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], Kyle Gaulton[3], Amy Barrett[2], Amanda J Bennett[2], Jason M Torres[1], Vibe Nylander[2], Anubha Mahajan[1], 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

39 increasingly-precise genetic and islet epigenomic information accelerates
40 definition of causal mechanisms implicated in T2D pathogenesis.
41
42 # 1. Introduction
43 T2D is a complex disease characterised by insulin resistance and reduced beta
44 cell function. Recent GWAS have identified a large number of T2D susceptibility
45 loci (Mahajan et al., 2014, Morris et al., 2012, Voight et al., 2010), the majority of
46 which affect insulin secretion and beta cell function (Dimas et al., 2014, Wood et
47 al., 2017). However, most GWAS signals map to the non-coding genome and
48 identification of the molecular mechanisms through which non-coding variants
49 exert their effect has proven challenging. Several studies have demonstrated that
50 T2D-associated variants map disproportionately to regulatory elements,
51 particularly those which influence RNA expression and cellular function of
52 human pancreatic islets. (Parker et al., 2013, Pasquali et al., 2014, van de Bunt et
53 al., 2015, Olsson et al., 2014, Dayeh et al., 2014, Volkov et al., 2017, Varshney et
54 al., 2017, Gaulton et al., 2015b, Gaulton et al., 2010).
55
56 Characterisation of the islet regulome has until now been limited in scope. The
57 use of DNA methylation and open chromatin data to further annotate ChIP-seq
58 derived chromatin states has successfully uncovered novel biology for other
59 diseases (Wang et al., 2016). Existing methylation studies in islets, however,
60 have either profiled a very small proportion of methylation sites using
61 methylation arrays (Olsson et al., 2014, Dayeh et al., 2014) or focused on T2D-
62 associated disease differentially methylated regions (dDMRs) rather than the
63 integration of DNA methylation status with T2D-relevant GWAS data (Volkov et
64 al., 2017). At the same time, assays of open chromatin in human islets have been
65 restricted to small sample numbers (limiting the potential to capture allelic
66 imbalance in chromatin accessibility for example): these have focussed
67 predominantly on the impact of clustered or "stretch" enhancers (Parker et al.,
68 2013, Pasquali et al., 2014, Gaulton et al., 2010, Varshney et al., 2017). Most
69 importantly, in part due to historical challenges in accessing human islet
70 material or authentic human cellular models, reference annotations of the islet
71 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 assays 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, Supplementary Figure S1A): 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).

105

106 Most previous studies of the relationship between GWAS data and tissue-specific
107 methylation patterns (including those interrogating the relationship between
108 islet methylation and T2D predisposition (Dayeh et al., 2014, Olsson et al.,
109 2014)) had used data generated on the Illumina 450k methylation array
110 (Hannon et al., 2016, Mitchell et al., 2016, Kato et al., 2015, Ventham et al., 2016).
111 For comparative purposes, we generated 450k array methylation data from 32
112 islet samples from non-diabetic donors of European descent (5 overlapping with
113 those from whom WGBS data were generated). As with the WGBS data,
114 methylation levels were highly correlated across individuals (mean CpG
115 methylation Spearman's rho across 32 individual 450k donor=0.98,
116 Supplementary Figure S1B). After pooling 450k array data across samples,
117 methylation profiles generated from the 450k array and WGBS were highly
118 correlated at the small subset of total CpG sites at which they overlap: this was
119 observed across pooled samples (pooled WGBS vs. 450k Spearman's rho=0.89,
120 Figure 1A) and across the 5 donors analysed by both methods (mean Spearman's
121 rho=0.80, not shown).
122

123 WGBS and 450k array data differed substantially in terms of genome-wide
124 coverage. The 450k array was designed to interrogate ~480k CpG sites
125 (approximately 2% of all sites in the genome), selected primarily because they
126 are located near gene promoters and CpG-island regions. This results in marked
127 differences in the distributions of genome-wide methylation values between
128 WGBS and the 450k array. Whilst the WGBS data revealed the expected pattern
129 of widespread high methylation levels with hypomethylation (<50%) restricted
130 to 11.2% (2.6M/23.3M CpG sites) of the genome, the array disproportionately
131 interrogated those hypomethylated sites (218k [46%] of all 450k CpG probes)
132 (Kolmogorov–Smirnov (KS) test for difference, D=0.40, $P<2.2x10^{-16}$) (Figure 1B).
133 These differences in methylation distribution were also evident within specific
134 islet regulatory elements from previously defined standard chromatin state
135 maps (Parker et al., 2013) (Figure 1C, Supplementary Figure 1C-D). We found
136 significant (FDR<0.05) differences between the methylation levels of CpG sites
137 accessed on the array, and those interrogated by WGBS, across most islet

138 chromatin states: the largest differences were observed for weak promoters
139 (median WGBS=0.71 vs. median 450k=0.11, KS test D=0.51, $P<2.2x10^{-16}$,) and
140 weak enhancers (WGBS=0.87 vs. 450k median=0.76, D=0.39, $P<2.2x10^{-16}$,
141 Supplementary Figure 1D).
142
143 In terms of coverage, most chromatin states, apart from promoters, were poorly
144 represented by CpG sites directly interrogated by the array: for example the
145 array assayed only ~2.9% of CpG sites in strong enhancer states (2.7-3.8%
146 depending on strong enhancer subtype, Figure 1C). Although methylation levels
147 were previously reported to be highly correlated across short (0.1-2kb) genomic
148 distances (Zhang et al., 2015, Bell et al., 2011, Eckhardt et al., 2006, Guo et al.,
149 2017), the observed significant differences in the methylation distribution
150 (Figure 1C, Supplementary Figure 1D) across chromatin states including weak
151 promoter (median size 600bp) and enhancer subtypes (median size ranges from
152 200-1200bp) indicate that these correlative effects are not strong enough to
153 counterbalance the low coverage of the 450k array. These findings are consistent
154 with 450k array content being focused towards CpG-dense hypomethylated and
155 permissive promoter regions. This highlights the limited capacity of the array to
156 comprehensively interrogate the global DNA methylome, in particular at distal
157 regulatory states such as enhancers.
158
159 To understand the value of these data to reveal molecular mechanisms at GWAS
160 loci, where we and others had shown enrichment for islet enhancer states
161 (Pasquali et al., 2014, Gaulton et al., 2015a, Parker et al., 2013), we were
162 interested to see how the selective coverage of the array might impact on its
163 ability to interrogate methylation in GWAS-identified regions. We used the
164 largest currently available T2D GWAS data set (from the DIAGRAM consortium
165 involving 26.7k cases and 132.5k controls of predominantly European origin,
166 imputed to the 1000G Mar 2012 reference panel) to identify the "credible sets"
167 of variants at each locus which collectively account for 99% of the posterior
168 probability of association (PPA, (Scott et al., 2017, Maller et al., 2012)).
169

170 To estimate the respective proportions of these T2D-associated variants
171 captured by CpG sites assayed by the 450k array and WGBS, we determined, for
172 each locus, the combined PPA of all 99% credible set variants mapping within
173 1000bp of any CpG site captured. This is based on evidence that across short
174 distances CpG methylation is highly correlated (Zhang et al., 2015, Bell et al.,
175 2011, Eckhardt et al., 2006, Guo et al., 2017) and may be influenced by genetic
176 variants associated with altered transcription factor binding (Do et al., 2016). We
177 found that coverage of this space of putative T2D GWAS variants by the 450k
178 array is low: across GWAS loci, the combined PPA attributable to variants within
179 regions assayed by the array ranged from 0-99% with a median PPA per locus of
180 16% (compared to a WGBS median PPA per locus=99%, KS-test $P<2.2x10^{-16}$,
181 Figure 1D, top). We estimated that the equivalent figure for a recently developed
182 upgrade of the 450k array, which captures ~850k CpG sites and aims to provide
183 better coverage of enhancer regions, would be ~39% (range 0%-99% Figure 1D,
184 top). For instance, at the *DGKB* T2D locus (centred on rs10276674), CpG sites
185 covered by the 450k array interrogated less than 1% of the PPA of associated
186 variants (vs. 99% captured by WGBS); the equivalent figure for the 850k array
187 would be 23% (Figure 1E). We obtained similar results when we performed
188 equivalent analyses using GWAS data for fasting glucose (FG, from the ENGAGE
189 consortium (Horikoshi et al., 2015)), another phenotype dominated by islet
190 dysfunction (Figure 1D, bottom).
191
192 These data indicate that available methylation arrays provide poor genome-wide
193 coverage of methylation status and are notably deficient in capturing
194 methylation status around the distal regulatory enhancer regions most relevant
195 to T2D predisposition. For this reason, we focused subsequent analyses on the
196 WGBS data.
197
198 **2.2 Integration islet methylation and other epigenomic annotations**
199
200 Studies in a variety of other tissues have shown that hypomethylation is a strong
201 indicator of regulatory function (Stadler et al., 2011). More specifically,
202 continuous stretches of CpG-poor Low-Methylated Regions (LMRs, with

