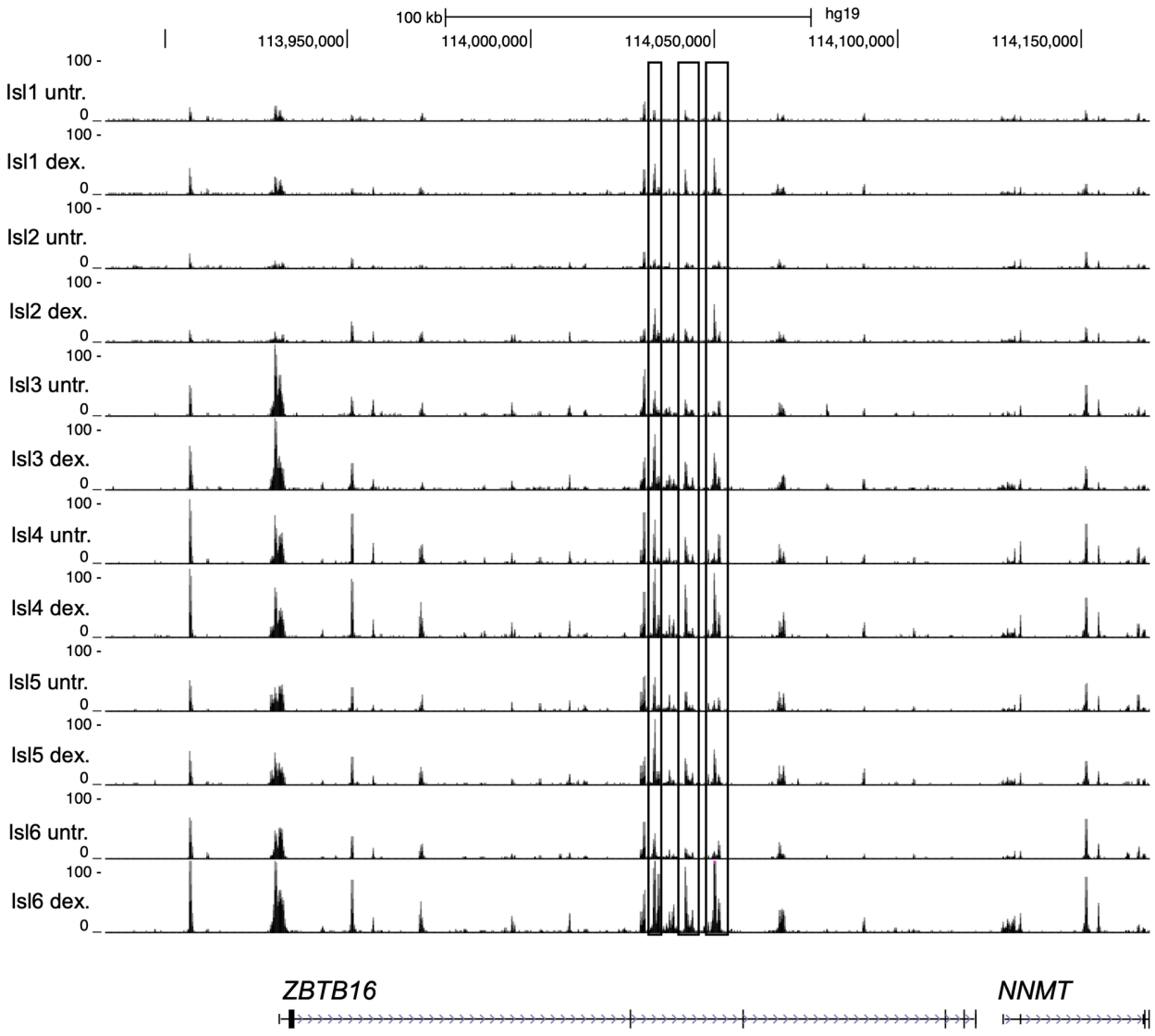


Figure S2.2. Islet accessible chromatin signal across replicate samples at *ZBTB16*. RPKM normalized ATAC-seq signal for individual islet sample in high-dose glucocorticoid treated and untreated islets. Sites with differences in chromatin accessibility across conditions are highlighted.

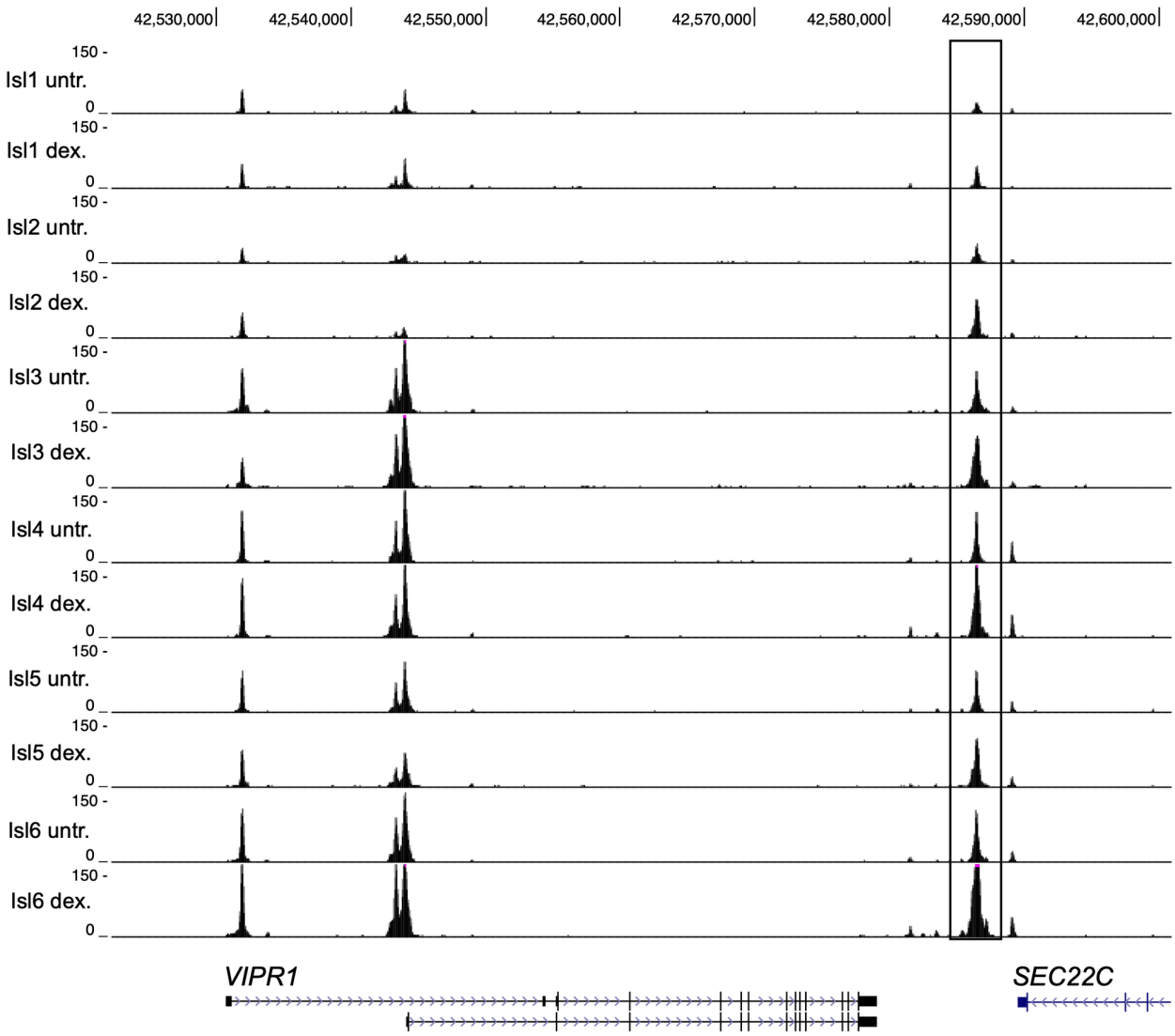


Figure S2.3. Islet accessible chromatin signal across replicate samples at *VIPR1*. RPKM normalized ATAC-seq signal for individual islet sample in high-dose glucocorticoid treated and untreated islets. Sites with differences in chromatin accessibility across conditions are highlighted.

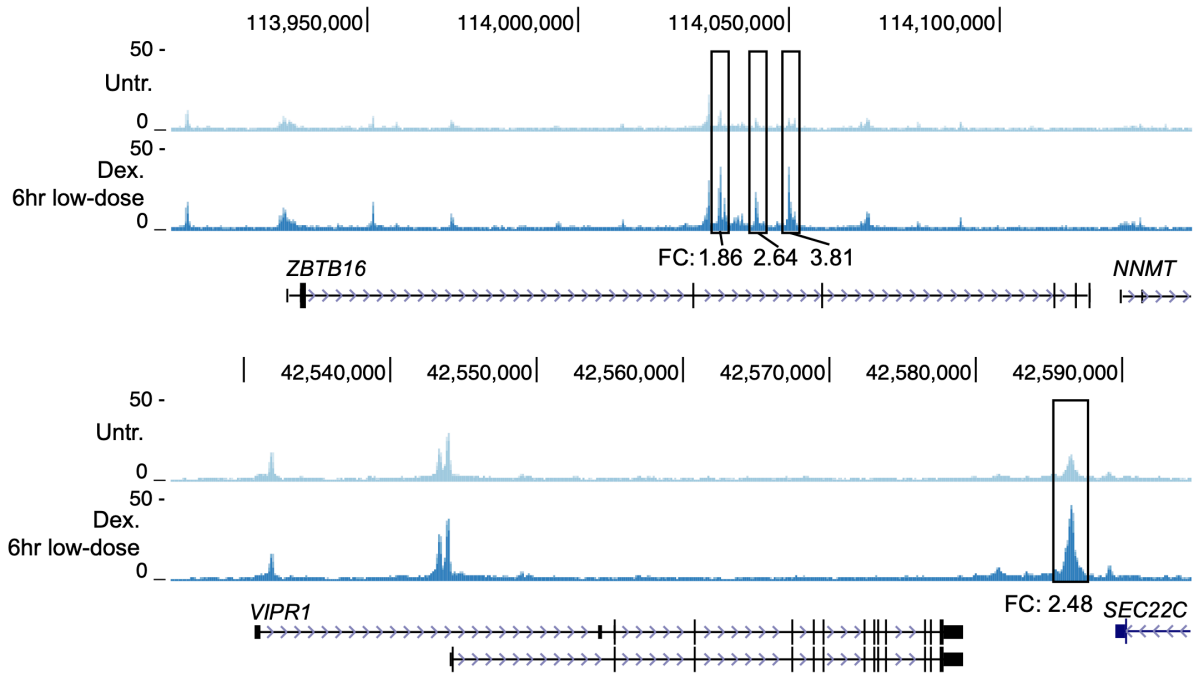


Figure S2.4. Accessible chromatin signal in islets in response to low dose glucocorticoid treatment. RPKM normalized ATAC-seq signal in low-dose (4ng/mL for 6hr) glucocorticoid treated and untreated islets at the (A) *ZBTB16* and (B) *VIPR1* loci. Sites induced by glucocorticoid treatment are highlighted.

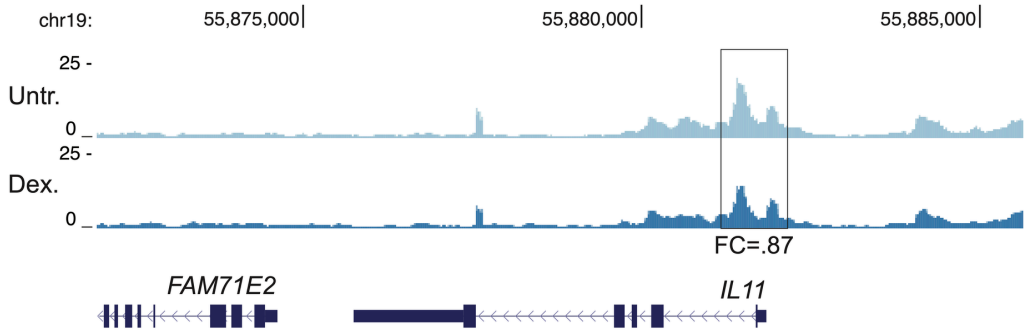


Figure S2.5. Islet accessible chromatin signal at *IL11*. RPKM normalized ATAC-seq signal in high-dose glucocorticoid treated and untreated islets at the *IL11* locus. The *IL11* promoter which has reduced accessibility in glucocorticoid treated islets at high dose is highlighted.

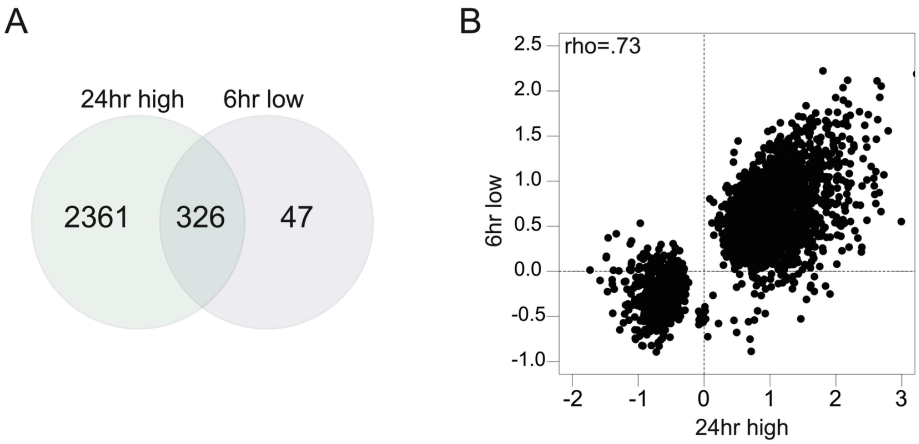


Figure S2.6. Differential chromatin accessibility in high- and low-dose glucocorticoid treatment. (A) Venn diagram of overlap in sites with differential activity in high-dose (100ng/mL for 24hr, n=6) and low-dose (4ng/mL for 6hr, n=3) glucocorticoid treatment. (B) Effects of high-dose and low-dose glucocorticoid treatment on sites with significant differential activity in either treatment.

