Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci

Matthias Thurner[1,2], Martijn van de Bunt[1,2], Jason M Torres[1], Anubha Mahajan[1], Vibe Nylander[2], Amanda J Bennett[2], Kyle Gaulton[3], Amy Barrett[2], Carla Burrows[2], Christopher G Bell[4,5], Robert Lowe[6], Stephan Beck[7], Vardhman K Rakyan[6], Anna L Gloyn[1,2,8], Mark I McCarthy*[1,2,8]

1. The Wellcome Trust Centre for Human Genetics, University of Oxford, UK
2. Oxford Centre for Diabetes, Endocrinology and Metabolism, University, of Oxford, Oxford, UK
3. Department of Pediatrics, University of California, San Diego, US
4. Department of Twin Research & Genetic Epidemiology, Kings College London, London, UK
5. MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK
6. Centre for Genomics and Child Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, London, UK
7. Department of Cancer Biology, UCL Cancer Institute, University College London, London, UK
8. Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK

*Corresponding author (mark.mccarthy@drl.ox.ac.uk)

Abstract

Human genetic studies have emphasised the dominant contribution of pancreatic islet dysfunction to development of Type 2 Diabetes (T2D). However, limited annotation of the islet epigenome has constrained efforts to define the molecular mechanisms mediating the, largely regulatory, signals revealed by Genome-Wide Association Studies (GWAS). We characterised patterns of chromatin accessibility (ATAC-seq, n=17) and DNA methylation (whole-genome bisulphite sequencing, n=10) in human islets, generating high-resolution chromatin state maps through integration with established ChIP-seq marks. We found enrichment of GWAS signals for T2D and fasting glucose was concentrated in subsets of islet enhancers characterised by open chromatin and hypomethylation, with the former annotation predominant. At several loci (including *CDC123, ADCY5, KLHDC5*) the combination of fine-mapping genetic data and chromatin state enrichment maps, supplemented by allelic imbalance in chromatin accessibility pinpointed likely causal variants. The combination of

increasingly-precise genetic and islet epigenomic information accelerates definition of causal mechanisms implicated in T2D pathogenesis.

1. Introduction

T2D is a complex disease characterised by insulin resistance and reduced beta cell function. Recent GWAS have identified a large number of T2D susceptibility loci (Scott et al., 2017, Mahajan et al., 2014, Morris et al., 2012, Voight et al., 2010), the majority of which affect insulin secretion and beta cell function (Dimas et al., 2014, Wood et al., 2017). However, most GWAS signals map to the non-coding genome and identification of the molecular mechanisms through which non-coding variants exert their effect has proven challenging. Several studies have demonstrated that T2D-associated variants map disproportionately to regulatory elements, particularly those which influence RNA expression and cellular function of human pancreatic islets. (Parker et al., 2013, Pasquali et al., 2014, van de Bunt et al., 2015, Olsson et al., 2014, Dayeh et al., 2014, Volkov et al., 2017, Varshney et al., 2017, Gaulton et al., 2015b, Gaulton et al., 2010).

Characterisation of the islet regulome has until now been limited in scope. The use of DNA methylation and open chromatin data to further annotate ChIP-seq derived chromatin states has successfully uncovered novel biology for other diseases (Wang et al., 2016). Existing methylation studies in islets, however, have either profiled a very small proportion of methylation sites using methylation arrays (Olsson et al., 2014, Dayeh et al., 2014) or focused on T2D-associated disease differentially methylated regions (dDMRs) rather than the integration of DNA methylation status with T2D-relevant GWAS data (Volkov et al., 2017). At the same time, assays of open chromatin in human islets have been restricted to small sample numbers (limiting the potential to capture allelic imbalance in chromatin accessibility for example): these have focussed predominantly on the impact of clustered or “stretch” enhancers (Parker et al., 2013, Pasquali et al., 2014, Gaulton et al., 2010, Varshney et al., 2017).

Most importantly, in part due to historical challenges in accessing human islet material or authentic human cellular models, reference annotations of the islet

epigenome and transcriptome (in the context of projects such as GTex, ENCODE and Epigenome Roadmap) have been largely absent. It is worth noting that islets constitute only ~1% of the pancreas, and islet epigenomes and transcriptomes cannot therefore be reliably assayed in analyses involving the entire organ. Previous islet epigenome studies have, therefore, had only limited ability to directly relate genetic variation to regulatory performance or to broadly characterise the role of DNA methylation in these processes.

In this study, we set out to expand upon previous studies of the islet regulome in several ways. First, we explored the human islet methylome in unprecedented depth using Whole-Genome Bisulphite Sequencing (WGBS) applied to a set of 10 human islet preparations. Second, we explored both basal and genotype-dependent variation in chromatin accessibility through ATAC-seq in 17 human islet preparations. Third, we integrated these genome-wide data with existing islet regulatory annotations to generate a high-resolution, epigenome map of this key tissue. Finally, we used this detailed map to interpret GWAS signals for T2D (and the related trait of fasting glucose) and deduce the molecular mechanisms through which some of these loci operate.

2. Results:

2.1 Characterising the DNA methylation landscape of human pancreatic islets

To characterise the human islet methylome and characterise the role of DNA methylation with respect to T2D genetic risk, we performed WGBS (mean coverage 13X) in human pancreatic islet DNA samples isolated from 10 non-diabetic cadaveric donors of European descent. Methylation levels across the genome were highly correlated across individual donors (mean CpG methylation Spearman's rho across 10 individual WGBS donors=0.71, Figure 1-figure supplement 1A): we pooled the WGBS results to generate a single high-pass (mean coverage 85X) set of pooled human pancreatic islet methylation data covering 23.3 million CpG sites (minimum pooled coverage 10X).

Most previous studies of the relationship between GWAS data and tissue-specific methylation patterns (including those interrogating the relationship between islet methylation and T2D predisposition (Dayeh et al., 2014, Olsson et al., 2014)) had used data generated on the Illumina 450k methylation array (Hannon et al., 2016, Mitchell et al., 2016, Kato et al., 2015, Ventham et al., 2016). For comparative purposes, we generated 450k array methylation data from 32 islet samples ascertained from non-diabetic donors of European descent (5 overlapping with those from whom WGBS data were generated). As with the WGBS data, methylation levels were highly correlated across individuals (mean CpG methylation Spearman's rho across 32 individual 450k donor=0.98, Figure 1-figure supplement 1B). After pooling 450k array data across samples, methylation profiles generated from the 450k array and WGBS were highly correlated at the small subset of total CpG sites for which they overlap: this was observed across pooled samples (pooled WGBS vs. 450k Spearman's rho=0.89, Figure 1A) and across the 5 donors analysed by both methods (mean Spearman's rho=0.80, not shown).

WGBS and 450k array data differed substantially in terms of genome-wide coverage. The 450k array was designed to interrogate with high precision and coverage ~480k CpG sites (approximately 2% of all sites in the genome), selected primarily because they are located near gene promoters and CpG-island regions. The focus of the 450k array on these regions, which tend to be less variable in terms of methylation, explains the high 450k array correlation levels between donors. In addition, this selective design results in marked differences in the distributions of genome-wide methylation values between WGBS and the 450k array. Whilst the WGBS data revealed the expected pattern of widespread high methylation levels with hypomethylation (<50%) restricted to 11.2% (2.6M/23.3M CpG sites) of the genome, the array disproportionately interrogated those hypomethylated sites (218k [46%] of all 450k CpG probes) (Kolmogorov–Smirnov (KS) test for difference, D=0.40, $P<2.2x10^{-16}$) (Figure 1B). These differences in methylation distribution were also evident within specific islet regulatory elements from previously defined standard chromatin state

maps (Parker et al., 2013) (Figure 1C, Figure 1-figure supplement 1C-D). We found significant (FDR<0.05) differences between the methylation levels of CpG sites accessed on the array, and those interrogated by WGBS, across most islet chromatin states: the largest differences were observed for weak promoters (median WGBS=0.71 vs. median 450k=0.11, KS test D=0.51, $P<2.2x10^{-16}$,) and weak enhancers (WGBS=0.87 vs. 450k median=0.76, D=0.39, $P<2.2x10^{-16}$, Figure 1-figure supplement 1D).

In terms of coverage, most chromatin states, apart from promoters, were poorly represented by CpG sites directly interrogated by the array: for example the array assayed only ~2.9% of CpG sites in strong enhancer states (2.7-3.8% depending on strong enhancer subtype, Figure 1C). Although methylation levels were previously reported to be highly correlated across short (0.1-2kb) genomic distances (Zhang et al., 2015, Bell et al., 2011, Eckhardt et al., 2006, Guo et al., 2017), the observed significant differences in the methylation distribution (Figure 1C, Figure 1-figure supplement 1D) across chromatin states including weak promoter (median size 600bp) and enhancer subtypes (median size ranges from 200-1200bp) indicate that these correlative effects are not strong enough to counterbalance the low coverage of the 450k array. These findings are consistent with 450k array content being focused towards CpG-dense hypomethylated and permissive promoter regions. This highlights the limited capacity of the array to comprehensively interrogate the global DNA methylome, in particular at distal regulatory states such as enhancers.

To understand the value of these data to reveal molecular mechanisms at GWAS loci, where we and others had shown enrichment for islet enhancer states (Pasquali et al., 2014, Gaulton et al., 2015a, Parker et al., 2013), we were interested to see how the selective coverage of the array might impact on its ability to interrogate methylation in GWAS-identified regions. We used the largest currently available T2D DIAGRAM GWAS data set (involving 26.7k cases and 132.5k controls of predominantly European origin, see dataset section for details) to identify the “credible sets” of variants at each locus which collectively

account for 99% of the posterior probability of association (PPA) (Scott et al., 2017, Maller et al., 2012).

To estimate the respective proportions of these T2D-associated variants captured by CpG sites assayed by the 450k array and WGBS, we determined, for each locus, the combined PPA of all 99% credible set variants mapping within 1000bp of any CpG site captured. This is based on evidence that across short distances CpG methylation is highly correlated (Zhang et al., 2015, Bell et al., 2011, Eckhardt et al., 2006, Guo et al., 2017) and may be influenced by genetic variants associated with altered transcription factor binding (Do et al., 2016). We found that coverage of this space of putative T2D GWAS variants by the 450k array is low: across GWAS loci, the combined PPA attributable to variants within regions assayed by the array ranged from 0-99% with a median PPA per locus of 16% (compared to a WGBS median PPA per locus=99%, KS-test $P<2.2x10^{-16}$, Figure 1D, top). We estimated that the equivalent figure for a recently developed upgrade of the 450k array, which captures ~850k CpG sites and aims to provide better coverage of enhancer regions, would be ~39% (range 0%-99% Figure 1D, top). For instance, at the *DGKB* T2D locus (centred on rs10276674), CpG sites covered by the 450k array interrogated less than 1% of the PPA of associated variants (vs. 99% captured by WGBS); the figure for the 850k array would be 23% (Figure 1E). We obtained similar results when we performed equivalent analyses using GWAS data for fasting glucose (FG, from the ENGAGE consortium (Horikoshi et al., 2015)), another phenotype dominated by islet dysfunction (Figure 1D, bottom).

These data indicate that available methylation arrays provide poor genome-wide coverage of methylation status and are notably deficient in capturing methylation status around the distal regulatory enhancer regions most relevant to T2D predisposition. For this reason, we focused subsequent analyses on the WGBS data.

2.2 Integration islet methylation and other epigenomic annotations

Studies in a variety of other tissues have shown that hypomethylation is a strong indicator of regulatory function (Stadler et al., 2011). More specifically, continuous stretches of CpG-poor Low-Methylated Regions (LMRs, with methylation ranging from 10-50% and containing fewer than 30 CpG sites) denote potential distal regulatory elements such as enhancers, while stretches of CpG-rich UnMethylated Regions (UMRs, containing more than 30 CpG sites) are more likely to represent proximal regulatory elements including promoters (Burger et al., 2013). We detected 37.1k LMRs, 13.6k UMRs (Figure 2A) and 10.7k Partially Methylated Domains (PMDs, Methods and Figure 2-figure supplement 1A-B). PMDs represent large regions of unordered methylation states associated with DNA sequence features (Gaidatzis et al., 2014). As anticipated, we found significant enrichment of LMRs with weak and strong enhancer states as defined by islet chromatin state maps derived from existing ChIP-seq data (Parker et al., 2013) (69.2% of islet LMRs overlapped islet strong and weak enhancer states, $\log_2$FE=2.2-2.9, Bonferroni P<0.05, (Figure 2B, Figure 1-figure supplement 1C). Similarly, UMRs were enriched for islet active promoter chromatin states (90.8% of UMRs overlapped islet active promoters, $\log_2$FE=3.9, FDR <0.05, Figure 2B).

To further characterise these hypomethylation domains, we overlapped information from analyses of islet cis-expression QTLs (eQTLs) (van de Bunt et al., 2015) and islet ChIP-seq transcription factor binding sites (TFBS) (Pasquali et al., 2014). We observed marked enrichment for eQTLs (LMR $\log_2$FE=1.1, UMR $\log_2$FE=2.7, Bonferroni P<0.05) and TFBS (LMR $\log_2$FE=4.1-4.6; UMR $\log_2$FE=2.4-3.9, Bonferroni P<0.05, Figure 2B). These observations confirm that islet LMRs and UMRs correspond to important tissue-specific regulatory regions, overlapping cis-regulatory annotations known to be enriched for T2D GWAS signals (Pasquali et al., 2014, Gaulton et al., 2015b).

We also considered the relationship between LMR and UMR regions defined in our non-diabetic islet WGBS, and a complementary set of methylation-based

annotations previously derived from WGBS of islets from 6 T2D and 8 control individuals (Volkov et al., 2017). In that study, comparisons between diabetic and non-diabetic islets had been used to define a set of 25,820 "disease differentially methylated regions" (dDMRs, min absolute methylation difference 5% and P<0.02). We found only limited overlap between these dDMRs and the UMRs and LMRs from our data: of the 25,820 dDMRs, 2.2% overlapped LMRs and 2.4% UMRs. This overlap was slightly greater than expected by chance (Bonferroni P<0.05, LMR $\log_2$FE=1.0 and promoter-like UMRs $\log_2$FE=1.1, Figure 2B) but more modest than seen for the other regulatory annotations. Similarly, we also observed that dDMRs showed more modest ($\log_2$FE=0.4-1.0), but still significant (Bonferroni P<0.05) levels of enrichment with respect to all other islet regulatory annotations (Figure 2B). The modest enrichment of dDMRs indicates that only a fraction of these regions correspond to islet genomic regulatory sites. Given that T2D risk variants preferentially map in islet regulatory sites, the corollary is that most dDMRs are unlikely to directly contribute to the mediation of genetic T2D risk.

2.3 Refining islet enhancer function using methylation and open chromatin data

To further characterise the regulatory potential of hypomethylated regions, including LMRs and UMRs, we combined the islet WGBS methylation data with chromatin accessibility data generated from ATAC-seq assays of 17 human islet samples (from non-diabetic donors of European descent; mean read count after filtering =130M, Figure 2-figure supplement 1C). We identified a total of 141k open chromatin regions based on read depth, peak width and signal-to-noise ratio (see Methods). These regions of islet open chromatin showed substantial overlap (78%) with equivalent regions described in a recent study of two human islets (Varshney et al., 2017) ($\log_2$FE=2.8 compared to random sites, not shown). In addition, our islet ATAC-Seq sites demonstrated substantial overlap with LMRs: 53% of LMRs overlapped 16% of all ATAC-seq peaks (LMR $\log_2$FE=3.8 compared to randomised sites, Figure 2B). Almost all UMRs (98%) were

contained within regions overlapping (13% of) ATAC-seq peaks (UMR $\log_2$FE=3.4 compared to randomised sites, Figure 2B).

To fully leverage information across multiple overlapping islet epigenome assays, we generated augmented chromatin state maps, using chromHMM (Ernst and Kellis, 2012). These maps combined the WGBS methylation and ATAC-Seq open chromatin data with previously generated ChIP-seq marks (Figure 3A, Figure 3-figure supplement 1A). For these analyses, we initially used a single definition for hypomethylated regions (methylation<60%) that captured both UMRs and LMRs (see Methods).

This augmented and larger set of 15 islet chromatin states retained the broad classification of regulatory elements that included promoters (positive for H3K4me3), transcribed and genic regions (H3K36me3), strong enhancers (H3K4me1; H3K27ac), weak enhancers (H3K4me1), insulators (CTCF) and repressed elements (H3K27me) (Figure 3A). The addition of islet methylation and open chromatin data expanded existing chromatin state definitions to provide new subclasses, particularly amongst enhancer elements. Here, we observed two subclasses of strong enhancers and three of weak enhancers (Figure 3A). We denote the strong enhancer subtypes as “open” (n=32k genome-wide), characterised by open chromatin and hypomethylation, and “closed” (n=110k) with closed chromatin and hypermethylation (Figure 3A). The three weak enhancer states we denote as “open” (n=38k: open chromatin, hypomethylation), “lowly-methylated” (n=78k; closed chromatin, hypomethylation) and “closed” (n=206k: closed chromatin, hypermethylation). No equivalent class of “lowly-methylated” strong enhancers was observed in the 15-state model. When comparing these chromatin states to those identified using only ChIP-seq marks ((Parker et al., 2013), Figure 1-figure supplement 1C,), the two strong enhancer subclasses we identified subdivided the “strong enhancer 1” state as described by Parker (defined by H3K27ac and H3K4me1). Additional comparison to “stretch” enhancer clusters (Parker et al., 2013), showed that there was considerable overlap between the “open” strong and weak enhancer states we identify here and previously-described “stretch”

enhancer states (16.1k out of 23k stretch enhancer overlapped 32k out of 70.1k "open" enhancers). Even so, most (55%) "open" enhancer states, and in particular "open weak enhancers" (70%), were not captured within "stretch" enhancer intervals, and we regard these as distinct islet enhancer subclasses.

To understand the relationship of these various state definitions to genetic variants influencing T2D risk, we applied the hierarchical modelling approach FGWAS to the same sets of large-scale GWAS data for T2D (from DIAGRAM (Scott et al., 2017)) and FG (ENGAGE (Horikoshi et al., 2015)) described in section 2.1. FGWAS allowed us to combine GWAS and genomic data to determine the genome-wide enrichment within islet regulatory features for variants associated with T2D risk. These enrichment priors were then used to generate credible variant sets that are informed by both GWAS and genomic data, as described in section 2.4.

In single-feature analyses, we found significant enrichment (lower limit of Confidence Interval (CI)>0) limited to four enhancer states (open weak enhancers, both types of strong enhancer and H3K36me3 marked genic enhancers) (Figure 3B, Table 1). To take into account protein-coding variant and conserved sequence effects, we also included CoDing exon Sequence (CDS) (Carlson and Maintainer, 2015) and CONServed sequence (CONS) (Lindblad-Toh et al., 2011) as additional annotations which were previously found to be strongly enriched for T2D GWAS signal (Finucane et al., 2015).We observed significant enrichment for CDS and CONS sequence in the single state results (Figure 3B, Table 1). FGWAS multi-feature analyses for T2D, incorporating all annotations positive in single-element analyses, retained both subclasses of strong enhancer, the subclass of open weak enhancers, genic enhancers and CDS in the joint model (Figure 3C and Methods). Conserved sequence annotations were not retained in the joint model.

We observed markedly different levels of enrichment for T2D association between and within open and closed enhancer states (Figure 3B-3C, Table 1). Using these augmented chromatin state maps, we demonstrated clear

enrichment for T2D association for the subset of "open" weak enhancers (12% of all weak enhancer sites) with no evidence of enrichment in the remaining subclasses ("closed" and "lowly-methylated") (Figure 3B and Table 1). This concentration of enrichment amongst a relatively small subset of the weak enhancers was consistent with the lack of enrichment across all weak enhancers defined solely on the basis of H3K4me1 signal ((Parker et al., 2013), single state log_2FE=0.9, CI=-2.5 to 2.0, Table 1, Figure 1-figure supplement 1C). We also saw differences in enrichment signal between open and closed strong enhancers, with the most marked enrichment amongst open strong enhancers (22% of the total, Figure 3B-C, Table 1). This effect was particularly obvious in the joint-analysis (open strong enhancer joint FE=4.1, CI=3.3 to 4.8 vs. closed strong enhancer joint log_2FE=2.4, CI=0.5 to 3.3, Figure 3C).

Hypomethylation and open chromatin are highly correlated, but the observed difference in T2D enrichment between the weak enhancer states (particularly between "lowly-methylated" and "open" which differ markedly with respect to chromatin status) points to a primary role for open chromatin. To test this further, we regenerated chromatin state maps using different subsets of the data (ChIP-only, with optional addition of methylation and/or open chromatin information, see Methods and Figure 3-figure supplement 1A-3B). These analyses confirmed that the T2D GWAS enrichment signal was predominantly driven by the distribution of islet open chromatin (Figure 3-figure supplement 1C).

We further evaluated the role of subclasses of DNA methylation regulatory region with respect to T2D GWAS enrichment. We divided hypomethylated (<60% methylated) sequence into enhancer-like LMRs (6.5% of all hypomethylated sequence), promoter-like UMRs (7.5% of hypomethylated sequence), as well as PMDs (61% of hypomethylated sequence). The remaining 25% of hypomethylated sequence did not fit any category. LMRs were significantly (CI>0) enriched (log_2FE=3.2, CI=2.3 to 3.9) for T2D association signals consistent with their co-localisation with distal regulatory elements, and displayed modestly increased enrichment compared to enhancer states derived

from ChIP-seq alone (Figure 3D, Figure 3-source data 1). In contrast, no significant enrichment was found for human islet (promoter-like) UMRs (log_2FE=1.4, CI=-0.6 to 2.5) or PMDs (log_2FE=-0.8, CI=-1.7 to -0.1). We also found no evidence that recently-described regions of T2D-associated differential methylation (dDMRs: derived from comparison of WGBS data from islets of diabetic and non-diabetic individuals) were enriched for genome-wide T2D association signals (log_2FE=-24.6, CI=-44.6 to 3.7) (Figure 3D, Figure 3-source data 1).

Finally, since the hypomethylation signal for T2D enrichment was concentrated in LMRs (Figure 3D, Figure 3-source data 1), we reran a FGWAS joint-analysis combining open chromatin peaks, LMRs and ChIP-only states using a nested model (Figure 3E, Figure 3-figure supplement 1D-E, see Methods). This confirmed that the improvement in enrichment was mainly driven by open chromatin but showed that LMRs also contributed significantly and independently to the enrichment (Figure 3E, Figure 3-source data 2).

FGWAS analysis for FG corroborated the observations from T2D analysis. Despite reduced power of the FG GWAS data due to a lower number of significantly associated FG GWAS loci, both single feature and joint-model analyses of human islet epigenome data found significant enrichment in strong enhancer states with the strongest enrichment in enhancers with open chromatin and hypomethylation (Figure 3-figure supplement 2A-B and Table 1). In addition, evaluation of the relative contributions of ATAC-seq open chromatin and DNA methylation to FG GWAS enrichment across both single-feature (Figure 3-figure supplement 2C-D) and joint-model analysis (Figure 3-figure supplement 2E-F and Figure 3-source data 3) indicated that open chromatin was primarily responsible for the enhanced enrichment.

Overall, these analyses demonstrate that the addition of open chromatin and DNA methylation data to ChIP-seq marks enhances the resolution of regulatory annotation for human islets. In particular, it defines subsets of weak and strong enhancers that differ markedly with respect to the impact of genetic variation on

T2D risk. Although DNA accessibility and hypomethylation status are strongly correlated and provide broadly similar enrichments, the effects of the former predominate. In line with the dominance of open chromatin status for T2D GWAS enrichment, we observed that T2D risk in relation to methylation status is primarily invested in hypomethylated LMRs (i.e. enhancers) rather than UMRs, dDMRs or PMDs.