203 methylation ranging from 10-50% and containing fewer than 30 CpG sites)
204 denote potential distal regulatory elements such as enhancers, while stretches of
205 CpG-rich UnMethylated Regions (UMRs, containing more than 30 CpG sites) are
206 more likely to represent proximal regulatory elements including promoters
207 (Burger et al., 2013). We detected 37.1k LMRs, 13.6k UMRs (Figure 2A) and
208 10.7k Partially Methylated Domains (PMDs, Methods and Supplementary Figure
209 2A-B). PMDs represent large regions of unordered methylation states associated
210 with DNA sequence features (Gaidatzis et al., 2014). As anticipated, we found
211 significant enrichment of LMRs with weak and strong enhancer states (69.2% of
212 islet LMRs overlapped islet strong and weak enhancer states, $\log_2$FE=2.2-2.9,
213 FDR<0.05, Figure 2B) as defined by islet chromatin state maps derived from
214 existing ChIP-seq data (Parker et al., 2013) (see Supplementary Figure 1C).
215 Similarly, UMRs were enriched for islet active promoter chromatin states (90.8%
216 of UMRs overlapped islet active promoters, $\log_2$FE=3.9, FDR <0.05, Figure 2B).
217
218 To further characterise these hypomethylation domains, we overlapped
219 information from analyses of islet cis-expression QTLs (eQTLs) (van de Bunt et
220 al., 2015) and islet ChIP-seq transcription factor binding sites (TFBS) (Pasquali
221 et al., 2014). We observed marked enrichment for eQTLs (LMR $\log_2$FE=1.1, UMR
222 $\log_2$FE=2.7, FDR<0.05) and TFBS (LMR $\log_2$FE=4.1-4.6; UMR $\log_2$FE=2.4-3.9,
223 FDR<0.05, Figure 2B). These observations confirm that islet LMRs and UMRs
224 correspond to important tissue-specific regulatory regions, overlapping cis-
225 regulatory annotations known to be enriched for T2D GWAS signals (Pasquali et
226 al., 2014, Gaulton et al., 2015b).
227
228 We also considered the relationship between LMR and UMR regions defined in
229 our non-diabetic islet WGBS, and a complementary set of methylation-based
230 annotations previously derived from WGBS of islets from 6 T2D and 8 control
231 individuals (Volkov et al., 2017). In that study, comparisons between diabetic
232 and non-diabetic islets had been used to define a set of 25,820 "disease
233 differentially methylated regions" (dDMRs, min absolute methylation difference
234 5% and P<0.02). We found only limited overlap between these dDMRs and the
235 UMRs and LMRs from our data: of the 25,820 dDMRs, only 2.2% overlapped

236 LMRs and 2.4% UMRs. This overlap was slightly greater than expected by chance
237 (FDR<0.05, LMR $\log_2$FE=1.0 and promoter-like UMRs $\log_2$FE=1.1, Figure 2B) but
238 more modest than seen for the other regulatory annotations. Similarly, we also
239 observed that dDMRs showed more modest ($\log_2$FE=0.4-1.0), but still significant
240 (FDR<0.05) levels of enrichment with respect to all other islet regulatory
241 annotations (Figure 2B). The modest enrichment of dDMRs indicates that only a
242 fraction of these regions correspond to islet genomic regulatory sites. Given that
243 T2D risk variants preferentially map in islet regulatory sites, the corollary is that
244 most dDMRs are unlikely to directly contribute to the mediation of genetic T2D
245 risk.
246
247 **2.3 Refining islet enhancer function using methylation and open chromatin**
248 **data**
249
250 To further characterise the regulatory potential of hypomethylated regions,
251 including LMRs and UMRs, we combined the islet WGBS methylation data with
252 chromatin accessibility data generated from ATAC-seq assays of 17 human islet
253 samples (from non-diabetic donors of European descent; mean read count after
254 filtering =130M, Supplementary Figure 2C). We identified a total of 141k open
255 chromatin regions based on read depth, peak width and signal to noise ratio (see
256 Methods). These regions of islet open chromatin showed substantial overlap
257 (78%) with equivalent regions described in a recent study of two human islets
258 (Varshney et al., 2017) ($\log_2$FE=2.8 compared to random sites, FDR<0.05, not
259 shown). In addition, our islet ATAC-Seq sites demonstrated substantial overlap
260 with LMRs: 53% of LMRs overlapped 16% of all ATAC-seq peaks (LMR
261 $\log_2$FE=3.8 compared to randomised sites, Figure 2B). Almost all UMRs (98%)
262 were contained within regions overlapping (13% of) ATAC-seq peaks (UMR
263 $\log_2$FE=3.4 compared to randomised sites, Figure 2B).
264
265 To fully leverage information across multiple overlapping islet epigenome
266 assays, we generated augmented chromatin state maps, using chromHMM (Ernst
267 and Kellis, 2012). These maps combined the WGBS methylation and ATAC-Seq
268 open chromatin data with previously generated ChIP-seq marks (Figure 3A,

269 Supplementary Figure 3A). For these analyses, we initially used a single
270 definition for hypomethylated regions (methylation<60%) that captured both
271 UMRs and LMRs (see Methods).
272
273 This augmented and larger set of 15 islet chromatin states retained the broad
274 classification of regulatory elements that included promoters (positive for
275 H3K4me3), transcribed and genic regions (H3K36me3), strong enhancers
276 (H3K4me1; H3K27ac), weak enhancers (H3K4me1), insulators (CTCF) and
277 repressed elements (H3K27me) (Figure 3A). The addition of islet methylation
278 and open chromatin data expanded existing chromatin state definitions to
279 provide new subclasses, particularly amongst enhancer elements. Here, we
280 observed two subclasses of strong enhancers and three of weak enhancers
281 (Figure 3A). We denote the strong enhancer subtypes as "open" (n=32k genome-
282 wide), characterised by open chromatin and hypomethylation, and "closed"
283 (n=110k) with closed chromatin and hypermethylation (Figure 3A). The three
284 weak enhancer states we denote as "open" (n=38k: open chromatin,
285 hypomethylation), "lowly-methylated" (n=78k; closed chromatin,
286 hypomethylation) and "closed" (n=206k: closed chromatin, hypermethylation).
287 No equivalent class of "lowly-methylated" strong enhancers was observed in the
288 15-state model. When comparing these chromatin states to those identified
289 using only ChIP-seq marks ((Parker et al., 2013), Supplementary Figure 1C,), the
290 two strong enhancer subclasses we identified subdivided the "strong enhancer
291 1" state as described by Parker (defined by H3K27ac and H3K4me1). Additional
292 comparison to "stretch" enhancer clusters (Parker et al., 2013), showed that
293 there was considerable overlap between the "open" strong and weak enhancer
294 states we identify here and previously-described "stretch" enhancer states
295 (16.1k out of 23k stretch enhancer overlapped 32k out of 70.1k "open"
296 enhancers). Even so, most (55%) "open" enhancer states, and in particular
297 "open weak enhancers" (70%), were not captured within "stretch" enhancer
298 intervals, and we regard these as distinct islet enhancer subclasses.
299
300 To understand the relationship of these various state definitions to genetic
301 variants influencing T2D risk, we applied the hierarchical modelling approach

302 FGWAS to the same sets of large-scale GWAS data for T2D (from DIAGRAM (Scott
303 et al., 2017)) and FG (ENGAGE (Horikoshi et al., 2015)) described in section 2.1.
304 FGWAS allowed us to combine GWAS and genomic data to determine the
305 genome-wide enrichment within islet regulatory features for variants associated
306 with T2D risk. These enrichment priors were then used to generate credible
307 variant sets that are informed by both GWAS and genomic data, as described in
308 section 2.4.
309
310 In single-feature analyses, we found significant enrichment (Confidence Interval
311 (CI)>0) limited to four enhancer states (open weak enhancers, both types of
312 strong enhancer and H3K36me3 marked genic enhancers) (Figure 3B, Table 1).
313 To take into account protein-coding variant effects, we also included CoDing
314 exon Sequence (CDS) as an additional annotation (Carlson and Maintainer, 2015)
315 and observed significant enrichment for CDS in the single state results (Figure
316 3B, Table 1). FGWAS multi-feature analyses for T2D, incorporating all
317 annotations positive in single-element analyses, retained both subclasses of
318 strong enhancer, the subclass of open weak enhancers, and CDS in the joint
319 model (Figure 3C and Methods).
320
321 We observed markedly different levels of enrichment for T2D association
322 between and within open and closed enhancer states (Figure 3B-3C, Table 1).
323 Using these augmented chromatin state maps, we demonstrated clear
324 enrichment for T2D association for the subset of "open" weak enhancers (12% of
325 all weak enhancer sites) with no evidence of enrichment in the remaining
326 subclasses ("closed" and "lowly-methylated") (Figure 3B and Table 1). This
327 concentration of enrichment amongst a relatively small subset of the weak
328 enhancers was consistent with the lack of enrichment across all weak enhancers
329 defined solely on the basis of H3K4me1 signal ((Parker et al., 2013), single state
330 $\log_2$FE=0.9, CI=-2.5 to 2.0, Table 1, Supplementary Figure 1C). We also saw
331 differences in enrichment signal between open and closed strong enhancers,
332 with the most marked enrichment amongst open strong enhancers (22% of the
333 total, Figure 3B-C, Table 1). This effect was particularly obvious in the joint-

334 analysis (open strong enhancer joint FE=4.1, CI=3.3 to 4.8 vs. closed strong
335 enhancer joint $\log_2$FE=2.2, CI=0.03 to 3.2, Figure 3C).
336 Hypomethylation and open chromatin are highly correlated, but the observed
337 difference in T2D enrichment between the weak enhancer states (particularly
338 between "lowly-methylated" and "open" which differ markedly with respect to
339 chromatin status) points to a primary role for open chromatin. To test this
340 further, we regenerated chromatin state maps using different subsets of the data
341 (ChIP-only, with optional addition of methylation and/or open chromatin
342 information, see Methods and Supplementary Figure 3A-3B). These analyses
343 confirmed that the T2D GWAS enrichment signal was predominantly driven by
344 the distribution of islet open chromatin (Supplementary Figure 3C).
345
346 We further evaluated the role of subclasses of DNA methylation regulatory
347 region with respect to T2D GWAS enrichment. We divided hypomethylated
348 (<60% methylated) sequence into enhancer-like LMRs (6.5% of all
349 hypomethylated sequence), promoter-like UMRs (7.5% of hypomethylated
350 sequence), as well as PMDs (61% of hypomethylated sequence). The remaining
351 25% of hypomethylated sequence did not fit any category. LMRs were
352 significantly (CI>0) enriched ($\log_2$FE=3.2, CI=2.3 to 3.9) for T2D association
353 signals consistent with their co-localisation with distal regulatory elements, and
354 displayed modestly increased enrichment compared to enhancer states derived
355 from ChIP-seq alone (Figure 3D, Supplementary Table 1). In contrast, no
356 significant enrichment was found for human islet (promoter-like) UMRs
357 ($\log_2$FE=1.4, CI=-0.6 to 2.5) or PMDs ($\log_2$FE=-0.8, CI=-1.7 to -0.1). We also found
358 no evidence that recently-described regions of T2D-associated differential
359 methylation (dDMRs: derived from comparison of WGBS data from islets of
360 diabetic and non-diabetic individuals) were enriched for genome-wide T2D
361 association signals ($\log_2$FE=-24.6, CI=-44.6 to 3.7) (Figure 3D, Supplementary
362 Table 1).
363
364 Finally, since the hypomethylation signal for T2D enrichment was concentrated
365 in LMRs (Figure 3D, Supplementary Table 1), we reran a FGWAS joint-analysis
366 combining open chromatin peaks, LMRs and ChIP-only states using a nested