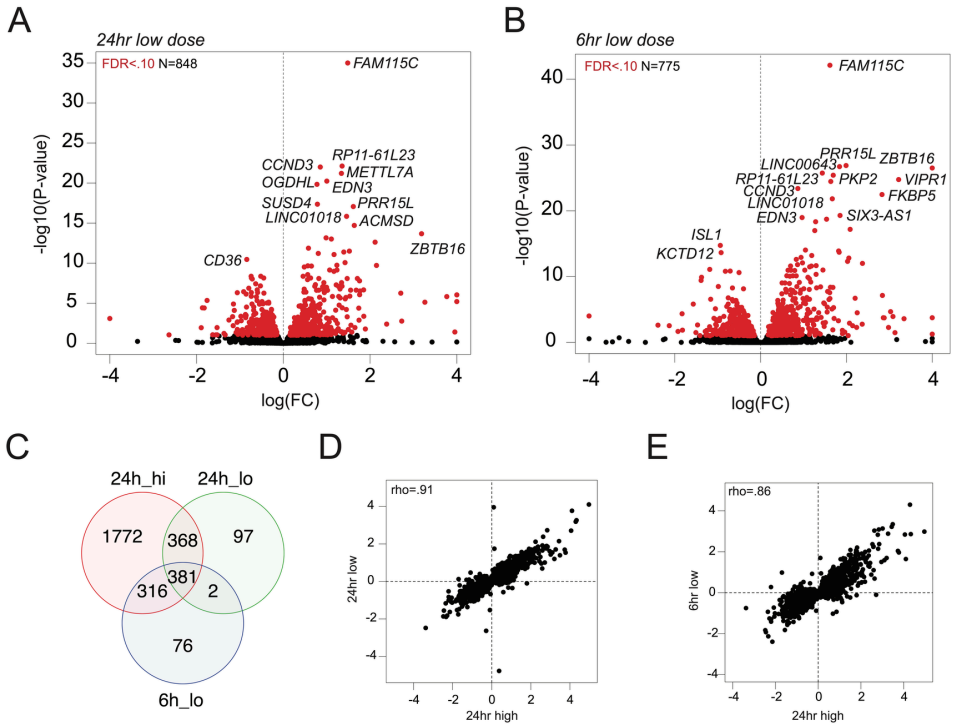


Figure S2.7. Differential gene expression in high- and low-dose glucocorticoid treatment. (A,B) Volcano plot of differential gene expression in glucocorticoid-treated islets at low dose for 24hr or 6hr compared to untreated islets. Genes with significant differential expression (FDR<.10) are highlighted in red, and genes with most pronounced changes in expression are listed. (C) Venn diagram of overlap between genes differentially expressed in 24hr high (n=6), 24hr low (n=3), 6hr low (n=3) glucocorticoid treatment. (D) Effects of 24hr high- and low-dose treatment on genes with significant differential expression in either treatment. (E) Effects of 24hr high- and 6hr low-dose treatment on genes with significant differential expression in either treatment.

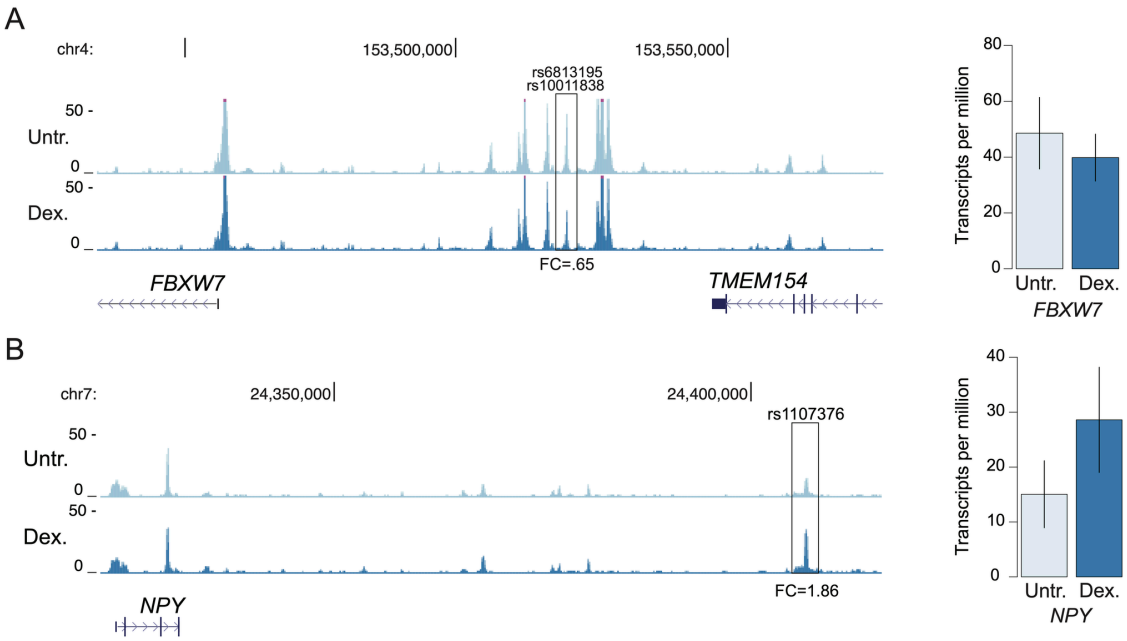


Figure S2.8. T2D-associated variants in differential chromatin sites. (A) Multiple variants at the *FBXW7/TMEM154* locus mapped in a site with decreased activity and *FBXW7* had decreased expression in glucocorticoid stimulation. (B) A variant at the *NPY* locus mapped in a site with increased activity and NPY had increased expression in glucocorticoid stimulation. Genome browser tracks represent RPKM normalized ATAC-seq signal, and expression bar plots represent mean expression and standard error. Values shown are from high-dose treatment. The fold-change (FC) in accessible chromatin signal in glucocorticoid treatment compared to untreated is indicated at highlighted sites.

2.7 Data Availability

The authors confirm that all data underlying the findings are fully available without restriction. All raw data are available from the GEO database (GSE167250). All data underlying graphs are provided in the main text or as Supporting Information in the online version of this article (https://doi.org/10.1371/journal.pgen.1009531).

2.8 Acknowledgements

We thank the UC San Diego IGM Genomics Center and the anonymous reviewers of the manuscript.

Chapter 2, in full, is a reformatted reprint of material as it appears in appears in Aylward, A., Okino, M., Benaglio, P., Chiou, J., Beebe, E., Padilla, J.A., Diep, S., Gaulton, K.J. Glucocorticoid signaling in pancreatic islets modulates gene regulatory programs and genetic risk of type 2 diabetes. *PLOS Genet.* **17**, e1009531 (2021). The dissertation author was a primary investigator and author of this paper.

2.9 Author Information

Anthony Aylward[1,#], Mei-Lin Okino[2,#], Paola Benaglio[2], Joshua Chiou[3], Elisha Beebe[2], Jose Andres Padilla[2], Sharlene Diep[2], Kyle J Gaulton[2,4,*]

1. Bioinformatics and Systems Biology graduate program, University of California San Diego, La Jolla, California, USA
2. Department of Pediatrics, University of California San Diego, La Jolla, California, USA
3. Biomedical Sciences graduate program, University of California San Diego, La Jolla, California, USA
4. Institute for Genomic Medicine, University of California San Diego, La Jolla, California, USA

# Authors contributed equally to this work

* kgaulton@ucsd.edu

K.J.G. designed the study; K.J.G, A.J.A., P.B., and J.C. wrote the manuscript and performed genetic and genomic analyses; M.O., E.B., J.A.P, and S.D. performed experiments and contributed to analyses.

2.10 References

1. Becker DE. Basic and clinical pharmacology of glucocorticosteroids. *Anesth Prog*. 2013;60(1):25-31; quiz 32. doi:10.2344/0003-3006-60.1.25

2. Coutinho AE, Chapman KE. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. *Mol Cell Endocrinol*. 2011;335(1):2-13. doi:10.1016/j.mce.2010.04.005

3. Patel R, Williams-Dautovich J, Cummins CL. Minireview: new molecular mediators of glucocorticoid receptor activity in metabolic tissues. *Mol Endocrinol*. 2014;28(7):999-1011. doi:10.1210/me.2014-1062

4. Bauerle KT, Harris C. Glucocorticoids and Diabetes. *Mo Med*. 2016;113(5):378-383.

5. Suh S, Park MK. Glucocorticoid-Induced Diabetes Mellitus: An Important but Overlooked Problem. *Endocrinol Metab (Seoul)*. 2017;32(2):180-189. doi:10.3803/EnM.2017.32.2.180

6. Colvin ES, Ma H-Y, Chen Y-C, Hernandez AM, Fueger PT. Glucocorticoid-induced suppression of β-cell proliferation is mediated by Mig6. *Endocrinology*. 2013;154(3):1039-1046. doi:10.1210/en.2012-1923

7. Fine NHF, Doig CL, Elhassan YS, Vierra NC, Marchetti P, Bugliani M, Nano R, Piemonti L, Rutter GA, Jacobson DA, Lavery GG, Hodson DJ. Glucocorticoids Reprogram β-Cell Signaling to Preserve Insulin Secretion. *Diabetes*. 2018;67(2):278-290. doi:10.2337/db16-1356

8. Gesina E, Tronche F, Herrera P, Duchene B, Tales W, Czernichow P, Breant B. Dissecting the role of glucocorticoids on pancreas development. *Diabetes*. 2004;53(9):2322-2329. doi:10.2337/diabetes.53.9.2322

9. Liu X, Turban S, Carter RN, Ahmad S, Ramage L, Webster SP, Walker BR, Seckl JR, Morton NM. β-Cell-Specific Glucocorticoid Reactivation Attenuates Inflammatory β-Cell Destruction. *Front Endocrinol (Lausanne)*. 2014;5:165. doi:10.3389/fendo.2014.00165

10. Lambillotte C, Gilon P, Henquin JC. Direct glucocorticoid inhibition of insulin secretion. An in vitro study of dexamethasone effects in mouse islets. *J Clin Invest*. 1997;99(3):414-423. doi:10.1172/JCI119175

11. Ullrich S, Berchtold S, Ranta F, Seebohm G, Henke G, Lupescu A, Mack AF, Chao C-M, Su J, Nitschke R, Alexander D, Friedrich B, Wulff P, Kuhl D, Lang F. Serum- and glucocorticoid-inducible kinase 1 (SGK1) mediates glucocorticoid-induced inhibition of insulin secretion. *Diabetes*. 2005;54(4):1090-1099. doi:10.2337/diabetes.54.4.1090

12. Oakley RH, Cidlowski JA. The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease. *J Allergy Clin Immunol*. 2013;132(5):1033-1044. doi:10.1016/j.jaci.2013.09.007

13. Reddy TE, Pauli F, Sprouse RO, Neff NF, Newberry KM, Garabedian MJ, Myers RM. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res*. 2009;19(12):2163-2171. doi:10.1101/gr.097022.109

14. McDowell IC, Barrera A, D'Ippolito AM, Vockley CM, Hong LK, Leichter SM, Bartelt LC, Majoros WH, Song L, Safi A, Koçak DD, Gersbach CA, Hartemink AJ, Crawford GE, Engelhardt BE, Reddy TE. Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding. *Genome Res*. 2018;28(9):1272-1284. doi:10.1101/gr.233346.117

15. Vockley CM, D'Ippolito AM, McDowell IC, Majoros WH, Safi A, Song L, Crawford GE, Reddy TE. Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. *Cell*. 2016;166(5):1269-1281.e19. doi:10.1016/j.cell.2016.07.049

16. Ling J, Kumar R. Crosstalk between NFkB and glucocorticoid signaling: a potential target of breast cancer therapy. *Cancer Lett*. 2012;322(2):119-126. doi:10.1016/j.canlet.2012.02.033

17. De Bosscher K, Vanden Berghe W, Vermeulen L, Plaisance S, Boone E, Haegeman G. Glucocorticoids repress NF-kappaB-driven genes by disturbing the interaction of p65 with the basal transcription machinery, irrespective of coactivator levels in the cell. *Proc Natl Acad Sci USA*. 2000;97(8):3919-3924. doi:10.1073/pnas.97.8.3919

18. Hazlehurst JM, Gathercole LL, Nasiri M, Armstrong MJ, Borrows S, Yu J, Wagenmakers AJM, Stewart PM, Tomlinson JW. Glucocorticoids fail to cause insulin resistance in human subcutaneous adipose tissue in vivo. *J Clin Endocrinol Metab*. 2013;98(4):1631-1640. doi:10.1210/jc.2012-3523

19. Jubb AW, Boyle S, Hume DA, Bickmore WA. Glucocorticoid Receptor Binding Induces Rapid and Prolonged Large-Scale Chromatin Decompaction at Multiple Target Loci. *Cell Rep*. 2017;21(11):3022-3031. doi:10.1016/j.celrep.2017.11.053

20. John S, Sabo PJ, Thurman RE, Sung M-H, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet*. 2011;43(3):264-268. doi:10.1038/ng.759

21. Grøntved L, John S, Baek S, Liu Y, Buckley JR, Vinson C, Aguilera G, Hager GL. C/EBP maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements. *EMBO J*. 2013;32(11):1568-1583. doi:10.1038/emboj.2013.106

22. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Mägi

R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss MH, Prins BP, Guo X, Bielak LF, Below JE, Bowden DW, Chambers JC, Kim YJ, Ng MCY, Petty LE, Sim X, Zhang W, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Ec Kardt K-U, Fischer K, Kardia SLR, Kronenberg F, Läll K, Liu C-T, Locke AE, Luan J, Ntalla I, Nylander V, Schönherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jørgensen ME, Jørgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stančáková A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50(11):1505-1513. doi:10.1038/s41588-018-0241-6

23. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, Rivas MA, Perry JRB, Sim X, Blackwell TW, Robertson NR, Rayner NW, Cingolani P, Locke AE, Tajes JF, Highland HM, Dupuis J, Chines PS, Lindgren CM, Hartl C, Jackson AU, Chen H, Huyghe JR, van de Bunt M, Pearson RD, Kumar A, Müller-Nurasyid M, Grarup N, Stringham HM, Gamazon ER, Lee J, Chen Y, Scott RA, Below JE, Chen P, Huang J, Go MJ, Stitzel ML, Pasko D, Parker SCJ, Varga TV, Green T, Beer NL, Day-Williams AG, Ferreira T, Fingerlin T, Horikoshi M, Hu C, Huh I, Ikram MK, Kim B-J, Kim Y, Kim YJ, Kwon M-S, Lee J, Lee S, Lin K-H, Maxwell TJ, Nagai Y, Wang X, Welch RP, Yoon J, Zhang W, Barzilai N, Voight BF, Han B-G, Jenkinson CP, Kuulasmaa T, Kuusisto J, Manning A, Ng MCY, Palmer ND, Balkau B, Stančáková A, Abboud HE, Boeing H, Giedraitis V, Prabhakaran D, Gottesman O, Scott J, Carey J, Kwan P, Grant G, Smith JD, Neale BM, Purcell S, Butterworth AS, Howson JMM, Lee HM, Lu Y, Kwak S-H, Zhao W, Danesh J, Lam VKL, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536(7614):41-47. doi:10.1038/nature18642

24. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, Mahajan A, Locke A, Rayner NW, Robertson N, Scott RA, Prokopenko I, Scott LJ, Green T, Sparso T, Thuillier D, Yengo L, Grallert H, Wahl S, Frånberg M, Strawbridge RJ, Kestler H, Chheda H, Eisele L, Gustafsson S, Steinthorsdottir V, Thorleifsson G, Qi L, Karssen LC, van Leeuwen EM, Willems SM, Li M, Chen H, Fuchsberger C, Kwan P, Ma C, Linderman M, Lu Y, Thomsen SK, Rundle JK, Beer NL, van de Bunt M, Chalisey A, Kang HM, Voight BF, Abecasis GR, Almgren P, Baldassarre D, Balkau B, Benediktsson R, Blüher M, Boeing H, Bonnycastle LL, Bottinger EP, Burtt NP, Carey J, Charpentier G, Chines PS, Cornelis MC, Couper DJ, Crenshaw AT, van Dam RM, Doney ASF, Dorkhan M, Edkins S, Eriksson JG, Esko T, Eury E, Fadista J, Flannick J, Fontanillas P, Fox C, Franks PW, Gertow K, Gieger C, Gigante B, Gottesman O, Grant GB, Grarup N, Groves CJ, Hassinen M, Have CT, Herder C, Holmen OL, Hreidarsson AB, Humphries SE, Hunter DJ, Jackson AU, Jonsson A, Jørgensen ME, Jørgensen T, Kao W-HL, Kerrison ND, Kinnunen L, Klopp N, Kong A, Kovacs P, Kraft P, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet*. 2015;47(12):1415-1425. doi:10.1038/ng.3437