2.4 Augmented chromatin maps and open chromatin allelic imbalance refine likely causal variants at *ADCY5* , *CDC123,* and *KLHDC5*

We next deployed the insights from the global FGWAS enrichment analyses to define the molecular mechanisms at individual T2D susceptibility loci, refining T2D causal variant localisation using the combination of genetic data (from fine-mapping) and the genome-wide patterns of epigenomic enrichment.

Specifically, we applied FGWAS to the T2D DIAGRAM GWAS data (Scott et al., 2017) under the joint model (Figure 3C) derived from the augmented chromatin state maps. We divided the genome into 2327 segments (average size 5004 SNPs or 1.2Mb) and identified 52 segments significantly associated with T2D genome-wide (segmental FGWAS PPA >=0.9 or single variant GWAS $P< 5x10^{-8}$, see Methods for details). These corresponded to 49 known T2D associated regions representing that subset of the ~120 known T2D GWAS loci which passed those significance/filtering criteria in this European-only dataset. We then calculated reweighted PPAs for each variant within each segment and generated reweighted 99% credible sets. (Of note, in line with traditional GWAS nomenclature, locus names were defined based on proximity between the lead variant and the closest gene and does not, of itself, indicate any causal role for the gene in T2D susceptibility).

Consistent with the increased T2D GWAS enrichment of states including open chromatin and DNA methylation information, we found that analyses using enrichments from the augmented chromatin state model (combining ChIP-seq, ATAC-seq and WGBS data) were associated with smaller 99% credible sets

(median of 17 SNPs) than those derived from FGWAS enrichment derived from ChIP-seq data alone (median 23). In parallel, the PPA for the best variant per locus increased (median 0.39 vs 0.31). Individual T2D GWAS locus results are shown in Figure 4A-B. We also expanded the FGWAS PPA analysis to investigate open chromatin and DNA methylation effects on fine-mapping and found that the reduction in 99% credible set size and increase in maximum variant PPA was driven predominantly by open chromatin (Figure 4-figure supplement 1, Figure 4-source data1). This demonstrates that the inclusion of open chromatin maps helps to improve prioritisation of causal variants at many T2D GWAS loci.

A subset of T2D GWAS signals are known to influence T2D risk through a primary effect on insulin secretion, whilst others act primarily through insulin resistance. We used previous categorisations of T2D GWAS loci based on the patterns of association with quantitative measurements of metabolic function and anthropometry (Wood et al., 2017, Dimas et al., 2014), to define a set of 15/48 loci most clearly associated with deficient insulin secretion (and therefore most likely to involve islet dysfunction). At 11 of these 15 loci, we found that islet "open strong enhancer" states, and to a lesser extent "open weak enhancer" and "closed strong enhancer", captured more than 60% of the PPA (median 92%, Figure 4C). Variants in these islet enhancer subclasses also captured at least 95% of the PPA at 4 T2D GWAS loci that could not be classified according to physiological association data but which have been previously implicated in human islet genome or functional regulation based on islet eQTL (van de Bunt et al., 2015) or mQTL (Olsson et al., 2014) data (Figure 4C, genes highlighted in bold). In contrast, at 3/6 of the insulin resistance and all but 5 unclassified loci, the PPA was mostly (>50%) attributable to other non-islet enhancer states (across all insulin resistance and unclassified loci, DNA not overlapping islet enhancers and defined as "Other" capture a median PPA of 64%). Thus, islet regulatory annotations are particularly useful for fine-mapping T2D GWAS loci that affect insulin secretion and beta-cell function.

To obtain additional evidence to support the localisation of causal variants, we tested for allelic imbalance in ATAC-seq open chromatin data. We selected 54

variants within 33 T2D-associated GWAS segments for testing of allelic imbalance on the basis of (a) a reweighted variant PPA >=10% and (b) overlap with an enriched regulatory state within the FGWAS T2D joint-model (Figure 4D, Figure 4-source data 2). Of these, 20 variants (at 16 loci) had sufficient numbers of heterozygous samples (>2) and ATAC_Seq read depth (depth>9 and at least 5 reads for each allele). After correcting for mapping bias using WASP, we observed the strongest evidence for allelic imbalance (FDR< 0.05) at 3 out of the 20 variants (rs11257655 near *CDC123* and *CAMK1D*, rs10842991 near *KLHDC5* and rs11708067 at *ADCY5*) (Table 2). All three overlapped refined islet open strong or open weak enhancer regions characterised by open chromatin and hypomethylation.

Variant rs11257655 accounts for 95% of the reweighted PPA (compared to a PPA of 20% from genetic data alone) at the *CDC123/CAMK1D* locus, overlaps an "open strong enhancer" region (Figure 5A) and the risk allele correlates with increased chromatin accessibility. The same variant is in high LD (r2=0.82) with the lead variant for a cis-eQTL for *CAMK1D* in islets (van de Bunt et al., 2015). In experimental assays (Fogarty et al., 2014), the T2D-risk allele has been shown to be associated with increased *CAMK1D* gene expression and enhanced binding of the FOXA1 and FOXA2 transcription factors. These data all point to rs11257655 as the causal variant at this locus.

At *KLHDC5*, no clear causal variant emerged based on genetic fine-mapping data alone as the credible set contained 23 variants in high mutual LD (r^2>0.8, top variant PPA<5%, Figure 5B). Of these, variants rs10771372 (genetic fine-mapping PPA= 5%), rs10842992 (genetic fine mapping PPA=5%) and rs10842991 (genetic fine-mapping PPA=3%) overlapped "open strong enhancer" regions (Figure 5B), such that their reweighted PPAs rose to 21% (rs10771372), 21% (rs10842992) and 13% (rs10842991), respectively. We observed allelic imbalance only at rs10842991 with the T2D-risk C allele showing greater chromatin accessibility (binomial P=4.1×10^{-3}, Table 2). This variant further overlapped a predicted TFBS motif for PAX6 as determined by the software tool FIMO (Grant et al., 2011): the T2D-risk allele was predicted to

499 enhance PAX6 transcription factor binding consistent with the allelic effects on
500 increasing chromatin accessibility (Figure 5-figure supplement 1A). This strong
501 enhancer region is almost exclusively found in islets, with strong enhancer
502 H3K27ac states overlapping rs10842991 in only two non-islet (heart and
503 smooth muscle) Epigenome Roadmap tissues (out of 99 tissues with 18-state
504 chromatin state information, Figure 5B)(Roadmap Epigenomics et al., 2015).
505 Islet eQTL data (Varshney et al., 2017) also links rs10842991 and close proxy
506 SNPs (including rs7960190) to islet transcription with the risk allele increasing
507 *KLHDC5* expression. These data prioritise rs10842991 as the likely causal
508 variant at the *KLHDC5* T2D GWAS locus, and indicate a likely molecular
509 mechanism involving modified PAX6 transcription factor binding and an impact
510 on *KLHDC5* expression and islet function.
511
512 The third example of allelic imbalance mapped to the *ADCY5* locus. Fine-mapping
513 based solely on genetic data could not prioritise a distinct causal variant due to
514 multiple variants in high LD (range for top 5 variants=12-26%, Figure 5C).
515 However, reweighting of variants based on epigenomic annotation clearly
516 prioritised variant rs11708067: this SNP overlapped an "open weak enhancer"
517 and captured most of the reweighted PPA (PPA=92%). Allelic imbalance analysis
518 also showed that the T2D-risk A allele was associated with decreased chromatin
519 accessibility (binomial $P=1.2x10^{-6}$, Table 2). The same lead variant maps to an
520 islet cis-eQTL and methylation QTL (Figure 5C, Figure 5-figure supplement 1B)
521 at which the T2D-risk allele is associated with reduced *ADCY5* expression and
522 increased *ADCY5* gene body DNA methylation.
523
524 To further understand the role of the rs11708067 variant, we performed ATAC-
525 seq and Next Generation Capture-C, in the glucose-responsive human beta-cell
526 line EndoC-βH1 (n=3). We targeted the *ADCY5* promoter to define distal regions
527 interacting with the promoter, and confirmed physical contact with the
528 hypomethylated open chromatin enhancer region harbouring rs11708067
529 (Figure 5C, Figure 5-figure supplement 1C). To resolve the significance of the
530 interaction between the restriction fragment encompassing rs11708067 and the
531 *ADCY5* promoter, we used the programme peakC(de Wit and Geeven, 2017)

(*https://github.com/deWitLab/peakC)* to evaluate the interactions of 12 fragments covering the lead SNP rs11708067 and 15 SNPs in high LD (r2 > 0.8) across a region of 47kb. After adjusting for multiple testing using FDR correction, only two fragments yielded a significant normalised read number over background. This included the open-chromatin overlapping fragment containing rs11708067 and another fragment harbouring rs2877716, rs6798189, rs56371916 (Figure 5-figure supplement 1D). These SNPs fall into a region that did not show evidence of open chromatin.

These findings support rs11708067 as the likely causal variant affecting islet accessible chromatin (in line with another recent study (Roman et al., 2017)), and link the open and hypomethylated enhancer element in which it sits to regulation of *ADCY5* expression in islets.

3. Discussion

A key challenge in the quest to describe the molecular mechanisms through which GWAS signals influence traits of interest, involves the identification of the causal variants responsible and, given that most lie in non-coding sequence, the characterisation of the regulatory elements which they perturb. This underpins efforts to define the effector genes through which these variants operate and to reconstruct the biological networks that are central to disease pathogenesis.

Genetic and physiological studies have highlighted the singular importance of pancreatic islet dysfunction in type 2 diabetes, but epigenomic characterisation of this tissue has been limited in large-scale community projects such as ENCODE and GTex. The present study seeks to address this deficit by describing, in unprecedented detail, genome-wide patterns of methylation and chromatin accessibility in human islet material. We have combined these data with existing islet epigenomic marks to generate a refined regulatory map which, based on the evidence of improved enrichment for T2D association signals, offers more granular annotation of functional impact.

Our data show that, for DNA methylation, the signal of T2D predisposition is primarily associated with enhancer-like LMRs rather than other categories of methylation elements including UMRs, dDMRs or PMDs. We highlight the strong correlation between islet methylation status and chromatin accessibility but demonstrate that open chromatin predominantly contributes to defining the regulatory impact associated with genetic T2D risk. Finally, we demonstrate how these enhanced epigenomic annotations, when analysed in concert with genetic fine-mapping data and information from allelic imbalance in chromatin accessibility allow us to home in on likely causal variants at T2D association signals such as those near *ADCY5, CDC123* and *KLHDC5*.

While previous studies had explored the candidacy of selected variants at the *CDC123* (Fogarty et al., 2014) and *ADCY5* (Olsson et al., 2014, Hodson et al., 2014, van de Bunt et al., 2015) loci with respect to islet regulation and T2D predisposition, our integrative analysis of T2D GWAS and epigenetic data has enabled a detailed and comprehensive analysis that considers the regulatory impact of all variants at these loci across multiple islet samples. Our analysis implicates the rs11257655 and rs11708067 variants as the most likely causal variants at the *CDC123* and *ADCY5* loci respectively and highlights their relationship to islet enhancer activity. The findings at *ADCY5* are supported by a recent paper that found allelic imbalance in H3K27 acetylation involving the rs11708067 variant in a single human islet sample, and which observed that deletion of the relevant enhancer element led to reduction in both *ADCY5* gene expression and insulin secretion (Roman et al., 2017).

At the *KLHDC5* locus, local LD frustrated efforts to define the causal variant using genetic data alone, but the integration of genetic and epigenetic data pinpointed rs10842991 as the likely culprit based on its impact on chromatin accessibility in an open enhancer region. Evidence that this variant co-localises with an islet cis-eQTL signal points to *KLHDC5* as the likely downstream target (Varshney et al., 2017). Overall, our integrative approach provides useful insights into the functional mechanisms through which T2D GWAS signals operate. Our findings mirror those from other studies, which have, in various ways, and for other

598 complex traits, combined diverse epigenomic annotations to explore the basis of
599 genetic risk (Wang et al., 2016).
600
601 The whole genome methylation data generated in the present study also allowed
602 us to evaluate the likely contribution of previously identified T2D-associated
603 dDMRs (Volkov et al., 2017) with respect to T2D predisposition. These dDMRs,
604 defined on the basis of observed differences in methylation between islets
605 recovered from diabetic and non-diabetic donors, cover a substantial part of the
606 genome, but we were able to show that only a small minority of these overlap
607 functional islet regulatory regions. As a consequence, dDMR regions as a whole
608 had no significant enrichment for T2D association signals. This suggests that
609 most of the dDMR signal involves stochastic effects and/or the secondary
610 consequences on methylation of the diabetic state. However, we cannot exclude
611 that some of the dDMR signals are causal contributors to the diabetic phenotype
612 either because they reflect environmental rather than genetic predisposition, or
613 because they accelerate further perturbation of islet dysfunction as diabetes
614 develops.
615
616 Although we provide highly detailed functional fine-mapping of T2D genetic
617 variants to uncover causal variants, the FGWAS approach applied in this study is
618 limited in its ability to determine the effect of multiple variants at individual loci.
619 Specifically, FGWAS relies on the assumption of a single causal variant within
620 each region, which may not necessarily be true for all loci. This assumption could
621 be violated where there are multiple independent signals at a given locus, or
622 where there are multiple (small effect size) variants on a single risk haplotype
623 which jointly impact the phenotype. Analysis methods that combine functional
624 fine-mapping with conditional analysis and consider LD and haplotype patterns
625 are likely to provide a more complete overview of the causal interactions at T2D
626 GWAS loci.
627
628 In addition, while the present study characterises islet epigenome status and
629 variability in chromatin accessibility in substantially larger numbers of islet
630 samples than those previously reported (Gaulton et al., 2015b, Parker et al.,

2013, Pasquali et al., 2014, Varshney et al., 2017), the number of islet preparations for which these data were available was still limited. As a result, our power to detect allelic imbalance in chromatin accessibility was restricted to sites with common variants and relatively large effects. We anticipate that expansion of these sample numbers will extend our capacity to detect such allelic imbalance, and offer more granular insights into the relationships between genetic variation and methylation status. A further limitation is that the genomic data we analysed was generated only from islet samples from non-diabetic donors. Whilst causal inference is possible through the integration of basal epigenomic annotations with genetic data, addition of epigenomic data from islets recovered from diabetic donors has the potential to add a further dimension to such analyses, and to unravel what are likely to be complex causal relationships between genetic variants, epigenomic phenotypes and disease states (Gutierrez-Arcelus et al., 2013). Finally, future work should also focus on experimental validation of likely causal variants and mechanisms e.g. differential binding of the TF PAX6 could be tested at the *KLHDC5* rs10842991 variant through electrophoretic mobility shift assays. Our ongoing research efforts are now concentrated on improving the fine-mapping analysis and expanding these genomic enrichment analyses in larger numbers of human islet samples from healthy and diabetic islets. By coupling the integration of these data with empirical functional studies, we expect to provide an increasingly complete description of the causal interactions between DNA methylation, chromatin state, RNA expression and T2D susceptibility.

4. Methods

4.1 Human Pancreatic islet samples

4.1.2 WGBS and 450k array human pancreatic islet sample collection

Human islets were retrieved from deceased Caucasian non-diabetic donors from the Oxford DRWF Human Islet Isolation Facility (n=34) and at the Alberta Diabetes Institute in Edmonton in Canada (n=10). For the analysis only samples with a purity >70% were used as determined by dithizone labeling. The Human

Research Ethics Board at the University of Alberta (Pro00001754), the University of Oxford's Oxford Tropical Research Ethics Committee (OxTREC Reference: 2–15), or the Oxfordshire Regional Ethics Committee B (REC reference: 09/H0605/2) approved the studies. All organ donors provided informed consent for use of pancreatic tissue in research.

For all WGBS (n=10) and a subset of 450k array samples (n=18) human pancreatic islet DNA was extracted from 100,000-150,000 islet cells using Trizol- (Ambion, UK or Sigma Aldrich, Canada) as described previously (van de Bunt et al., 2015). For the remaining 23 samples islet DNA was extracted using the ReliaPrep gDNA Tissue Miniprep system (Promega). Extracted DNA was stored at -80°C before further use.

4.1.2 ATAC-seq human pancreatic islet sample collection

Human pancreatic islets preparations (n=18) were retrieved from 17 deceased non-diabetic donors of European descent from the Oxford DRWF Human Islet Isolation Facility and stored for 1-3 days in CMRL or UW media. The latter were reactivated in CMRL for 1h before processing them further. Approximately 50,000 islet cells per sample were hand-picked and immediately processed for ATAC-seq as described previously (Buenrostro et al., 2013), however, an additional round of purification was performed using Agencourt AMPure XP magnetic beads (Beckman Coulter).

4.2. WGBS data generation

4.2.1 Bisulphite conversion

400ng of DNA per human islet samples (n=10) were sent as part of a collaborative effort to the Blizard Institute, Queen Mary University, London, UK and bisulphite- converted using the Ovation Ultralow Methyl-Seq DR Multiplex System 1-8 (Nugen) and purified using Agencourt AMPure beads (Beckman Coulter) as described previously (Lowe et al., 2013).

4.2.2 Library generation and processing of reads

The libraries were sequenced by members of the High-Throughput Genomics group at the Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. Samples were sequenced as multiplex libraries across 3 HiSeq2000 lanes with 100bp paired-end read length (including a PhIX spike-in of 5%) to obtain high-coverage read data. The obtained reads were trimmed using a customized python3 script (10bp at the start and 15bp at the end) and aligned to hg19 using the software Bismark (settings: L,0,-0.6, version 0.12.5, RRID:SCR_005604)(Krueger and Andrews, 2011). Specifically, paired-end alignment of trimmed reads was performed and unmapped reads from read 1 were realigned using Bismark and merged with the paired-end alignment using samtools (Li et al., 2009) (version 0.1.19, RRID:SCR_002105) in order to increase mapping efficiency. Coverage for the merged paired-end and realigned HiSeq read alignments was estimated for the human mappable genome (NCBI hg19 2.8 billion base pairs excluding gaps and unmappable and blacklisted regions according to UCSC and Encode(EncodeProjectConsortium, 2012)) using bedtools (version v2.21.0) (Quinlan, 2014).

4.2.3 WGBS DNA methylation quantification and prediction of hypomethylated regulatory regions

CpG methylation levels were determined for each sample by calculating the ratio of unmodified C (methylated) and bisulphite converted T (unmethylated) alleles using BiFAST (first described here (Lowe et al., 2013)). High-pass pooled WGBS data was generated by adding methylated and unmethylated read counts across individual low-pass samples to then estimate the average beta methylation levels.

Regulatory regions were identified using the R package methylseek (RRID:SCR_006513) (Burger et al., 2013). After removing PMDs, which represent highly heterogenous methylation states determined by DNA sequence features (Gaidatzis et al., 2014), LMRs (<30 CpGs) and UMRs (>30 CpGs) were predicted in hypomethylated regions (<50%) at an FDR of 0.05. The methylation level and

FDR parameter was inferred from the data as suggested by the methylseek workflow (Burger et al., 2013).

4.3 450k DNA methylation array data generation

In total, 41 samples were processed for the Illumina Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA). Of these 18 samples were bisulphite-converted and processed as part of a collaboration at the UCL Cancer Institute, University College London, London, UK while the remaining 23 samples were processed in OCDEM, University of Oxford, Oxford, UK. The DNA was bisulphite converted using the EZ DNA Methylation™ Kit (© Zymogen Research Corp, Irvine, CA) and hybridised to the Illumina 450k array and scanned with iScan (Illumina) according to the manufacturer's protocol.

The resulting data was analysed using the Package minfi (RRID:SCR_012830) (Aryee et al., 2014) and custom R scripts ((RCoreTeam), R version 3.0.2, RRID:SCR_001905). Specifically, CpG sites with a detection P-value >0.01 were removed from the analysis and samples with >5% of CpG sites failing this threshold (n=9) were also removed from the analysis.

Following separate quantile normalisation of signal intensities derived from methylated and unmethylated Type I probes and Type II probes, methylation levels (ß) were estimated, based on the intensities of the methylated (M) and unmethylated (U) signal in the following way: $\beta = M/(M+U+100)$. To correct for batch effects the ComBat function implemented in the sva (Johnson et al., 2007, Leek et al.) package was used (Figure 1-figure supplement 2).

4.4 ATAC-seq data generation

4.4.1 Sequencing of ATAC-seq reads

ATAC-seq libraries were sequenced at the High-Throughput Genomics group which is part of the Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. Samples were sequenced as 4-6plex libraries across 1-3 Hiseq2500 lanes with 50bp paired-end read length.

4.4.2 Processing of ATAC-seq reads

Raw FASTQ reads were processed with an in-house pipeline (first described in (Hay et al., 2016) and on the website http://userweb.molbiol.ox.ac.uk/public/telenius/PipeSite.html). Specifically, library and sequencing quality was checked with FASTQC (RRID:SCR_014583) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and reads were mapped to the human genome (hg19) via bowtie (Langmead et al., 2009) (version 1.1.0, RRID:SCR_005476) with default settings but -m 2, and maxins 2000 which allows mapping of reads with a maximum number of 2 alignments and a maximum insert size of 2000bp. For reads that could not be aligned the first time, adapters were removed with Trim Galore at the 3 prime end (RRID:SCR_011847, settings -length 10, -qualFilter 20, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to enhance the chance of mapping. The resulting trimmed reads were then mapped again with bowtie. Any remaining unmapped and trimmed reads were processed with FLASH (Magoc and Salzberg, 2011) (version 1.2.8, RRID:SCR_005531, settings -m 9 -x 0.125) which combines overlapping read pairs and reconstructs read pairs without overlap. These are then realigned a third time using bowtie. PCR duplicates are then removed from the mapped bam files using samtools rmdup function (Li et al., 2009). Additionally, all reads overlapping any of the "unmappable" UCSC Duke blacklisted hg19 regions (EncodeProjectConsortium, 2012) are also removed from the final bam file.

Open chromatin peaks were called through the aforementioned in-house pipeline by applying sample-specific read depth and width parameters, which were chosen based on the signal to noise ratio of a given sample.

4.5 ChIP-seq data and identification of chromatin states

4.5.1 Processing of available ChIP-seq data

Human islet ChIP-seq histone mark and TFBS data were obtained from various sources: H3K4me1, *CTCF* and H3K27ac (Pasquali et al., 2014), H3K36me3 and H3K4me3 (Moran et al., 2012) and H3K27me3 (Roadmap Epigenomics et al.,

2015). Available raw fastq files were aligned to hg19 using bowtie1 (version 1.1.1) with modified default settings (-m 1 which removes reads with more than 1 valid alignment and -n 1 which defines the maximum number of mismatches in seed) and PCR duplicates were removed from the aligned bam files with Picard tools (RRID:SCR_006525, v1.119, http://broadinstitute.github.io/picard/). The resulting reads were converted into bed format using the bedtools bamToBed function (Quinlan, 2014) (RRID:SCR_006646, version v2.21.0) and extended by 200bp towards the 3' end to account for fragment size.

4.5.2 Identification of chromatin states using chromHMM

Binarised 200bp density maps from the bed files of the 6 ChIP-seq marks were created using a Poisson distribution implemented in the BinaryBed function of the ChromHMM software as described in (Ernst and Kellis, 2012, Ernst et al., 2011). From these epigenomic density maps, 11 ChIP-only chromatin states were derived using a multivariate Hidden Markov Model implemented in the Learnmodel function (standard settings, h19 genome) of the software ChromHMM (Ernst and Kellis, 2012).