367 model (Figure 3E, Supplementary Figure 3D-E, see Methods). This confirmed
368 that the improvement in enrichment was mainly driven by open chromatin but
369 showed that LMRs also contributed significantly and independently to the
370 enrichment (Figure 3E, Supplementary Table 2).
371
372 FGWAS analysis for FG corroborated the observations from T2D analysis. Both
373 single feature and joint-model analyses of human islet epigenome data found
374 significant enrichment in strong and weak enhancer states with the strongest
375 enrichment in enhancers with open chromatin and hypomethylation
376 (Supplementary Figure 4A-B and Table 1). In addition, evaluation of the relative
377 contributions of ATAC-seq open chromatin and DNA methylation to FG GWAS
378 enrichment across both single-feature (Supplementary Figure 4C-D) and joint-
379 model analysis (Supplementary Figure 4E-F and Supplementary Table 3)
380 indicated that open chromatin was primarily responsible for the enhanced
381 enrichment.
382
383 Overall, these analyses demonstrate that the addition of open chromatin and
384 DNA methylation data to ChIP-seq marks enhances the resolution of regulatory
385 annotation for human islets. In particular, it defines subsets of weak and strong
386 enhancers that differ markedly with respect to the impact of genetic variation on
387 T2D risk. Although DNA accessibility and hypomethylation status are strongly
388 correlated and provide broadly similar enrichments, the effects of the former
389 predominate. In line with the dominance of open chromatin status for T2D GWAS
390 enrichment, we observed that T2D risk in relation to methylation status is
391 primarily invested in hypomethylated LMRs (i.e. enhancers) rather than UMRs,
392 dDMRs or PMDs.
393
394 **2.4 Augmented chromatin maps and open chromatin allelic imbalance**
395 **refine likely causal variants at *ADCY5* , *CDC123*, and *KLHDC5***
396
397 We next deployed the insights from the global FGWAS enrichment analyses to
398 define the molecular mechanisms at individual T2D susceptibility loci, refining

399 T2D causal variant localisation using the combination of genetic data (from fine-
400 mapping) and the genome-wide patterns of epigenomic enrichment.
401
402 Specifically, we applied FGWAS to the T2D DIAGRAM GWAS data (Scott et al.,
403 2017) under the joint model (Figure 3C) derived from the augmented chromatin
404 state maps. We divided the genome into 2327 segments (average size 5004 SNPs
405 or 1.2Mb) and identified 51 segments significantly associated with T2D genome-
406 wide (segmental FGWAS PPA >=0.9 or single variant GWAS P< $5x10^{-8}$, see
407 Methods for details). These corresponded to 48 known T2D associated regions
408 representing that subset of the ~120 known T2D GWAS loci which passed those
409 significance/filtering criteria in this European-only dataset. We then calculated
410 reweighted PPAs for each variant within each segment and generated
411 reweighted 99% credible sets.
412
413 Consistent with the increased T2D GWAS enrichment of states including open
414 chromatin and DNA methylation information, we found that analyses using
415 enrichments from the augmented chromatin state model (combining ChIP-seq,
416 ATAC-seq and WGBS data) were associated with smaller 99% credible sets
417 (median of 18 SNPs) than those derived from FGWAS enrichment derived from
418 ChIP-seq data alone (median 23). In parallel, the PPA for the best variant per
419 locus increased (median 0.33 vs 0.26). Individual T2D GWAS locus results are
420 shown in Figure 4A-B. This demonstrates that the inclusion of open chromatin
421 and DNA methylation in epigenome maps helps improve prioritisation of causal
422 variants at many T2D GWAS loci.
423
424 A subset of T2D GWAS signals are known to influence T2D risk through a
425 primary effect on insulin secretion, whilst others act primarily through insulin
426 resistance. We used previous categorisations of T2D GWAS loci based on the
427 patterns of association with quantitative measurements of metabolic function
428 and anthropometry (Wood et al., 2017, Dimas et al., 2014), to define a set of
429 15/48 loci most clearly associated with deficient insulin secretion (and therefore
430 most likely to involve islet dysfunction). At 7 of these 15 loci, we found that islet
431 "open strong enhancer" states, and to a lesser extent "open weak enhancer" and

432 “closed strong enhancer”, captured more than 60% of the PPA (median 91%,
433 Figure 4C). Variants in these islet enhancer subclasses also captured at least 95%
434 of the PPA at 4 T2D GWAS loci that could not be classified according to
435 physiological association data but which have been previously implicated in
436 human islet genome or functional regulation based on islet eQTL (van de Bunt et
437 al., 2015) or mQTL (Olsson et al., 2014) data (Figure 4C, genes highlighted in
438 bold). In contrast, at 3/6 of the insulin resistance and all but 6 unclassified loci,
439 the PPA was mostly (>50%) attributable to other non-islet enhancer states
440 (across all insulin resistance and unclassified loci, DNA not overlapping islet
441 enhancers and defined as “Other” capture a median PPA of 67%). Thus, islet
442 regulatory annotations are particularly useful for fine-mapping T2D GWAS loci
443 that affect insulin secretion and beta-cell function.
444
445 To obtain additional evidence to support the localisation of causal variants, we
446 tested for allelic imbalance in ATAC-seq open chromatin data. We selected 43
447 variants within 27 T2D-associated GWAS segments for testing of allelic
448 imbalance on the basis of (a) a reweighted variant PPA >=10% and (b) overlap
449 with an enriched regulatory state within the FGWAS T2D joint-model (Figure
450 4D). Of these, 16 variants (at 13 loci) had sufficient numbers of heterozygous
451 samples (>2) and ATAC_Seq read depth (depth>9 and at least 5 reads for each
452 allele). After correcting for mapping bias using WASP, we observed the strongest
453 evidence for allelic imbalance (FDR< 0.05) at 3 out of the 16 variants
454 (rs11257655 near *CDC123* and *CAMK1D*, rs10842991 near *KLHDC5* and
455 rs11708067 at *ADCY5*) (Table 2). All three overlapped refined islet open strong
456 or open weak enhancer regions characterised by open chromatin and
457 hypomethylation.
458
459 Variant rs11257655 accounts for 95% of the reweighted PPA (compared to a
460 PPA of 20% from genetic data alone) at the *CDC123/CAMK1D* locus, overlaps an
461 “open strong enhancer” region (Figure 5A) and the risk allele correlates with
462 increased chromatin accessibility. The same variant is in high LD (r2=0.82) with
463 the lead variant for a cis-eQTL for *CAMK1D* in islets (van de Bunt et al., 2015). In
464 experimental assays (Fogarty et al., 2014), the T2D-risk allele has been shown to

465 be associated with increased *CAMK1D* gene expression and enhanced binding of
466 the FOXA1 and FOXA2 transcription factors. These data all point to rs11257655
467 as the causal variant at this locus.
468
469 At *KLHDC5*, no clear causal variant emerged based on genetic fine-mapping data
470 alone as the credible set contained 23 variants in high mutual LD (r^2>0.8, top
471 variant PPA<5%, Figure 5B). Of these, variants rs10771372 (genetic fine-
472 mapping PPA= 5%), rs10842992 (genetic fine mapping PPA=5%) and
473 rs10842991 (genetic fine-mapping PPA=3%) overlapped "open strong
474 enhancer" regions (Figure 5B), such that their reweighted PPAs rose to 23%
475 (rs10771372), 22% (rs10842992) and 13% (rs10842991), respectively. We
476 observed allelic imbalance only at rs10842991 with the T2D-risk C allele
477 showing greater chromatin accessibility (binomial P=2.9×10^{-4}, Table 2). This
478 variant further overlapped a predicted TFBS motif for PAX6 as determined by
479 the software tool FIMO (Grant et al., 2011): the T2D-risk allele was predicted to
480 enhance PAX6 transcription factor binding consistent with the allelic effects on
481 increasing chromatin accessibility (Supplementary Figure 5A). This strong
482 enhancer region is almost exclusively found in islets, with strong enhancer
483 H3K27ac states overlapping rs10842991 in only two non-islet (heart and
484 smooth muscle) Epigenome Roadmap tissues (out of 99 tissues with 18-state
485 chromatin state information, Figure 5B)(Roadmap Epigenomics et al., 2015).
486 Islet eQTL data (Varshney et al., 2017) also links rs10842991 and close proxy
487 SNPs (including rs7960190) to islet transcription with the risk allele increasing
488 *KLHDC5* expression. These data prioritise rs10842991 as the likely causal
489 variant at the *KLHDC5* T2D GWAS locus, and indicate a likely molecular
490 mechanism involving modified PAX6 transcription factor binding and an impact
491 on *KLHDC5* expression and islet function.
492
493 The third example of allelic imbalance mapped to the *ADCY5* locus. Fine-mapping
494 based solely on genetic data could not prioritise a distinct causal variant due to
495 multiple variants in high LD (range for top 5 variants=12-26%, Figure 5C).
496 However, reweighting of variants based on epigenomic annotation clearly
497 prioritised variant rs11708067: this SNP overlapped an "open weak enhancer"