25. Gaulton KJ. Mechanisms of Type 2 Diabetes Risk Loci. *Curr Diab Rep*. 2017;17(9):72. doi:10.1007/s11892-017-0908-x

26. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman İ, Tena JJ, Morán I, Gómez-Marín C, van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T,

Cebola I, García-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Skarmeta JLG, Müller F, McCarthy MI, Ferrer J. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet*. 2014;46(2):136-143. doi:10.1038/ng.2870

27. Chiou J, Zeng C, Cheng Z, Han JY, 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 KJ. *Single Cell Chromatin Accessibility Reveals Pancreatic Islet Cell Type- and State-Specific Regulatory Programs of Diabetes Risk*. Genomics; 2019. doi:10.1101/693671

28. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, Berney T, Montanya E, Mohlke KL, Lieb JD, Ferrer J. A map of open chromatin in human pancreatic islets. *Nat Genet*. 2010;42(3):255-259. doi:10.1038/ng.530

29. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, Welch R, Parker SCJ, Boyle AP, Scott LJ, NISC Comparative Sequencing Program, Margulies EH, Boehnke M, Furey TS, Crawford GE, Collins FS. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab*. 2010;12(5):443-455. doi:10.1016/j.cmet.2010.09.012

30. Varshney A, Scott LJ, Welch RP, Erdos MR, Chines PS, Narisu N, Albanus RD, Orchard P, Wolford BN, Kursawe R, Vadlamudi S, Cannon ME, Didion JP, Hensley J, Kirilusha A, NISC Comparative Sequencing Program, Bonnycastle LL, Taylor DL, Watanabe R, Mohlke KL, Boehnke M, Collins FS, Parker SCJ, Stitzel ML. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci USA*. 2017;114(9):2301-2306. doi:10.1073/pnas.1621192114

31. Parker SCJ, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, van Bueren KL, Chines PS, Narisu N, NISC Comparative Sequencing Program, Black BL, Visel A, Pennacchio LA, Collins FS, 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 USA*. 2013;110(44):17921-17926. doi:10.1073/pnas.1317023110

32. Greenwald WW, Chiou J, Yan J, Qiu Y, Dai N, Wang A, Nariai N, Aylward A, Han JY, Kadakia N, Regue L, Okino M-L, Drees F, Kramer D, Vinckier N, Minichiello L, Gorkin D, 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. *Nat Commun*. 2019;10(1):2078. doi:10.1038/s41467-019-09975-4

33. Miguel-Escalada I, Bonàs-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Atla G, Javierre BM, Rolando DMY, Farabella I, Morgan CC, García-Hurtado J, Beucher A, Morán I, Pasquali L, Ramos-Rodríguez M, Appel EVR, Linneberg A, Gjesing AP, Witte DR, Pedersen O, Grarup N, Ravassard P, Torrents D, Mercader JM, Piemonti L, Berney T, de Koning EJP, Kerr-Conte J, Pattou F, Fedko IO, Groop L, Prokopenko I, Hansen T, Marti-Renom MA, Fraser P, Ferrer J. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat Genet*. 2019;51(7):1137-1148. doi:10.1038/s41588-019-0457-0

34. Ramos-Rodríguez M, Raurell-Vila H, Colli ML, Alvelos MI, Subirana-Granés M, Juan-Mateu J, Norris R, Turatsinze J-V, Nakayasu ES, Webb-Robertson B-JM, Inshaw JRJ, Marchetti P, Piemonti L, Esteller M, Todd JA, Metz TO, Eizirik DL, Pasquali L. The impact of proinflammatory cytokines on the β-cell regulatory landscape provides insights into the genetics of type 1 diabetes. *Nat Genet*. 2019;51(11):1588-1595. doi:10.1038/s41588-019-0524-6

35. Wasim M, Carlet M, Mansha M, Greil R, Ploner C, Trockenbacher A, Rainer J, Kofler R. PLZF/ZBTB16, a glucocorticoid response gene in acute lymphoblastic leukemia, interferes with glucocorticoid-induced apoptosis. *J Steroid Biochem Mol Biol*. 2010;120(4-5):218-227. doi:10.1016/j.jsbmb.2010.04.019

36. Wasim M, Mansha M, Kofler A, Awan AR, Babar ME, Kofler R. Promyelocytic leukemia zinc finger protein (PLZF) enhances glucocorticoid-induced apoptosis in leukemic cell line NALM6. *Pak J Pharm Sci*. 2012;25(3):617-621.

37. Fahnenstich J, Nandy A, Milde-Langosch K, Schneider-Merck T, Walther N, Gellersen B. Promyelocytic leukaemia zinc finger protein (PLZF) is a glucocorticoid- and progesterone-induced transcription factor in human endometrial stromal cells and myometrial smooth muscle cells. *Mol Hum Reprod*. 2003;9(10):611-623. doi:10.1093/molehr/gag080

38. Breen MS, Bierer LM, Daskalakis NP, Bader HN, Makotkine I, Chattopadhyay M, Xu C, Buxbaum Grice A, Tocheva AS, Flory JD, Buxbaum JD, Meaney MJ, Brennand K, Yehuda R. Differential transcriptional response following glucocorticoid activation in cultured blood immune cells: a novel approach to PTSD biomarker development. *Transl Psychiatry*. 2019;9(1):201. doi:10.1038/s41398-019-0539-x

39. Wang J, Zhu Z, Nolfo R, Elias JA. Dexamethasone regulation of lung epithelial cell and fibroblast interleukin-11 production. *Am J Physiol*. 1999;276(1):L175-185. doi:10.1152/ajplung.1999.276.1.L175

40. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet*. 2014;94(4):559-573. doi:10.1016/j.ajhg.2014.03.004

41. Suzuki K, Akiyama M, Ishigaki K, Kanai M, Hosoe J, Shojima N, Hozawa A, Kadota A, Kuriki K, Naito M, Tanno K, Ishigaki Y, Hirata M, Matsuda K, Iwata N, Ikeda M, Sawada N, Yamaji T, Iwasaki M, Ikegawa S, Maeda S, Murakami Y, Wakai K, Tsugane S, Sasaki M, Yamamoto M, Okada Y, Kubo M, Kamatani Y, Horikoshi M, Yamauchi T, Kadowaki T. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat Genet*. 2019;51(3):379-386. doi:10.1038/s41588-018-0332-4

42. Spracklen CN, Horikoshi M, Kim YJ, Lin K, Bragg F, Moon S, Suzuki K, Tam CH, Tabara Y, Kwak S-H, Takeuchi F, Long J, Lim VJ, Chai J-F, Chen C-H, Nakatochi M, Yao J, Sun Choi H, Iyengar AK, Perrin HJ, Brotman SM, van de Bunt M, Gloyn AL, Below JE, Boehnke M, Bowden DW, Chambers JC, Mahajan A, McCarthy MI, Ng MC, Petty LE, Zhang W, Morris AP, Adair LS, Bian Z, Chan JC, Chang L-C, Chee M-L, Ida Chen Y-D, Chen Y-T, Chen Z, Chuang L-M, Du S, Gordon-Larsen P, Gross M, Guo X, Guo Y, Han S, Howard A-G, Huang W, Hung Y-J, Yeong Hwang M, Hwu C-M, Ichihara S, Isono M, Jang H-M, Jiang G, Jonas JB, Kamatani Y, Katsuya T, Kawaguchi T, Khor C-C, Kohara K, Lee M-S, Lee NR, Li L, Liu J, Luk AO, Lv J, Okada Y, Pereira MA, Sabanayagam C, Shi J, Mun Shin D, Yee So W, Takahashi A, Tomlinson B, Tsai F-J, van Dam RM, Xiang Y-B, Yamamoto K, Yamauchi T,

Yoon K, Yu C, Yuan J-M, Zhang L, Zheng W, Igase M, Shin Cho Y, Rotter JI, Wang Y-X, Sheu WH, Yokota M, Wu J-Y, Cheng C-Y, Wong T-Y, Shu X-O, Kato N, et al. *Identification of Type 2 Diabetes Loci in 433,540 East Asian Individuals*. Genetics; 2019. doi:10.1101/685172

43. Rajan AS, Aguilar-Bryan L, Nelson DA, Yaney GC, Hsu WH, Kunze DL, Boyd AE. Ion channels and insulin secretion. *Diabetes Care*. 1990;13(3):340-363. doi:10.2337/diacare.13.3.340

44. Jacobson DA, Shyng S-L. Ion Channels of the Islets in Type 2 Diabetes. *J Mol Biol*. Published online August 30, 2019. doi:10.1016/j.jmb.2019.08.014

45. Ye R, Gordillo R, Shao M, Onodera T, Chen Z, Chen S, Lin X, SoRelle JA, Li X, Tang M, Keller MP, Kuliawat R, Attie AD, Gupta RK, Holland WL, Beutler B, Herz J, Scherer PE. Intracellular lipid metabolism impairs β cell compensation during diet-induced obesity. *J Clin Invest*. 2018;128(3):1178-1189. doi:10.1172/JCI97702

46. Imai Y, Cousins RS, Liu S, Phelps BM, Promes JA. Connecting pancreatic islet lipid metabolism with insulin secretion and the development of type 2 diabetes. *Ann N Y Acad Sci*. 2020;1461(1):53-72. doi:10.1111/nyas.14037

47. Dickmeis T. Glucocorticoids and the circadian clock. *J Endocrinol*. 2009;200(1):3-22. doi:10.1677/JOE-08-0415

48. King EM, Chivers JE, Rider CF, Minnich A, Giembycz MA, Newton R. Glucocorticoid repression of inflammatory gene expression shows differential responsiveness by transactivation- and transrepression-dependent mechanisms. *PLoS ONE*. 2013;8(1):e53936. doi:10.1371/journal.pone.0053936

49. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, Kubo M, Okada Y, Kamatani Y. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet*. 2018;50(3):390-400. doi:10.1038/s41588-018-0047-6

50. Spracklen CN, Shi J, Vadlamudi S, Wu Y, Zou M, Raulerson CK, Davis JP, Zeynalzadeh M, Jackson K, Yuan W, Wang H, Shou W, Wang Y, Luo J, Lange LA, Lange EM, Popkin BM, Gordon-Larsen P, Du S, Huang W, Mohlke KL. Identification and functional analysis of glycemic trait loci in the China Health and Nutrition Survey. *PLoS Genet*. 2018;14(4):e1007275. doi:10.1371/journal.pgen.1007275

51. Carl M, Loosli F, Wittbrodt J. Six3 inactivation reveals its essential role for the formation and patterning of the vertebrate eye. *Development*. 2002;129(17):4057-4063.

52. Brodbeck S, Besenbeck B, Englert C. The transcription factor Six2 activates expression of the Gdnf gene as well as its own promoter. *Mech Dev*. 2004;121(10):1211-1222. doi:10.1016/j.mod.2004.05.019

53. Lagutin OV, Zhu CC, Kobayashi D, Topczewski J, Shimamura K, Puelles L, Russell HRC, McKinnon PJ, Solnica-Krezel L, Oliver G. Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev*. 2003;17(3):368-379. doi:10.1101/gad.1059403

54. He G, Tavella S, Hanley KP, Self M, Oliver G, Grifone R, Hanley N, Ward C, Bobola N. Inactivation of Six2 in mouse identifies a novel genetic mechanism controlling development and growth of the cranial base. *Dev Biol*. 2010;344(2):720-730. doi:10.1016/j.ydbio.2010.05.509

55. Samuel A, Rubinstein AM, Azar TT, Ben-Moshe Livne Z, Kim S-H, Inbal A. Six3 regulates optic nerve development via multiple mechanisms. *Sci Rep*. 2016;6:20267. doi:10.1038/srep20267

56. Steinmetz PR, Urbach R, Posnien N, Eriksson J, Kostyuchenko RP, Brena C, Guy K, Akam M, Bucher G, Arendt D. Six3 demarcates the anterior-most developing brain region in bilaterian animals. *Evodevo*. 2010;1(1):14. doi:10.1186/2041-9139-1-14

57. Arda HE, Li L, Tsai J, Torre EA, Rosli Y, Peiris H, Spitale RC, Dai C, Gu X, Qu K, Wang P, Wang J, Grompe M, Scharfmann R, Snyder MS, Bottino R, Powers AC, Chang HY, Kim SK. Age-Dependent Pancreatic Gene Regulation Reveals Mechanisms Governing Human β Cell Function. *Cell Metab*. 2016;23(5):909-920. doi:10.1016/j.cmet.2016.04.002

58. Reynolds MS, Hancock CR, Ray JD, Kener KB, Draney C, Garland K, Hardman J, Bikman BT, Tessem JS. β-Cell deletion of Nr4a1 and Nr4a3 nuclear receptors impedes mitochondrial respiration and insulin secretion. *Am J Physiol Endocrinol Metab*. 2016;311(1):E186-201. doi:10.1152/ajpendo.00022.2016

59. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324

60. 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. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352

61. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160-165. doi:10.1093/nar/gkw257

62. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137. doi:10.1186/gb-2008-9-9-r137

63. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res*. 2019;47(8):e47. doi:10.1093/nar/gkz114

64. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8

65. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-140. doi:10.1093/bioinformatics/btp616

66. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29. doi:10.1186/gb-2014-15-2-r29

67. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007

68. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-589. doi:10.1016/j.molcel.2010.05.004

69. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247

70. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

71. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, 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 JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012;22(9):1760-1774. doi:10.1101/gr.135350.111

72. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323. doi:10.1186/1471-2105-12-323

73. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102

74. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033

75. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795

76. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki A-E, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K,

Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PIW, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

77. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods*. 2015;12(11):1061-1063. doi:10.1038/nmeth.3582

78. Zhang Q, Keles S. An empirical Bayes test for allelic-imbalance detection in ChIP-seq. *Biostatistics*. 2018;19(4):546-561. doi:10.1093/biostatistics/kxx060

79. Aylward A, Chiou J, Okino M-L, Kadakia N, Gaulton KJ. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Hum Mol Genet*. Published online November 7, 2018. doi:10.1093/hmg/ddy314

80. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet*. 2007;81(2):208-227. doi:10.1086/519024

81. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

82. Viñuela A, Varshney A, van de Bunt M, Prasad RB, Asplund O, Bennett A, Boehnke M, Brown AA, Erdos MR, Fadista J, Hansson O, Hatem G, Howald C, Iyengar AK, Johnson P, Krus U, MacDonald PE, Mahajan A, Manning Fox JE, Narisu N, Nylander V, Orchard P, Oskolkov N, Panousis NI, Payne A, Stitzel ML, Vadlamudi S, Welch R, Collins FS, Mohlke KL, Gloyn AL, Scott LJ, Dermitzakis ET, Groop L, Parker SCJ, McCarthy MI. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat Commun*. 2020;11(1):4912. doi:10.1038/s41467-020-18581-8

83. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383. doi:10.1371/journal.pgen.1004383

CHAPTER 3: Allelic imbalance mapping improves fine-mapping of diabetes risk loci

3.1 Abstract

Disease risk loci identified in genome-wide association studies (GWAS) largely map to non-coding regions of the genome. Biological interpretation of non-coding signals requires determining the functional effects of risk variants on the epigenome such as chromatin accessibility and transcription factor binding. Allelic imbalance mapping (AIM) offers a resource-efficient approach for determining genetic effects on the epigenome as, unlike QTL mapping, it can be applied to data generated from few samples. In this work we develop a novel framework for AIM from epigenomic data across multiple experiments with shallow sequencing data with or without existing genotype data. We then demonstrate how the resulting AIM statistics can be used to interpret eQTLs and diabetes risk signals. First, we applied this approach to ChIP-seq data for 10 TFs from the ENCODE project and identified 84 fine-mapped liver eQTLs with evidence for AIM in at least one TF. Second, we applied this approach to ATAC-seq data generated from 5 pancreatic islet samples. Incorporating AIM statistics identified 298 pancreatic islet eQTLs and 31 T2D risk signals with evidence for AIM in chromatin accessibility, and we further demonstrated that T2D risk variants with AIM have higher causal probabilities in a larger, independent fine-mapping dataset. Together this work provides a flexible strategy for generating AIM from epigenomic data that can be applied to many existing datasets and that can interpret the molecular mechanisms of complex disease risk variants.