To generate additional sets of chromatin states based on ChIP-seq, ATAC-seq and DNA methylation data, ATAC-seq open chromatin and DNA methylation status were binarised. Specifically, ATAC-seq peaks (presence/absence) and whole-genome CpG methylation status (hypermethylation/hypomethylation based on a threshold of 60% methylation) were binarised across 200bp windows of the genome.

These binarised 200bp ChIP-seq, ATAC-seq and DNA methylation maps were combined and used to generate 3 sets of chromatin states derived from ChIP and DNA methylation data (ChIP+Meth), ChIP and ATAC-seq data (ChIP+ATAC) or ChIP, ATAC-seq and DNA methylation data (ChIP+ATAC+Meth) using the Learnmodel ChromHMM function (Figure 3A and Figure 3-figure supplement 1A-B) . As suggested by (Ernst et al., 2011), after evaluating models with up to 20

chromatin states, a 15 state model was chosen based on the resolution provided by identified states

4.6 *ADCY5* Capture C analysis and ATAC-seq in EndoC-ßH1

Next-generation Capture-C was performed in order to map physical chromatin interactions with the *ADCY5* promoter in EndoC-ßH1 (RRID:CVCL_L909) cell lines (n = 3) (see protocol in Methods in (Davies et al., 2016)).
In brief, chromatin conformation capture (3C) libraries were generated by formaldehyde fixation prior to DpnII restriction enzyme digestion and subsequent DNA ligation. Following cross-link reversal, DNA extraction and sonication, sequencing adapters were added to sonicated fragments (~200bp). Library fragments were subjected to a double capture through hybridisation with a biotinylated oligonucleotide probes encompassing the *ADCY5* promoter and enriched using streptavidin bead pulldown. PCR amplified fragments were then sequenced through paired-end sequencing (Illumina Next-Seq). An *in silico* restriction enzyme digestion was performed on the set of reconstructed fragments (from paired-end sequenced reads) using the DpnII2E.pl script (Davies, 2015)(https://github.com/Hughes-Genome-Group/captureC).
Uncaptured reads and PCR duplicates were removed prior to mapping to the human genome (hg19) with Bowtie (Langmead et al., 2009)(v 1.1.0). Removal of PCR duplicates and classification of fragments as 'capture' (i.e. including the *ADCY5* promoter) or 'reporter' (outside the capture fragment on exclusion region) was performed with the CCanalyser2.pl wrapper (Davies, 2015)(https://github.com/Hughes-Genome-Group/captureC). Unique mapped interactions were normalized to the total number of *cis* interactions (i.e. same chromosome) per 100,000 interactions. Significant chromatin interactions were determined from a rank-sum test implemented in the program peakC (de Wit and Geeven, 2017)(https://github.com/deWitLab/peakC). Specifically, we evaluated interactions involving all SNPs in high LD (r2 > 0.8) with the lead rs11708067. The lead variant (rs11708067) was in high LD with 15 SNPs (mapping to 12 DpnII fragments) that spanned a region of 47kb. We applied the Benjamini-Hochberg correction to control the false discovery rate for the set of

p-values corresponding to each restriction fragment within the 47kb region at the ADCY5 locus.

In addition, ATAC-seq was performed in 50,000 cells of EndoC-ßH1 cell lines (n=3) and the data was analysed in the same way as described above for human islet samples.

Endo-βH1 cells were obtained from Endocells and have been previously authenticated (Ravassard et al., 2011). In addition, the cell line was tested and found negative for mycoplasma contamination.

4.7 Overlaying generated epigenomic datasets generated here with other genomic regulatory regions

CpG sites and/or hypomethylated regulatory regions identified from the WGBS and/or 450k array data were overlapped with existing islet chromatin state maps (Parker et al., 2013), islet transcription factor binding sites (*FOXA2, MAFB, NKX2.2, NKX6.1, PDX1*), T2D-associated islet dDMRs (Dayeh et al., 2014) and eQTLs (van de Bunt et al., 2015). Similarly, ATAC-seq open chromatin peaks generated here were overlapped with publicly available ATAC-seq peaks (Varshney et al., 2017).

In addition, we also obtained the 850k array manifest file to determine overlap of 850k array CpG sites with GWAS credible set regions (https://support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html).

4.8 Genetic datasets used in this study

Credible sets from the DIAGRAM (Scott et al., 2017)(involving 26.7k cases and 132.5k controls of predominantly European origin, imputed to the 1000G March 2012 reference panel) and ENGAGE (Horikoshi et al., 2015)(including 46.7k individuals, imputed to the 1000G March 2012 reference panel) consortium

were used to compare the ability of the 450k, 850k and WGBS methylation array to interrogate T2D and FG GWAS regions.

The DIAGRAM and ENGAGE GWAS SNP summary level data was used for the FGWAS analysis to determine enrichment of regulatory annotations in T2D and FG GWAS signal.

Furthermore, data from (Wood et al., 2017) and (Dimas et al., 2014) were used to categories T2D GWAS loci into physiological groups of insulin secretion, insulin resistance or unclassified loci.

4.9 Statistical and computational analysis

4.9.1 Enrichment analysis of identified regulatory annotations in other genomic annotations

Enrichment of hypomethylated regulatory regions (LMRs and UMRs, result section 2.2.) and ATAC-seq open chromatin peaks (result section 2.3) in the aforementioned genomic annotations (method section 4.6) was determined through 100,000 random permutations. P-values and fold enrichment was determined by comparing the true overlap results to the permuted overlap results. The resulting P-values were multiple testing corrected using Bonferroni correction (an adjusted P-value < 0.05 was considered significant).

4.9.2 FGWAS enrichment analysis

FGWAS (Pickrell, 2014) (version 0.3.6) applied a hierarchical model that determined shared properties of loci affecting a trait. The FGWAS model used SNP-based GWAS summary level data and divided the genome into windows (setting "k"=5000 which represents the number of SNPs per window), which are larger than the expected LD patterns in the population. The model assumed that each window either contained a single SNP that affected the trait or that there was no SNP in the window that influenced the trait. The model estimated the prior probability of a window to contain an association and the conditional prior

probability that a SNP within the window was the causal variant. These prior probabilities were variable, dependent on regional annotations and estimated based on enrichment patterns of annotations across the genome using a Bayes approach.

4.9.2.1 FGWAS Single state analysis

FGWAS was used with standard settings to determine enrichment of individual islet chromatin states, LMRs, UMRs, PMDS and ATAC-seq open chromatin peaks, CDS and CONS sequence in DIAGRAM (setting "cc" was applied for use with T2D-case-control GWAS data) and ENGAGE GWAS SNP summary level data.

For each individual annotation, the model provided maximum likelihood enrichment parameters and annotations were considered as significantly enriched if the parameter estimate and 95% CI was above zero.

4.9.2.2 FGWAS Joint model analysis

To determine the maximum likelihood model the following approach suggested by (Pickrell, 2014) was used for each set of chromatin states (ChIP-only, ChIP+ATAC, ChIP+Meth and ChIP+ATAC+Meth), separately. In addition, CDS and CONS sequenced were used as well for each set of chromatin states in the joint analysis. Firstly, a model was fitted for each annotation individually to identify all annotations that were significantly enriched with the trait. Secondly, the annotation with the highest increase (and enrichment) in the maximum log-likelihood was added to the model and the analysis was repeated with the additional annotation. Thirdly, annotations were added as long as they increase the maximum log-likelihood of the newly tested model. Fourthly, a 10-fold cross-validation approach was used after determining a penalty parameter based on the maximum likelihood of a penalised log-likelihood function to avoid overfitting. Fifthly, each annotation was dropped from the model and the maximum cross-validation likelihood was evaluated. If a reduced model has a higher cross-validation maximum likelihood, additional annotations are dropped

until the model cannot be further improved. This model was described as the best fitted model and used for the remaining analysis. The maximum likelihood enrichment parameters and 95% CI for each annotation of the best model were reported (independent of significance).

4.9.2.3 Comparing FGWAS enrichment parameter across chromatin states

Initially, similar enhancer chromatin states derived from the 4 different ChromHMM analyses (ChIP-only, ChIP+ATAC, ChIP+Meth, ChIP+ATAC+Meth) were compared. Similarity was determined based on shared histone chromatin marks according to the chromHMM emission parameters. Further comparisons between the ChIP-only and ChIP+ATAC+Meth model were performed based on the reweighted FGWAS maximum variant PPA and the number of reweighted 99% credible set variants per T2D locus (for details regarding FGWAS PPA see next section).

However, considering that the chromatin states were derived from distinct sets of annotations across different analyses of ChromHMM, a direct comparison was not fully possible. Hence, a nested model approach was used to further dissect the contribution of open chromatin and DNA methylation to the enrichment. Specifically, an FGWAS analysis was performed that combined the ChIP-only chromHMM states with raw LMRs (representing DNA methylation) and ATAC-seq peaks (representing open chromatin). After determining the best maximum-likelihood cross-validation model (combining ChIP-only, ATAC-seq and LMR states) a nested model and log-likelihood ratio test were used to determine the contribution of each annotation to the model (Figure 3-figure supplement 1D).

4.9.3 Reweighting of variant PPA and testing of allelic imbalance

The enrichment priors derived from the FGWAS maximum likelihood model were used as a basis for evaluating both the significance and functional impact of associated variants in GWAS regions; allowing variants that map into annotations that show global enrichment to be afforded extra weight.

Specifically, variants at significant GWAS regions with a high FGWAS PPA (PPA >=10%) and overlapping open enhancer states were prioritised for further follow-up. Genome-wide significance of loci was determined based on P-values ($P < 5x10^{-8}$) or a regional FGWAS PPA >= 90% (representing the sum of the PPAs of all SNPs in a given region). The latter threshold is based on a recommendation from (Pickrell, 2014) who observed that a regional PPA of 90% or above can be used to identify sub-threshold GWAS loci.

Of the prioritised variants, only variants with at least 2 heterozygous samples and ATAC_Seq read depth of at least 9 reads (minimum 5 reads for each allele) were tested for allelic imbalance.

To avoid read-mapping and reference allele bias the software WASP (van de Geijn et al., 2015)(Version 0.2261) was used to remove reads associated with mapping bias. In short, reads of the unfiltered bam file that overlapped the variant of interest were identified. For each read overlapping an SNP, the genotype of that SNP was changed to the alternative allele and the read was remapped using bwa (Li and Durbin, 2009) (version 0.5.8c). Any read that failed to realign in the same position in the genome was discarded. Ultimately, PCR duplicates were filtered using the WASP "rmdup_pe.py" script, which removed duplicated reads randomly (independent of the mapping score) to avoid any bias.

Allelic imbalance was determined using a binomial test as implemented in R.

4.9.4 Identification of TFBS at SNPs that display allelic imbalance

The tool "Fimo"(Grant et al., 2011) implemented in the "meme" software package (RRID:SCR_001783) was applied to identify TF motifs that significantly (FDR <0.05) matched the sequence overlapping a SNP variant showing allelic imbalance (20bp up and downstream).

4.9.5 Overlap of regulatory regions

Overlap between genomic regulatory regions was performed using bedtools intersectBed function (Quinlan, 2014) (version 2.21.0). Summary statistics across 200bp windows were determined using bedtools mapBed function. Random permutations of regulatory regions were performed by applying the bedtools shuffleBed function.

4.9.6 Statistical analysis

All statistical analysis (unless otherwise stated) was performed using R (version 3.0.2) including Spearman's correlation analysis to compare the 450k and WGBS array, the KS-test to compare 450k and WGBS DNA methylation distributions, the binomial test to evaluate allelic imbalance and principal component analysis to identify batch effects in the 450k data. Significance is defined as $P<0.05$ unless otherwise stated.

4.9.7 Visualisation and figure generation

All figures unless otherwise stated were generated using R (version 3.0.2) and/or ggplot2(Wickham, 2009). Figure 1E was generated using locuszoom (Pruim et al., 2010). Chromatin state CHiP-seq enrichment maps (Figure 3A, Figure 3-figure supplement 1A-B) were generated using chromHMM (Ernst and Kellis, 2012). The genome-browser views (Figure 5) were generated using the UCSC genome browser tool (Kent et al., 2002).

5. Acknowledgements

We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532) for the generation of the Sequencing data.

MT was supported by a Wellcome Trust Doctoral Studentship**.** MvdB was supported by a Novo Nordisk postdoctoral fellowship run in partnership with the University of Oxford. JMT is supported by Wellcome Trust Strategic Award. VN is funded by the European Union Horizon 2020 Programme (T2DSYSTEMS). SB was supported by a Royal Society Wolfson Research Merit Award (WM100023). ALG is a Wellcome Trust Senior Fellow in Basic Biomedical Science (095101/Z/10/Z and 200837/Z/16/Z) and MIM is a Wellcome Trust

Senior Investigator. The research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). This work was also supported by EU (HEALTH-F4-2007-201413), Wellcome Trust (090367, 090532, 106130, 098381 and 099673/Z/12/Z) and NIH (U01-DK105535, U01-DK085545, R01-DK098032 and R01-MH090941) grants. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

6. Competing interests

The authors do not have any competing interest.

7. Main table information and associated titles/legends.

Chromatin States	Total number of states	T2D log2FE (CI)	FG log2FE (CI)
1. Active Promoter	20k	1.6 (-0.8 to 2.7)	2.7 (0 to 4.1)
2. Weak Promoter	33k	1.7 (-4.8 to 2.9)	2.7 (-0.1 to 4.2)
3. Transcriptional Elongation	71k	-0.4 (-20 to 1.1)	-26.1 (-46.1 to 1)
4. Low Methylation	73k	-1.5 (-3.1 to -0.6)	-1.7 (-4.2 to -0.3)
5. Closed Weak Enhancer	206k	1.2 (-0.1 to 2)	1.7 (0 to 2.9)
6. Lowly-methylated Weak Enhancer	78k	-0.5 (-20 to 1.6)	-26.7 (-46.7 to 1.6)
7. Open Weak Enhancer	38k	**3.4 (2.5 to 4.2)**	3.1 (-0.6 to 4.6)
8. Closed Strong Enhancer	110k	**2.7 (1.8 to 3.4)**	**3.3 (2 to 4.4)**
9. Open Strong Enhancer	32k	**3.8 (3.1 to 4.5)**	**4.3 (2.8 to 5.5)**
10. Genic Enhancer	39k	**2.5 (1.3 to 3.4)**	**2.9 (0.8 to 4.3)**
11. Accessible chromatin	14k	-25.2 (-45.2 to 2.5)	-28.4 (-48.4 to 3.7)
12. Insulator	31k	0.9 (-20 to 2.6)	-0.6 (-20 to 3.6)
13. Heterochromatin	216k	2.3 (-20 to 3.9)	1.8 (-1.5 to 4)
14. Polycomb Repressed	71k	-25.5 (-45.5 to 0.9)	-33.2 (-53.2 to 1.5)
15. Quiescent State	1.7k	-1 (-2.2 to -0.1)	-28.6 (-48.6 to -0.6)
CDS	NA	**2.6 (1.2 to 3.5)**	2.7 (-0.2 to 4.3)
CONS	NA	**2.1 (1.1 to 2.9)**	**1.9 (0.2 to 3.2)**
Parker Weak Enhancer	119k	0.9 (-2.5 to 2.0)	-2.0 (-20.0 to 2.4)
Parker Strong Enhancer (all)	123k	**2.7 (2.0 to 3.3)**	**3.1 (2.0 to 4.4)**
Parker Strong Enhancer (open)	64k	**3.1 (2.4 to 3.7)**	**3.6 (2.3 to 4.8)**
Parker Strong Enhancer (closed)	59k	**1.9 (0.8 to 2.7)**	**2.3 (0.5 to 3.5)**

TABLE 1. Single FGWAS annotation enrichment in T2D and FG GWAS data. For each annotation (chromatin state and CDS) the total number of sites and the single state FGWAS log2 Fold Enrichment (log2FE) in T2D and FG is shown. 95% Confidence Intervals (CI) for log2FE are shown in brackets and significantly enriched states are highlighted in bold (CI>0).

Variant	Locus	DIAGRAM P-value	FGWAS T2D PPA	Allelic imbalance Allele Ratio (Allele #)	Allelic imbalance WASP P-value	Direction of effect (T2D)
rs11708067	ADCY5	8.8E-13	0.92	0.29 (38 A VS 94 G alleles)	1.2E-06	risk allele A closed
rs11257655	CDC123	4.0E-08	0.95	0.39 (278 C VS 435 T alleles)	4.5E-09	risk allele T open
rs10842991	KLHDC5	7.3E-07	0.13	0.64 (75 C VS 43 T alleles)	4.1E-03	risk C allele open

TABLE 2. T2D-associated variants with allelic imbalance in open chromatin.

8. Legends and titles for main figures.

FIGURE 1. Comparison of human pancreatic islet WGBS and 450k methylation data across the genome. A) Smooth Scatter plot shows Spearman's rho correlation between the 450k array (x-axis) and WGBS (y-axis) at overlapping sites. Darker colour indicates higher density of sites. B) Comparison of the 450k array (orange) and WGBS (yellow) methylation levels (x-axis) of all CpGs genome-wide assayed by either method (y-axis shows density). The P-value shown is derived using a Kolmogorov-Smirnov (KS) test. C) For each chromatin state from Parker et al 2013 the methylation levels of all CpG sites independent of overlap (diamond indicates the median) are shown as violin plots (left y-axis) and the CpG probe percentage per state for the 450k array (orange) and WGBS (yellow) are shown as bar-plot (right y-axis). The 450k probes represent the percentage of the total number of CpG sites which is determined by the number of WGBS CpG sites detected (WGBS=100%). D) Distribution of GWAS Posterior Probabilities (Type 2 Diabetes and Fasting Glucose) captured by CpG sites on the 450k array (orange), 850k array (green) and WGBS (yellow/black line). E) Locuszoom plot showing CpG density and credible set SNPs. SNPs are shown with P-values (dots, y-axis left), recombination rate (line, y-axis right) and chromosome positions (x-axis) while CpG and gene annotations are shown below. These annotations include CpGs identified from WGBS (yellow strips), 450k CpG probes (orange stripes), 850k CpG probes (green stripes) and gene overlap (*DGKB* label). The highlighted region in blue captures the 99% credible set region plus additional 1000bp on either side. At the very bottom the position on chromosome 7 is shown in Megabases (Mb).

FIGURE 2. Overlap of WGBS hypomethylation and ATAC-seq open chromatin peaks with regulatory annotation . A) Methylation levels in percent (y-axis) and $\log_2$ CpG density (x-axis) of UMR and LMR regulatory regions with the dashed line indicating the CpG-number (30 CpGs) that distinguishes LMRs and UMRs. B) Log_2 Fold Enrichment ($\log_2$FE) of LMRs (green shape), UMRs (blue shape) in various islet annotations is shown. These annotations include islet chromatin states, islet relevant TFBS (FOXA2, MAFB, NKX2.2, NKX.61, PDX1), islet eQTLs, WGBS derived T2D-associated islet disease DMRs (dDMRs) and ATAC-seq open chromatin peaks. The dDMRs were derived from 6 T2D and 8 non-diabetic individuals by Volkov et al 2017 and dDMRs (orange shape) were also tested for enrichment in the aforementioned islet regulatory annotations. For all annotations, the empirically determined Bonferroni adjusted P-value is ≤0.00032 unless otherwise indicated by the

shape: a dot corresponds to an Bonferroni adjusted P-value <0.00032 while the 3 triangles indicates Bonferroni adjusted P-values > 0.00032: UMR enrichment adjusted P-value for weak enhancers=1; dDMR enrichment adjusted P-value for MAFB=0.006 and dDMR enrichment adjusted P-value for islet eQTLs=0.01.

FIGURE 3. Integration of islet epigenetic data to refine chromatin regulatory states and enrichment of these states in T2D GWAS data. A) 15 chromatin states (y-axis) were derived from ChIP histone marks, DNA methylation and ATAC-seq open chromatin annotations (x-axis) using chromHMM. For each state the relevant marks characterising the state are shown. The colour is based on the chromHMM emission parameters and a darker colour indicates a higher frequency of a mark at a given state. Weak enhancers (marked by H3K4me1 alone, red) and strong enhancers (marked by H3K27ac and H3K4me1, green) were subdivided by the chromHMM analysis according to methylation and ATAC-seq status (highlighted in red and green box). The black bar at the x-axis highlights the most important marks for characterising enhancer subtypes. B-C) FGWAS Log_2 Fold Enrichment including 95% CI (log_2FE, x-axis) of all chromatin states (y-axis) in T2D GWAS regions is shown which demonstrate differential enrichment amongst enhancer subclasses in single-feature enrichment analysis. In addition, log2FE of Coding Sequence (CDS) and Conserved Sequence (CONS) annotations are shown to include the effect of protein-coding and conserved regions. Significantly enriched annotations are shown in black while non-siginificant annotations are shown in grey. C) T2D FGWAS maximum likelihood model determined through cross-validation. Log_2FE and 95% CI (x-axis) of annotations included in the maximum likelihood model (y-axis) also demonstrate differential enrichment amongst enhancer subclasses. *Analysis for Genic Enhancers (state 10) did not converge and hence, only a point log2FE estimate is provided. D) Single feature log2FE including 95% CI (x-axis) results are shown highlighting the differences in T2D GWAS enrichment of various annotations. These include ATAC-seq open chromatin peaks (red), WGBS methylation regions (including enhancer-like LMRs, promoter-like UMRs and Partially Methylated Domains, blue), ChiP-seq chromatin states (orange) and CDS (green). E) Chi-square distribution (curved black line) with the indicated results of a maximum likelihood ratio test based on the maximum likelihood difference between a model including LMRs or ATAC-seq peaks compared to the ChIP-only model. The dashed red line indicates significance (P-value<0.05). For all FGWAS enrichment plots the axis has been truncated at -6 to facilitate visualisation and accurate values are provided in the supplementary tables.

FIGURE 4. Evaluating Posterior Probabilities (PP) derived from the FGWAS maximum likelihood model at significant T2D GWAS loci. (A) Per locus the difference in the number of 99% credible set variants between ChIP+ATAC+Meth and ChIP-only model is shown (positive values indicate a reduction in the number of 99% credible set variants in the ChIP_ATAC_Meth model). B) Per locus the difference in the maximum single variant PPA between the ChIP+ATAC+Meth and ChIP-only model is shown (positive values indicate an increase in the maximum single variant PPA in the ChIP+ATAC+Meth model). C) T2D GWAS loci were classified into insulin secretion (ISR), insulin resistance (IR) or unclassified loci based on genetic association with physiological traits derived

from Dimas et al 2014 and Wood et al 2017. In addition, loci with known role in islet genomic regulation or function are highlighted in bold. These include loci with islet eQTLs (*ZMIZ1, CDC123*) and mQTLs (*WFS1, KCNJ11*). D) Identification of T2D GWAS loci and variants enriched for enhancer chromatin states using FGWAS PP. Per locus the highest PPA variant is shown (y-axis) and the number of variants with PPA >0.01 (x-axis). Loci with high PPA variants (min PPA >0.1, dashed horizontal line) that overlap one of the enhancer states (green) are highlighted and the high PPA variants (PPA>0.1) were tested for allelic imbalance in open chromatin.