498 and captured most of the reweighted PPA (PPA=90%). Allelic imbalance analysis
499 also showed that the T2D-risk A allele was associated with decreased chromatin
500 accessibility (binomial $P=2.9x10^{-6}$, Table 2). The same lead variant maps to an
501 islet cis-eQTL and methylation QTL (Figure 5C, Supplementary Figure 5B) at
502 which the T2D-risk allele is associated with reduced *ADCY5* expression and
503 increased *ADCY5* gene body DNA methylation.
504
505 To further understand the role of the rs11708067 variant, we performed ATAC-
506 seq and Next Generation Capture-C, in the glucose-responsive human beta-cell
507 line EndoC-βH1 (n=3). We targeted the *ADCY5* promoter to define distal regions
508 interacting with the promoter, and confirmed physical contact with the
509 hypomethylated open chromatin enhancer region harbouring rs11708067
510 (Figure 5C, Supplementary Figure 5C). These findings support rs11708067 as
511 the likely causal variant affecting islet accessible chromatin (in line with another
512 recent study (Roman et al., 2017)), and link the open and hypomethylated
513 enhancer element in which it sits to regulation of *ADCY5* expression in islets.
514
515 **3. Discussion**
516
517 A key challenge in the quest to describe the molecular mechanisms through
518 which GWAS signals influence traits of interest, involves the identification of the
519 causal variants responsible and, given that most lie in non-coding sequence, the
520 characterisation of the regulatory elements which they perturb. This underpins
521 efforts to define the effector genes through which these variants operate and to
522 reconstruct the biological networks that are central to disease pathogenesis.
523
524 Genetic and physiological studies have highlighted the singular importance of
525 pancreatic islet dysfunction in type 2 diabetes, but epigenomic characterisation
526 of this tissue has been limited in large-scale community projects such as ENCODE
527 and GTex. The present study seeks to address this deficit by describing, in
528 unprecedented detail, genome-wide patterns of methylation and chromatin
529 accessibility in human islet material. We have combined these data with existing
530 islet epigenomic marks to generate a refined regulatory map which, based on the

531 evidence of improved enrichment for T2D association signals, offers more
532 granular annotation of functional impact.
533
534 Our data show that, for DNA methylation, the signal of T2D predisposition is
535 primarily associated with enhancer-like LMRs rather than other categories of
536 methylation elements including UMRs, dDMRs or PMDs. We highlight the strong
537 correlation between islet methylation status and chromatin accessibility but
538 demonstrate that both genomic annotations independently contribute to
539 defining the regulatory impact associated with genetic T2D risk. Having said that,
540 comparative analyses reveal that assays of DNA accessibility capture more of
541 that genetic risk than whole genome methylation data. Finally, we demonstrate
542 how these enhanced epigenomic annotations, when analysed in concert with
543 genetic fine-mapping data and information from allelic imbalance in chromatin
544 accessibility allow us to home in on likely causal variants at T2D association
545 signals such as those near *ADCY5, CDC123* and *KLHDC5*.
546
547 While previous studies had explored the candidacy of selected variants at the
548 *CDC123* (Fogarty et al., 2014) and *ADCY5* (Olsson et al., 2014, van de Bunt et al.,
549 2015) loci with respect to islet regulation and T2D predisposition, our
550 integrative analysis of T2D GWAS and epigenetic data has enabled a detailed and
551 comprehensive analysis that considers the regulatory impact of all variants at
552 these loci across multiple islet samples. Our analysis implicates the rs11257655
553 and rs11708067 variants as the most likely causal variants at the *CDC123* and
554 *ADCY5* loci respectively and highlights their relationship to islet enhancer
555 activity. The findings at *ADCY5* are supported by a recent paper that found allelic
556 imbalance in H3K27 acetylation involving the rs11708067 variant in a single
557 human islet sample, and which observed that deletion of the relevant enhancer
558 element led to reduction in both *ADCY5* gene expression and insulin secretion
559 (Roman et al., 2017).
560
561 At the *KLHDC5* locus, local LD frustrated efforts to define the causal variant using
562 genetic data alone, but the integration of genetic and epigenetic data pinpointed
563 rs10842991 as the likely culprit based on its impact on chromatin accessibility in

564 an open enhancer region. Evidence that this variant co-localises with an islet cis-
565 eQTL signal points to *KLHDC5* as the likely downstream target (Varshney et al.,
566 2017). Overall, our integrative approach provides useful insights into the
567 functional mechanisms through which T2D GWAS signals operate. Our findings
568 mirror those from other studies, which have, in various ways, and for other
569 complex traits, combined diverse epigenomic annotations to explore the basis of
570 genetic risk (Wang et al., 2016).
571
572 The whole genome methylation data generated in the present study also allowed
573 us to evaluate the likely contribution of previously identified T2D-associated
574 dDMRs (Volkov et al., 2017) with respect to T2D predisposition. These dDMRs,
575 defined on the basis of observed differences in methylation between islets
576 recovered from diabetic and non-diabetic donors, cover a substantial part of the
577 genome, but we were able to show that only a small minority of these overlap
578 functional islet regulatory regions. As a consequence, dDMR regions as a whole
579 had no significant enrichment for T2D association signals. This suggests that
580 most of the dDMR signal involves stochastic effects and/or the secondary
581 consequences on methylation of the diabetic state. However, we cannot exclude
582 that some of the dDMR signals are causal contributors to the diabetic phenotype
583 either because they reflect environmental rather than genetic predisposition, or
584 because they accelerate further perturbation of islet dysfunction as diabetes
585 develops.
586
587 Although the present study characterises islet epigenome status and variability
588 in chromatin accessibility in substantially larger numbers of islet samples than
589 those previously reported (Gaulton et al., 2015b, Parker et al., 2013, Pasquali et
590 al., 2014, Varshney et al., 2017), the number of islet preparations for which these
591 data were available was still limited. As a result, our power to detect allelic
592 imbalance in chromatin accessibility was restricted to sites with common
593 variants and relatively large effects. We anticipate that expansion of these
594 sample numbers will extend our capacity to detect such allelic imbalance, and
595 offer more granular insights into the relationships between genetic variation and
596 methylation status. A further limitation is that the genomic data we analysed was

597 generated only from islet samples from non-diabetic donors. Whilst causal
598 inference is possible through the integration of basal epigenomic annotations
599 with genetic data, addition of epigenomic data from islets recovered from
600 diabetic donors has the potential to add a further dimension to such analyses,
601 and to unravel what are likely to be complex causal relationships between
602 genetic variants, epigenomic phenotypes and disease states (Gutierrez-Arcelus et
603 al., 2013). Our on-going research efforts are now concentrated on expanding
604 these analyses in larger numbers of human islet samples from healthy and
605 diabetic islets so as to provide an increasingly complete description of the causal
606 interactions between DNA methylation, chromatin state expression and T2D
607 susceptibility.
608
609

610 **4. Methods**

611

612 **4.1 Human Pancreatic islet samples**

613

614 **4.1.2 WGBS and 450k array human pancreatic islet sample collection**

615 Human islets were retrieved from deceased Caucasian non-diabetic donors from
616 the Oxford DRWF Human Islet Isolation Facility (n=34) and at the Alberta
617 Diabetes Institute in Edmonton in Canada (n=10). For the analysis only samples
618 with a purity >70% were used as determined by dithizone labeling. The Human
619 Research Ethics Board at the University of Alberta (Pro00001754), the
620 University of Oxford's Oxford Tropical Research Ethics Committee (OxTREC
621 Reference: 2–15), or the Oxfordshire Regional Ethics Committee B (REC
622 reference: 09/H0605/2) approved the studies. All organ donors provided
623 informed consent for use of pancreatic tissue in research.

624

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

631

632 **4.1.2 ATAC-seq human pancreatic islet sample collection**

633 Human pancreatic islets preparations (n=18) were retrieved from 17 deceased
634 non-diabetic donors of European descent from the Oxford DRWF Human Islet
635 Isolation Facility and stored for 1-3 days in CMRL or UW media. The latter were
636 reactivated in CMRL for 1h before processing them further. Approximately
637 50,000 islet cells per sample were hand-picked and immediately processed for
638 ATAC-seq as described previously (Buenrostro et al., 2013), however, an
639 additional round of purification was performed using Agencourt AMPure XP
640 magnetic beads (Beckman Coulter).

641

642

643 **4.2. WGBS data generation**

644

645 **4.2.1 Bisulphite conversion**

646 400ng of DNA per human islet samples (n=10) were sent as part of a
647 collaborative effort to the Blizard Institute, Queen Mary University, London, UK
648 and bisulphite- converted using the Ovation Ultralow Methyl-Seq DR Multiplex
649 System 1-8 (Nugen) and purified using Agencourt AMPure beads (Beckman
650 Coulter) as described previously (Lowe et al., 2013).

651

652 **4.2.2 Library generation and processing of reads**

653 The libraries were sequenced by members of the High-Throughput Genomics
654 group at the Wellcome Trust Centre for Human Genetics, University of Oxford,
655 Oxford, UK. Samples were sequenced as multiplex libraries across 3 HiSeq2000
656 lanes with 100bp paired-end read length (including a PhIX spike-in of 5%) to
657 obtain high-coverage read data. The obtained reads were trimmed using a
658 customized python3 script (10bp at the start and 15bp at the end) and aligned to
659 hg19 using the software Bismark (settings: L,0,-0.6, version 0.12.5)(Krueger and
660 Andrews, 2011). Specifically, paired-end alignment of trimmed reads was
661 performed and unmapped reads from read 1 were realigned using Bismark and
662 merged with the paired-end alignment using samtools (Li et al., 2009) (version
663 0.1.19) in order to increase mapping efficiency. Coverage for the merged paired-
664 end and realigned HiSeq read alignments was estimated for the human
665 mappable genome (NCBI hg19 2.8 billion base pairs excluding gaps and
666 unmappable and blacklisted regions according to UCSC and
667 Encode(EncodeProjectConsortium, 2012)) using bedtools (version v2.21.0)
668 (Quinlan, 2014).