3.2 Introduction

To interpret the results of genome-wide association studies (GWAS), genetic variants associated with a phenotypic trait must be functionally characterized to establish the biological effect of variant on trait. However, due to linkage disequilibrium (LD) in the human genome, GWAS identify genomic loci associated with disease risk but often cannot precisely locate causal variants underlying the association[1]. Fine-mapping the specific causal variants underlying association signals is therefore a subject of considerable interest[2–4]. One advance in fine-mapping was the discovery that trait-associated variants are enriched with *cis*-regulatory elements of the genome, and integrating GWAS data with functional genomic data can greatly improve the resolution of causal variants[5]. Large initiatives such as the ENCODE and NIH Epigenome Roadmap project provide a wealth of epigenomic datasets and *cis*-regulatory elements in different tissues and cells that can be used for these analyses[6]. However, variants overlapping *cis*-regulatory element annotations might not necessarily directly affect the activity of the element, and therefore additional information is often needed to prioritize causal variants.

One phenotype for which GWAS have been especially fruitful is type 2 diabetes (T2D)[7–9], which is a highly prevalent complex disease affecting 400 million individuals worldwide. Over 400 genomic loci have been identified that influence risk of T2D. The majority (>90%) of associated variants at T2D risk loci map to non-coding regions, suggesting they affect gene regulation. T2D associated variants are enriched with *cis*-regulatory elements, and integration of T2D associated variants with epigenomic data has identified enrichment in key tissues such as pancreatic islets and refined the resolution of causal variants[7,8,10,11]. Furthermore, T2D associated variants at specific loci have been shown to have functional effects of regulatory activity in pancreatic islets and other relevant tissues[12–19]. Fine-mapping causal variants at these loci and understanding their function is a critical step in translating GWAS results into biomedical insights and novel

therapies[7,9]. However, the specific causal variants and biological mechanisms of most T2D-associated loci to date remain unknown.

Sequencing data from epigenomic experiments can be used to identify variants with functional effects on *cis*-regulatory activity by performing quantitative trait locus (QTL) mapping or allelic imbalance mapping. Mapping QTLs or allelic imbalance for molecular phenotypes such as transcription factor binding, accessible chromatin or gene expression can help determine the mechanisms of risk loci[20], since variants influencing complex phenotypes are likely to also affect these molecular traits[21]. QTL mapping measures a molecular trait across different samples and tests for correlation with variant genotypes[22]. For example, researchers may use RNA-seq to map gene expression QTLs (eQTLs)[23,24] or ATAC-seq to map chromatin QTLs (caQTLs)[25,26]. However, QTL mapping often requires large numbers of samples which may be prohibitive in studying molecular traits with limited available data such as transcription factor (TF) binding. Conversely, allelic imbalance mapping uses heterozygous variants to measure within-sample differences in allelic effects on a molecular trait[27–33]. When there is a significant difference in the observed number of reads from the two alleles, the variant is said to be imbalanced which implies a functional difference in the molecular activity of the alleles. Allelic imbalance mapping can be performed using fewer samples than QTL studies, even from a single sample. However, only variants that are heterozygous in the assayed samples can be tested and they suffer technical challenges including overdispersion and reference mapping bias. It is also of interest to quantify the magnitude of the effect of specific variant alleles on molecular traits. Efforts have been made to define generally applicable effect size statistics for QTL and imbalance studies[32,33]. Accurate estimation of allelic effects can likely help further improve fine-mapping of causal variants, but is challenging especially when using shallow sequence data.

A wide variety of published epigenomic sequencing data are available, including hundreds of datasets from sources such as the ENCODE project or ROADMAP epigenomics project[6,34]. These datasets are amenable to allelic imbalance analysis and could be used to better understand the function of trait-associated variants, but there are still obstacles. First, sequencing depth is often low, which together with overdispersion limits statistical power for calling imbalanced variants in a genome-wide context and accurately estimating their effects. Second, genotypes are often not available for these datasets, so heterozygous variants are not known. This has led to the development of methods for inferring genotypes directly from epigenomic data, including QuASAR[29]. Previous studies have circumvented this challenge by focusing on datasets with known genotypes or performing genotyping experiments on cell lines[30,35]. An additional consideration for cancer-derived cell lines is that some chromosomes may be polyploid[36,37]. Since diploidy is a critical assumption of standard imbalance analysis, care should be taken that only variants on diploid chromosomes are studied.

In this study, we demonstrate accurate inference of heterozygous genotypes from ChIP-seq data from a liver cell line HepG2 and ATAC-seq data from primary pancreatic islet samples using QuASAR. We test heterozygous variants for allelic imbalance, accounting for mapping bias and overdispersion, and show that imbalanced variants called from a given epigenomic dataset are enriched for relevant TF binding motifs. We then calculate imbalance effect sizes using a novel bayes estimator and show how the effect size statistics can be leveraged to interpret fine-mapping data for T2D risk signals. By comparing two independent fine-mapping datasets, we show that large imbalance effects are predictive of likely causal variants at T2D loci. Finally, we experimentally validate the functional effect of a fine-mapped variant with allelic imbalance in pancreatic islet accessible chromatin at the *HMG20A* locus suggesting it is a likely causal variant

for T2D risk. Together these methods and results should provide valuable insight into the genetic basis of complex traits and disease.

3.3 Results

3.3.1 Confidently inferring genotypes from epigenome sequencing data

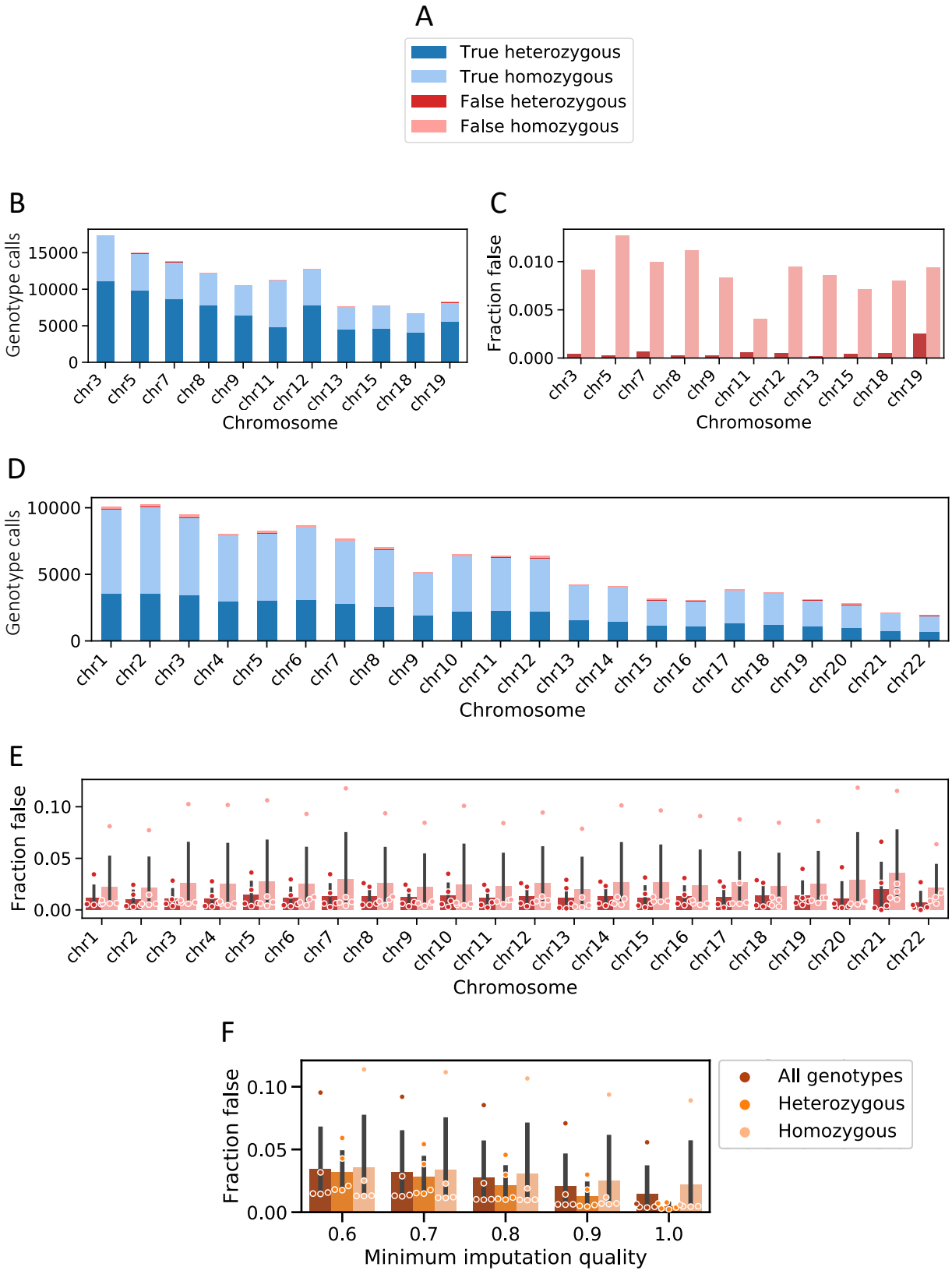
We inferred genotypes from ChIP-seq and ATAC-seq data using QuASAR[29]. The method starts with inputs consisting of a predefined set of variants with reference and alternate allele counts for each. It uses an EM-algorithm based procedure to compute three genotype probabilities for each variant (probability homozygous ref, homozygous alt, or heterozygous). For our subsequent analysis, we considered all variants genotyped with confidence >99% and discarded those genotyped with lower confidence (see methods for additional details).

To assess the accuracy of QuASAR, we applied it to two datasets with gold-standard genotypes available for comparison. First, we genotyped HepG2 using 15 high-quality ChIP-seq datasets from ENCODE (**Table S3.1**). For HepG2, all analysis was restricted to 11 diploid chromosomes (3, 5, 7, 8, 9, 11, 12, 13, 15, 18, 19). While chromosome 22 is also diploid in HepG2, we excluded it because of its loss of heterozygosity. We inferred genotypes for 286,944 SNPs and compared the results to 2,112,748 independent genotypes from whole-genome sequencing data of HepG2 cells[37]. There were 123,206 SNPs genotyped in common between QuASAR and WGS data. In addition, we analyzed ATAC-seq data from our previous study and compared the results to array-based genotypes[38]. We applied QuASAR to 5 pancreatic islet ATAC-seq datasets and inferred 139,048-346,685 genotypes per sample. To confirm the accuracy of these genotypes, we checked them against 508,598 genotypes obtained by arrays and others obtained by imputation into the HRC reference panel r1.1 using the Michigan Imputation Server (**see**

methods)[39]. There were 13,839-29,567 SNPs per sample genotyped in common between QuASAR and array results.

We declared QuASAR genotypes “false” when they did not match the corresponding gold-standard genotypes. We found that genotyping HepG2 with QuASAR had an overall false call rate of 0.4%, with a false heterozygous call rate of 0.06% and a false homozygous call rate of 0.9% (**Figure 3.1B-C**). In primary islets, when compared to direct array genotypes, QuASAR had a mean overall false call rate of 1.5% (95% CI: 0.4%-3.5%), false heterozygous call rate of 0.5% (0.3%-0.7%), and false homozygous call rate of 2.2% (0.5%-5.6%). When compared to imputed genotypes with minimum imputation quality 0.9, the mean overall false call rate was 2.1% (0.6%-4.7%), false heterozygous 1.3% (0.5%-2.3%), false homozygous 2.5% (0.7%-5.9%) (**Figure 3.1D-F**).

Figure 3.1: QuASAR accurately infers genotypes from ChIP-seq and ATAC-seq data. (A) legend for barplots in panels B-E. (B) Barplot of the number of true heterozygous, true homozygous, false heterozygous, and false homozygous genotype calls made by QuASAR in HepG2. (C) Barplot showing the rate of false heterozygous and false homozygous calls at each chromosome in HepG2. (D) barplot of the number of true heterozygous, true homozygous, false heterozygous, and false homozygous genotype calls made by QuASAR in 5 primary pancreatic islet samples (mean value shown). (E) Rates of false heterozygous calls and false homozygous calls across 5 islet samples. (F) Overall rate of false genotype calls from Quasar according to several minimum imputation quality levels for the array genotypes.


A
True heterozygous
True homozygous
False heterozygous
False homozygous
B
Genotype calls
15000
10000
5000
0
chr3 chr5 chr7 chr8 chr9 chr11 chr12 chr13 chr15 chr18 chr19
Chromosome
C
Fraction false
0.010
0.005
0.000
chr3 chr5 chr7 chr8 chr9 chr11 chr12 chr13 chr15 chr18 chr19
Chromosome
D
Genotype calls
10000
5000
0
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
Chromosome
E
Fraction false
0.10
0.05
0.00
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22
Chromosome
F
Fraction false
0.10
0.05
0.00
0.6 0.7 0.8 0.9 1.0
Minimum imputation quality
All genotypes
Heterozygous
Homozygous

3.3.2 Calling and quantifying allelic imbalance

To prepare the HepG2 TF ChIP-seq and islet ATAC-seq datasets for allelic imbalance mapping, we first applied WASP to correct for reference bias at reads mapping to each heterozygous variant[40]. We then obtained read counts for each variant allele. Next, to account for overdispersion, we used NPBin to estimate a beta-binomial null distribution using read counts for each dataset. ChIP-seq data were slightly more over-dispersed than ATAC-seq data. Estimated overdispersion in HepG2 ChIP-seq datasets had a mean of 0.016 (95% CI: 0.013-0.019), while estimates from primary islet ATAC-seq datasets had a mean of 0.011 (0.010-0.013) (**Figure 3.2A**). We called imbalanced variants from the merged counts using a beta-binomial test, calculated q-values from the resulting beta-binomial p-values, and considered variants significant at FDR < .10.

To call imbalanced variants we assumed a beta-binomial distribution for reference allele counts, estimated the beta parameters using NPBin, and performed beta-binomial tests (see methods)[31]. We considered variants significantly imbalanced if they had a false discovery rate <.1. In HepG2 ChIP-seq data, detection of imbalanced variants was dependent on the number of TF-bound sites identified in the data (**Figure 3.2B, Table 3.1**). Only variants which were in a ChIP-seq site and had sufficient sequencing coverage (10x coverage) were used for imbalance analysis. The number of TF-bound sites was highly correlated both with the total number of variants tested (spearman r=0.68, p=0.029) and the number of imbalanced variants detected (spearman r=0.79, p=0.006). Among individual islet ATAC-seq samples, the number of open chromatin sites identified per sample was also highly correlated with the number of SNPs tested (spearman r=0.60, p=0.210) and the number of imbalanced SNPs detected (spearman r=0.60, p=0.210), although these estimates were not significant (**Figure 3.2C, Table 3.2**).