FIGURE 5. Epigenome Landscape of selected loci with allelic imbalance. For each locus A) *CDC123*, B) *KLHDC5* and C) *ADCY5* the following information is shown: 3 ATAC-seq Endoß tracks (green, top), variant level information (depending on the region GWAS lead SNP red, credible set black, eQTL blue and high LD SNPs with r2>0.8 black), WGBS methylation data (black, middle), 4 human islet ATAC-seq tracks (green, middle), islet chromatin states (from this study as well as Parker et al 2013 and Pasquali et al 2014) and Encode chromatin states from 9 cell types (bottom). For *ADCY5* the Capture C results in the Endoß cell line are shown as well (middle blue). Abbreviation for cell types: B-lymphoblastoid cells (GM12878), embryonic stem cells (H1 ES), erythrocytic leukaemia cells (K562), hepatocellular carcinoma cells (HepG2), umbilical vein endothelial cells (HUVEC), mammary epithelial cells (HMEC), skeletal muscle myoblasts (HSMM),normal epidermal keratinocytes (NHEK) and normal lung fibroblasts (NHLF).

9. Data accessibility and supplementary file information:

9.1 Sequencing data:

ATAC-seq and WGBS sequencing data has been deposited at the EBI hosted European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/) and is accessible via the EGA accession number: EGAS00001002592.

9.2 Supplementary figure information:

Figure 1-figure supplement 1 is associated with primary figure 1 (uploaded on eLife submission website with label: "Figure 1-figure supplement 1")

FIGURE 1-FIGURE SUPPLEMENT 1. Correlation of DNA methylation across WGBS and 450k sites and comparison of WGBS and 450k methylation levels across chromatin states A-B) Spearman's rho correlation of DNA methylation across 10 individual (A) WGBS and (B)10 selected (out of 32) 450k samples on the x-axis and y-axis. C) Islet chromatin state definitions based on ChIP-seq data reproduced from Parker et al 2013. TSS: Transcription Start Site D) The differences in the 450k and WGBS methylation level distribution measured as D statistic, which represents the difference in the cumulative distributions and is derived from the Kolmogorov-Smirnov test, are shown for each chromatin state separately.

Figure 2-figure supplement 1 is associated with primary figure 2 (uploaded on eLife submission website with label: "Figure 2-figure supplement 1")

FIGURE 2-FIGURE SUPPLEMENT 1. Identification and removal of Partially Methylated Domains (PMDs) A-B) Density distribution of the alpha value (A) before and (B) after removing PMDs (green curve in (A)) on chromosome 22. Alpha values represent a summary statistic derived from DNA methylation of windows of 100 CpGs and represents an indication of the polarisation status of methylation values in the genome which is expected to contain either highly methylated or unmethylated regions. Distributions with alpha <1 indicate methylation levels that are bimodal with either 0 or 1 methylation. Alpha=1 corresponds to a uniform distribution of methylation; and distributions with alpha>1 tend to have primarily intermediate methylation levels. The red and green curve in (A) represent the non-PMD (red) and PMD regions (green) in the genome. C) Number of peaks (x-axis) and mapped and filtered reads (y-axis) per ATAC-seq islet preparation. The dashed line indicates the mean read number. D) Log_2 Fold Enrichment (log_2FE, x-axis) and associated -log10 Bonferroni adjusted P-values (y-axis) of LMRs (circle), UMRs (triangle)in various islet annotations (colours) is shown. These annotations include islet chromatin states, islet relevant TFBS (FOXA2, MAFB, NKX2.2, NKX.61, PDX1), islet eQTLs, WGBS derived T2D-associated islet disease DMRs (dDMRs) and ATAC-seq open chromatin peaks. dDMRs (square) were also tested for enrichment in the aforementioned islet regulatory annotations. The results cluster near -log10 P-value of 3.5 since most Bonferroni adjusted P-values were more extreme than 0.00032.

Figure 3-figure supplement 1 is associated with primary figure 3 (uploaded on eLife submission website with label: "Figure 3-figure supplement 1")

FIGURE 3-FIGURE SUPPLEMENT 1. Prediction of regulatory regions using WGBS data and testing these regions for enrichment in T2D GWAS regions. A) Different combinations of epigenomic data (top) were combined to generate different sets of refined chromatin states (middle, 11 ChIP-only and 15 ChIP+Meth, ChIP+ATAC and ChIP+ATAC-Meth states, see figure S3B and 3A-B for actual states) using chromHMM. These sets of chromatin states were then tested for enrichment in T2D-related GWAS traits using FGWAS to compare enrichment across states (bottom). B) ChromHMM (I) 11 ChIP-only and 15 state (II) ChIP+ATAC state and (III) ChIP+Meth models. C) Single feature log2FE (x-axis) for different enhancer states (grey panels) defined from different combinations of epigenetic marks (y-axis) including ChIP+ATAC+Meth, ChIP+ATAC, ChIP+Meth and ChIP+only. The grey-dashed line indicates the enrichment value of CDS as reference. Enhancers are defined as follows: Strong enhancers are marked by both H3K4me1 and H3K27ac, weak Enhancers are defined by H3K4me1 only, gene enhancers are marked by H3K4me1 and H3K36me3, other enhancers are marked by H3K4me1, H3K4me3 and H3K27ac and are often referred to as TSS upstream regions (only included in the FGWAS T2D model for ChIP-only and ChIP+Meth chromatin states). D) Since chromatin states defined from a different set of epigenomic marks (ChIP-only, ChIP+Meth, ChIP+ATAC and ChIP+ATAC+Meth), as described in S3A-B, are not equivalent and the enrichment can not be easily compared across models, a nested model approach was applied. That is, ChIP-only chromatin states were generated and after evaluating the individual enrichment of each annotation (see Figure 3D), FGWAS maximum likelihood models were defined using ChIP-only, hypomethylated and/or ATAC-seq peak regulatory regions. The combination of all these annotations represented a nested linear model and the changes in maximum likelihood by adding/removing hypomethylated regulatory and ATAC-seq states could be statistically evaluated using a Loglikelihood Ratio Test (LRT) as shown in Figure 3E. E) Maximum likelihood FGWAS nested model combining ChIP-only, ATAC-peaks and LMR states (y-axis) showing log_2FE enrichment (x-axis) which was used for the LRT in Figure 3E. For all FGWAS enrichment plots the axis has been truncated at -6 to facilitate visualisation and accurate values are provided in the supplementary tables.

Figure 3-figure supplement 2 is associated with primary figure 3 (uploaded on eLife submission website with label: "Figure 3-figure supplement 2)

FIGURE 3-FIGURE SUPPLEMENT 2. Enrichment of refined islet regulatory states in FG GWAS data A) FGWAS Log2 Fold Enrichment including 95% CI (log2FE, x-axis) of all chromatin states (y-axis) in FG GWAS regions. In addition, CDS is shown to also include the effect of protein-coding regions. Significantly enriched annotations are shown in black. B) FG FGWAS maximum likelihood model determined through cross-validation. log_2FE and 95% CI (x-axis) of annotations included in the maximum likelihood model (y-axis) are shown. C)

Single feature log_2FE (x-axis) for different enhancer states (grey panels) defined from different combinations of epigenetic marks (y-axis) including ChIP+ATAC+Meth, ChIP+ATAC, ChIP+Meth and ChIP-only. Enhancers are defined as follows: Strong enhancers are marked by both H3K4me1 and H3K27ac, weak Enhancers are defined by H3K4me1 only, gene enhancers are marked by H3K4me1 and H3K36me3, other enhancers are marked by H3K4me1, H3K4me3 and H3K27ac and are often referred to as TSS upstream regions (only included in the FGWAS T2D model for ChIP-only and ChIP+Meth chromatin states). D) Single feature log_2FE including 95% CI (x-axis) results of various annotations derived from ChiP-seq (ChIP-only), ATAC-seq, WGBS methylation status and CDS are shown.E) Maximum likelihood FGWAS nested model combining ChIP-only, ATAC-peaks and LMR states (y-axis) showing log2FE enrichment (x-axis) which was used for the LRT in Supplementary Figure S3F. F) Chi-square distribution (black curved line) with the indicated results of a maximum likelihood ratio test based on the maximum likelihood difference between a model including LMRs or ATAC-seq peaks compared to the ChIP-only model. The dashed line indicates significance (P-value<0.05). For all FGWAS enrichment plots the axis has been truncated at -6 to facilitate visualisation and accurate values are provided in the supplementary tables.

Figure 4-figure supplement 1 is associated with primary Figure 4 (uploaded on eLife submission website with label: "Figure 4-figure supplement 1")

Figure 4-figure supplement 1. Evaluating annotation effect on Posterior Probabilities (PPA) derived from the FGWAS maximum likelihood model at significant T2D GWAS loci. (A) Violin plot showing the distribution of 99% credible set variant size (y-axis, log10 scale) of different annotation types used (x-axis, ChIP-only, ChIP+Meth, ChIP+ATAC, ChIP+ATAC+Meth, ATAC-only and LMR-only model). B) Violin plot showing the distribution in the maximum single variant PPA (y-axis) of different annotation types used (x-axis, ChIP-only, ChIP+Meth, ChIP+ATAC, ChIP+ATAC+Meth, ATAC-only and LMR-only model). Dots indicate mean value. C) Median 99% credible set variant size (x-axis) and median top variant PPA (y-axis) information for ChIP-only, ChIP+Meth, ChIP+ATAC, ChIP+ATAC+Meth, ATAC-only and LMR-only models.

Figure 5-figure supplement 1 is associated with primary Figure 6 (uploaded on eLife submission website with label: "Figure 5-figure supplement 1")

FIGURE 5-FIGURE SUPPLEMENT 1. A) Predicted *PAX6* Transcription factor binding motif likely affected by allelic imbalance of the variant rs10842991 (highlighted in purple). B) The *ADCY5* rs11708067 risk A allele was associated with increased methylation levels (y-axis, while genotypes are shown on the x-axis) . C) Chromatin Capture (Capture C) in the human beta-cell line EndoßH1 showed interactions between the *ADCY5* promoter (peak) and the flanking regions of the promoter. The x-axis shows the position on the chromosome in Mb while the y-axis indicates mapped reads per fragment. D) Chromatin Capture (Capture C) in the human beta-cell line EndoßH1 focussed at the genomic region (~47kb) near the variant rs11708067 (highlighted) and variants in high LD

(r2>0.8) with it (variants are depicted as black dots). Fragments containing rs11708067 (red) or other high LD variants (dark grey) are highlighted. The x-axis shows the position on the chromosome in bp while the y-axis indicates normalised mapped reads per fragment. The two fragments with P-values have a significant (FDR <0.05) number of normalised read counts over background: The fragment with the P-value on the left (in red) contains rs11708067 while the fragment with the P-value on the right harbours rs2877716, rs6798189, rs56371916.

Figure 1-figure supplement 2 is mentioned in the methods section and associated with primary Figure 1 (uploaded on eLife submission website with label: "Figure 1-figure supplement 2")

FIGURE 1-FIGURE SUPPLEMENT 2. A-B) PCA analysis of 450k DNA methylation data of 32 human islet samples coloured according to the location of origin and processing (A) before correction for Sample-location and (B) after correction for Sample-location using the ComBat function included in the sva package. The shape indicates sex. Sample location **EDM_OX**: samples obtained from the Alberta Diabetes Institute in Edmonton (Canada) and processed at the University of Oxford. **OX_OX**: samples obtained from Oxford DRWF Human Islet Isolation Facility and processed at the University of Oxford. **OX_UCL**: samples obtained from Oxford DRWF Human Islet Isolation Facility and processed at University College London.

9.3 Supplementary table information (uploaded as source files):

Figure 3-source data 1 is associated with primary Figure 3 (uploaded on eLife submission website with label: "Figure 3- source data 1")

FIGURE 3-SOURCE DATA 1. Annotation enrichment in T2D GWAS data. For each annotation the data source and the log2 Fold Enrichment (log_2FE) in T2D is shown. 95% Confidence Intervals (CI) for log_2FE are shown in brackets and significantly enriched states are highlighted in bold (CI>0).

Figure 3-source data 2 is associated with primary Figure 3 (uploaded on eLife submission website with label: "Figure 3-source data 2")

FIGURE 3-SOURCE DATA 2. Evaluating enrichment in T2D GWAS data. For each annotation the single feature and joint-model log2 Fold Enrichment (log_2FE) in T2D is shown. 95% Confidence Intervals (CI) for log_2FE are shown in brackets. In addition, the LRT statistic and P-value of a nested joint-model excluding a given annotation is shown.

Figure 3-source data 3 is associated with primary Figure 3 (uploaded on eLife submission website with label: "Figure 3-source data 3")

FIGURE 3-SOURCE DATA 3. Evaluating enrichment in FG GWAS data. For each annotation the single feature and joint-model log2 Fold enrichment (log_2FE) in FG is shown. 95% Confidence Intervals (CI) for log_2FE are shown in brackets. In addition, the LRT statistic and P-value of a nested joint-model excluding a given annotation is shown.

Figure 4-source data 1 is associated with primary Figure 4 (uploaded on eLife submission website with label: "Figure 4-source data 1")

FIGURE 4-SOURCE DATA 1. Comparison of variant variant PPA and 99% credible set size across annotations. For each set of annotations used the median segment top variant PPA (thigher values indicate better performance), the median segment 99% credible set size (lower values indicate better performance) and the number of significant segments (higher number indicates better performance) is shown. Significant loci were defined solely on a combined segmental PPA of at least 0.90.

Figure 4-source data 2 is associated with primary Figure 4 (uploaded on eLife submission website with label: "Figure 4-source data 2")

Figure 4-source data 2. Information for variants overlapping a genomic annotation included in the FGWAS T2D-joint model. For each variant that overlaps a genomic annotation included in FGWAS T2D-joint model the following information is provided: rsID; FGWAS PPA; T2D GWAS P-value; FGWAS segment number; T2D locus name; tested for allelic imbalance (Yes/No). If available, the following eQTL information from Varshney et al 2017 is shown as well: eQTL allele1 (effector), eQTL allele 2, eQTL q-value, eQTL effect and eQTL gene.

9.4 Source data bed file information:

Figure 2-source data 1. LMR_UMR_source_MThurner_Oct_2017.tds is associated with primary Figure 2 Bed file providing coordinates of WGBS hypomethylated regulatory regions defined as UMRs and LMRs.

Figure 3-source data 4. Merged_ATAC_seq_peaks_MThurner_Oct_2017.tds is associated with primary Figure 3 (uploaded on eLife submission website with label: "Figure 3-source data bed file 1"). Bed file providing coordinates of ATAC-seq open chromatin peaks merged across all samples.

Figure 3-source data 5.
Pancreatic_islet_15_chromatin_states_MThurner_Oct_2017.tds.zip is associated with primary Figure 3. Zipped bed file providing coordinates of human pancreatic islet chromatin states.

10. List of abbreviations:

450k array	Infinium Human Methylation 450K BeadChip
850 array	Infinium MethylationEPIC BeadChip
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
CDS	CoDing Sequence
ChIP	Chromatin ImmunoPrecipitation
CI	Confidence Interval
CONS	CONServed sequence
dDMRs	disease Differentially Methylated Regions
DIAGRAM	DIAbetes Genetics Replication And Meta-analysis
DNA	DeoxyriboNucleic Acid
ENGAGE	European Network for Genetic and Genomic Epidemiology
eQTL	expression Quantitative Trait Locus
FE	Fold Enrichment
FG	Fasting Glucose
GWAS	Genome-Wide Association Studies
KS-test	Kolmogorov–Smirnov test
LD	Linkage Disequilibrium
LMRs	Low-Methylated Regions
log2FE	log2 Fold Enrichment
mQTL	methylation Quantitative Trait Locus
P	P-value
PMDs	Partially Methylated Domains
PPA	Posterior Probability of Association
RNA	RiboNucleic Acid
SNP	Single-Nucleotide Polymorphism
T2D	Type 2 Diabetes
TFBS	Transcription Factor Binding Site
UMRs	UnMethylated Regions
WGBS	Whole-Genome Bisulphite Sequencing

11. References

ARYEE, M. J., JAFFE, A. E., CORRADA-BRAVO, H., LADD-ACOSTA, C., FEINBERG, A. P., HANSEN, K. D. & IRIZARRY, R. A. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics,* 30**,** 1363-9.

BELL, J. T., PAI, A. A., PICKRELL, J. K., GAFFNEY, D. J., PIQUE-REGI, R., DEGNER, J. F., GILAD, Y. & PRITCHARD, J. K. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol,* 12**,** R10.

BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods,* 10**,** 1213-8.

BURGER, L., GAIDATZIS, D., SCHUBELER, D. & STADLER, M. B. 2013. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res,* 41**,** e155.

CARLSON, M. & MAINTAINER, B. 2015. TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s). *R package version 3.2.2*.

DAVIES, J. O., TELENIUS, J. M., MCGOWAN, S. J., ROBERTS, N. A., TAYLOR, S., HIGGS, D. R. & HUGHES, J. R. 2016. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods,* 13**,** 74-80.

DAVIES, J. O. J. 2015. captureC. *Github.* https://github.com/Hughes-Genome-Group/captureC . *aae8046*

DAYEH, T., VOLKOV, P., SALO, S., HALL, E., NILSSON, E., OLSSON, A. H., KIRKPATRICK, C. L., WOLLHEIM, C. B., ELIASSON, L., RONN, T., BACOS, K. & LING, C. 2014. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet,* 10**,** e1004160.

DE WIT, E. & GEEVEN, G. 2017. peakC. *Github.* https://github.com/deWitLab/peakC . *a2ac8c1*.

DIMAS, A. S., LAGOU, V., BARKER, A., KNOWLES, J. W., MAGI, R., HIVERT, M. F., BENAZZO, A., RYBIN, D., JACKSON, A. U., STRINGHAM, H. M., SONG, C., FISCHER-ROSINSKY, A., BOESGAARD, T. W., GRARUP, N., ABBASI, F. A., ASSIMES, T. L., HAO, K., YANG, X., LECOEUR, C., BARROSO, I., BONNYCASTLE, L. L., BOTTCHER, Y., BUMPSTEAD, S., CHINES, P. S., ERDOS, M. R., GRAESSLER, J., KOVACS, P., MORKEN, M. A., NARISU, N., PAYNE, F., STANCAKOVA, A., SWIFT, A. J., TONJES, A., BORNSTEIN, S. R., CAUCHI, S., FROGUEL, P., MEYRE, D., SCHWARZ, P. E., HARING, H. U., SMITH, U., BOEHNKE, M., BERGMAN, R. N., COLLINS, F. S., MOHLKE, K. L., TUOMILEHTO, J., QUERTEMOUS, T., LIND, L., HANSEN, T., PEDERSEN, O., WALKER, M., PFEIFFER, A. F., SPRANGER, J., STUMVOLL, M., MEIGS, J. B., WAREHAM, N. J., KUUSISTO, J., LAAKSO, M., LANGENBERG, C., DUPUIS, J., WATANABE, R. M., FLOREZ, J. C., INGELSSON, E., MCCARTHY, M. I., PROKOPENKO, I. & INVESTIGATORS, M. 2014. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes,* 63**,** 2158-71.

DO, C., LANG, C. F., LIN, J., DARBARY, H., KRUPSKA, I., GABA, A., PETUKHOVA, L., VONSATTEL, J. P., GALLAGHER, M. P., GOLAND, R. S., CLYNES, R. A., DWORK, A., KRAL, J. G., MONK, C., CHRISTIANO, A. M. & TYCKO, B. 2016. Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation. *Am J Hum Genet,* 98**,** 934-55.

ECKHARDT, F., LEWIN, J., CORTESE, R., RAKYAN, V. K., ATTWOOD, J., BURGER, M., BURTON, J., COX, T. V., DAVIES, R., DOWN, T. A., HAEFLIGER, C., HORTON, R., HOWE, K., JACKSON, D. K., KUNDE, J., KOENIG, C., LIDDLE, J., NIBLETT, D., OTTO, T., PETTETT, R., SEEMANN, S., THOMPSON, C., WEST, T., ROGERS, J., OLEK, A., BERLIN, K. & BECK, S. 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet,* 38**,** 1378-85.

ENCODEPROJECTCONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

ERNST, J. & KELLIS, M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods,* 9**,** 215-6.

ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R., COYNE, M., KU, M.,

DURHAM, T., KELLIS, M. & BERNSTEIN, B. E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473**,** 43-9.

FINUCANE, H. K., BULIK-SULLIVAN, B., GUSEV, A., TRYNKA, G., RESHEF, Y., LOH, P. R., ANTTILA, V., XU, H., ZANG, C., FARH, K., RIPKE, S., DAY, F. R., REPROGEN, C., SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C., CONSORTIUM, R., PURCELL, S., STAHL, E., LINDSTROM, S., PERRY, J. R., OKADA, Y., RAYCHAUDHURI, S., DALY, M. J., PATTERSON, N., NEALE, B. M. & PRICE, A. L. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet,* 47**,** 1228-35.

FOGARTY, M. P., CANNON, M. E., VADLAMUDI, S., GAULTON, K. J. & MOHLKE, K. L. 2014. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet,* 10**,** e1004633.

GAIDATZIS, D., BURGER, L., MURR, R., LERCH, A., DESSUS-BABUS, S., SCHUBELER, D. & STADLER, M. B. 2014. DNA sequence explains seemingly disordered methylation levels in partially methylated domains of Mammalian genomes. *PLoS Genet,* 10**,** e1004143.

GAULTON, K. J., FERREIRA, T., LEE, Y., RAIMONDO, A., MAGI, R., RESCHEN, M. E., MAHAJAN, A., LOCKE, A., RAYNER, N. W., ROBERTSON, N., SCOTT, R. A., PROKOPENKO, I., SCOTT, L. J., GREEN, T., SPARSO, T., THUILLIER, D., YENGO, L., GRALLERT, H., WAHL, S., FRANBERG, M., STRAWBRIDGE, R. J., KESTLER, H., CHHEDA, H., EISELE, L., GUSTAFSSON, S., STEINTHORSDOTTIR, V., THORLEIFSSON, G., QI, L., KARSSEN, L. C., VAN LEEUWEN, E. M., WILLEMS, S. M., LI, M., CHEN, H., FUCHSBERGER, C., KWAN, P., MA, C., LINDERMAN, M., LU, Y., THOMSEN, S. K., RUNDLE, J. K., BEER, N. L., VAN DE BUNT, M., CHALISEY, A., KANG, H. M., VOIGHT, B. F., ABECASIS, G. R., ALMGREN, P., BALDASSARRE, D., BALKAU, B., BENEDIKTSSON, R., BLUHER, M., BOEING, H., BONNYCASTLE, L. L., BOTTINGER, E. P., BURTT, N. P., CAREY, J., CHARPENTIER, G., CHINES, P. S., CORNELIS, M. C., COUPER, D. J., CRENSHAW, A. T., VAN DAM, R. M., DONEY, A. S., DORKHAN, M., EDKINS, S., ERIKSSON, J. G., ESKO, T., EURY, E., FADISTA, J., FLANNICK, J., FONTANILLAS, P., FOX, C., FRANKS, P. W., GERTOW, K., GIEGER, C., GIGANTE, B., GOTTESMAN, O., GRANT, G. B., GRARUP, N., GROVES, C. J., HASSINEN, M., HAVE, C. T., HERDER, C., HOLMEN, O. L., HREIDARSSON, A. B., HUMPHRIES, S. E., HUNTER, D. J., JACKSON, A. U., JONSSON, A., JORGENSEN, M. E., JORGENSEN, T., KAO, W. H., KERRISON, N. D., KINNUNEN, L., KLOPP, N., KONG, A., KOVACS, P., KRAFT, P., KRAVIC, J., LANGFORD, C., et al. 2015a. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet,* 47**,** 1415-25.