669

670 **4.2.3 WGBS DNA methylation quantification and prediction of**
671 **hypomethylated regulatory regions**

672 CpG methylation levels were determined for each sample by calculating the ratio
673 of unmodified C (methylated) and bisulphite converted T (unmethylated) alleles
674 using BiFAST (first described here (Lowe et al., 2013)). High-pass pooled WGBS
675 data was generated by adding methylated and unmethylated read counts across

676 individual low-pass samples to then estimate the average beta methylation
677 levels.
678
679 Regulatory regions were identified using the R package methylseek (Burger et
680 al., 2013). After removing PMDs, which represent highly heterogenous
681 methylation states determined by DNA sequence features (Gaidatzis et al., 2014),
682 LMRs (<30 CpGs) and UMRs (>30 CpGs) were predicted in hypomethylated
683 regions (<50%) at an FDR of 0.05. The methylation level and FDR parameter was
684 inferred from the data as suggested by the methylseek workflow (Burger et al.,
685 2013).
686
687 **4.3 450k DNA methylation array data generation**
688 In total, 41 samples were processed for the Illumina Infinium
689 HumanMethylation450 BeadChip (Illumina, San Diego, CA). Of these 18 samples
690 were bisulphite-converted and processed as part of a collaboration at the UCL
691 Cancer Institute, University College London, London, UK while the remaining 23
692 samples were processed in OCDEM, University of Oxford, Oxford, UK. The DNA
693 was bisulphite converted using the EZ DNA Methylation™ Kit (© Zymogen
694 Research Corp, Irvine, CA) and hybridised to the Illumina 450k array and
695 scanned with iScan (Illumina) according to the manufacturer's protocol.
696
697 The resulting data was analysed using the Package minfi (Aryee et al., 2014) and
698 custom R scripts ((RCoreTeam), R version 3.0.2). Specifically, CpG sites with a
699 detection P-value >0.01 were removed from the analysis and samples with >5%
700 of CpG sites failing this threshold (n=9) were also removed from the analysis.
701
702 Following separate quantile normalisation of signal intensities derived from
703 methylated and unmethylated Type I probes and Type II probes, methylation
704 levels (ß) were estimated, based on the intensities of the methylated (M) and
705 unmethylated (U) signal in the following way: $\beta = M/(M+U*100)$. To correct for
706 batch effects the ComBat function implemented in the sva (Johnson et al., 2007,
707 Leek et al.) package was used (Supplementary Figure 6).
708

709 **4.4 ATAC-seq data generation**

710

711 **4.4.1 Sequencing of ATAC-seq reads**

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

716

717 **4.4.2 Processing of ATAC-seq reads**

718 Raw FASTQ reads were processed with an in-house pipeline (first described in
719 (Hay et al., 2016) and on the website
720 http://userweb.molbiol.ox.ac.uk/public/telenius/PipeSite.html). Specifically,
721 library and sequencing quality was checked with FASTQC
722 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and reads were
723 mapped to the human genome (hg19) via bowtie (Langmead et al., 2009)
724 (version 1.1.0) with default settings but -m 2, and maxins 2000 which allows
725 mapping of reads with a maximum number of 2 alignments and a maximum
726 insert size of 2000bp. For reads that could not be aligned the first time, adapters
727 were removed with Trim Galore at the 3 prime end (settings -length 10, -
728 qualFilter 20,
729 http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to enhance
730 the chance of mapping. The resulting trimmed reads were then mapped again
731 with bowtie. Any remaining unmapped and trimmed reads were processed with
732 FLASH (Magoc and Salzberg, 2011) (version 1.2.8 settings -m 9 -x 0.125) which
733 combines overlapping read pairs and reconstructs read pairs without overlap.
734 These are then realigned a third time using bowtie. PCR duplicates are then
735 removed from the mapped bam files using samtools rmdup function (Li et al.,
736 2009). Additionally, all reads overlapping any of the "unmappable" UCSC Duke
737 blacklisted hg19 regions (EncodeProjectConsortium, 2012) are also removed
738 from the final bam file.
739 Open chromatin peaks were called through the aforementioned in-house
740 pipeline by applying sample-specific read depth and width parameters, which
741 were chosen based on the signal to noise ratio of a given sample.

742 **4.5 ChIP-seq data and identification of chromatin states**

743

744 **4.5.1 Processing of available ChIP-seq data**

745 Human islet ChIP-seq histone mark and TFBS data were obtained from various
746 sources: H3K4me1, *CTCF* and H3K27ac (Pasquali et al., 2014), H3K36me3 and
747 H3K4me3 (Moran et al., 2012) and H3K27me3 (Roadmap Epigenomics et al.,
748 2015). Available raw fastq files were aligned to hg19 using bowtie1 (version
749 1.1.1) with modified default settings (-m 1 which removes reads with more than
750 1 valid alignment and -n 1 which defines the maximum number of mismatches in
751 seed) and PCR duplicates were removed from the aligned bam files with Picard
752 tools (v1.119, http://broadinstitute.github.io/picard/). The resulting reads were
753 converted into bed format using the bedtools bamToBed function (Quinlan,
754 2014) (version v2.21.0) and extended by 200bp towards the 3' end to account
755 for fragment size.

756

757 **4.5.2 Identification of chromatin states using chromHMM**

758

759 Binarised 200bp density maps from the bed files of the 6 ChIP-seq marks were
760 created using a Poisson distribution implemented in the BinaryBed function of
761 the ChromHMM software as described in (Ernst and Kellis, 2012, Ernst et al.,
762 2011). From these epigenomic density maps, 11 ChIP-only chromatin states
763 were derived using a multivariate Hidden Markov Model implemented in the
764 Learnmodel function (standard settings, h19 genome) of the software
765 ChromHMM (Ernst and Kellis, 2012).

766

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

773

774 These binarised 200bp ChIP-seq, ATAC-seq and DNA methylation maps were
775 combined and used to generate 3 sets of chromatin states derived from ChIP and
776 DNA methylation data (ChIP+Meth), ChIP and ATAC-seq data (ChIP+ATAC) or
777 ChIP, ATAC-seq and DNA methylation data (ChIP+ATAC+Meth) using the
778 Learnmodel ChromHMM function (Figure 3A and Supplementary Figure 3A-B) .
779 As suggested by (Ernst et al., 2011), after evaluating models with up to 20
780 chromatin states, a 15 state model was chosen based on the resolution provided
781 by identified states
782
783 **4.6 *ADCY5* Capture C analysis and ATAC-seq in EndoC-ßH1**
784 Next-generation Capture-C was performed in order to map physical chromatin
785 interactions with the *ADCY5* promoter in EndoC-ßH1 cell lines (n = 3) (see
786 protocol in Methods in (Davies et al., 2016)).
787 In brief, chromatin conformation capture (3C) libraries were generated by
788 formaldehyde fixation prior to DpnII restriction enzyme digestion and
789 subsequent DNA ligation. Following cross-link reversal, DNA extraction and
790 sonication, sequencing adapters were added to sonicated fragments (~200bp).
791 Library fragments were subjected to a double capture through hybridisation
792 with a biotinylated oligonucleotide probes encompassing the *ADCY5* promoter
793 and enriched using streptavidin bead pulldown. PCR amplified fragments were
794 then sequenced through paired-end sequencing (Illumina Next-Seq). An *in silico*
795 restriction enzyme digestion was performed on the set of reconstructed
796 fragments (from paired-end sequenced reads) using the DpnII2E.pl script
797 (https://github. com/telenius/captureC/releases). Uncaptured reads and PCR
798 duplicates were removed prior to mapping to the human genome (hg19) with
799 Bowtie (Langmead et al., 2009)(v 1.1.0). Removal of PCR duplicates and
800 classification of fragments as 'capture' (i.e. including the *ADCY5* promoter) or
801 'reporter' (outside the capture fragment on exclusion region) was performed
802 with the CCanalyser2.pl wrapper
803 (https://github.com/telenius/captureC/releases). Unique mapped interactions
804 were normalized to the total number of *cis* interactions (i.e. same chromosome)
805 per 100,000 interactions.
806

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

810 **4.7 Overlaying generated epigenomic datasets generated here with other**
811 **genomic regulatory regions**

812

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

820

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

825

826 **4.8 Genetic datasets used in this study**

827 Credible sets from the DIAGRAM (Scott et al., 2017)(involving 26.7k cases and
828 132.5k controls of predominantly European origin, imputed to the 1000G March
829 2012 reference panel) and ENGAGE (Horikoshi et al., 2015)(including 46.7k
830 individuals, imputed to the 1000G March 2012 reference panel) consortium
831 were used to compare the ability of the 450k, 850k and WGBS methylation array
832 to interrogate T2D and FG GWAS regions.

833

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

837

838 Furthermore, data from (Wood et al., 2017) and (Dimas et al., 2014) were used
839 to categories T2D GWAS loci into physiological groups of insulin secretion,
840 insulin resistance or unclassified loci.
841
842 **4.9 Statistical and computational analysis**
843
844 **4.9.1 Enrichment analysis of identified regulatory annotations in other**
845 **genomic annotations**
846
847 Enrichment of hypomethylated regulatory regions (LMRs and UMRs, result
848 section 2.2.) and ATAC-seq open chromatin peaks (result section 2.3) in the
849 aforementioned genomic annotations (method section 4.6) was determined
850 through 1000 random permutations. P-values and fold enrichment was
851 determined by comparing the true overlap results to the permuted overlap
852 results. The resulting P-values were multiple testing corrected using an FDR (an
853 FDR< 0.05 was considered significant).
854
855 **4.9.2 FGWAS enrichment analysis**
856 FGWAS (Pickrell, 2014) (version 0.3.6) applied a hierarchical model that
857 determined shared properties of loci affecting a trait. The FGWAS model used
858 SNP-based GWAS summary level data and divided the genome into windows
859 (setting "k"=5000 which represents the number of SNPs per window), which are
860 larger than the expected LD patterns in the population. The model assumed that
861 each window either contained a single SNP that affected the trait or that there
862 was no SNP in the window that influenced the trait. The model estimated the
863 prior probability of a window to contain an association and the conditional prior
864 probability that a SNP within the window was the causal variant. These prior
865 probabilities were variable, dependent on regional annotations and estimated
866 based on enrichment patterns of annotations across the genome using a Bayes
867 approach.
868
869
870