For three experiments, we identified at least 50 variants with significantly imbalance (FDR<.10): CEBPB ChIP-seq in HepG2, CREB1 ChIP-seq in HepG2, and islet ATAC-seq. In these experiments, we tested imbalanced SNPs (FDR<.10): for enrichment with TF motifs relative to background SNPs with balanced allele counts (FDR≥.10). In the HepG2 ChIP-seq experiments, imbalanced SNPs were enriched with motifs relevant to the ChIP-seq target such as CEBPA/B/G for CEBPG ChIP-seq (**Figure 3.2D**) and ATF2/3 for CREB1 ChIP-seq (**Figure 3.2E**). For islet ATAC-Seq, imbalanced variants were enriched with motifs for TFs with functions in pancreatic beta cells, such as SP1[41,42(p3)] and RREB1[43(p1),44(p1)]. Other strongly enriched motifs included ZNF281, ZNF740, ZSCAN4, and SP2 (**Figure 3.2F**).

We identified few variants with significant allelic imbalance in most datasets, which is not surprising given the shallow sequencing of the assays and the small number of samples tested. Therefore, we next sought to quantify the magnitude of imbalance for all tested variants for each experiment regardless of whether the variant reached significance. For each variant tested in a given experiment, we calculated two related measures of effect size: log allelic fold-change (laFC) and Bayes-estimated log allelic fold change (B-laFC) (**Figure 3.3A-C**). laFC is a previously published statistic designed to provide a universally applicable effect size for allelic imbalance[32]. B-laFC is a novel bayes estimator of laFC (**see methods**). Where laFC uses only the ratio of alternate to reference alleles to estimate the effect, B-laFC assumes a beta prior distribution on the alt/ref ratio. The prior is empirically estimated from the data across all variants and incorporated into the estimator. laFC and B-laFC are roughly equivalent at SNPs where read depth is high, but when read depth is low B-laFC values are reduced in magnitude. This makes B-laFC more robust to overdispersion and ensures that extreme B-laFC values are correlated with *bona fide* imbalance effects. In our observations, the distribution and properties of B-laFC

were similar across diverse datasets, suggesting that like IaFC, B-IaFC offers a useful measure that can be easily interpreted in a variety of contexts.

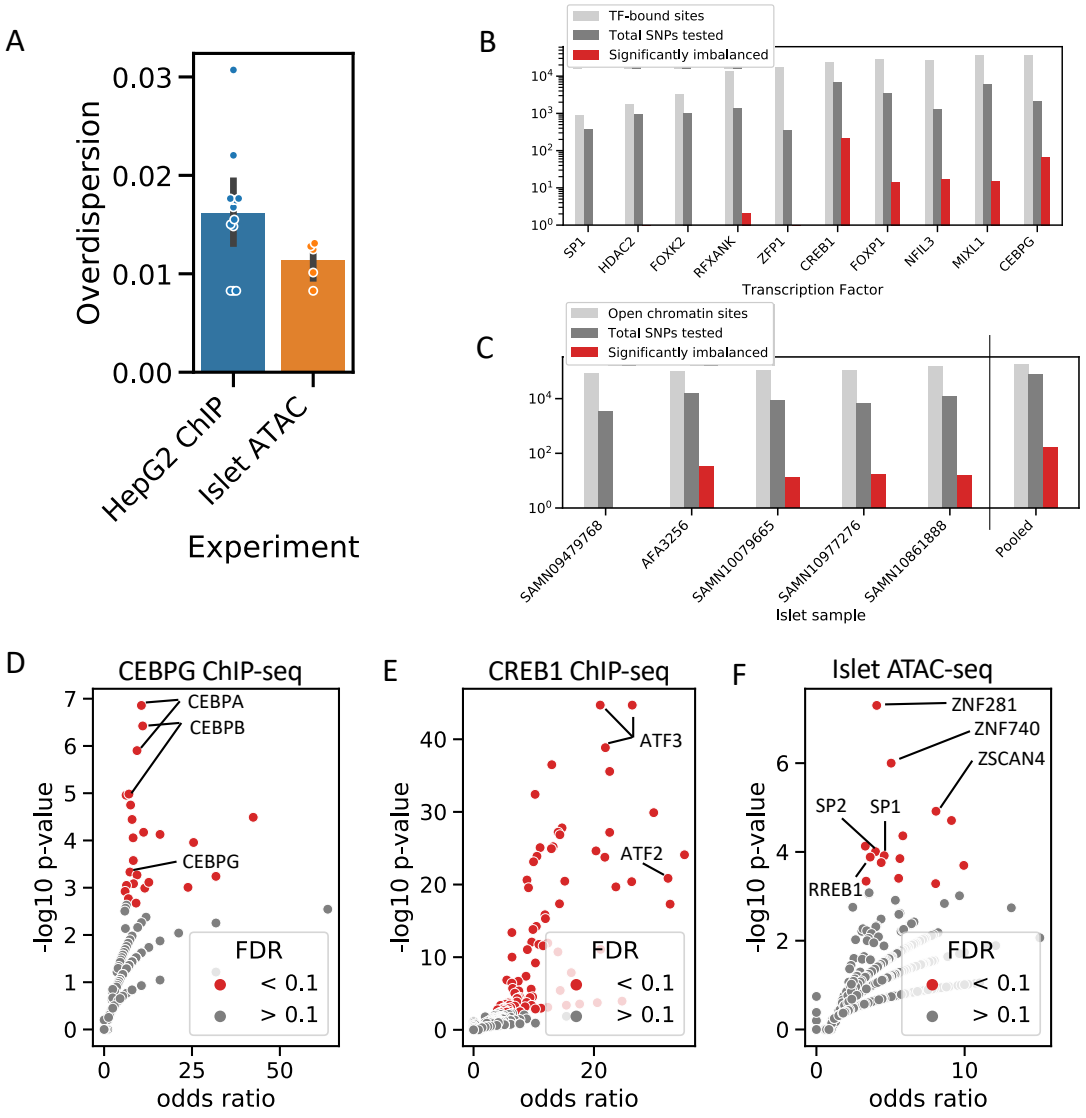


Figure 3.2: Calling and quantifying allelic imbalance. (A) Overdispersion parameters for null distributions estimated from HepG2 ChIP-seq and islet ATAC-seq experiments by NPBin. (B) Total TF-bound sites, SNPs tested for imbalance, and significantly imbalanced SNPs for HepG2 ChIP-seq experiments. (C) Total open chromatin sites, SNPs tested for imbalance, and significantly imbalanced SNPs for islet ATAC-seq experiments. (D-F) enrichment of imbalanced variants with motifs in HepG2 ChIP-seq and islet ATAC-seq data.

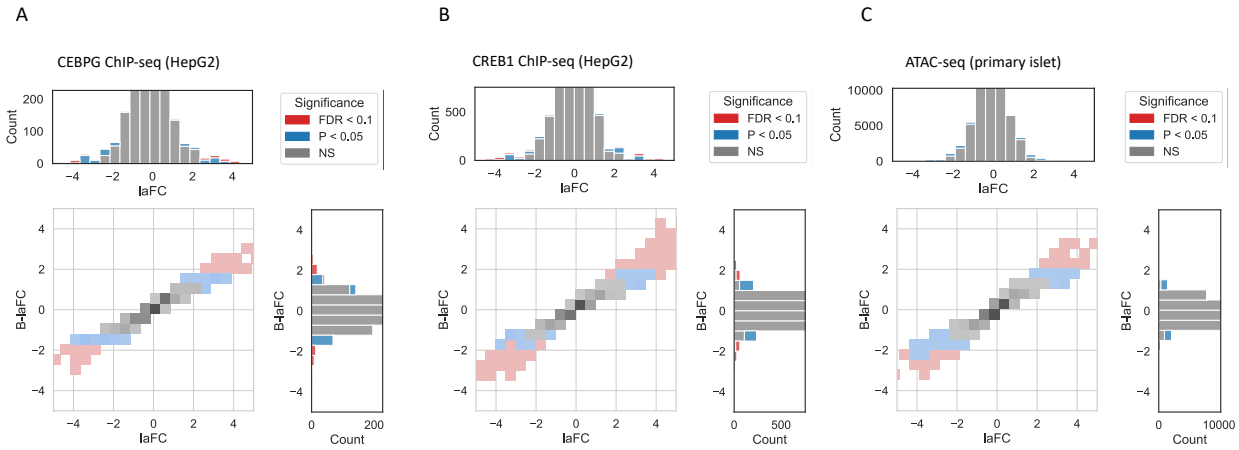


Figure 3.3: Comparison of IaFC vs B-IaFC. (A-C) Bivariate histograms of IaFC vs. B-IaFC for HepG2 ChIP-seq of CEBPG (A) and CREB1 (B) and primary islet ATAC-seq (C).

Table 3.1: Basic imbalance statistics for HepG2 ChIP-seq datasets.

ChIP-seq peaks	Total SNPs tested	Significantly imbalanced SNPs	Transcription Factor	Replicates
889	373	0	SP1	1
1768	945	1	HDAC2	1
3195	1028	0	FOXK2	1
13722	1371	2	RFXANK	2
16982	341	1	ZFP1	1
23030	6621	208	CREB1	1
28127	3398	14	FOXP1	2
27151	1268	17	NFIL3	1
36215	6019	15	MIXL1	2
36581	2199	64	CEBPG	1

Table 3.2: Basic imbalance statistics for pancreatic islet ATAC-seq datasets.

ATAC-seq peaks	Total SNPs tested	Significantly imbalanced SNPs	Sample ID	Replicates
80854	3528	0	SAMN09479768	1
94647	15171	35	AFA3256	1
103020	8660	13	SAMN10079665	1
109933	6906	18	SAMN10977276	1
154059	12181	16	SAMN10861888	1
182970	74915	174	Pooled	5

3.3.3 Imbalance mapping augments fine-mapping and interpretation of eQTLs and diabetes risk loci

To evaluate the utility of allelic fold-change (B-IaFC) statistics in fine-mapping causal variants, we first intersected them with eQTL fine-mapping data in human liver (GTEx) and pancreatic islets (Inspire; Vineula et al). Since we were concerned with the magnitude of effects but not with their direction, we compared PPA to the absolute value of B-IaFC for all subsequent analyses. For each locus, we considered variants to have "high PPA" if they were included in a 50% credible set. In liver, we identified SNPs with high fine-mapping PPA and moderate (|B-IaFC|=0.5-1.5, 73 SNPs) or strong (|B-IaFC| >1.5, 8 SNPs) imbalance effects. In islets, fine-mapped eQTLs included 284 moderately imbalanced and 14 strongly imbalanced SNPs.

For liver eQTLs, imbalance in TF ChIP-seq data in HepG2 cells is useful because it suggests specific mechanisms of transcriptional regulation potentially driving effects on gene expression. For example, expression of non-coding RNA *LINC00265* may be regulated by binding of CREB1 at variant rs386712410 (PPA=.87, |B-IaFC|=2.10) and expression of TF *ZNF780A* may be regulated by binding of CEBPG at variant rs337799 (PPA=.25, |B-IaFC|=2.85) (**Figure 3.4A-B**).

For islet eQTLs, in multiple cases we observed that well-known eQTLs fine-mapped to a single SNP had allelic imbalance in islet ATAC-seq data, for example rs11257655 (PPA=.53, |B-IaFC|=.62) at the *CDC123* locus and rs11708067 at the *ADCY5* locus (PPA=1.0, |B-IaFC|=1.13) (**Figure 3.4C-D**). In the case of rs11257655 this variant has been previously shown to affect regulatory activity in pancreatic islets cells, supporting a functional role in gene

expression[45(p123),46(p5)]. At other loci, where fine-mapping credible sets included multiple candidate causal variants, B-IaFC values added additional information that could be used to prioritize specific SNPs for experimental validation, for example variant rs1999068 (PPA=.0.33, |B-IAFC|=1.53) at the *TDRD5* locus and rs709062 (PPA=.34, |B-IaFC|=1.13) at the *ZNF467* locus both had evidence for allelic imbalance (**Figure 3.4E-F**).

We next compared B-IaFC statistics to fine-mapping credible sets of 239 T2D-associated loci from DIAMANTE GWAS data. A set of 481 variants were present in credible sets and had B-IaFC values available for islet ATAC-seq data. We first sought to determine whether or not evidence for allelic imbalance (high B-IAFC value) was correlated with the causal probability of variants. We observed that average T2D PPA values were higher for variants with high absolute B-IaFC (**Figure 3.5A**). Furthermore, variants with high B-IaFC magnitude were significantly more likely to have high causal probabilities (PPA > 0.5) compared to other tested variants (logistic regression P=0.001) (**Figure 3.5B**). We performed a similar analysis to compare B-IaFC values to credible sets of islet eQTLs, and found a less pronounced but still significant correlation (P=0.007) (**Figure S3.1**).

Next, we asked if B-IaFC values could be used effectively to augment fine-mapping to improve identification of causal variants. We used previous fine-mapping of 107 T2D loci from Metabochip, GoT2D and DIAGRAM 1000G GWAS data, all precursors to the DIAMANTE GWAS dataset generated from smaller sample sizes (**Table 3.3**)[7,19,47,48]. We identified 22 T2D-associated loci at which we could easily draw a one-to-one correspondence between fine-mapping signals in Metabochip/GoT2D/DIAGRAM vs. DIAMANTE. Across those loci there were 82 variants with data available for B-IaFC and PPA in Metabochip/GoT2D/DIAGRAM fine-mapping. Of those, 20 had increased PPA in DIAMANTE fine-mapping relative to Metabochip/GoT2D/DIAGRAM fine-

mapping, while 62 had decreased PPA in the larger fine-mapping. We determined that variants with high absolute B-IaFC were significantly more likely to have their PPA increased in the DIAMANTE fine-mapping (logistic regression P=0.022) (**Figure 3.6A-C**).

Finally, we determined if allelic imbalance defined using B-IaFC could inform prioritization of variants likely causal for T2D at specific loci. At the *HMG20A* locus fine-mapping resolved 24 variants all with modest PPA at best (PPA=0.007-0.075). Among these variants, we identified a single variant rs34591043 with evidence for allelic imbalance in islets (|B-IaFC|=1.01), suggesting this variant is potentially causal for the T2D risk signal (**Figure 3.6D-F**). We therefore tested this variant for effects on islet enhancer activity using gene reporter assays in the MIN6 cell line. We identified a significant effect of rs34591043 alleles on enhancer activity where the alternate allele G had increased activity (**Figure 3.6G**). The G allele was also associated with increased expression of *HMG20A* in islet eQTL data and was predicted to bind an ETV6 TF motif.

Together these results reveal that allelic imbalance mapping of shallow sequence data from epigenomic assays of few samples can help prioritize variants affecting *cis*-regulatory activity and risk of T2D and other complex disease.