GAULTON, K. J., FERREIRA, T., LEE, Y., RAIMONDO, A., MAGI, R., RESCHEN, M. E., MAHAJAN, A., LOCKE, A., WILLIAM RAYNER, N., ROBERTSON, N., SCOTT, R. A., PROKOPENKO, I., SCOTT, L. J., GREEN, T., SPARSO, T., THUILLIER, D., YENGO, L., GRALLERT, H., WAHL, S., FRANBERG, M., STRAWBRIDGE, R. J., KESTLER, H., CHHEDA, H., EISELE, L., GUSTAFSSON, S., STEINTHORSDOTTIR, V., THORLEIFSSON, G., QI, L., KARSSEN, L. C., VAN LEEUWEN, E. M., WILLEMS, S. M., LI, M., CHEN, H., FUCHSBERGER, C., KWAN, P., MA, C., LINDERMAN, M., LU, Y., THOMSEN, S. K., RUNDLE, J. K.,

BEER, N. L., VAN DE BUNT, M., CHALISEY, A., KANG, H. M., VOIGHT, B. F., ABECASIS, G. R., ALMGREN, P., BALDASSARRE, D., BALKAU, B., BENEDIKTSSON, R., BLUHER, M., BOEING, H., BONNYCASTLE, L. L., BOTTINGER, E. P., BURTT, N. P., CAREY, J., CHARPENTIER, G., CHINES, P. S., CORNELIS, M. C., COUPER, D. J., CRENSHAW, A. T., VAN DAM, R. M., DONEY, A. S., DORKHAN, M., EDKINS, S., ERIKSSON, J. G., ESKO, T., EURY, E., FADISTA, J., FLANNICK, J., FONTANILLAS, P., FOX, C., FRANKS, P. W., GERTOW, K., GIEGER, C., GIGANTE, B., GOTTESMAN, O., GRANT, G. B., GRARUP, N., GROVES, C. J., HASSINEN, M., HAVE, C. T., HERDER, C., HOLMEN, O. L., HREIDARSSON, A. B., HUMPHRIES, S. E., HUNTER, D. J., JACKSON, A. U., JONSSON, A., JORGENSEN, M. E., JORGENSEN, T., KAO, W. H., KERRISON, N. D., KINNUNEN, L., KLOPP, N., KONG, A., KOVACS, P., KRAFT, P., KRAVIC, J., LANGFORD, C., et al. 2015b. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet,* 47**,** 1415-25.

GAULTON, K. J., NAMMO, T., PASQUALI, L., SIMON, J. M., GIRESI, P. G., FOGARTY, M. P., PANHUIS, T. M., MIECZKOWSKI, P., SECCHI, A., BOSCO, D., BERNEY, T., MONTANYA, E., MOHLKE, K. L., LIEB, J. D. & FERRER, J. 2010. A map of open chromatin in human pancreatic islets. *Nat Genet,* 42**,** 255-9.

GRANT, C. E., BAILEY, T. L. & NOBLE, W. S. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics,* 27**,** 1017-8.

GUO, S., DIEP, D., PLONGTHONGKUM, N., FUNG, H. L., ZHANG, K. & ZHANG, K. 2017. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet,* 49**,** 635-642.

GUTIERREZ-ARCELUS, M., LAPPALAINEN, T., MONTGOMERY, S. B., BUIL, A., ONGEN, H., YUROVSKY, A., BRYOIS, J., GIGER, T., ROMANO, L., PLANCHON, A., FALCONNET, E., BIELSER, D., GAGNEBIN, M., PADIOLEAU, I., BOREL, C., LETOURNEAU, A., MAKRYTHANASIS, P., GUIPPONI, M., GEHRIG, C., ANTONARAKIS, S. E. & DERMITZAKIS, E. T. 2013. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife,* 2**,** e00523.

HANNON, E., SPIERS, H., VIANA, J., PIDSLEY, R., BURRAGE, J., MURPHY, T. M., TROAKES, C., TURECKI, G., O'DONOVAN, M. C., SCHALKWYK, L. C., BRAY, N. J. & MILL, J. 2016. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci,* 19**,** 48-54.

HAY, D., HUGHES, J. R., BABBS, C., DAVIES, J. O. J., GRAHAM, B. J., HANSSEN, L., KASSOUF, M. T., MARIEKE OUDELAAR, A. M., SHARPE, J. A., SUCIU, M. C., TELENIUS, J., WILLIAMS, R., RODE, C., LI, P. S., PENNACCHIO, L. A., SLOANE-STANLEY, J. A., AYYUB, H., BUTLER, S., SAUKA-SPENGLER, T., GIBBONS, R. J., SMITH, A. J. H., WOOD, W. G. & HIGGS, D. R. 2016. Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat Genet,* 48**,** 895-903.

HODSON, D. J., MITCHELL, R. K., MARSELLI, L., PULLEN, T. J., GIMENO BRIAS, S., SEMPLICI, F., EVERETT, K. L., COOPER, D. M., BUGLIANI, M., MARCHETTI, P., LAVALLARD, V., BOSCO, D., PIEMONTI, L., JOHNSON, P. R., HUGHES, S. J., LI, D., LI, W. H., SHAPIRO, A. M. & RUTTER, G. A. 2014. ADCY5 couples glucose to insulin secretion in human islets. *Diabetes,* 63**,** 3009-21.

HORIKOSHI, M., MGI, R., VAN DE BUNT, M., SURAKKA, I., SARIN, A. P., MAHAJAN, A., MARULLO, L., THORLEIFSSON, G., HGG, S., HOTTENGA, J. J., LADENVALL, C., RIED, J. S., WINKLER, T. W., WILLEMS, S. M., PERVJAKOVA, N., ESKO, T., BEEKMAN, M., NELSON, C. P., WILLENBORG, C., WILTSHIRE, S., FERREIRA, T., FERNANDEZ, J., GAULTON, K. J., STEINTHORSDOTTIR, V., HAMSTEN, A., MAGNUSSON, P. K., WILLEMSEN, G., MILANESCHI, Y., ROBERTSON, N. R., GROVES, C. J., BENNETT, A. J., LEHTIMKI, T., VIIKARI, J. S., RUNG, J., LYSSENKO, V., PEROLA, M., HEID, I. M., HERDER, C., GRALLERT, H., MULLER-NURASYID, M., RODEN, M., HYPPONEN, E., ISAACS, A., VAN LEEUWEN, E. M., KARSSEN, L. C., MIHAILOV, E., HOUWING-DUISTERMAAT, J. J., DE CRAEN, A. J., DEELEN, J., HAVULINNA, A. S., BLADES, M., HENGSTENBERG, C., ERDMANN, J., SCHUNKERT, H., KAPRIO, J., TOBIN, M. D., SAMANI, N. J., LIND, L., SALOMAA, V., LINDGREN, C. M., SLAGBOOM, P. E., METSPALU, A., VAN DUIJN, C. M., ERIKSSON, J. G., PETERS, A., GIEGER, C., JULA, A., GROOP, L., RAITAKARI, O. T., POWER, C., PENNINX, B. W., DE GEUS, E., SMIT, J. H., BOOMSMA, D. I., PEDERSEN, N. L., INGELSSON, E., THORSTEINSDOTTIR, U., STEFANSSON, K., RIPATTI, S., PROKOPENKO, I., MCCARTHY, M. I., MORRIS, A. P. & CONSORTIUM, E. 2015. Discovery and Fine-Mapping of Glycaemic and Obesity-Related Trait Loci Using High-Density Imputation. *PLoS Genet,* 11**,** e1005230.

JOHNSON, W. E., LI, C. & RABINOVIC, A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics,* 8**,** 118-27.

KATO, N., LOH, M., TAKEUCHI, F., VERWEIJ, N., WANG, X., ZHANG, W., KELLY, T. N., SALEHEEN, D., LEHNE, B., MATEO LEACH, I., DRONG, A. W., ABBOTT, J., WAHL, S., TAN, S. T., SCOTT, W. R., CAMPANELLA, G., CHADEAU-HYAM, M., AFZAL, U., AHLUWALIA, T. S., BONDER, M. J., CHEN, P., DEHGHAN, A., EDWARDS, T. L., ESKO, T., GO, M. J., HARRIS, S. E., HARTIALA, J., KASELA, S., KASTURIRATNE, A., KHOR, C. C., KLEBER, M. E., LI, H., MOK, Z. Y., NAKATOCHI, M., SAPARI, N. S., SAXENA, R., STEWART, A. F., STOLK, L., TABARA, Y., TEH, A. L., WU, Y., WU, J. Y., ZHANG, Y., AITS, I., DA SILVA COUTO ALVES, A., DAS, S., DORAJOO, R., HOPEWELL, J. C., KIM, Y. K., KOIVULA, R. W., LUAN, J., LYYTIKAINEN, L. P., NGUYEN, Q. N., PEREIRA, M. A., POSTMUS, I., RAITAKARI, O. T., SCANNELL BRYAN, M., SCOTT, R. A., SORICE, R., TRAGANTE, V., TRAGLIA, M., WHITE, J., YAMAMOTO, K., ZHANG, Y., ADAIR, L. S., AHMED, A., AKIYAMA, K., ASIF, R., AUNG, T., BARROSO, I., BJONNES, A., BRAUN, T. R., CAI, H., CHANG, L. C., CHEN, C. H., CHENG, C. Y., CHONG, Y. S., COLLINS, R., COURTNEY, R., DAVIES, G., DELGADO, G., DO, L. D., DOEVENDANS, P. A., GANSEVOORT, R. T., GAO, Y. T., GRAMMER, T. B., GRARUP, N., GREWAL, J., GU, D., WANDER, G. S., HARTIKAINEN, A. L., HAZEN, S. L., HE, J., HENG, C. K., HIXSON, J. E., HOFMAN, A., HSU, C., HUANG, W., HUSEMOEN, L. L., HWANG, J. Y., et al. 2015. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet,* 47**,** 1282-93.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

KRUEGER, F. & ANDREWS, S. R. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics,* 27**,** 1571-2.
LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol,* 10**,** R25.
LEEK, J. T., JOHNSON, W. E., PARKER, H. S., JAFFE, A. E. & STOREY, J. D. sva: Surrogate Variable Analysis. *R package version 3.8.0.*
LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.
LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.
LINDBLAD-TOH, K., GARBER, M., ZUK, O., LIN, M. F., PARKER, B. J., WASHIETL, S., KHERADPOUR, P., ERNST, J., JORDAN, G., MAUCELI, E., WARD, L. D., LOWE, C. B., HOLLOWAY, A. K., CLAMP, M., GNERRE, S., ALFOLDI, J., BEAL, K., CHANG, J., CLAWSON, H., CUFF, J., DI PALMA, F., FITZGERALD, S., FLICEK, P., GUTTMAN, M., HUBISZ, M. J., JAFFE, D. B., JUNGREIS, I., KENT, W. J., KOSTKA, D., LARA, M., MARTINS, A. L., MASSINGHAM, T., MOLTKE, I., RANEY, B. J., RASMUSSEN, M. D., ROBINSON, J., STARK, A., VILELLA, A. J., WEN, J., XIE, X., ZODY, M. C., BROAD INSTITUTE SEQUENCING, P., WHOLE GENOME ASSEMBLY, T., BALDWIN, J., BLOOM, T., CHIN, C. W., HEIMAN, D., NICOL, R., NUSBAUM, C., YOUNG, S., WILKINSON, J., WORLEY, K. C., KOVAR, C. L., MUZNY, D. M., GIBBS, R. A., BAYLOR COLLEGE OF MEDICINE HUMAN GENOME SEQUENCING CENTER SEQUENCING, T., CREE, A., DIHN, H. H., FOWLER, G., JHANGIANI, S., JOSHI, V., LEE, S., LEWIS, L. R., NAZARETH, L. V., OKWUONU, G., SANTIBANEZ, J., WARREN, W. C., MARDIS, E. R., WEINSTOCK, G. M., WILSON, R. K., GENOME INSTITUTE AT WASHINGTON, U., DELEHAUNTY, K., DOOLING, D., FRONIK, C., FULTON, L., FULTON, B., GRAVES, T., MINX, P., SODERGREN, E., BIRNEY, E., MARGULIES, E. H., HERRERO, J., GREEN, E. D., HAUSSLER, D., SIEPEL, A., GOLDMAN, N., POLLARD, K. S., PEDERSEN, J. S., LANDER, E. S. & KELLIS, M. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature,* 478**,** 476-82.
LOWE, R., GEMMA, C., BEYAN, H., HAWA, M. I., BAZEOS, A., LESLIE, R. D., MONTPETIT, A., RAKYAN, V. K. & RAMAGOPALAN, S. V. 2013. Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics,* 8**,** 445-54.
MAGOC, T. & SALZBERG, S. L. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics,* 27**,** 2957-63.
MAHAJAN, A., GO, M. J., ZHANG, W., BELOW, J. E., GAULTON, K. J., FERREIRA, T., HORIKOSHI, M., JOHNSON, A. D., NG, M. C., PROKOPENKO, I., SALEHEEN, D., WANG, X., ZEGGINI, E., ABECASIS, G. R., ADAIR, L. S., ALMGREN, P., ATALAY, M., AUNG, T., BALDASSARRE, D., BALKAU, B., BAO, Y., BARNETT, A. H., BARROSO, I., BASIT, A., BEEN, L. F., BEILBY, J., BELL, G. I., BENEDIKTSSON, R., BERGMAN, R. N., BOEHM, B. O., BOERWINKLE, E., BONNYCASTLE, L. L., BURTT, N., CAI, Q., CAMPBELL, H., CAREY, J., CAUCHI, S., CAULFIELD, M., CHAN, J. C., CHANG, L. C., CHANG, T. J., CHANG, Y. C., CHARPENTIER, G., CHEN, C. H., CHEN, H., CHEN, Y. T., CHIA, K. S.,

CHIDAMBARAM, M., CHINES, P. S., CHO, N. H., CHO, Y. M., CHUANG, L. M., COLLINS, F. S., CORNELIS, M. C., COUPER, D. J., CRENSHAW, A. T., VAN DAM, R. M., DANESH, J., DAS, D., DE FAIRE, U., DEDOUSSIS, G., DELOUKAS, P., DIMAS, A. S., DINA, C., DONEY, A. S., DONNELLY, P. J., DORKHAN, M., VAN DUIJN, C., DUPUIS, J., EDKINS, S., ELLIOTT, P., EMILSSON, V., ERBEL, R., ERIKSSON, J. G., ESCOBEDO, J., ESKO, T., EURY, E., FLOREZ, J. C., FONTANILLAS, P., FOROUHI, N. G., FORSEN, T., FOX, C., FRASER, R. M., FRAYLING, T. M., FROGUEL, P., FROSSARD, P., GAO, Y., GERTOW, K., GIEGER, C., GIGANTE, B., GRALLERT, H., GRANT, G. B., GRROP, L. C., GROVES, C. J., GRUNDBERG, E., GUIDUCCI, C., HAMSTEN, A., HAN, B. G., HARA, K., HASSANALI, N., et al. 2014. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet,* 46**,** 234-44.

MALLER, J. B., MCVEAN, G., BYRNES, J., VUKCEVIC, D., PALIN, K., SU, Z., HOWSON, J. M., AUTON, A., MYERS, S., MORRIS, A., PIRINEN, M., BROWN, M. A., BURTON, P. R., CAULFIELD, M. J., COMPSTON, A., FARRALL, M., HALL, A. S., HATTERSLEY, A. T., HILL, A. V., MATHEW, C. G., PEMBREY, M., SATSANGI, J., STRATTON, M. R., WORTHINGTON, J., CRADDOCK, N., HURLES, M., OUWEHAND, W., PARKES, M., RAHMAN, N., DUNCANSON, A., TODD, J. A., KWIATKOWSKI, D. P., SAMANI, N. J., GOUGH, S. C., MCCARTHY, M. I., DELOUKAS, P. & DONNELLY, P. 2012. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet,* 44**,** 1294-301.

MITCHELL, J. S., LI, N., WEINHOLD, N., FORSTI, A., ALI, M., VAN DUIN, M., THORLEIFSSON, G., JOHNSON, D. C., CHEN, B., HALVARSSON, B. M., GUDBJARTSSON, D. F., KUIPER, R., STEPHENS, O. W., BERTSCH, U., BRODERICK, P., CAMPO, C., EINSELE, H., GREGORY, W. A., GULLBERG, U., HENRION, M., HILLENGASS, J., HOFFMANN, P., JACKSON, G. H., JOHNSSON, E., JOUD, M., KRISTINSSON, S. Y., LENHOFF, S., LENIVE, O., MELLQVIST, U. H., MIGLIORINI, G., NAHI, H., NELANDER, S., NICKEL, J., NOTHEN, M. M., RAFNAR, T., ROSS, F. M., DA SILVA FILHO, M. I., SWAMINATHAN, B., THOMSEN, H., TURESSON, I., VANGSTED, A., VOGEL, U., WAAGE, A., WALKER, B. A., WIHLBORG, A. K., BROYL, A., DAVIES, F. E., THORSTEINSDOTTIR, U., LANGER, C., HANSSON, M., KAISER, M., SONNEVELD, P., STEFANSSON, K., MORGAN, G. J., GOLDSCHMIDT, H., HEMMINKI, K., NILSSON, B. & HOULSTON, R. S. 2016. Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nat Commun,* 7**,** 12050.

MORAN, I., AKERMAN, I., VAN DE BUNT, M., XIE, R., BENAZRA, M., NAMMO, T., ARNES, L., NAKIC, N., GARCIA-HURTADO, J., RODRIGUEZ-SEGUI, S., PASQUALI, L., SAUTY-COLACE, C., BEUCHER, A., SCHARFMANN, R., VAN ARENSBERGEN, J., JOHNSON, P. R., BERRY, A., LEE, C., HARKINS, T., GMYR, V., PATTOU, F., KERR-CONTE, J., PIEMONTI, L., BERNEY, T., HANLEY, N., GLOYN, A. L., SUSSEL, L., LANGMAN, L., BRAYMAN, K. L., SANDER, M., MCCARTHY, M. I., RAVASSARD, P. & FERRER, J. 2012. Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab,* 16**,** 435-48.

MORRIS, A. P., VOIGHT, B. F., TESLOVICH, T. M., FERREIRA, T., SEGRE, A. V., STEINTHORSDOTTIR, V., STRAWBRIDGE, R. J., KHAN, H., GRALLERT, H.,

MAHAJAN, A., PROKOPENKO, I., KANG, H. M., DINA, C., ESKO, T., FRASER, R. M., KANONI, S., KUMAR, A., LAGOU, V., LANGENBERG, C., LUAN, J., LINDGREN, C. M., MULLER-NURASYID, M., PECHLIVANIS, S., RAYNER, N. W., SCOTT, L. J., WILTSHIRE, S., YENGO, L., KINNUNEN, L., ROSSIN, E. J., RAYCHAUDHURI, S., JOHNSON, A. D., DIMAS, A. S., LOOS, R. J., VEDANTAM, S., CHEN, H., FLOREZ, J. C., FOX, C., LIU, C. T., RYBIN, D., COUPER, D. J., KAO, W. H., LI, M., CORNELIS, M. C., KRAFT, P., SUN, Q., VAN DAM, R. M., STRINGHAM, H. M., CHINES, P. S., FISCHER, K., FONTANILLAS, P., HOLMEN, O. L., HUNT, S. E., JACKSON, A. U., KONG, A., LAWRENCE, R., MEYER, J., PERRY, J. R., PLATOU, C. G., POTTER, S., REHNBERG, E., ROBERTSON, N., SIVAPALARATNAM, S., STANCAKOVA, A., STIRRUPS, K., THORLEIFSSON, G., TIKKANEN, E., WOOD, A. R., ALMGREN, P., ATALAY, M., BENEDIKTSSON, R., BONNYCASTLE, L. L., BURTT, N., CAREY, J., CHARPENTIER, G., CRENSHAW, A. T., DONEY, A. S., DORKHAN, M., EDKINS, S., EMILSSON, V., EURY, E., FORSEN, T., GERTOW, K., GIGANTE, B., GRANT, G. B., GROVES, C. J., GUIDUCCI, C., HERDER, C., HREIDARSSON, A. B., HUI, J., JAMES, A., JONSSON, A., RATHMANN, W., KLOPP, N., KRAVIC, J., KRJUTSKOV, K., LANGFORD, C., LEANDER, K., LINDHOLM, E., LOBBENS, S., MANNISTO, S., et al. 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet,* 44**,** 981-90.

OLSSON, A. H., VOLKOV, P., BACOS, K., DAYEH, T., HALL, E., NILSSON, E. A., LADENVALL, C., RONN, T. & LING, C. 2014. Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. *PLoS Genet,* 10**,** e1004735.

PARKER, S. C., STITZEL, M. L., TAYLOR, D. L., OROZCO, J. M., ERDOS, M. R., AKIYAMA, J. A., VAN BUEREN, K. L., CHINES, P. S., NARISU, N., PROGRAM, N. C. S., BLACK, B. L., VISEL, A., PENNACCHIO, L. A., COLLINS, F. S., NATIONAL INSTITUTES OF HEALTH INTRAMURAL SEQUENCING CENTER COMPARATIVE SEQUENCING PROGRAM, A. & AUTHORS, N. C. S. P. 2013. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A,* 110**,** 17921-6.

PASQUALI, L., GAULTON, K. J., RODRIGUEZ-SEGUI, S. A., MULARONI, L., MIGUEL-ESCALADA, I., AKERMAN, I., TENA, J. J., MORAN, I., GOMEZ-MARIN, C., VAN DE BUNT, M., PONSA-COBAS, J., CASTRO, N., NAMMO, T., CEBOLA, I., GARCIA-HURTADO, J., MAESTRO, M. A., PATTOU, F., PIEMONTI, L., BERNEY, T., GLOYN, A. L., RAVASSARD, P., GOMEZ-SKARMETA, J. L., MULLER, F., MCCARTHY, M. I. & FERRER, J. 2014. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet,* 46**,** 136-43.

PICKRELL, J. K. 2014. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet,* 94**,** 559-73.

PRUIM, R. J., WELCH, R. P., SANNA, S., TESLOVICH, T. M., CHINES, P. S., GLIEDT, T. P., BOEHNKE, M., ABECASIS, G. R. & WILLER, C. J. 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics,* 26**,** 2336-7.

QUINLAN, A. R. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics,* 47**,** 11 12 1-34.

RAVASSARD, P., HAZHOUZ, Y., PECHBERTY, S., BRICOUT-NEVEU, E., ARMANET, M., CZERNICHOW, P. & SCHARFMANN, R. 2011. A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. *J Clin Invest,* 121**,** 3589-97.

RCORETEAM R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.

ROADMAP EPIGENOMICS, C., KUNDAJE, A., MEULEMAN, W., ERNST, J., BILENKY, M., YEN, A., HERAVI-MOUSSAVI, A., KHERADPOUR, P., ZHANG, Z., WANG, J., ZILLER, M. J., AMIN, V., WHITAKER, J. W., SCHULTZ, M. D., WARD, L. D., SARKAR, A., QUON, G., SANDSTROM, R. S., EATON, M. L., WU, Y. C., PFENNING, A. R., WANG, X., CLAUSSNITZER, M., LIU, Y., COARFA, C., HARRIS, R. A., SHORESH, N., EPSTEIN, C. B., GJONESKA, E., LEUNG, D., XIE, W., HAWKINS, R. D., LISTER, R., HONG, C., GASCARD, P., MUNGALL, A. J., MOORE, R., CHUAH, E., TAM, A., CANFIELD, T. K., HANSEN, R. S., KAUL, R., SABO, P. J., BANSAL, M. S., CARLES, A., DIXON, J. R., FARH, K. H., FEIZI, S., KARLIC, R., KIM, A. R., KULKARNI, A., LI, D., LOWDON, R., ELLIOTT, G., MERCER, T. R., NEPH, S. J., ONUCHIC, V., POLAK, P., RAJAGOPAL, N., RAY, P., SALLARI, R. C., SIEBENTHALL, K. T., SINNOTT-ARMSTRONG, N. A., STEVENS, M., THURMAN, R. E., WU, J., ZHANG, B., ZHOU, X., BEAUDET, A. E., BOYER, L. A., DE JAGER, P. L., FARNHAM, P. J., FISHER, S. J., HAUSSLER, D., JONES, S. J., LI, W., MARRA, M. A., MCMANUS, M. T., SUNYAEV, S., THOMSON, J. A., TLSTY, T. D., TSAI, L. H., WANG, W., WATERLAND, R. A., ZHANG, M. Q., CHADWICK, L. H., BERNSTEIN, B. E., COSTELLO, J. F., ECKER, J. R., HIRST, M., MEISSNER, A., MILOSAVLJEVIC, A., REN, B., STAMATOYANNOPOULOS, J. A., WANG, T. & KELLIS, M. 2015. Integrative analysis of 111 reference human epigenomes. *Nature,* 518**,** 317-30.