871 **4.9.2.1 FGWAS Single state analysis**

872 FGWAS was used with standard settings to determine enrichment of individual
873 islet chromatin states, LMRs, UMRs, PMDS and ATAC-seq open chromatin peaks
874 in DIAGRAM (setting "cc" was applied for use with T2D-case-control GWAS data)
875 and ENGAGE GWAS SNP summary level data.
876
877 For each individual annotation, the model provided maximum likelihood
878 enrichment parameters and annotations were considered as significantly
879 enriched if the parameter estimate and 95% CI was above zero.
880

881 **4.9.2.2 FGWAS Joint model analysis**

882 To determine the maximum likelihood model the following approach suggested
883 by (Pickrell, 2014) was used for each set of chromatin states (ChIP-only,
884 ChIP+ATAC, ChIP+Meth and ChIP+ATAC+Meth), separately. Firstly, a model was
885 fitted for each annotation individually to identify all annotations that were
886 significantly enriched with the trait. Secondly, the annotation with the highest
887 increase (and enrichment) in the maximum log-likelihood was added to the
888 model and the analysis was repeated with the additional annotation. Thirdly,
889 annotations were added as long as they increase the maximum log-likelihood of
890 the newly tested model. Fourthly, a 10-fold cross-validation approach was used
891 after determining a penalty parameter based on the maximum likelihood of a
892 penalised log-likelihood function to avoid overfitting. Fifthly, each annotation
893 was dropped from the model and the maximum cross-validation likelihood was
894 evaluated. If a reduced model has a higher cross-validation maximum likelihood,
895 additional annotations are dropped until the model cannot be further improved.
896 This model was described as the best fitted model and used for the remaining
897 analysis. The maximum likelihood enrichment parameters and 95% CI for each
898 annotation of the best model were reported (independent of significance).
899

900 **4.9.2.3 Comparing FGWAS enrichment parameter across chromatin states**

901 Initially, similar enhancer chromatin states derived from the 4 different
902 ChromHMM analyses (ChIP-only, ChIP+ATAC, ChIP+Meth, ChIP+ATAC+Meth)
903 were compared. Similarity was determined based on shared histone chromatin

904 marks according to the chromHMM emission parameters. Further comparisons
905 between the ChIP-only and ChIP+ATAC+Meth model were performed based on
906 the reweighted FGWAS maximum variant PPA and the number of reweighted
907 99% credible set variants per T2D locus (for details regarding FGWAS PPA see
908 next section).
909
910 However, considering that the chromatin states were derived from distinct sets
911 of annotations across different analyses of ChromHMM, a direct comparison was
912 not fully possible. Hence, a nested model approach was used to further dissect
913 the contribution of open chromatin and DNA methylation to the enrichment.
914 Specifically, an FGWAS analysis was performed that combined the ChIP-only
915 chromHMM states with raw LMRs (representing DNA methylation) and ATAC-
916 seq peaks (representing open chromatin). After determining the best maximum-
917 likelihood cross-validation model (combining ChIP-only, ATAC-seq and LMR
918 states) a nested model and log-likelihood ratio test were used to determine the
919 contribution of each annotation to the model (Supplementary Figure 3D).
920
921 **4.9.3 Reweighting of variant PPA and testing of allelic imbalance**
922 The enrichment priors derived from the FGWAS maximum likelihood model
923 were used as a basis for evaluating both the significance and functional impact of
924 associated variants in GWAS regions; allowing variants that map into
925 annotations that show global enrichment to be afforded extra weight.
926
927
928 Specifically, variants at significant GWAS regions with a high FGWAS PPA (PPA
929 >=10%) and overlapping open enhancer states were prioritised for further
930 follow-up. Genome-wide significance of loci was determined based on P-values
931 ($P < 5x10^{-8}$) or a regional FGWAS PPA >= 90% (representing the sum of the PPAs
932 of all SNPs in a given region). The latter threshold is based on a recommendation
933 from (Pickrell, 2014) who observed that a regional PPA of 90% or above can be
934 used to identify sub-threshold GWAS loci.
935

936 Of the prioritised variants, only variants with at least 2 heterozygous samples
937 and ATAC_Seq read depth of at least 9 reads (minimum 5 reads for each allele)
938 were tested for allelic imbalance.
939
940 To avoid read-mapping and reference allele bias the software WASP (van de
941 Geijn et al., 2015)(Version 0.2261) was used to remove reads associated with
942 mapping bias. In short, reads of the unfiltered bam file that overlapped the
943 variant of interest were identified. For each read overlapping an SNP, the
944 genotype of that SNP was changed to the alternative allele and the read was
945 remapped using bwa (Li and Durbin, 2009) (version 0.5.8c). Any read that failed
946 to realign in the same position in the genome was discarded. Ultimately, PCR
947 duplicates were filtered using the WASP "rmdup_pe.py" script, which removed
948 duplicated reads randomly (independent of the mapping score) to avoid any
949 bias.
950
951 Allelic imbalance was determined using a binomial test as implemented in R.
952
953 **4.9.4 Identification of TFBS at SNPs that display allelic imbalance**
954
955 The tool "Fimo"(Grant et al., 2011) implemented in the "meme" software
956 package was applied to identify TF motifs that significantly (FDR <0.05) matched
957 the sequence overlapping a SNP variant showing allelic imbalance (20bp up and
958 downstream).
959 .
960 **4.9.5 Overlap of regulatory regions**
961 Overlap between genomic regulatory regions was performed using bedtools
962 intersectBed function (Quinlan, 2014) (version 2.21.0). Summary statistics
963 across 200bp windows were determined using bedtools mapBed function.
964 Random permutations of regulatory regions were performed by applying the
965 bedtools shuffleBed function.
966
967
968

969 **4.9.6 Statistical analysis**

970 All statistical analysis (unless otherwise stated) was performed using R (version
971 3.0.2) including Spearman's correlation analysis to compare the 450k and WGBS
972 array, the KS-test to compare 450k and WGBS DNA methylation distributions,
973 the binomial test to evaluate allelic imbalance and principal component analysis
974 to identify batch effects in the 450k data. Significance is defined as $P<0.05$ unless
975 otherwise stated.

976

977 **4.9.7 Visualisation and figure generation**

978 All figures unless otherwise stated were generated using R (version 3.0.2)
979 and/or ggplot2(Wickham, 2009). Figure 1E was generated using locuszoom
980 (Pruim et al., 2010). Chromatin state CHiP-seq enrichment maps (Figure 3A,
981 Supplementary Figure 3A-B) were generated using chromHMM (Ernst and Kellis,
982 2012). The genome-browser views (Figure 5) were generated using the UCSC
983 genome browser tool (Kent et al., 2002).

984

985 **5. Acknowledgements**

986 We thank the High-Throughput Genomics Group at the Wellcome Trust Centre
987 for Human Genetics (funded by Wellcome Trust grant reference 090532) for the
988 generation of the Sequencing data.
989 MT was supported by the Wellcome Trust**.** MvdB was supported by a Novo
990 Nordisk postdoctoral fellowship run in partnership with the University of
991 Oxford. JMT is supported by Wellcome Trust Strategic Award. VN is funded by
992 the European Union Horizon 2020 Programme (T2DSYSTEMS). SB was
993 supported by a Royal Society Wolfson Research Merit Award (WM100023). ALG
994 is a Wellcome Trust Senior Fellow in Basic Biomedical Science (095101/Z/10/Z
995 and 200837/Z/16/Z) and MIM is a Wellcome Trust Senior Investigator. The
996 research was supported by the National Institute for Health Research (NIHR)
997 Oxford Biomedical Research Centre (BRC). This work was also supported by EU
998 (HEALTH-F4-2007-201413), Wellcome Trust (090367, 090532, 106130 and
999 098381) and NIH (U01-DK105535, U01-DK085545, R01-DK098032 and R01-
1000 MH090941) grants. The views expressed are those of the author(s) and not
1001 necessarily those of the NHS, the NIHR or the Department of Health.

6. Competing interests

The authors do not have any competing interest.