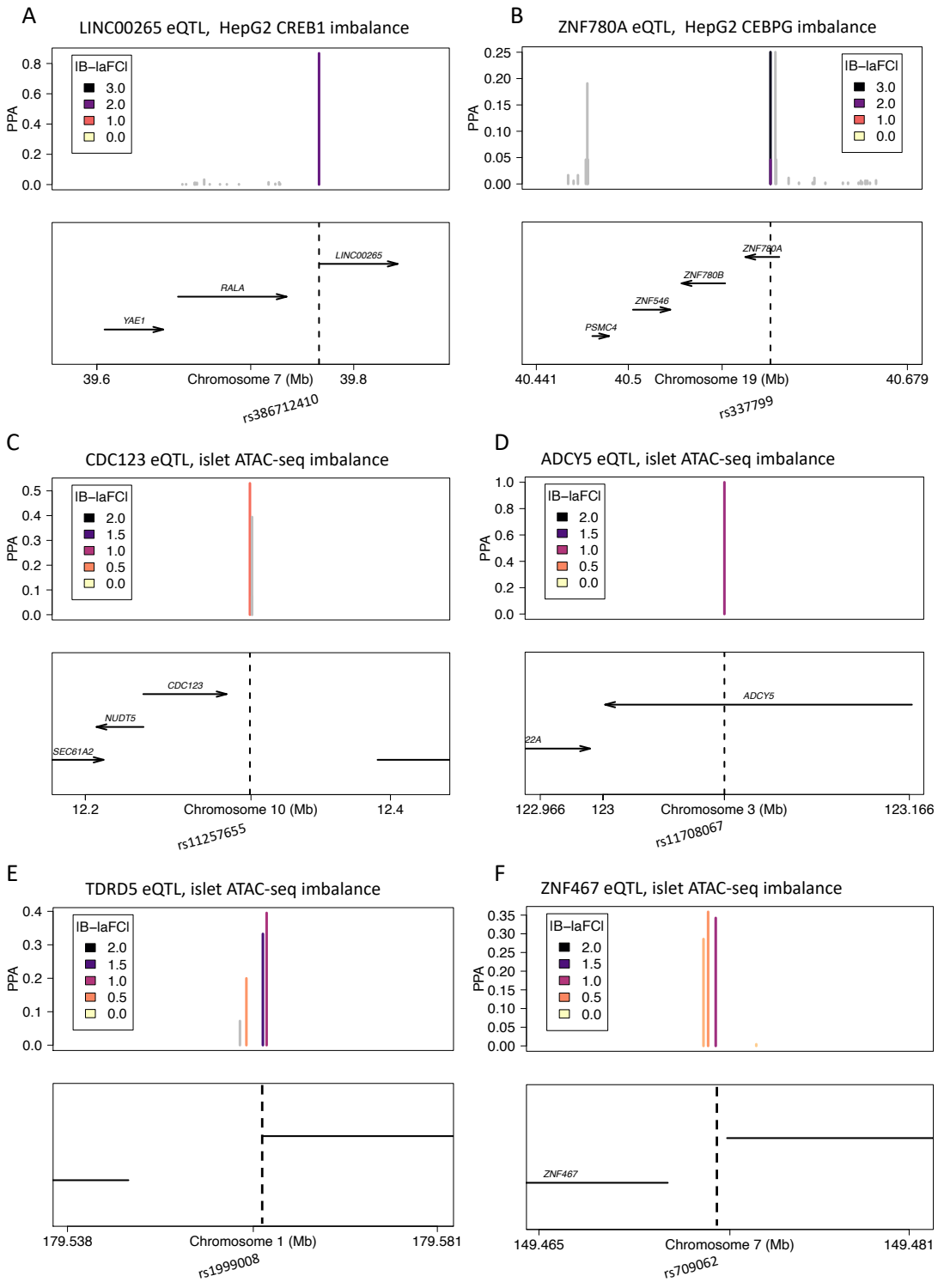


Figure 3.4: Likely causal SNPs with evidence for allelic imbalance. (A-B) fine-mapping of Liver eQTLs for expression of: (A) LINC00265, SNP with imbalanced CREB1 binding in HepG2 shown (B) for ZNF780A in HepG2, SNP with imbalanced CEBPG binding in HepG2 shown. (C-F) fine-mapping of pancreatic islet eQTLs, SNPs with imbalanced chromatin accessibility shown: (C) CDC123, (D) ADCY5, (E) TDRD5, (F) ZNF467

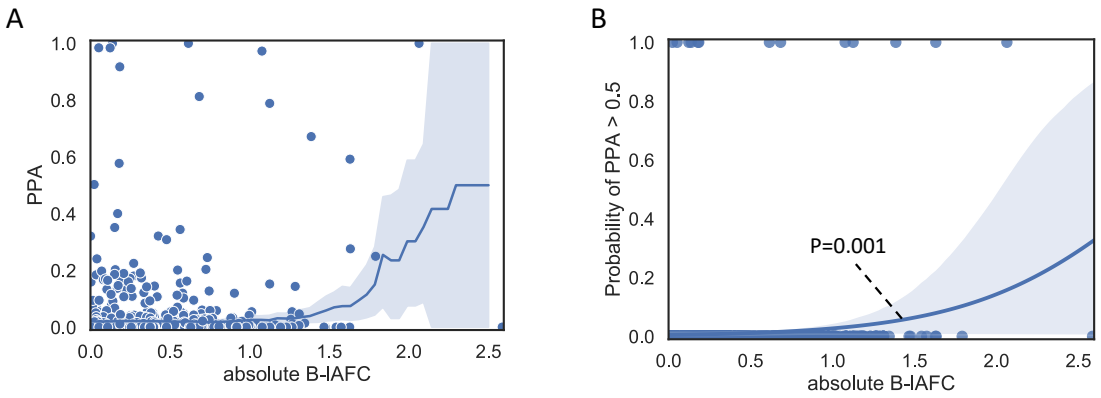


Figure 3.5: High absolute B-laFC values are correlated with likely causal T2D variants. (A) Scatterplot of absolute B-laFC values against PPA for SNPs across fine-mapped T2D loci. Smoothed mean values with 95% confidence intervals shown. (B) logistic regression of absolute B-laFC against the probability of PPA>0.5. Bootstrapping-based 95% confidence intervals shown.

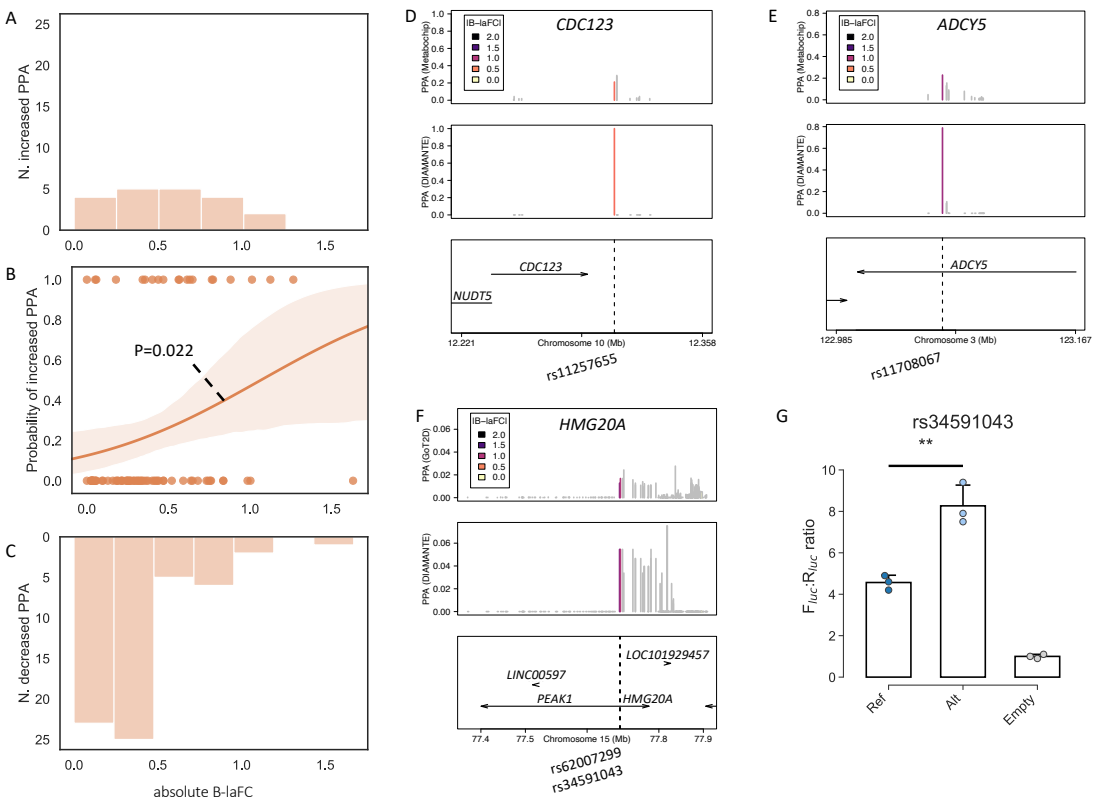


Figure 3.6: High absolute B-laFC values predict likely causal variants. (A) Histogram of B-laFC values for SNPs with PPA (DIAMANTE) > PPA (DIAGRAM) (B) logistic regression of B-laFC against the probability of PPA (DIAMANTE) > PPA (DIAGRAM) for SNPs across loci. (C) Histogram of B-laFC values for SNPs with PPA (DIAMANTE) < PPA (DIAGRAM). (D-F) comparison of DIAGRAM fine-mapping to DIAMANTE fine mapping at (D) CDC123 (E) *ADCY5* (F) *HMG20A*. (G) Luciferase reporter assay of *HMG20A* SNP rs34591043 in MIN6 cells.

Table 3.3: Sample sizes of T2D GWAS studies

Study	Sample size
Metabochip	84,780
GoT2D	111,548
DIAGRAM	159,208
DIAMANTE	898,130

3.4 Discussion

In this study we have developed a framework for identifying imbalanced SNPs in ChIP-seq, ATAC-seq, or other forms of epigenomic data with shallow sequencing and with or without genotyping data. Independent genotypes may be used to identify heterozygous SNPs, or they may be inferred directly from the epigenomic data using QuASAR[29]. Reference mapping bias in allele counts is corrected using WASP[40], and then SNPs are statistically tested for imbalance using a beta-binomial test based on NPBin[31]. Finally, we estimate effect sizes for imbalanced SNPs using a novel bayes estimator of allelic fold-change (B-IAFC)[32].

In the original publication of QuASAR, the authors showed its application to RNA-seq data and suggested that it could also be applied to other types of data including ChIP-seq, ATAC-seq and others. Here we have demonstrated QuASAR's application to ChIP-seq and ATAC-seq data. We validated the results by comparing genotypes inferred by QuASAR from epigenomic data to independent genotypes from whole-genome sequencing[37(p2)] or genotyping arrays. We genotyped hundreds of thousands of SNPs from a small collection of HepG2 ChIP-seq datasets or primary pancreatic islet ATAC-seq datasets. The default settings of QuASAR are strict when making heterozygous calls and relatively liberal for homozygous calls. While this means that some truly heterozygous SNPs will be missed during genotyping, it satisfies an important prerequisite for subsequent imbalance analysis in ensuring that variants passed for subsequent imbalance mapping are truly heterozygous with high confidence.

Allelic imbalance analysis cannot replace molecular QTL studies, but it is a useful complementary approach. Importantly, imbalance analysis is effective when sample sizes are small, while QTL studies are challenging if fewer than 20 samples are available. The low sample size requirement makes imbalance analysis especially attractive for sequencing experiments

involving multiple cell types or conditions, or for multiple tested transcription factors, where collecting enough samples in each condition for a QTL study would be more difficult[38,49]. While only small samples are required for allelic imbalance mapping, increasing the sample size does offer increasing returns. Multiple samples with identical genotypes (e.g. from cell lines) can increase power by providing additional reads at each SNP, while samples with distinct genotypes (e.g. from multiple donors of primary cells) provide access to a greater number of heterozygous SNPs.

To estimate effect sizes of allelic imbalance, we have proposed B-laFC, a bayes estimator of the previously published log allelic fold-change (laFC)[32]. B-laFC inherits from laFC its convenient biological interpretation and mathematical properties. It differs from laFC in that effect sizes are subject to a shrinkage factor when read counts are small, and therefore provides more robust estimation of allelic effects from assays with shallow sequencing. Given the substantial number of molecular assays with shallow sequence data in ENCODE, NIH Roadmap and GEO, this metric should be widely applicable. B-laFC values have a comparable distribution across multiple data types, and they prioritize the accuracy of large effects. Given these properties future studies can leverage B-laFC values for multiple additional analyses. First, as the global properties of each individual assay are used in calculating B-laFC values, these values can be more naturally combined across different assays for the same TF or histone mark in a meta-analysis. Second, accurate estimates of allelic imbalance can be used directly in genome-wide disease enrichment analyses to help identify TFs relevant to disease risk, including those with directional effects for example where decreased binding of a TF may increased disease risk.

Expression QTL and GWAS studies have identified thousands of loci associated with gene expression, disease, or other traits. In particular, the GTEx project has identified eQTLs in a

variety of tissues, and large GWAS studies of T2D have been generated[9,50]. Integrating allelic imbalance effects with fine-mapping data can prioritize likely causal SNPs for functional validation experiments. By combining HepG2 ChIP-seq data with GTEx liver eQTL data and connecting eQTLs to allelic effects on specific transcription factors, we can uncover elements of the biological mechanisms of gene regulation. By combining primary islet ATAC-seq data with islet eQTL data, we have shown how imbalance effects from several samples can augment fine-mapping across many loci. Finally, we have demonstrated that imbalance statistics can predict likely causal variants at T2D risk loci and functionally validated one such example at the *HMG20A* locus. This method can therefore effectively leverage existing epigenomic data to facilitate the identification of causal variants underlying complex disease risk and understand their molecular mechanisms.

3.5 Methods

ChIP-seq and ATAC-seq datasets

We downloaded 15 HepG2 ChIP-seq datasets representing 10 transcription factors from ENCODE (**Table S3.1**), and we reprocessed 5 primary pancreatic islet ATAC-seq datasets from our previous work (**Table S3.2**)[6,38]. All data were aligned to GRCh37/hg19 using BWA mem with default parameters. After sequence alignment, reads were filtered by several steps: (i) Low-quality reads were removed using “samtools view -bh -F 1804 -q 10”, (2) Supplementary alignments were removed using “samtools view -bh -F 2048”, (3) Reads overlapping ENCODE blacklisted regions were filtered out, (4) for HepG2 ChIP-seq datasets, reads were filtered to only those from the 11 diploid chromosomes: 3, 5, 7, 8, 9, 11, 12, 13, 15, 18, 19. For primary islet ATAC-seq, reads were filtered to only those from autosomes. Finally, duplicate reads were removed using the unbiased method from WASP[40].

Independent genotypes

HepG2 genotypes from whole-genome sequencing were downloaded from ENCODE[6]. For pancreatic islets, we used genotyping array data corresponding to samples from our previous study[38]. We called genotypes using GenomeStudio (v.2.0.4) with default settings. We then used PLINK to filter out variants with 1) minor allele frequency (MAF) less than 0.01 in the Haplotype Reference Consortium (HRC) panel r1.1 and 2) ambiguous A/T or G/C alleles with MAF greater than 0.4[51]. For variants that passed these filters, we imputed genotypes into the HRC reference panel r1.1 using the Michigan Imputation Server with minimac4[39]. Post imputation, we removed imputed genotypes with low imputation quality (R2 < .9).