ROMAN, T. S., CANNON, M. E., VADLAMUDI, S., BUCHKOVICH, M. L., WOLFORD, B. N., WELCH, R. P., MORKEN, M. A., KWON, G. J., VARSHNEY, A., KURSAWE, R., WU, Y., JACKSON, A. U., SEQUENCING PROGRAM, N. C., ERDOS, M. R., KUUSISTO, J., LAAKSO, M., SCOTT, L. J., BOEHNKE, M., COLLINS, F. S., PARKER, S. C. J., STITZEL, M. L. & MOHLKE, K. L. 2017. A Type 2 Diabetes-Associated Functional Regulatory Variant in a Pancreatic Islet Enhancer at the Adcy5 Locus. *Diabetes*.

SCOTT, R. A., SCOTT, L. J., MAGI, R., MARULLO, L., GAULTON, K. J., KAAKINEN, M., PERVJAKOVA, N., PERS, T. H., JOHNSON, A. D., EICHER, J. D., JACKSON, A. U., FERREIRA, T., LEE, Y., MA, C., STEINTHORSDOTTIR, V., THORLEIFSSON, G., QI, L., VAN ZUYDAM, N. R., MAHAJAN, A., CHEN, H., ALMGREN, P., VOIGHT, B. F., GRALLERT, H., MULLER-NURASYID, M., RIED, J. S., RAYNER, W. N., ROBERTSON, N., KARSSEN, L. C., VAN LEEUWEN, E. M., WILLEMS, S. M., FUCHSBERGER, C., KWAN, P., TESLOVICH, T. M., CHANDA, P., LI, M., LU, Y., DINA, C., THUILLIER, D., YENGO, L., JIANG, L., SPARSO, T., KESTLER, H. A., CHHEDA, H., EISELE, L., GUSTAFSSON, S., FRANBERG, M., STRAWBRIDGE, R. J., BENEDIKTSSON, R., HREIDARSSON, A. B., KONG, A., SIGURETHSSON, G., KERRISON, N. D., LUAN, J., LIANG, L., MEITINGER, T., RODEN, M., THORAND, B., ESKO, T., MIHAILOV, E., FOX, C., LIU, C. T., RYBIN, D., ISOMAA, B., LYSSENKO, V., TUOMI, T., COUPER, D. J., PANKOW, J. S., GRARUP, N., HAVE, C. T., JORGENSEN, M. E., JORGENSEN, T., LINNEBERG, A., CORNELIS, M. C., VAN DAM, R. M., HUNTER, D. J., KRAFT, P., SUN, Q., EDKINS, S., OWEN, K. R.,

PERRY, J. R., WOOD, A. R., ZEGGINI, E., TAJES-FERNANDES, J., ABECASIS, G. R., BONNYCASTLE, L. L., CHINES, P. S., STRINGHAM, H. M., KOISTINEN, H. A., KINNUNEN, L., SENNBLAD, B., MUHLEISEN, T. W., NOTHEN, M. M., PECHLIVANIS, S., BALDASSARRE, D., GERTOW, K., HUMPHRIES, S. E., TREMOLI, E., KLOPP, N., MEYER, J., STEINBACH, G., et al. 2017. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes*.

STADLER, M. B., MURR, R., BURGER, L., IVANEK, R., LIENERT, F., SCHOLER, A., VAN NIMWEGEN, E., WIRBELAUER, C., OAKELEY, E. J., GAIDATZIS, D., TIWARI, V. K. & SCHUBELER, D. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature,* 480**,** 490-5.

VAN DE BUNT, M., MANNING FOX, J. E., DAI, X., BARRETT, A., GREY, C., LI, L., BENNETT, A. J., JOHNSON, P. R., RAJOTTE, R. V., GAULTON, K. J., DERMITZAKIS, E. T., MACDONALD, P. E., MCCARTHY, M. I. & GLOYN, A. L. 2015. 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,* 11**,** e1005694.

VAN DE GEIJN, B., MCVICKER, G., GILAD, Y. & PRITCHARD, J. K. 2015. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods,* 12**,** 1061-3.

VARSHNEY, A., SCOTT, L. J., WELCH, R. P., ERDOS, M. R., CHINES, P. S., NARISU, N., ALBANUS, R. D., ORCHARD, P., WOLFORD, B. N., KURSAWE, R., VADLAMUDI, S., CANNON, M. E., DIDION, J. P., HENSLEY, J., KIRILUSHA, A., PROGRAM, N. C. S., BONNYCASTLE, L. L., TAYLOR, D. L., WATANABE, R., MOHLKE, K. L., BOEHNKE, M., COLLINS, F. S., PARKER, S. C. & STITZEL, M. L. 2017. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci U S A,* 114**,** 2301-2306.

VENTHAM, N. T., KENNEDY, N. A., ADAMS, A. T., KALLA, R., HEATH, S., O'LEARY, K. R., DRUMMOND, H., CONSORTIUM, I. B., CONSORTIUM, I. C., WILSON, D. C., GUT, I. G., NIMMO, E. R. & SATSANGI, J. 2016. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun,* 7**,** 13507.

VOIGHT, B. F., SCOTT, L. J., STEINTHORSDOTTIR, V., MORRIS, A. P., DINA, C., WELCH, R. P., ZEGGINI, E., HUTH, C., AULCHENKO, Y. S., THORLEIFSSON, G., MCCULLOCH, L. J., FERREIRA, T., GRALLERT, H., AMIN, N., WU, G., WILLER, C. J., RAYCHAUDHURI, S., MCCARROLL, S. A., LANGENBERG, C., HOFMANN, O. M., DUPUIS, J., QI, L., SEGRE, A. V., VAN HOEK, M., NAVARRO, P., ARDLIE, K., BALKAU, B., BENEDIKTSSON, R., BENNETT, A. J., BLAGIEVA, R., BOERWINKLE, E., BONNYCASTLE, L. L., BENGTSSON BOSTROM, K., BRAVENBOER, B., BUMPSTEAD, S., BURTT, N. P., CHARPENTIER, G., CHINES, P. S., CORNELIS, M., COUPER, D. J., CRAWFORD, G., DONEY, A. S., ELLIOTT, K. S., ELLIOTT, A. L., ERDOS, M. R., FOX, C. S., FRANKLIN, C. S., GANSER, M., GIEGER, C., GRARUP, N., GREEN, T., GRIFFIN, S., GROVES, C. J., GUIDUCCI, C., HADJADJ, S., HASSANALI, N., HERDER, C., ISOMAA, B., JACKSON, A. U., JOHNSON, P. R., JORGENSEN, T., KAO, W. H., KLOPP, N., KONG, A., KRAFT, P., KUUSISTO, J., LAURITZEN, T., LI, M., LIEVERSE, A., LINDGREN, C. M., LYSSENKO, V., MARRE, M., MEITINGER, T., MIDTHJELL, K., MORKEN, M. A., NARISU, N., NILSSON, P.,

OWEN, K. R., PAYNE, F., PERRY, J. R., PETERSEN, A. K., PLATOU, C., PROENCA, C., PROKOPENKO, I., RATHMANN, W., RAYNER, N. W., ROBERTSON, N. R., ROCHELEAU, G., RODEN, M., SAMPSON, M. J., SAXENA, R., SHIELDS, B. M., SHRADER, P., SIGURDSSON, G., SPARSO, T., STRASSBURGER, K., STRINGHAM, H. M., SUN, Q., SWIFT, A. J., THORAND, B., et al. 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet,* 42**,** 579-89.

VOLKOV, P., BACOS, K., OFORI, J. K., ESGUERRA, J. L., ELIASSON, L., RONN, T. & LING, C. 2017. Whole-Genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis. *Diabetes,* 66**,** 1074-1085.

WANG, X., TUCKER, N. R., RIZKI, G., MILLS, R., KRIJGER, P. H., DE WIT, E., SUBRAMANIAN, V., BARTELL, E., NGUYEN, X. X., YE, J., LEYTON-MANGE, J., DOLMATOVA, E. V., VAN DER HARST, P., DE LAAT, W., ELLINOR, P. T., NEWTON-CHEH, C., MILAN, D. J., KELLIS, M. & BOYER, L. A. 2016. Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *Elife,* 5.

WICKHAM, H. 2009. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.

WOOD, A. R., JONSSON, A., JACKSON, A. U., WANG, N., VAN LEEWEN, N., PALMER, N. D., KOBES, S., DEELEN, J., BOQUETE-VILARINO, L., PAANANEN, J., STANCAKOVA, A., BOOMSMA, D. I., DE GEUS, E. J., EEKHOFF, E. M., FRITSCHE, A., KRAMER, M., NIJPELS, G., SIMONIS-BIK, A., VAN HAEFTEN, T. W., MAHAJAN, A., BOEHNKE, M., BERGMAN, R. N., TUOMILEHTO, J., COLLINS, F. S., MOHLKE, K. L., BANASIK, K., GROVES, C. J., MCCARTHY, M. I. D., PEARSON, E. R., NATALI, A., MARI, A., BUCHANAN, T. A., TAYLOR, K. D., XIANG, A. H., GJESING, A. P., GRARUP, N., EIBERG, H., PEDERSEN, O., CHEN, Y. D., LAAKSO, M., NORRIS, J. M., SMITH, U., WAGENKNECHT, L. E., BAIER, L., BOWDEN, D. W., HANSEN, T., WALKER, M., WATANABE, R. M., T HART, L. M., HANSON, R. L. & FRAYLING, T. M. 2017. A Genome-Wide Association Study of IVGTT-Based Measures of First Phase Insulin Secretion Refines the Underlying Physiology of Type 2 Diabetes Variants. *Diabetes*.

ZHANG, W., SPECTOR, T. D., DELOUKAS, P., BELL, J. T. & ENGELHARDT, B. E. 2015. Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. *Genome Biol,* 16**,** 14.

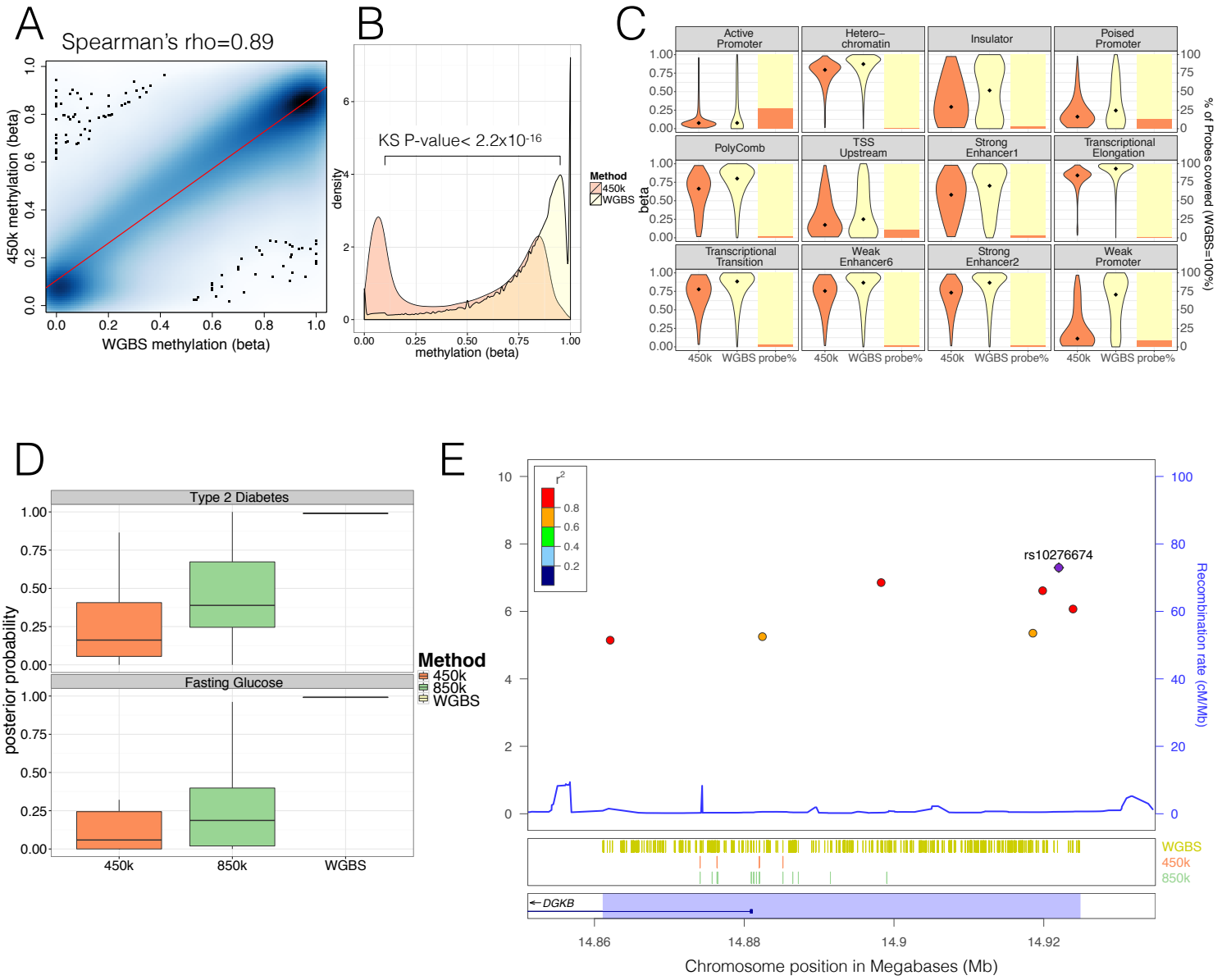


FIGURE 1. Comparison of human pancreatic islet WGBS and 450k methylation data across the genome. A) Smooth Scatter plot shows Spearman's rho correlation between the 450k array (x-axis) and WGBS (y-axis) at overlapping sites. Darker colour indicates higher density of sites. B) Comparison of the 450k array (orange) and WGBS (yellow) methylation levels (x-axis) of all CpGs genome-wide assayed by either method (y-axis shows density). The P-value shown is derived using a Kolmogorov-Smirnov (KS) test. C) For each chromatin state from Parker et al 2013 the methylation levels of all CpG sites independent of overlap (diamond indicates the median) are shown as violin plots (left y-axis) and the CpG probe percentage per state for the 450k array (orange) and WGBS (yellow) are shown as bar-plot (right y-axis). The 450k probes represent the percentage of the total number of CpG sites which is determined by the number of WGBS CpG sites detected (WGBS=100%). D) Distribution of GWAS Posterior Probabilities (Type 2 Diabetes and Fasting Glucose) captured by CpG sites on the 450k array (orange), 850k array (green) and WGBS (yellow/black line). E) Locuszoom plot showing CpG density and credible set SNPs. SNPs are shown with P-values (dots, y-axis left), recombination rate (line, y-axis right) and chromosome positions (x-axis) while CpG and gene annotations are shown below. These annotations include CpGs identified from WGBS (yellow strips), 450k CpG probes (orange stripes), 850k CpG probes (green stripes) and gene overlap (*DGKB* label). The highlighted region in blue captures the 99% credible set region plus additional 1000bp on either side. At the very bottom the position on chromosome 7 is shown in Megabases (Mb).

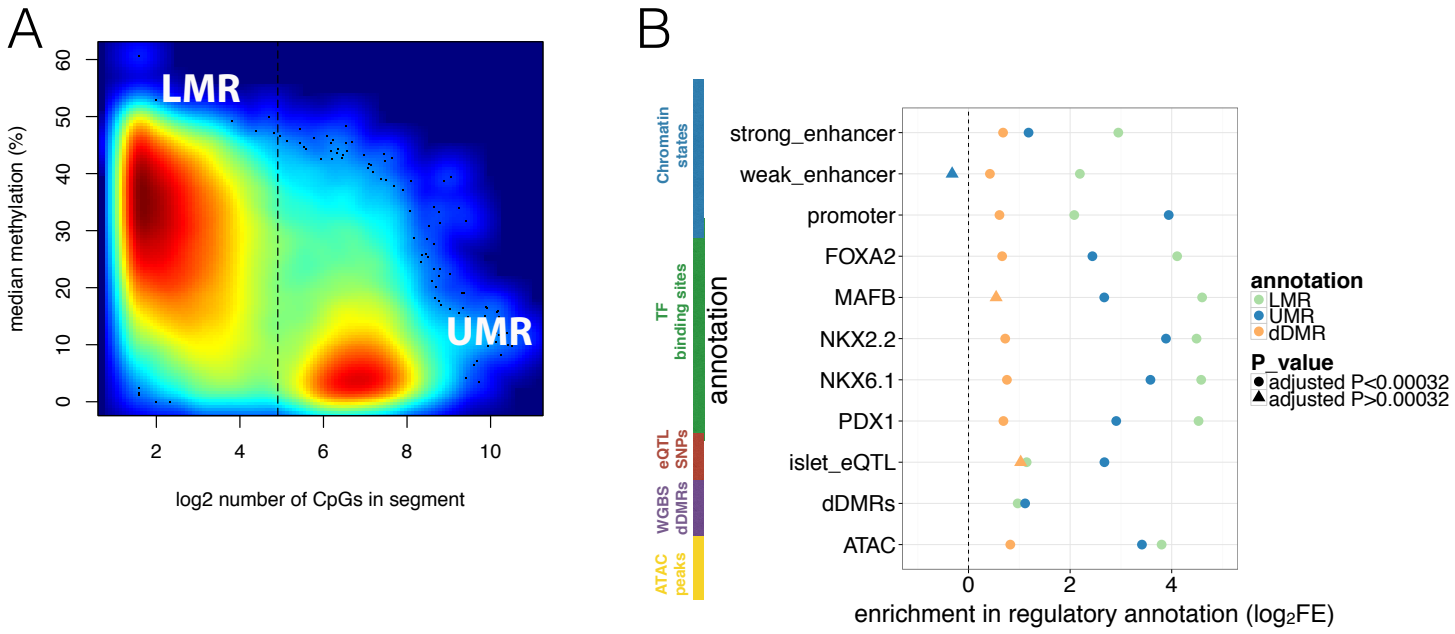


FIGURE 2. Overlap of WGBS hypomethylation and ATAC-seq open chromatin peaks with regulatory annotation . A) Methylation levels in percent (y-axis) and $\log_2$ CpG density (x-axis) of UMR and LMR regulatory regions with the dashed line indicating the CpG-number (30 CpGs) that distinguishes LMRs and UMRs. B) Log_2 Fold Enrichment ($\log_2$FE) of LMRs (green shape), UMRs (blue shape) in various islet annotations is shown. These annotations include islet chromatin states, islet relevant TFBS (FOXA2, MAFB, NKX2.2, NKX.61, PDX1), islet eQTLs, WGBS derived T2D-associated islet disease DMRs (dDMRs) and ATAC-seq open chromatin peaks. The dDMRs were derived from 6 T2D and 8 non-diabetic individuals by Volkov et al 2017 and dDMRs (orange shape) were also tested for enrichment in the aforementioned islet regulatory annotations. For all annotations, the empirically determined Bonferroni adjusted P-value is ≤0.00032 unless otherwise indicated by the shape: a dot corresponds to an Bonferroni adjusted P-value <0.00032 while the 3 triangles indicates Bonferroni adjusted P-values > 0.00032: UMR enrichment adjusted P-value for weak enhancers=1; dDMR enrichment adjusted P-value for MAFB=0.006 and dDMR enrichment adjusted P-value for islet eQTLs=0.01.

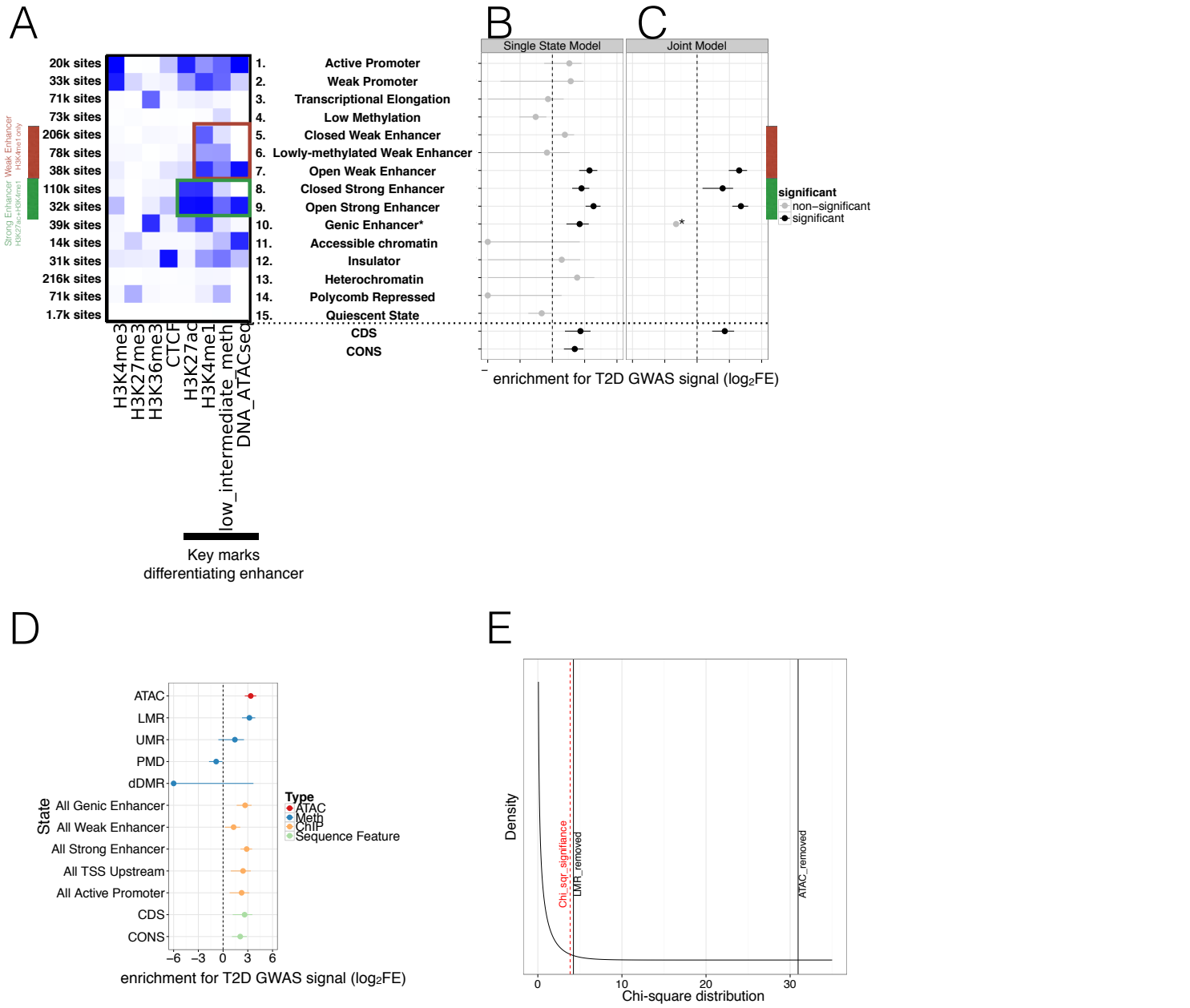


FIGURE 3. Integration of islet epigenetic data to refine chromatin regulatory states and enrichment of these states in T2D GWAS data. A) 15 chromatin states (y-axis) were derived from ChIP histone marks, DNA methylation and ATAC-seq open chromatin annotations (x-axis) using chromHMM. For each state the relevant marks characterising the state are shown. The colour is based on the chromHMM emission parameters and a darker colour indicates a higher frequency of a mark at a given state. Weak enhancers (marked by H3K4me1 alone, red) and strong enhancers (marked by H3K27ac and H3K4me1, green) were subdivided by the chromHMM analysis according to methylation and ATAC-seq status (highlighted in red and green box). The black bar at the x-axis highlights the most important marks for characterising enhancer subtypes. B-C) FGWAS Log_2 Fold Enrichment including 95% CI (log_2FE, x-axis) of all chromatin states (y-axis) in T2D GWAS regions is shown which demonstrate differential enrichment amongst enhancer subclasses in single-feature enrichment analysis. In addition, log2FE of Coding Sequence (CDS) and Conserved Sequence (CONS) annotations are shown to include the effect of protein-coding and conserved regions. Significantly enriched annotations are shown in black while non-siginificant annotations are shown in grey. C) T2D FGWAS maximum likelihood model determined through cross-validation. Log_2FE and 95% CI (x-axis) of annotations included in the maximum likelihood model (y-axis) also demonstrate differential enrichment amongst enhancer subclasses. *Analysis for Genic Enhancers (state 10) did not converge and hence, only a point log2FE estimate is provided. D) Single feature log2FE including 95% CI (x-axis) results are shown highlighting the differences in T2D GWAS enrichment of various annotations. These include ATAC-seq open chromatin peaks (red), WGBS methylation regions (including enhancer-like LMRs, promoter-like UMRs and Partially Methylated Domains, blue), ChiP-seq chromatin states (orange) and CDS (green). E) Chi-square distribution (curved black line) with the indicated results of a maximum likelihood ratio test based on the maximum likelihood difference between a model including LMRs or ATAC-seq peaks compared to the ChIP-only model. The dashed red line indicates significance (P-value<0.05). For all FGWAS enrichment plots the axis has been truncated at -6 to facilitate visualisation and accurate values are provided in the supplementary tables.