1005 **7. References**

1006

1007 ARYEE, M. J., JAFFE, A. E., CORRADA-BRAVO, H., LADD-ACOSTA, C., FEINBERG, A.
1008 P., HANSEN, K. D. & IRIZARRY, R. A. 2014. Minfi: a flexible and
1009 comprehensive Bioconductor package for the analysis of Infinium DNA
1010 methylation microarrays. *Bioinformatics,* 30**,** 1363-9.
1011 BELL, J. T., PAI, A. A., PICKRELL, J. K., GAFFNEY, D. J., PIQUE-REGI, R., DEGNER, J.
1012 F., GILAD, Y. & PRITCHARD, J. K. 2011. DNA methylation patterns
1013 associate with genetic and gene expression variation in HapMap cell lines.
1014 *Genome Biol,* 12**,** R10.
1015 BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J.
1016 2013. Transposition of native chromatin for fast and sensitive epigenomic
1017 profiling of open chromatin, DNA-binding proteins and nucleosome
1018 position. *Nat Methods,* 10**,** 1213-8.
1019 BURGER, L., GAIDATZIS, D., SCHUBELER, D. & STADLER, M. B. 2013.
1020 Identification of active regulatory regions from DNA methylation data.
1021 *Nucleic Acids Res,* 41**,** e155.
1022 CARLSON, M. & MAINTAINER, B. 2015. TxDb.Hsapiens.UCSC.hg19.knownGene:
1023 Annotation package for TxDb object(s). *R package version 3.2.2*.
1024 DAVIES, J. O., TELENIUS, J. M., MCGOWAN, S. J., ROBERTS, N. A., TAYLOR, S.,
1025 HIGGS, D. R. & HUGHES, J. R. 2016. Multiplexed analysis of chromosome
1026 conformation at vastly improved sensitivity. *Nat Methods,* 13**,** 74-80.
1027 DAYEH, T., VOLKOV, P., SALO, S., HALL, E., NILSSON, E., OLSSON, A. H.,
1028 KIRKPATRICK, C. L., WOLLHEIM, C. B., ELIASSON, L., RONN, T., BACOS, K.
1029 & LING, C. 2014. Genome-wide DNA methylation analysis of human
1030 pancreatic islets from type 2 diabetic and non-diabetic donors identifies
1031 candidate genes that influence insulin secretion. *PLoS Genet,* 10**,**
1032 e1004160.
1033 DIMAS, A. S., LAGOU, V., BARKER, A., KNOWLES, J. W., MAGI, R., HIVERT, M. F.,
1034 BENAZZO, A., RYBIN, D., JACKSON, A. U., STRINGHAM, H. M., SONG, C.,
1035 FISCHER-ROSINSKY, A., BOESGAARD, T. W., GRARUP, N., ABBASI, F. A.,
1036 ASSIMES, T. L., HAO, K., YANG, X., LECOEUR, C., BARROSO, I.,
1037 BONNYCASTLE, L. L., BOTTCHER, Y., BUMPSTEAD, S., CHINES, P. S.,
1038 ERDOS, M. R., GRAESSLER, J., KOVACS, P., MORKEN, M. A., NARISU, N.,
1039 PAYNE, F., STANCAKOVA, A., SWIFT, A. J., TONJES, A., BORNSTEIN, S. R.,
1040 CAUCHI, S., FROGUEL, P., MEYRE, D., SCHWARZ, P. E., HARING, H. U.,
1041 SMITH, U., BOEHNKE, M., BERGMAN, R. N., COLLINS, F. S., MOHLKE, K. L.,
1042 TUOMILEHTO, J., QUERTEMOUS, T., LIND, L., HANSEN, T., PEDERSEN, O.,
1043 WALKER, M., PFEIFFER, A. F., SPRANGER, J., STUMVOLL, M., MEIGS, J. B.,
1044 WAREHAM, N. J., KUUSISTO, J., LAAKSO, M., LANGENBERG, C., DUPUIS, J.,
1045 WATANABE, R. M., FLOREZ, J. C., INGELSSON, E., MCCARTHY, M. I.,
1046 PROKOPENKO, I. & INVESTIGATORS, M. 2014. Impact of type 2 diabetes
1047 susceptibility variants on quantitative glycemic traits reveals mechanistic
1048 heterogeneity. *Diabetes,* 63**,** 2158-71.
1049 DO, C., LANG, C. F., LIN, J., DARBARY, H., KRUPSKA, I., GABA, A., PETUKHOVA, L.,
1050 VONSATTEL, J. P., GALLAGHER, M. P., GOLAND, R. S., CLYNES, R. A.,
1051 DWORK, A., KRAL, J. G., MONK, C., CHRISTIANO, A. M. & TYCKO, B. 2016.

1052 Mechanisms and Disease Associations of Haplotype-Dependent Allele-
1053 Specific DNA Methylation. *Am J Hum Genet,* 98**,** 934-55.
1054 ECKHARDT, F., LEWIN, J., CORTESE, R., RAKYAN, V. K., ATTWOOD, J., BURGER, M.,
1055 BURTON, J., COX, T. V., DAVIES, R., DOWN, T. A., HAEFLIGER, C., HORTON,
1056 R., HOWE, K., JACKSON, D. K., KUNDE, J., KOENIG, C., LIDDLE, J., NIBLETT,
1057 D., OTTO, T., PETTETT, R., SEEMANN, S., THOMPSON, C., WEST, T.,
1058 ROGERS, J., OLEK, A., BERLIN, K. & BECK, S. 2006. DNA methylation
1059 profiling of human chromosomes 6, 20 and 22. *Nat Genet,* 38**,** 1378-85.
1060 ENCODEPROJECTCONSORTIUM 2012. An integrated encyclopedia of DNA
1061 elements in the human genome. *Nature,* 489**,** 57-74.
1062 ERNST, J. & KELLIS, M. 2012. ChromHMM: automating chromatin-state discovery
1063 and characterization. *Nat Methods,* 9**,** 215-6.
1064 ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D.,
1065 EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R., COYNE, M., KU, M.,
1066 DURHAM, T., KELLIS, M. & BERNSTEIN, B. E. 2011. Mapping and analysis
1067 of chromatin state dynamics in nine human cell types. *Nature,* 473**,** 43-9.
1068 FOGARTY, M. P., CANNON, M. E., VADLAMUDI, S., GAULTON, K. J. & MOHLKE, K. L.
1069 2014. Identification of a regulatory variant that binds FOXA1 and FOXA2
1070 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLoS Genet,* 10**,**
1071 e1004633.
1072 GAIDATZIS, D., BURGER, L., MURR, R., LERCH, A., DESSUS-BABUS, S.,
1073 SCHUBELER, D. & STADLER, M. B. 2014. DNA sequence explains
1074 seemingly disordered methylation levels in partially methylated domains
1075 of Mammalian genomes. *PLoS Genet,* 10**,** e1004143.
1076 GAULTON, K. J., FERREIRA, T., LEE, Y., RAIMONDO, A., MAGI, R., RESCHEN, M. E.,
1077 MAHAJAN, A., LOCKE, A., RAYNER, N. W., ROBERTSON, N., SCOTT, R. A.,
1078 PROKOPENKO, I., SCOTT, L. J., GREEN, T., SPARSO, T., THUILLIER, D.,
1079 YENGO, L., GRALLERT, H., WAHL, S., FRANBERG, M., STRAWBRIDGE, R. J.,
1080 KESTLER, H., CHHEDA, H., EISELE, L., GUSTAFSSON, S.,
1081 STEINTHORSDOTTIR, V., THORLEIFSSON, G., QI, L., KARSSEN, L. C., VAN
1082 LEEUWEN, E. M., WILLEMS, S. M., LI, M., CHEN, H., FUCHSBERGER, C.,
1083 KWAN, P., MA, C., LINDERMAN, M., LU, Y., THOMSEN, S. K., RUNDLE, J. K.,
1084 BEER, N. L., VAN DE BUNT, M., CHALISEY, A., KANG, H. M., VOIGHT, B. F.,
1085 ABECASIS, G. R., ALMGREN, P., BALDASSARRE, D., BALKAU, B.,
1086 BENEDIKTSSON, R., BLUHER, M., BOEING, H., BONNYCASTLE, L. L.,
1087 BOTTINGER, E. P., BURTT, N. P., CAREY, J., CHARPENTIER, G., CHINES, P.
1088 S., CORNELIS, M. C., COUPER, D. J., CRENSHAW, A. T., VAN DAM, R. M.,
1089 DONEY, A. S., DORKHAN, M., EDKINS, S., ERIKSSON, J. G., ESKO, T., EURY,
1090 E., FADISTA, J., FLANNICK, J., FONTANILLAS, P., FOX, C., FRANKS, P. W.,
1091 GERTOW, K., GIEGER, C., GIGANTE, B., GOTTESMAN, O., GRANT, G. B.,
1092 GRARUP, N., GROVES, C. J., HASSINEN, M., HAVE, C. T., HERDER, C.,
1093 HOLMEN, O. L., HREIDARSSON, A. B., HUMPHRIES, S. E., HUNTER, D. J.,
1094 JACKSON, A. U., JONSSON, A., JORGENSEN, M. E., JORGENSEN, T., KAO, W.
1095 H., KERRISON, N. D., KINNUNEN, L., KLOPP, N., KONG, A., KOVACS, P.,
1096 KRAFT, P., KRAVIC, J., LANGFORD, C., et al. 2015a. Genetic fine mapping
1097 and genomic annotation defines causal mechanisms at type 2 diabetes
1098 susceptibility loci. *Nat Genet,* 47**,** 1415-25.
1099 GAULTON, K. J., FERREIRA, T., LEE, Y., RAIMONDO, A., MAGI, R., RESCHEN, M. E.,
1100 MAHAJAN, A., LOCKE, A., WILLIAM RAYNER, N., ROBERTSON, N., SCOTT,