ChIP-seq and ATAC-seq -based genotyping

We used QuASAR to infer genotypes from ChIP-seq or ATAC-seq data as described in the QuASAR documentation[29]. Since QuASAR requires a predefined list of SNPs for genotyping, we used the "core set of 1KGP SNPs" recommended by the authors. For each dataset, we prepared a QuASAR input file, and we inferred genotypes using the quasar functions "fitAseNull()" (for individual primary islet samples) and "fitAseNullMulti()" (for grouped HepG2 samples). We filtered the QuASAR results to include only genotypes with at least 99% confidence and converted them to VCF format for subsequent analysis.

Allelic imbalance mapping

To maximize the number of SNPs available for analysis, we generated a merged set of heterozygous sites from both QuASAR and independent genotyping results. We then used WASP to correct mapping bias across this set[40]. For each ChIP-seq or ATAC-seq dataset, we identified variants with read depth of at least 10 and limited subsequent analysis to these variants. For HepG2 ChIP-seq datasets with two replicates available, we mapped imbalance on both replicates

individually and pooled. For islet ATAC-seq datasets, we mapped imbalance on individual samples as well as on merged read counts from all 5 samples. To identify imbalanced variants, we fit a beta-binomial model to the observed allele counts using the method of NPBin[31]. See table S3.3 for estimated parameter values. The formula for calculating overdispersion from the beta-binomial parameters was:

$$\frac{1}{1+\alpha+\beta}$$

We called imbalanced variants from the merged counts using a beta-binomial test, and then calculated q-values from the resulting beta-binomial p-values using the Benjamini-Hochberg procedure. We considered variants significant at FDR < .10.

Motif enrichment analysis

For the three experiments with more than 50 imbalanced variants (HepG2 CEBPG ChIP-seq, HepG2 CREB1 ChIP-seq, islet ATAC-seq), we divided variants into two groups: imbalanced (imbalance FDR < .10) and background (FDR > .10). For each group, we used FIMO from the MEME suite to find TF motifs in a 30bp window around each variant[52,53]. We discarded motifs that did not directly overlap a variant. For each motif found, we counted how many times the motif overlapped a variant in the imbalanced set and how many times in the background set. We compared the counts to the total number of variants each set and used a Fisher exact test to determine enrichment of the motif in the imbalanced set. We calculated false discovery rates using the Benjamini-Hochberg procedure, and considered motifs significantly enriched in the imbalanced set at FDR < .10.

Estimating imbalance effect sizes

We adopt log allelic fold change (laFC) as the effect size, as previously proposed[32]. Its formula for SNP *i* is:

$$\text{laFC} = \log_2 \frac{n_i - x_i}{x_i}$$

Where n_i is the read coverage at SNP *i* and x_i is the reference allele count. To derive a bayes estimator form for laFC, we assume that the ref allele fraction x_i/n_i approximates a "ground truth" value of p_i:

$$\frac{x_i}{n_i} \approx p_i$$

laFC is then an approximation of its own ground truth value which is:

$$laFC_{true} = \log_2 \frac{1 - p_i}{p_i}$$

We assume a beta-distributed prior for *pi* with parameters α, β

. We can then write a bayes estimator $\hat{p}_i$ for p_i:

$$\hat{p}_i = \frac{\alpha + x_i}{\alpha + \beta + n_i} \approx p_i$$

The estimator B-laFC is then:

$$\text{B-laFC} = \log_2 \frac{1 - \hat{p}_i}{\hat{p}_i}$$

Or, in terms of x_i and n_i:

$$\text{B-laFC} = \log_2 \frac{\beta + n_i - x_i}{\alpha + x_i}$$

In practice, we determine α and β by fitting a beta distribution to the distribution of ref allele fractions across all SNPs, using the method of alleledb[35]. This allows for one final improvement: We add a shift term which accounts for any residual reference bias that may be present in the data. The final B-laFC is then:

$$\text{B-laFC} = \log_2 \frac{\beta + n_i - x_i}{\alpha + x_i} - \log_2 \frac{\beta}{\alpha}$$

Fine-mapping eQTL and GWAS data

We downloaded liver eQTL mapping data from GTEx and pancreatic islet eQTL data from InsPIRE[50,54]. For each locus in each eQTL dataset, we converted beta and standard error statistics for each variant to approximate Bayes Factors[55]. We calculated the posterior probability of association (PPA) for each variant by dividing its Bayes Factor by the sum of Bayes Factors across the entire locus and determined 99% credible sets by taking the set of variants with PPA values summing to 99%. For the primary T2D GWAS data analysis, we downloaded a published collection of credible sets from the DIAMANTE consortium[9]. To compare the DIAMANTE sets against older data we used a fine mapping dataset compiled for a previous study, which consisted of 107 uniformly-processed loci drawn from the Metabochip, GoT2D, and DIAGRAM 1000 Genomes studies[7,19,47,48].

Gene reporter assay

To test for allelic differences in enhancer activity at the *HMG20A* locus, we cloned human DNA sequences (Integrated DNA Technologies) containing the reference allele upstream of the minimal promoter in the luciferase reporter vector pGL4.23 (Promega) using the enzymes Sac I and Kpn I. A construct containing the alternate allele was then created using the NEB Q5 SDM kit (New England Biolabs). The primer sequences used were as follows:

- Cloning FWD GATGCCCTTCACCCCTTGAA
- Cloning REV GCACCAAGCACCACCTTTTC
- SDM FWD TCCAGTGCTGGTGAGGGCTTG
- SDM REV AGTGGGTCGGACACCCCC

MIN6 cells were seeded at approximately 2.5E05 cells/cm^2 into a 48-well plate. The day after passaging into the 48-well plate, cells were co-transfected with 250ng of experimental firefly luciferase vector pGL4.23 containing the alt or ref allele in the forward direction or an empty pGL4.23 vector, and 15ng pRL-SV40 Renilla luciferase vector (Promega) using the Lipofectamine

3000 reagent. Cells were lysed 48 hours post transfection and assayed using the Dual-Luciferase Reporter system (Promega). Firefly activity was normalized to Renilla activity and normalized results were expressed as fold change compared to the luciferase activity of the empty vector. A two-sided t-test was used to compare the luciferase activity between the reference and alternative alleles.

3.6 Supplementary Figures

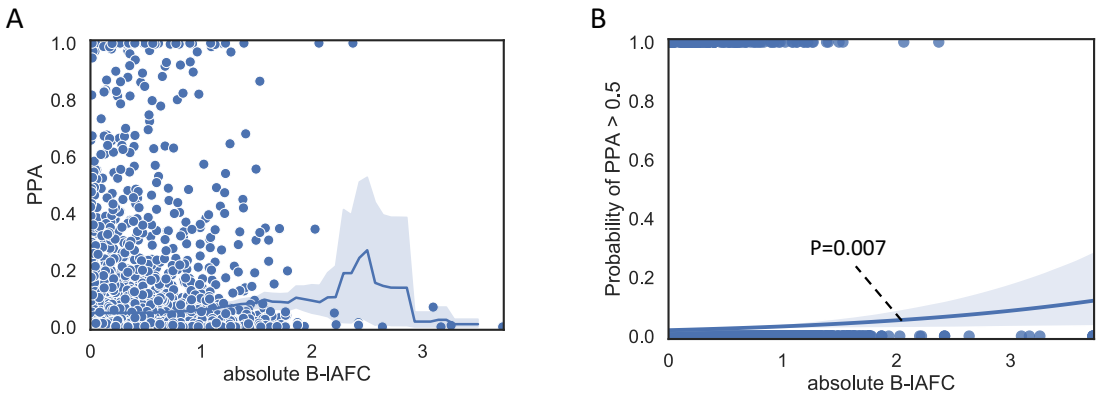


Figure S1: High absolute B-IaFC values are correlated with likely causal islet eQTL variants. (A) Scatterplot of absolute B-IaFC values against PPA for SNPs across fine-mapped islet eQTL loci. Smoothed mean values with 95% confidence intervals shown. (B) logistic regression of B-IaFC against the probability of PPA>0.5. Bootstrapping-based 95% confidence intervals shown.

3.7 Supplementary Tables

Table S3.1: HepG2 ChIP-seq datasets.

Experiment	Library	Type	Transcription Factor
ENCSR112ALD	ENCLB955KHX	Paired-end	CREB1
ENCSR171FUX	ENCLB729WHK	Paired-end	FOXK2
ENCSR337NWW	ENCLB492HKP	Paired-end	HDAC2
ENCSR334KIQ	ENCLB382UET	Paired-end	SP1
ENCSR639IIZ	ENCLB582SNX	Single-end	CEBPG
ENCSR369YUK	ENCLB433JGB	Single-end	FOXP1
ENCSR369YUK	ENCLB126INR	Single-end	FOXP1
ENCSR966PJY	ENCLB275ULF	Single-end	MIXL1
NCSR966PJY	ENCLB128AOL	Single-end	MIXL1
ENCSR201GGK	ENCLB523LQM	Single-end	NFIL3
ENCSR823ADL	ENCLB695ZQY	Single-end	RFXANK
ENCSR823ADL	ENCLB835RDE	Single-end	RFXANK
ENCSR586DEH	ENCLB455NSX	Single-end	ZFP1

Table S3.2: Pancreatic islet ATAC-seq datasets.

GEO accession	Sample ID
GSM5100317	SAMN10079665
GSM5100319	AFA3256
GSM5100321	SAMN09479768
GSM5100323	SAMN10861888
GSM5100325	SAMN10977276

Table S3.3: Beta-binomial parameters estimated by NPBin.

Experiment	Cell	Sample	Alpha	Beta	overdispersion
CEBPG ChIP-seq	HepG2	ENCFF430LJV	22.4238	21.9565	0.0220
CREB1 ChIP-seq	HepG2	ENCFF199RYE	33.3769	32.2108	0.0150
FOXK2 ChIP-seq	HepG2	ENCFF009QNH	28.2289	27.4473	0.0176
FOXP1 ChIP-seq	HepG2	ENCFF452XLY	16.2135	15.3497	0.0307
FOXP1 ChIP-seq	HepG2	ENCFF734HBO	28.6279	27.5632	0.0175
FOXP1 ChIP-seq	HepG2	Pooled	23.6746	22.5184	0.0212
HDAC2 ChIP-seq	HepG2	ENCFF489LNL	59.9625	59.9625	0.0083
MIXL1 ChIP-seq	HepG2	ENCFF616WVB	28.6279	27.5632	0.0175
MIXL1 ChIP-seq	HepG2	ENCFF790RMU	32.758	30.7095	0.0155
MIXL1 ChIP-seq	HepG2	Pooled	32.0755	30.6234	0.0157
NFIL3 ChIP-seq	HepG2	ENCFF170AOL	29.9153	28.7488	0.0168
RFXANK ChIP-seq	HepG2	ENCFF895HBS	59.9625	59.9625	0.0083
RFXANK ChIP-seq	HepG2	ENCFF958ONK	59.9625	59.9625	0.0083
RFXANK ChIP-seq	HepG2	Pooled	59.9625	59.9625	0.0083
SP1 ChIP-seq	HepG2	ENCFF124RBK	29.1143	26.538	0.0177
ZFP1 ChIP-seq	HepG2	ENCFF868BZO	34.1671	32.3468	0.0148
ATAC-seq	Islet	AFA3256	39.427	37.801	0.0128
ATAC-seq	Islet	SAMN09479768	38.4432	36.9096	0.0131
ATAC-seq	Islet	SAMN10079665	50.619	47.1217	0.0101
ATAC-seq	Islet	SAMN10861888	40.266	38.9502	0.0125
ATAC-seq	Islet	SAMN10977276	59.9625	59.9625	0.0083
ATAC-seq	Islet	Pooled	37.0134	35.9381	0.0135

3.8 Data and software availability

The HepG2 ChIP-seq and whole-genome sequencing datasets used in this study are available from the ENCODE project (https://www.encodeproject.org). Table S3.1 contains accession numbers for the ChIP-seq datasets. The Whole-genome sequencing data were from biosample ENCBS760ISV, experiment ENCSR319QHO, dataset ENCFF713BPG. The pancreatic islet ATAC-seq and genotyping datasets are available from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo) as series GSE167250. The liver cis-eQTL mapping data are available from the GTEx project (https://gtexportal.org/home) from GTEx Analysis V7. All T2D GWAS summary statistic data are available from the DIAGRAM consortium (https://www.diagram-consortium.org).

The code used to perform QuASAR-based genotyping is available in the python package pyQuASAR-genotype on github (https://github.com/anthony-aylward/pyQuASAR_genotype) and PyPI (https://pypi.org/project/pyQuASAR-genotype). The code used for calling imbalance with NPBin is available on github (https://github.com/anthony-aylward/npbin)

3.9 Acknowledgements

Chapter 3, in part, is currently being prepared for submission for publication of the material. Anthony Aylward, Mei-Lin Okino, Joshua Chiou, Jaspreet Kaur, Yleia Sanchez, Kyle J Gaulton. Allelic imbalance mapping improves fine-mapping of diabetes risk loci. The dissertation author was a primary investigator of this paper.

3.10 Author information

Anthony Aylward[1], Mei-Lin Okino[2], Joshua Chiou[3], Jaspreet Kaur[2], Yleia Sanchez[2], Kyle J Gaulton[2,4,*]. *In preparation.*

1. Bioinformatics and Systems Biology graduate program, University of California San Diego, La Jolla, California, USA
2. Department of Pediatrics, University of California San Diego, La Jolla, California, USA
3. Biomedical Sciences graduate program, University of California San Diego, La Jolla, California, USA
4. Institute for Genomic Medicine, University of California San Diego, La Jolla, California, USA

A.J.A and Y.S. collected published sequencing datasets for this study. K.J.G and A.J.A., wrote the manuscript. A.J.A, and J.C. performed genetic and genomic analyses. M.O. and performed experiments and contributed to analyses.