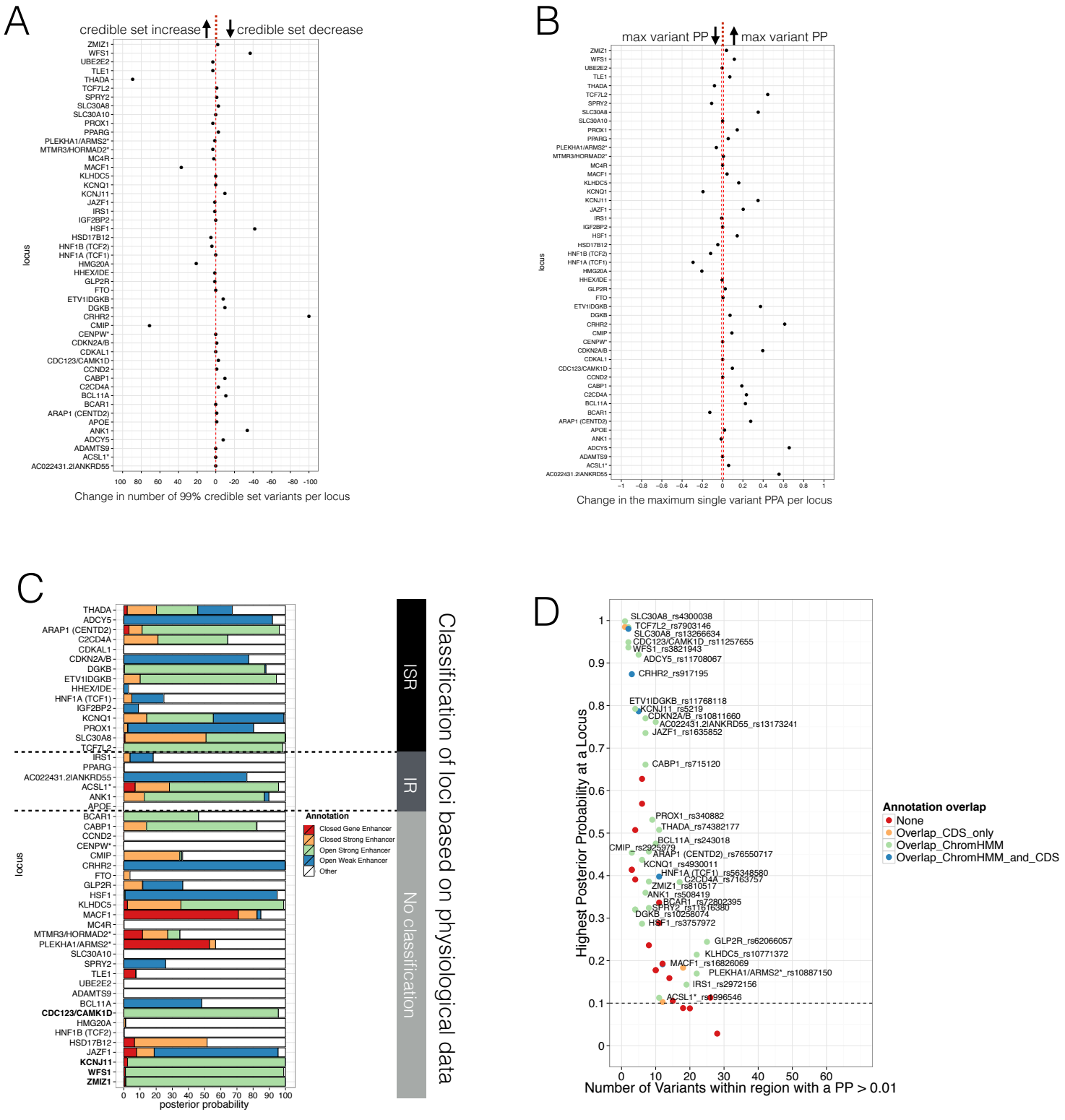


FIGURE 4. Evaluating Posterior Probabilities (PP) derived from the FGWAS maximum likelihood model at significant T2D GWAS loci. (A) Per locus the difference in the number of 99% credible set variants between ChIP+ATAC+Meth and ChIP-only model is shown (positive values indicate a reduction in the number of 99% credible set variants in the ChIP_ATAC_Meth model). B) Per locus the difference in the maximum single variant PPA between the ChIP+ATAC+Meth and ChIP-only model is shown (positive values indicate an increase in the maximum single variant PPA in the ChIP+ATAC+Meth model). C) T2D GWAS loci were classified into insulin secretion (ISR), insulin resistance (IR) or unclassified loci based on genetic association with physiological traits derived from Dimas et al 2014 and Wood et al 2017. In addition, loci with known role in islet genomic regulation or function are highlighted in bold. These include loci with islet eQTLs (*ZMIZ1, CDC123*) and mQTLs (*WFS1, KCNJ11*). D) Identification of T2D GWAS loci and variants enriched for enhancer chromatin states using FGWAS PP. Per locus the highest PPA variant is shown (y-axis) and the number of variants with PPA >0.01 (x-axis). Loci with high PPA variants (min PPA >0.1, dashed horizontal line) that overlap one of the enhancer states (green) are highlighted and the high PPA variants (PPA>0.1) were tested for allelic imbalance in open chromatin.

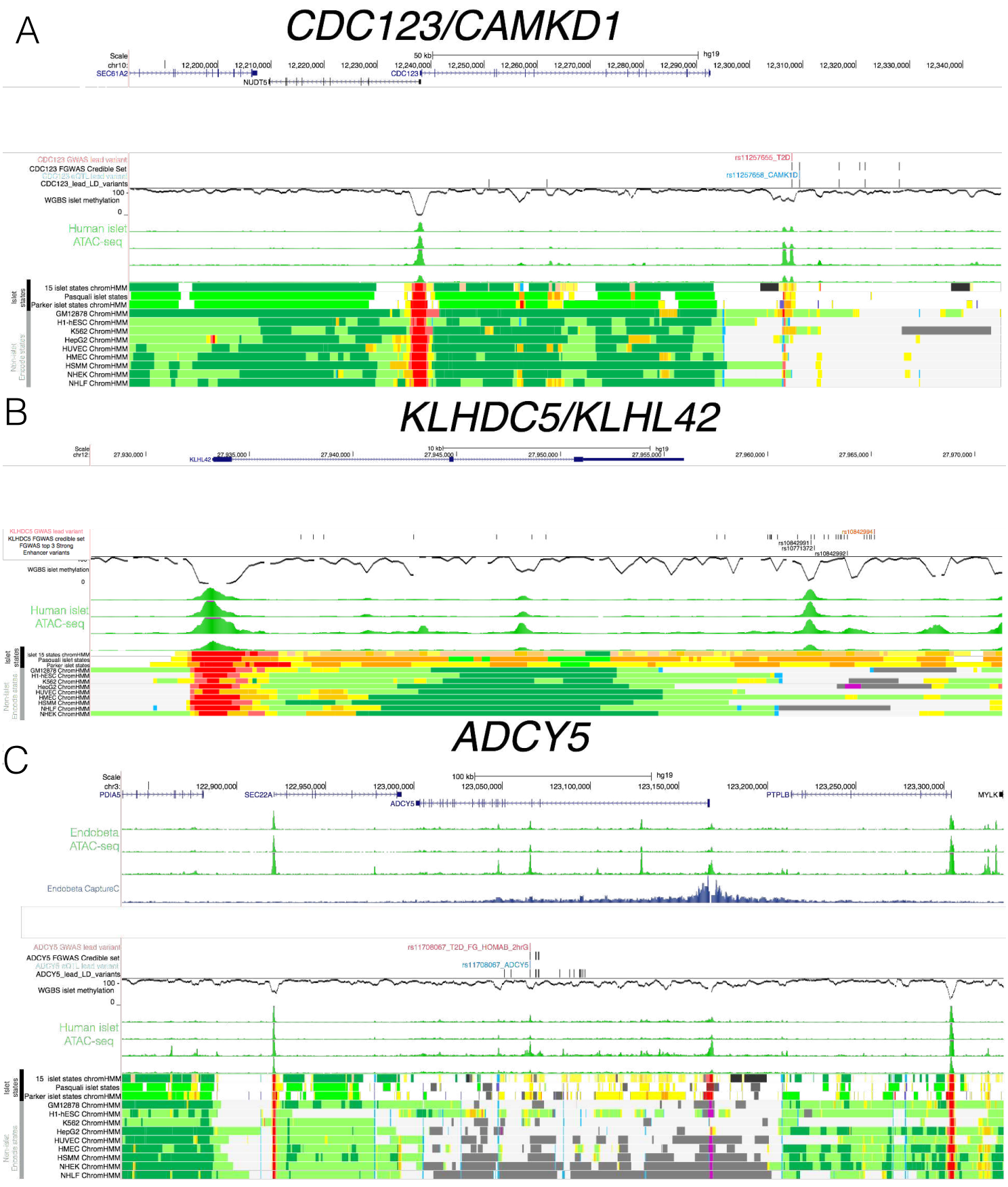
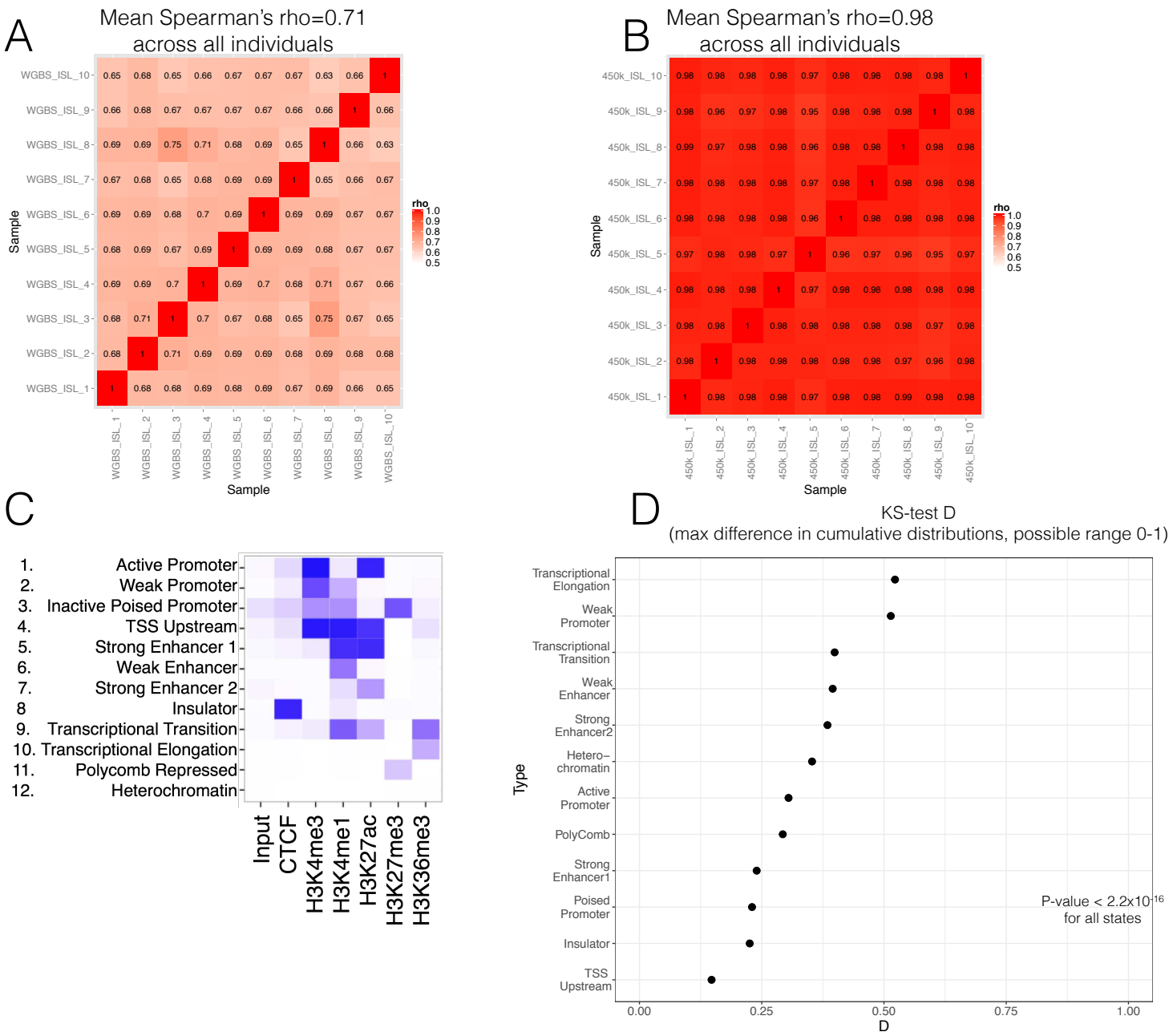
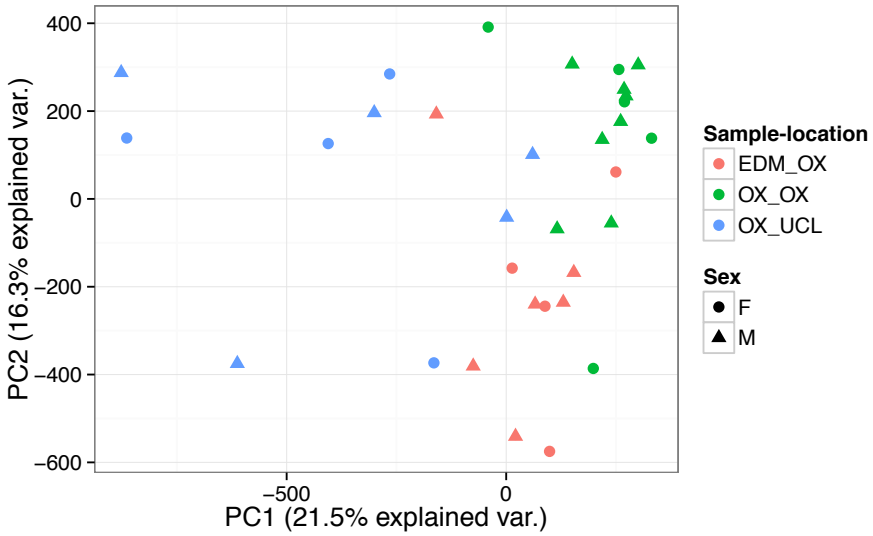
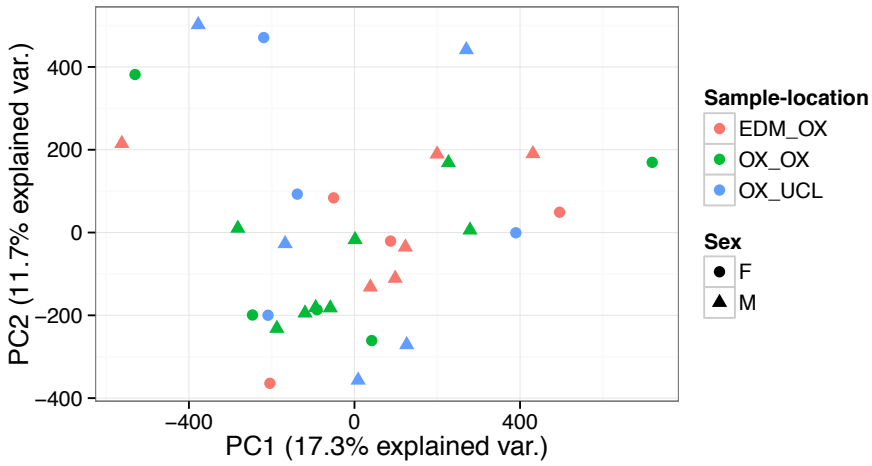


FIGURE 5. Epigenome Landscape of selected loci with allelic imbalance. For each locus A) *CDC123*, B) *KLHDC5* and C) *ADCY5* the following information is shown: 3 ATAC-seq Endoß tracks (green, top), variant level information (depending on the region GWAS lead SNP red, credible set black, eQTL blue and high LD SNPs with r2>0.8 black), WGBS methylation data (black, middle), 4 human islet ATAC-seq tracks (green, middle), islet chromatin states (from this study as well as Parker et al 2013 and Pasquali et al 2014) and Encode chromatin states from 9 cell types (bottom). For *ADCY5* the Capture C results in the Endoß cell line are shown as well (middle blue). Abbreviation for cell types: B-lymphoblastoid cells (GM12878), embryonic stem cells (H1 ES), erythrocytic leukaemia cells (K562), hepatocellular carcinoma cells (HepG2), umbilical vein endothelial cells (HUVEC), mammary epithelial cells (HMEC), skeletal muscle myoblasts (HSMM),normal epidermal keratinocytes (NHEK) and normal lung fibroblasts (NHLF).

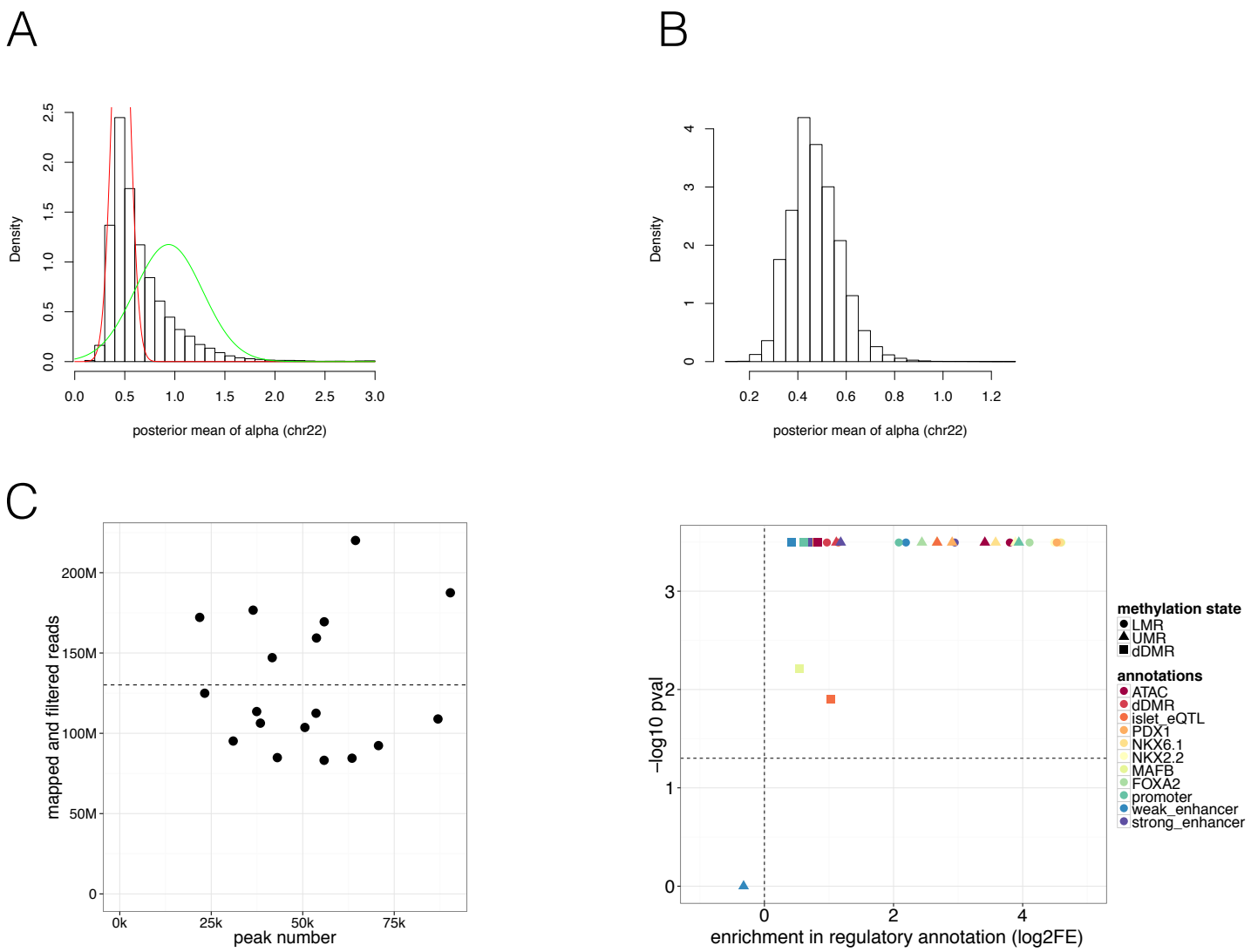


SUPPLEMENTARY FIGURE 1. Correlation of DNA methylation across WGBS and 450k sites and comparison of WGBS and 450k methylation levels across chromatin states A-B) Spearman's rho correlation of DNA methylation across 10 individual (A) WGBS and (B)10 selected (out of 32) 450k samples on the x-axis and y-axis. C) Islet chromatin state definitions based on ChIP-seq data reproduced from Parker et al 2013. TSS: Transcription Start Site D) The differences in the 450k and WGBS methylation level distribution measured as D statistic, which represents the difference in the cumulative distributions and is derived from the Kolmogorov-Smirnov test, are shown for each chromatin state separately.