3.11 References

1. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res*. 2012;22(9):1748-1759. doi:10.1101/gr.136127.111

2. Broekema RV, Bakker OB, Jonkers IH. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biology*. 10(1):190221. doi:10.1098/rsob.190221

3. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 2018;19(8):491-504. doi:10.1038/s41576-018-0016-z

4. Tehranchi A, Hie B, Dacre M, Kaplow I, Pettie K, Combs P, Fraser HB. Fine-mapping cis-regulatory variants in diverse human populations. Morris AP, Wittkopp PJ, eds. *eLife*. 2019;8:e39595. doi:10.7554/eLife.39595

5. Cano-Gamez E, Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet*. 2020;11. doi:10.3389/fgene.2020.00424

6. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature*. 2012;489(7414):57-74. doi:10.1038/nature11247

7. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, Mahajan A, Locke A, Rayner NW, Robertson N, Scott RA, Prokopenko I, Scott LJ, Green T, Sparso T, Thuillier D, Yengo L, Grallert H, Wahl S, Frånberg M, Strawbridge RJ, Kestler H, Chheda H, Eisele L, Gustafsson S, Steinthorsdottir V, Thorleifsson G, Qi L, Karssen LC, Leeuwen EM van, Willems SM, Li M, Chen H, Fuchsberger C, Kwan P, Ma C, Linderman M, Lu Y, Thomsen SK, Rundle JK, Beer NL, Bunt M van de, Chalisey A, Kang HM, Voight BF, Abecasis GR, Almgren P, Baldassarre D, Balkau B, Benediktsson R, Blüher M, Boeing H, Bonnycastle LL, Bottinger EP, Burtt NP, Carey J, Charpentier G, Chines PS, Cornelis MC, Couper DJ, Crenshaw AT, Dam RM van, Doney ASF, Dorkhan M, Edkins S, Eriksson JG, Esko T, Eury E, Fadista J, Flannick J, Fontanillas P, Fox C, Franks PW, Gertow K, Gieger C, Gigante B, Gottesman O, Grant GB, Grarup N, Groves CJ, Hassinen M, Have CT, Herder C, Holmen OL, Hreidarsson AB, Humphries SE, Hunter DJ, Jackson AU, Jonsson A, Jørgensen ME, Jørgensen T, Kao W-HL, Kerrison ND, Kinnunen L, Klopp N, Kong A, Kovacs P, Kraft P, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet*. 2015;47(12):1415-1425. doi:10.1038/ng.3437

8. Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, Yengo L, Lloyd-Jones LR, Sidorenko J, Wu Y, McRae AF, Visscher PM, Zeng J, Yang J. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications*. 2018;9(1):2941. doi:10.1038/s41467-018-04951-w

9. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Mägi R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss MH, Prins BP, Guo X, Bielak LF, Below JE, Bowden DW, Chambers JC, Kim YJ, Ng MCY, Petty LE, Sim X, Zhang W, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Ec Kardt K-U, Fischer K, Kardia SLR, Kronenberg

F, Läll K, Liu C-T, Locke AE, Luan J, Ntalla I, Nylander V, Schönherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jørgensen ME, Jørgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stančáková A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50(11):1505-1513. doi:10.1038/s41588-018-0241-6

10. Thurner M, van de Bunt M, Torres JM, Mahajan A, Nylander V, Bennett AJ, Gaulton KJ, Barrett A, Burrows C, Bell CG, Lowe R, Beck S, Rakyan VK, Gloyn AL, McCarthy MI. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *Elife*. 2018;7. doi:10.7554/eLife.31977

11. Gaulton KJ. Mechanisms of Type 2 Diabetes Risk Loci. *Curr Diab Rep*. 2017;17(9):72. doi:10.1007/s11892-017-0908-x

12. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman İ, Tena JJ, Morán I, Gómez-Marín C, Bunt M van de, Ponsa-Cobas J, Castro N, Nammo T, Cebola I, García-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Skarmeta JLG, Müller F, McCarthy MI, Ferrer J. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet*. 2014;46(2):136-143. doi:10.1038/ng.2870

13. Ndiaye FK, Ortalli A, Canouil M, Huyvaert M, Salazar-Cardozo C, Lecoeur C, Verbanck M, Pawlowski V, Boutry R, Durand E, Rabearivelo I, Sand O, Marselli L, Kerr-Conte J, Chandra V, Scharfmann R, Poulain-Godefroy O, Marchetti P, Pattou F, Abderrahmani A, Froguel P, Bonnefond A. Expression and functional assessment of candidate type 2 diabetes susceptibility genes identify four new genes contributing to human insulin secretion. *Mol Metab*. 2017;6(6):459-470. doi:10.1016/j.molmet.2017.03.011

14. Cebola I. Pancreatic Islet Transcriptional Enhancers and Diabetes. *Curr Diab Rep*. 2019;19(12):145. doi:10.1007/s11892-019-1230-6

15. Chen Z, Yuan W, Liu T, Huang D, Xiang L. Bioinformatics analysis of hepatic gene expression profiles in type 2 diabetes mellitus. *Experimental and Therapeutic Medicine*. 2019;18(6):4303-4312. doi:10.3892/etm.2019.8092

16. Christodoulou M-I, Avgeris M, Kokkinopoulou I, Maratou E, Mitrou P, Kontos CK, Pappas E, Boutati E, Scorilas A, Fragoulis EG. Blood-based analysis of type-2 diabetes mellitus susceptibility genes identifies specific transcript variants with deregulated expression and association with disease risk. *Scientific Reports*. 2019;9(1):1512. doi:10.1038/s41598-018-37856-1

17. Miguel-Escalada I, Bonàs-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Atla G, Javierre BM, Rolando DMY, Farabella I, Morgan CC, García-Hurtado J, Beucher A, Morán I, Pasquali L, Ramos-Rodríguez M, Appel EVR, Linneberg A, Gjesing AP, Witte DR, Pedersen O, Grarup N, Ravassard P, Torrents D, Mercader JM, Piemonti L, Berney T, de

Koning EJP, Kerr-Conte J, Pattou F, Fedko IO, Groop L, Prokopenko I, Hansen T, Marti-Renom MA, Fraser P, Ferrer J. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nature Genetics*. 2019;51(7):1137-1148. doi:10.1038/s41588-019-0457-0

18. Chiou J, Zeng C, Cheng Z, Han JY, Schlichting M, Miller M, Mendez R, Huang S, Wang J, Sui Y, Deogaygay A, Okino M-L, Qiu Y, Sun Y, Kudtarkar P, Fang R, Preissl S, Sander M, Gorkin DU, Gaulton KJ. Single-cell chromatin accessibility identifies pancreatic islet cell type– and state-specific regulatory programs of diabetes risk. *Nature Genetics*. Published online April 1, 2021:1-12. doi:10.1038/s41588-021-00823-0

19. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, Rivas MA, Perry JRB, Sim X, Blackwell TW, Robertson NR, Rayner NW, Cingolani P, Locke AE, Tajes JF, Highland HM, Dupuis J, Chines PS, Lindgren CM, Hartl C, Jackson AU, Chen H, Huyghe JR, van de Bunt M, Pearson RD, Kumar A, Müller-Nurasyid M, Grarup N, Stringham HM, Gamazon ER, Lee J, Chen Y, Scott RA, Below JE, Chen P, Huang J, Go MJ, Stitzel ML, Pasko D, Parker SCJ, Varga TV, Green T, Beer NL, Day-Williams AG, Ferreira T, Fingerlin T, Horikoshi M, Hu C, Huh I, Ikram MK, Kim B-J, Kim Y, Kim YJ, Kwon M-S, Lee J, Lee S, Lin K-H, Maxwell TJ, Nagai Y, Wang X, Welch RP, Yoon J, Zhang W, Barzilai N, Voight BF, Han B-G, Jenkinson CP, Kuulasmaa T, Kuusisto J, Manning A, Ng MCY, Palmer ND, Balkau B, Stančáková A, Abboud HE, Boeing H, Giedraitis V, Prabhakaran D, Gottesman O, Scott J, Carey J, Kwan P, Grant G, Smith JD, Neale BM, Purcell S, Butterworth AS, Howson JMM, Lee HM, Lu Y, Kwak S-H, Zhao W, Danesh J, Lam VKL, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536(7614):41-47. doi:10.1038/nature18642

20. Ye Y, Zhang Z, Liu Y, Diao L, Han L. A Multi-Omics Perspective of Quantitative Trait Loci in Precision Medicine. *Trends in Genetics*. 2020;36(5):318-336. doi:10.1016/j.tig.2020.01.009

21. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLOS Genetics*. 2010;6(4):e1000888. doi:10.1371/journal.pgen.1000888

22. Dermitzakis ET. Cellular genomics for complex traits. *Nature Reviews Genetics*. 2012;13(3):215-220. doi:10.1038/nrg3115

23. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci*. 2013;368(1620). doi:10.1098/rstb.2012.0362

24. Sun W, Hu Y. eQTL Mapping Using RNA-seq Data. *Stat Biosci*. 2013;5(1):198-219. doi:10.1007/s12561-012-9068-3

25. Kumasaka N, Knights AJ, Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nature Genetics*. 2016;48(2):206-213. doi:10.1038/ng.3467

26. Khetan S, Kursawe R, Youn A, Lawlor N, Jillette A, Marquez EJ, Ucar D, Stitzel ML. Type 2 Diabetes-Associated Genetic Variants Regulate Chromatin Accessibility in Human Islets. *Diabetes*. 2018;67(11):2466-2477. doi:10.2337/db18-0393

27. Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nature Reviews Genetics*. 2010;11(8):533-538. doi:10.1038/nrg2815

28. Hasin-Brumshtein Y, Hormozdiari F, Martin L, van Nas A, Eskin E, Lusis AJ, Drake TA. Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics*. 2014;15(1):471. doi:10.1186/1471-2164-15-471

29. Harvey CT, Moyerbrailean GA, Davis GO, Wen X, Luca F, Pique-Regi R. QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics*. 2015;31(8):1235-1242. doi:10.1093/bioinformatics/btu802

30. Cavalli M, Pan G, Nord H, Wallén Arzt E, Wallerman O, Wadelius C. Allele-specific transcription factor binding in liver and cervix cells unveils many likely drivers of GWAS signals. *Genomics*. 2016;107(6):248-254. doi:10.1016/j.ygeno.2016.04.006

31. Zhang Q, Keles S. An empirical Bayes test for allelic-imbalance detection in ChIP-seq. *Biostatistics*. 2018;19(4):546-561. doi:10.1093/biostatistics/kxx060

32. Mohammadi P, Castel SE, Brown AA, Lappalainen T. Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. *Genome Res*. 2017;27(11):1872-1884. doi:10.1101/gr.216747.116

33. Zou J, Hormozdiari F, Jew B, Castel SE, Lappalainen T, Ernst J, Sul JH, Eskin E. Leveraging allelic imbalance to refine fine-mapping for eQTL studies. *PLoS Genet*. 2019;15(12):e1008481. doi:10.1371/journal.pgen.1008481

34. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, Pfenning AR, Wang X, Claussnitzer M, Yaping Liu, Coarfa C, Alan Harris R, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, David Hawkins R, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Scott Hansen R, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh K-H, Feizi S, Karlic R, Kim A-R, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, Jager PLD, Farnham PJ, Fisher SJ, Haussler D, Jones SJM, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai L-H, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317-330. doi:10.1038/nature14248

35. Chen J, Rozowsky J, Galeev TR, Harmanci A, Kitchen R, Bedford J, Abyzov A, Kong Y, Regan L, Gerstein M. A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat Commun*. 2016;7(1):11101. doi:10.1038/ncomms11101

36. Zhou B, Ho SS, Greer SU, Zhu X, Bell JM, Arthur JG, Spies N, Zhang X, Byeon S, Pattni R, Ben-Efraim N, Haney MS, Haraksingh RR, Song G, Ji HP, Perrin D, Wong WH, Abyzov A, Urban AE. Comprehensive, integrated, and phased whole-genome analysis of the primary ENCODE cell line K562. *Genome Res*. 2019;29(3):472-484. doi:10.1101/gr.234948.118

37. Zhou B, Ho SS, Greer SU, Spies N, Bell JM, Zhang X, Zhu X, Arthur JG, Byeon S, Pattni R, Saha I, Huang Y, Song G, Perrin D, Wong WH, Ji HP, Abyzov A, Urban AE. Haplotype-resolved and integrated genome analysis of the cancer cell line HepG2. *Nucleic Acids Res*. 2019;47(8):3846-3861. doi:10.1093/nar/gkz169

38. Aylward A, Okino M-L, 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. *PLOS Genetics*. 2021;17(5):e1009531. doi:10.1371/journal.pgen.1009531

39. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki A-E, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PIW, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

40. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nature Methods*. 2015;12(11):1061-1063. doi:10.1038/nmeth.3582

41. Li T, Bai L, Li J, Igarashi S, Ghishan FK. Sp1 is required for glucose-induced transcriptional regulation of mouse vesicular glutamate transporter 2 gene. *Gastroenterology*. 2008;134(7):1994-2003. doi:10.1053/j.gastro.2008.02.076

42. Boylan MO, Jepeal LI, Wolfe MM. Sp1/Sp3 binding is associated with cell-specific expression of the glucose-dependent insulinotropic polypeptide receptor gene. *American Journal of Physiology-Endocrinology and Metabolism*. 2006;290(6):E1287-E1295. doi:10.1152/ajpendo.00535.2005

43. Mattis KK, Abaitua F, Grotz A, Wesolowska-Andersen A, Perez-Alcantara M, Mccarthy M, Mahajan A, Davies B, Gloyn AL. 3-LB: The Type 2 Diabetes-Associated Transcription Factor RREB1 Affects Beta-Cell Function and Development. *Diabetes*. 2019;68(Supplement 1). doi:10.2337/db19-3-LB

44. Deng Y-N, Xia Z, Zhang P, Ejaz S, Liang S. Transcription Factor RREB1: from Target Genes towards Biological Functions. *Int J Biol Sci*. 2020;16(8):1463-1473. doi:10.7150/ijbs.40834

45. Fogarty MP, Cannon ME, Vadlamudi S, Gaulton KJ, Mohlke KL. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet*. 2014;10(9):e1004633. doi:10.1371/journal.pgen.1004633

46. Roman TS, Cannon ME, Vadlamudi S, Buchkovich ML, Wolford BN, Welch RP, Morken MA, Kwon GJ, Varshney A, Kursawe R, Wu Y, Jackson AU, Program NI of HISC (NISC) CS, Erdos MR, Kuusisto J, Laakso M, Scott LJ, Boehnke M, Collins FS, Parker SCJ, Stitzel ML, Mohlke KL. A Type 2 Diabetes–Associated Functional Regulatory Variant in a Pancreatic

Islet Enhancer at the ADCY5 Locus. *Diabetes*. 2017;66(9):2521-2530. doi:10.2337/db17-0464

47. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans | Diabetes. Accessed June 10, 2021. https://diabetes.diabetesjournals.org/content/66/11/2888

48. Greenwald WW, Chiou J, Yan J, Qiu Y, Dai N, Wang A, Nariai N, Aylward A, Han JY, Kadakia N, Regue L, Okino M-L, Drees F, Kramer D, Vinckier N, Minichiello L, Gorkin D, 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. *Nat Commun*. 2019;10(1):2078. doi:10.1038/s41467-019-09975-4

49. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V, Lescano N, Wu B, Trombetta J, Ribado JV, Knowles DA, Gao Z, Blaeschke F, Parent AV, Burt TD, Anderson MS, Criswell LA, Greenleaf WJ, Marson A, Pritchard JK. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nature Genetics*. 2019;51(10):1494-1505. doi:10.1038/s41588-019-0505-9

50. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project | Biopreservation and Biobanking. Accessed May 27, 2021. https://www.liebertpub.com/doi/full/10.1089/bio.2015.0032

51. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW de, Daly MJ, Sham PC. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. 2007;81(3):559-575. doi:10.1086/519795

52. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research*. 2009;37(suppl_2):W202-W208. doi:10.1093/nar/gkp335

53. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27(7):1017-1018. doi:10.1093/bioinformatics/btr064

54. Viñuela A, Varshney A, van de Bunt M, Prasad RB, Asplund O, Bennett A, Boehnke M, Brown AA, Erdos MR, Fadista J, Hansson O, Hatem G, Howald C, Iyengar AK, Johnson P, Krus U, MacDonald PE, Mahajan A, Manning Fox JE, Narisu N, Nylander V, Orchard P, Oskolkov N, Panousis NI, Payne A, Stitzel ML, Vadlamudi S, Welch R, Collins FS, Mohlke KL, Gloyn AL, Scott LJ, Dermitzakis ET, Groop L, Parker SCJ, McCarthy MI. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat Commun*. 2020;11(1):4912. doi:10.1038/s41467-020-18581-8

55. Bayes factors for genome-wide association studies: comparison with P-values - Wakefield - 2009 - Genetic Epidemiology - Wiley Online Library. Accessed June 10, 2021. https://onlinelibrary.wiley.com/doi/abs/10.1002/gepi.20359