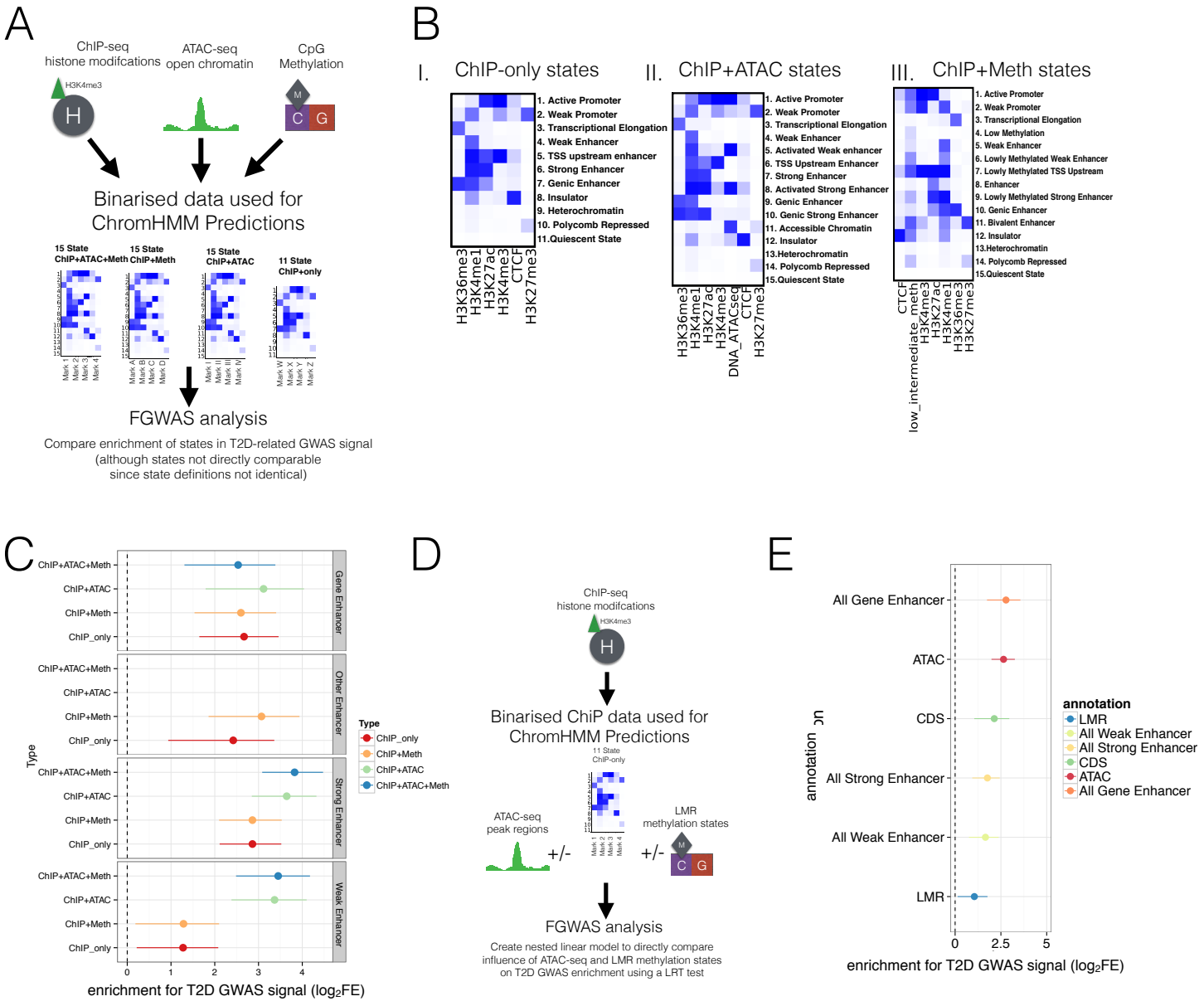




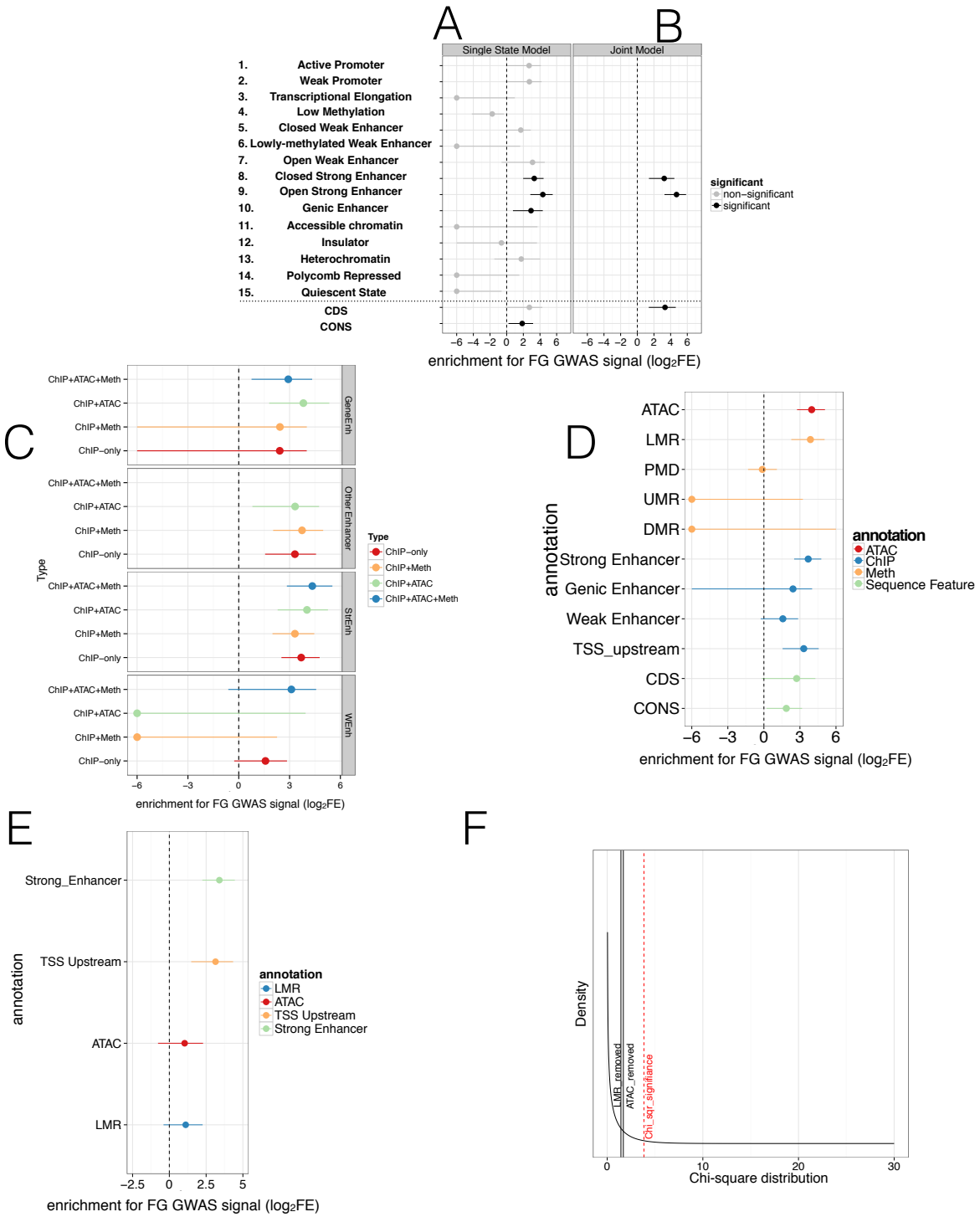
SUPPLEMENTARY FIGURE 7. A-B) PCA analysis of 450k DNA methylation data of 32 human islet samples coloured according to the location of origin and processing (A) before correction for Sample-location and (B) after correction for Sample-location using the ComBat function included in the sva package. The shape indicates sex. Sample location **EDM_OX**: samples obtained from the Alberta Diabetes Institute in Edmonton (Canada) and processed at the University of Oxford. **OX_OX**: samples obtained from Oxford DRWF Human Islet Isolation Facility and processed at the University of Oxford. **OX_UCL**: samples obtained from Oxford DRWF Human Islet Isolation Facility and processed at University College London.



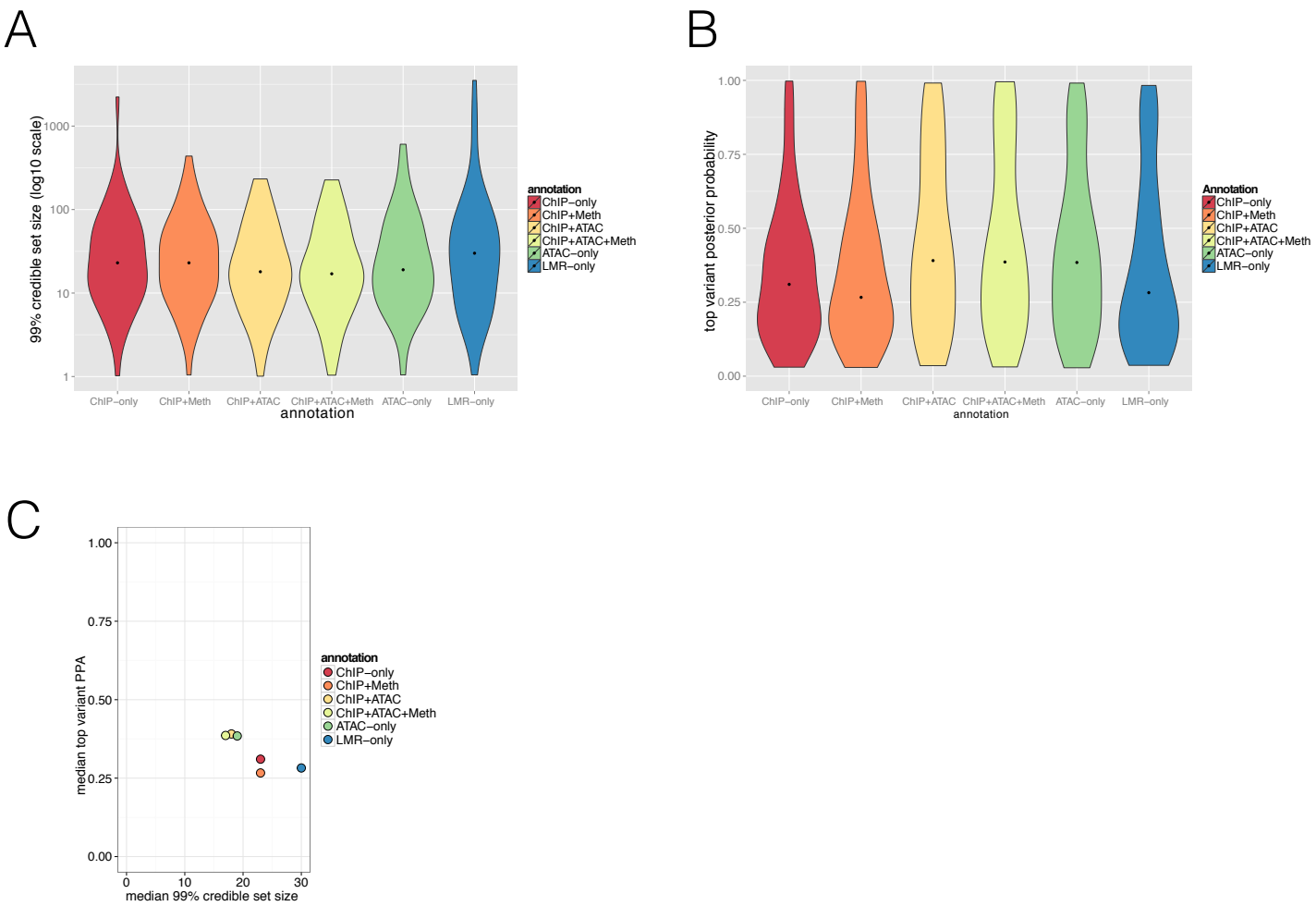
SUPPLEMENTARY FIGURE 2. Identification and removal of Partially Methylated Domains (PMDs) A-B) Density distribution of the alpha value (A) before and (B) after removing PMDs (green curve in (A)) on chromosome 22. Alpha values represent a summary statistic derived from DNA methylation of windows of 100 CpGs and represents an indication of the polarisation status of methylation values in the genome which is expected to contain either highly methylated or unmethylated regions. Distributions with alpha <1 indicate methylation levels that are bimodal with either 0 or 1 methylation. Alpha=1 corresponds to a uniform distribution of methylation; and distributions with alpha>1 tend to have primarily intermediate methylation levels. The red and green curve in (A) represent the non-PMD (red) and PMD regions (green) in the genome. C) Number of peaks (x-axis) and mapped and filtered reads (y-axis) per ATAC-seq islet preparation. The dashed line indicates the mean read number. D) Log_2 Fold Enrichment (log_2FE, x-axis) and associated -log10 Bonferroni adjusted P-values (y-axis) of LMRs (circle), UMRs (triangle)in various islet annotations (colours) is shown. These annotations include islet chromatin states, islet relevant TFBS (FOXA2, MAFB, NKX2.2, NKX.61, PDX1), islet eQTLs, WGBS derived T2D-associated islet disease DMRs (dDMRs) and ATAC-seq open chromatin peaks. dDMRs (square) were also tested for enrichment in the aforementioned islet regulatory annotations. The results cluster near -log10 P-value of 3.5 since most Bonferroni adjusted P-values were more extreme than 0.00032.



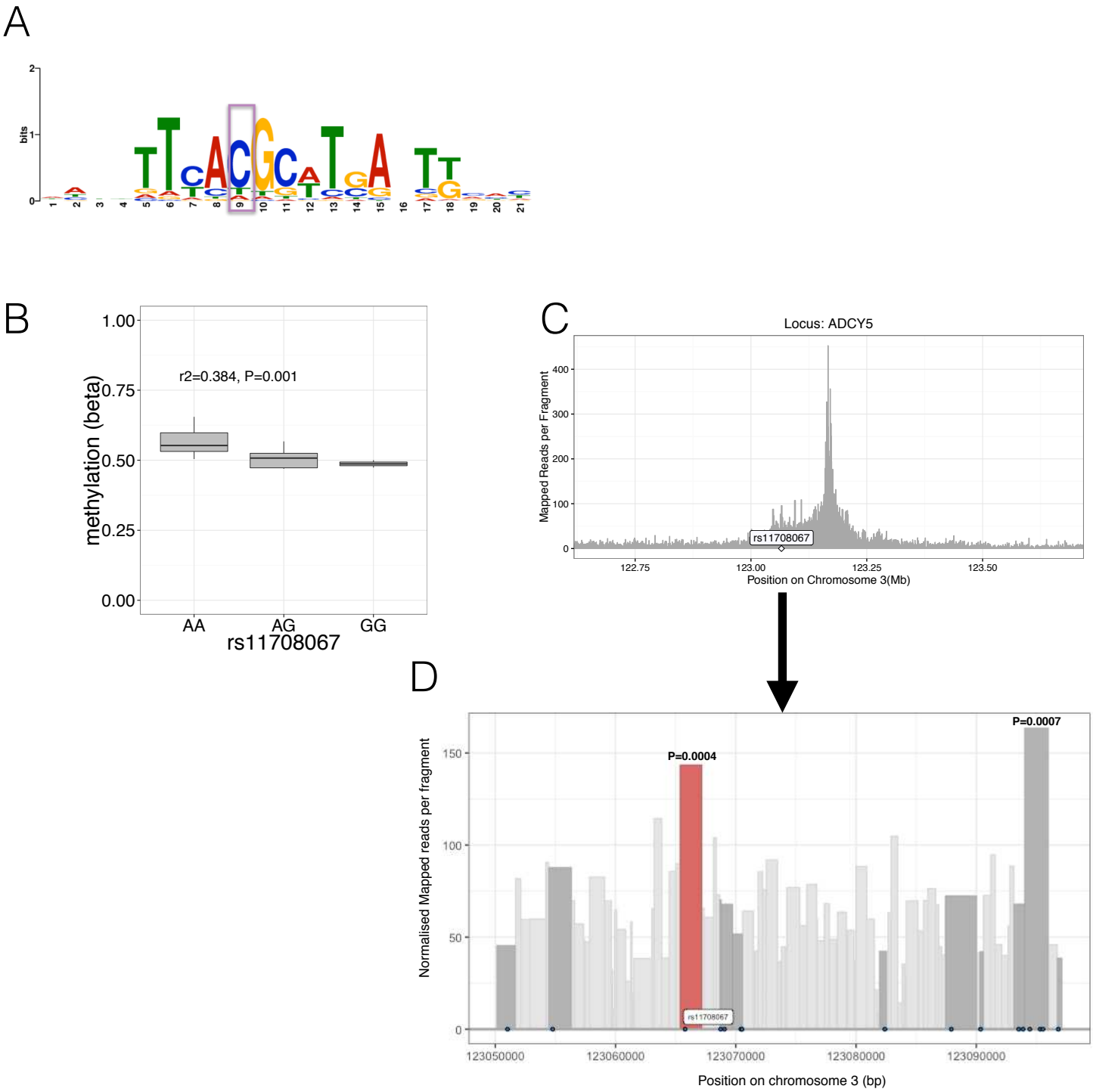
SUPPLEMENTARY FIGURE 3. Prediction of regulatory regions using WGBS data and testing these regions for enrichment in T2D GWAS regions. A) Different combinations of epigenomic data (top) were combined to generate different sets of refined chromatin states (middle, 11 ChIP-only and 15 ChIP+Meth, ChIP+ATAC and ChIP+ATAC-Meth states, see figure S3B and 3A-B for actual states) using chromHMM. These sets of chromatin states were then tested for enrichment in T2D-related GWAS traits using FGWAS to compare enrichment across states (bottom). B) ChromHMM (I) 11 ChIP-only and 15 state (II) ChIP+ATAC state and (III) ChIP+Meth models. C) Single feature log2FE (x-axis) for different enhancer states (grey panels) defined from different combinations of epigenetic marks (y-axis) including ChIP+ATAC+Meth, ChIP+ATAC, ChIP+Meth and ChIP+only. The grey-dashed line indicates the enrichment value of CDS as reference. Enhancers are defined as follows: Strong enhancers are marked by both H3K4me1 and H3K27ac, weak Enhancers are defined by H3K4me1 only, gene enhancers are marked by H3K4me1 and H3K36me3, other enhancers are marked by H3K4me1, H3K4me3 and H3K27ac and are often referred to as TSS upstream regions (only included in the FGWAS T2D model for ChIP-only and ChIP+Meth chromatin states). D) Since chromatin states defined from a different set of epigenomic marks (ChIP-only, ChIP+Meth, ChIP+ATAC and ChIP+ATAC+Meth), as described in S3A-B, are not equivalent and the enrichment can not be easily compared across models, a nested model approach was applied. That is, ChIP-only chromatin states were generated and after evaluating the individual enrichment of each annotation (see Figure 3D), FGWAS maximum likelihood models were defined using ChIP-only, hypomethylated and/or ATAC-seq peak regulatory regions. The combination of all these annotations represented a nested linear model and the changes in maximum likelihood by adding/removing hypomethylated regulatory and ATAC-seq states could be statistically evaluated using a Loglikelihood Ratio Test (LRT) as shown in Figure 3E. E) Maximum likelihood FGWAS nested model combining ChIP-only, ATAC-peaks and LMR states (y-axis) showing log_2FE enrichment (x-axis) which was used for the LRT in Figure 3E. For all FGWAS enrichment plots the axis has been truncated at -6 to facilitate visualisation and accurate values are provided in the supplementary tables.



SUPPLEMENTARY FIGURE 4. Enrichment of refined islet regulatory states in FG GWAS data A) FGWAS Log2 Fold Enrichment including 95% CI (log2FE, x-axis) of all chromatin states (y-axis) in FG GWAS regions. In addition, CDS is shown to also include the effect of protein-coding regions. Significantly enriched annotations are shown in black. B) FG FGWAS maximum likelihood model determined through cross-validation. log_2FE and 95% CI (x-axis) of annotations included in the maximum likelihood model (y-axis) are shown. C) Single feature log_2FE (x-axis) for different enhancer states (grey panels) defined from different combinations of epigenetic marks (y-axis) including ChIP+ATAC+Meth, ChIP+ATAC, ChIP+Meth and ChIP-only. Enhancers are defined as follows: Strong enhancers are marked by both H3K4me1 and H3K27ac, weak Enhancers are defined by H3K4me1 only, gene enhancers are marked by H3K4me1 and H3K36me3, other enhancers are marked by H3K4me1, H3K4me3 and H3K27ac and are often referred to as TSS upstream regions (only included in the FGWAS T2D model for ChIP-only and ChIP+Meth chromatin states). D) Single feature log_2FE including 95% CI (x-axis) results of various annotations derived from ChiP-seq (ChIP-only), ATAC-seq, WGBS methylation status and CDS are shown.E) Maximum likelihood FGWAS nested model combining ChIP-only, ATAC-peaks and LMR states (y-axis) showing log2FE enrichment (x-axis) which was used for the LRT in Supplementary Figure S3F. F) Chi-square distribution (black curved line) with the indicated results of a maximum likelihood ratio test based on the maximum likelihood difference between a model including LMRs or ATAC-seq peaks compared to the ChIP-only model. The dashed line indicates significance (P-value<0.05). For all FGWAS enrichment plots the axis has been truncated at -6 to facilitate visualisation and accurate values are provided in the supplementary tables.



Supplementary Figure 5. Evaluating annotation effect on Posterior Probabilities (PPA) derived from the FGWAS maximum likelihood model at significant T2D GWAS loci. (A) Violin plot showing the distribution of 99% credible set variant size (y-axis, log10 scale) of different annotation types used (x-axis, ChIP-only, ChIP+Meth, ChIP+ATAC, ChIP+ATAC+Meth, ATAC-only and LMR-only model). B) Violin plot showing the distribution in the maximum single variant PPA (y-axis) of different annotation types used (x-axis, ChIP-only, ChIP+Meth, ChIP+ATAC, ChIP+ATAC+Meth, ATAC-only and LMR-only model). Dots indicate mean value. C) Median 99% credible set variant size (x-axis) and median top variant PPA (y-axis) information for ChIP-only, ChIP+Meth, ChIP +ATAC, ChIP+ATAC+Meth, ATAC-only and LMR-only models.



SUPPLEMENTARY FIGURE 6. A) Predicted *PAX6* Transcription factor binding motif likely affected by allelic imbalance of the variant rs10842991 (highlighted in purple). B) The *ADCY5* rs11708067 risk A allele was associated with increased methylation levels (y-axis, while genotypes are shown on the x-axis) . C) Chromatin Capture (Capture C) in the human beta-cell line EndoßH1 showed interactions between the *ADCY5* promoter (peak) and the flanking regions of the promoter. The x-axis shows the position on the chromosome in Mb while the y-axis indicates mapped reads per fragment. D) Chromatin Capture (Capture C) in the human beta-cell line EndoßH1 focussed at the genomic region (~47kb) near the variant rs11708067 (highlighted) and variants in high LD (r2>0.8) with it (variants are depicted as black dots). Fragments containing rs11708067 (red) or other high LD variants (dark grey) are highlighted. The x-axis shows the position on the chromosome in bp while the y-axis indicates normalised mapped reads per fragment. The two fragments with P-values have a significant (FDR <0.05) number of normalised read counts over background: The fragment with the P-value on the left (in red) contains rs11708067 while the fragment with the P-value on the right harbours rs2877716, rs6798189, rs56371916.