1101 R. A., PROKOPENKO, I., SCOTT, L. J., GREEN, T., SPARSO, T., THUILLIER, D.,
1102 YENGO, L., GRALLERT, H., WAHL, S., FRANBERG, M., STRAWBRIDGE, R. J.,
1103 KESTLER, H., CHHEDA, H., EISELE, L., GUSTAFSSON, S.,
1104 STEINTHORSDOTTIR, V., THORLEIFSSON, G., QI, L., KARSSEN, L. C., VAN
1105 LEEUWEN, E. M., WILLEMS, S. M., LI, M., CHEN, H., FUCHSBERGER, C.,
1106 KWAN, P., MA, C., LINDERMAN, M., LU, Y., THOMSEN, S. K., RUNDLE, J. K.,
1107 BEER, N. L., VAN DE BUNT, M., CHALISEY, A., KANG, H. M., VOIGHT, B. F.,
1108 ABECASIS, G. R., ALMGREN, P., BALDASSARRE, D., BALKAU, B.,
1109 BENEDIKTSSON, R., BLUHER, M., BOEING, H., BONNYCASTLE, L. L.,
1110 BOTTINGER, E. P., BURTT, N. P., CAREY, J., CHARPENTIER, G., CHINES, P.
1111 S., CORNELIS, M. C., COUPER, D. J., CRENSHAW, A. T., VAN DAM, R. M.,
1112 DONEY, A. S., DORKHAN, M., EDKINS, S., ERIKSSON, J. G., ESKO, T., EURY,
1113 E., FADISTA, J., FLANNICK, J., FONTANILLAS, P., FOX, C., FRANKS, P. W.,
1114 GERTOW, K., GIEGER, C., GIGANTE, B., GOTTESMAN, O., GRANT, G. B.,
1115 GRARUP, N., GROVES, C. J., HASSINEN, M., HAVE, C. T., HERDER, C.,
1116 HOLMEN, O. L., HREIDARSSON, A. B., HUMPHRIES, S. E., HUNTER, D. J.,
1117 JACKSON, A. U., JONSSON, A., JORGENSEN, M. E., JORGENSEN, T., KAO, W.
1118 H., KERRISON, N. D., KINNUNEN, L., KLOPP, N., KONG, A., KOVACS, P.,
1119 KRAFT, P., KRAVIC, J., LANGFORD, C., et al. 2015b. Genetic fine mapping
1120 and genomic annotation defines causal mechanisms at type 2 diabetes
1121 susceptibility loci. *Nat Genet,* 47**,** 1415-25.
1122 GAULTON, K. J., NAMMO, T., PASQUALI, L., SIMON, J. M., GIRESI, P. G., FOGARTY,
1123 M. P., PANHUIS, T. M., MIECZKOWSKI, P., SECCHI, A., BOSCO, D., BERNEY,
1124 T., MONTANYA, E., MOHLKE, K. L., LIEB, J. D. & FERRER, J. 2010. A map of
1125 open chromatin in human pancreatic islets. *Nat Genet,* 42**,** 255-9.
1126 GRANT, C. E., BAILEY, T. L. & NOBLE, W. S. 2011. FIMO: scanning for occurrences
1127 of a given motif. *Bioinformatics,* 27**,** 1017-8.
1128 GUO, S., DIEP, D., PLONGTHONGKUM, N., FUNG, H. L., ZHANG, K. & ZHANG, K.
1129 2017. Identification of methylation haplotype blocks aids in
1130 deconvolution of heterogeneous tissue samples and tumor tissue-of-
1131 origin mapping from plasma DNA. *Nat Genet,* 49**,** 635-642.
1132 GUTIERREZ-ARCELUS, M., LAPPALAINEN, T., MONTGOMERY, S. B., BUIL, A.,
1133 ONGEN, H., YUROVSKY, A., BRYOIS, J., GIGER, T., ROMANO, L., PLANCHON,
1134 A., FALCONNET, E., BIELSER, D., GAGNEBIN, M., PADIOLEAU, I., BOREL, C.,
1135 LETOURNEAU, A., MAKRYTHANASIS, P., GUIPPONI, M., GEHRIG, C.,
1136 ANTONARAKIS, S. E. & DERMITZAKIS, E. T. 2013. Passive and active DNA
1137 methylation and the interplay with genetic variation in gene regulation.
1138 *Elife,* 2**,** e00523.
1139 HANNON, E., SPIERS, H., VIANA, J., PIDSLEY, R., BURRAGE, J., MURPHY, T. M.,
1140 TROAKES, C., TURECKI, G., O'DONOVAN, M. C., SCHALKWYK, L. C., BRAY,
1141 N. J. & MILL, J. 2016. Methylation QTLs in the developing brain and their
1142 enrichment in schizophrenia risk loci. *Nat Neurosci,* 19**,** 48-54.
1143 HAY, D., HUGHES, J. R., BABBS, C., DAVIES, J. O. J., GRAHAM, B. J., HANSSEN, L.,
1144 KASSOUF, M. T., MARIEKE OUDELAAR, A. M., SHARPE, J. A., SUCIU, M. C.,
1145 TELENIUS, J., WILLIAMS, R., RODE, C., LI, P. S., PENNACCHIO, L. A.,
1146 SLOANE-STANLEY, J. A., AYYUB, H., BUTLER, S., SAUKA-SPENGLER, T.,
1147 GIBBONS, R. J., SMITH, A. J. H., WOOD, W. G. & HIGGS, D. R. 2016. Genetic
1148 dissection of the alpha-globin super-enhancer in vivo. *Nat Genet,* 48**,** 895-
1149 903.

1150 HORIKOSHI, M., MGI, R., VAN DE BUNT, M., SURAKKA, I., SARIN, A. P., MAHAJAN,
1151 A., MARULLO, L., THORLEIFSSON, G., HGG, S., HOTTENGA, J. J.,
1152 LADENVALL, C., RIED, J. S., WINKLER, T. W., WILLEMS, S. M.,
1153 PERVJAKOVA, N., ESKO, T., BEEKMAN, M., NELSON, C. P., WILLENBORG,
1154 C., WILTSHIRE, S., FERREIRA, T., FERNANDEZ, J., GAULTON, K. J.,
1155 STEINTHORSDOTTIR, V., HAMSTEN, A., MAGNUSSON, P. K., WILLEMSEN,
1156 G., MILANESCHI, Y., ROBERTSON, N. R., GROVES, C. J., BENNETT, A. J.,
1157 LEHTIMKI, T., VIIKARI, J. S., RUNG, J., LYSSENKO, V., PEROLA, M., HEID, I.
1158 M., HERDER, C., GRALLERT, H., MULLER-NURASYID, M., RODEN, M.,
1159 HYPPONEN, E., ISAACS, A., VAN LEEUWEN, E. M., KARSSEN, L. C.,
1160 MIHAILOV, E., HOUWING-DUISTERMAAT, J. J., DE CRAEN, A. J., DEELEN, J.,
1161 HAVULINNA, A. S., BLADES, M., HENGSTENBERG, C., ERDMANN, J.,
1162 SCHUNKERT, H., KAPRIO, J., TOBIN, M. D., SAMANI, N. J., LIND, L.,
1163 SALOMAA, V., LINDGREN, C. M., SLAGBOOM, P. E., METSPALU, A., VAN
1164 DUIJN, C. M., ERIKSSON, J. G., PETERS, A., GIEGER, C., JULA, A., GROOP, L.,
1165 RAITAKARI, O. T., POWER, C., PENNINX, B. W., DE GEUS, E., SMIT, J. H.,
1166 BOOMSMA, D. I., PEDERSEN, N. L., INGELSSON, E., THORSTEINSDOTTIR,
1167 U., STEFANSSON, K., RIPATTI, S., PROKOPENKO, I., MCCARTHY, M. I.,
1168 MORRIS, A. P. & CONSORTIUM, E. 2015. Discovery and Fine-Mapping of
1169 Glycaemic and Obesity-Related Trait Loci Using High-Density Imputation.
1170 *PLoS Genet,* 11**,** e1005230.
1171 JOHNSON, W. E., LI, C. & RABINOVIC, A. 2007. Adjusting batch effects in
1172 microarray expression data using empirical Bayes methods. *Biostatistics,*
1173 8**,** 118-27.
1174 KATO, N., LOH, M., TAKEUCHI, F., VERWEIJ, N., WANG, X., ZHANG, W., KELLY, T.
1175 N., SALEHEEN, D., LEHNE, B., MATEO LEACH, I., DRONG, A. W., ABBOTT, J.,
1176 WAHL, S., TAN, S. T., SCOTT, W. R., CAMPANELLA, G., CHADEAU-HYAM, M.,
1177 AFZAL, U., AHLUWALIA, T. S., BONDER, M. J., CHEN, P., DEHGHAN, A.,
1178 EDWARDS, T. L., ESKO, T., GO, M. J., HARRIS, S. E., HARTIALA, J., KASELA,
1179 S., KASTURIRATNE, A., KHOR, C. C., KLEBER, M. E., LI, H., MOK, Z. Y.,
1180 NAKATOCHI, M., SAPARI, N. S., SAXENA, R., STEWART, A. F., STOLK, L.,
1181 TABARA, Y., TEH, A. L., WU, Y., WU, J. Y., ZHANG, Y., AITS, I., DA SILVA
1182 COUTO ALVES, A., DAS, S., DORAJOO, R., HOPEWELL, J. C., KIM, Y. K.,
1183 KOIVULA, R. W., LUAN, J., LYYTIKAINEN, L. P., NGUYEN, Q. N., PEREIRA, M.
1184 A., POSTMUS, I., RAITAKARI, O. T., SCANNELL BRYAN, M., SCOTT, R. A.,
1185 SORICE, R., TRAGANTE, V., TRAGLIA, M., WHITE, J., YAMAMOTO, K.,
1186 ZHANG, Y., ADAIR, L. S., AHMED, A., AKIYAMA, K., ASIF, R., AUNG, T.,
1187 BARROSO, I., BJONNES, A., BRAUN, T. R., CAI, H., CHANG, L. C., CHEN, C. H.,
1188 CHENG, C. Y., CHONG, Y. S., COLLINS, R., COURTNEY, R., DAVIES, G.,
1189 DELGADO, G., DO, L. D., DOEVENDANS, P. A., GANSEVOORT, R. T., GAO, Y.
1190 T., GRAMMER, T. B., GRARUP, N., GREWAL, J., GU, D., WANDER, G. S.,
1191 HARTIKAINEN, A. L., HAZEN, S. L., HE, J., HENG, C. K., HIXSON, J. E.,
1192 HOFMAN, A., HSU, C., HUANG, W., HUSEMOEN, L. L., HWANG, J. Y., et al.
1193 2015. Trans-ancestry genome-wide association study identifies 12
1194 genetic loci influencing blood pressure and implicates a role for DNA
1195 methylation. *Nat Genet,* 47**,** 1282-93.
1196 KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER,
1197 A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC.
1198 *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.

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.,

1247 DELOUKAS, P. & DONNELLY, P. 2012. Bayesian refinement of association
1248 signals for 14 loci in 3 common diseases. *Nat Genet,* 44**,** 1294-301.
1249 MITCHELL, J. S., LI, N., WEINHOLD, N., FORSTI, A., ALI, M., VAN DUIN, M.,
1250 THORLEIFSSON, G., JOHNSON, D. C., CHEN, B., HALVARSSON, B. M.,
1251 GUDBJARTSSON, D. F., KUIPER, R., STEPHENS, O. W., BERTSCH, U.,
1252 BRODERICK, P., CAMPO, C., EINSELE, H., GREGORY, W. A., GULLBERG, U.,
1253 HENRION, M., HILLENGASS, J., HOFFMANN, P., JACKSON, G. H., JOHNSSON,
1254 E., JOUD, M., KRISTINSSON, S. Y., LENHOFF, S., LENIVE, O., MELLQVIST, U.
1255 H., MIGLIORINI, G., NAHI, H., NELANDER, S., NICKEL, J., NOTHEN, M. M.,
1256 RAFNAR, T., ROSS, F. M., DA SILVA FILHO, M. I., SWAMINATHAN, B.,
1257 THOMSEN, H., TURESSON, I., VANGSTED, A., VOGEL, U., WAAGE, A.,
1258 WALKER, B. A., WIHLBORG, A. K., BROYL, A., DAVIES, F. E.,
1259 THORSTEINSDOTTIR, U., LANGER, C., HANSSON, M., KAISER, M.,
1260 SONNEVELD, P., STEFANSSON, K., MORGAN, G. J., GOLDSCHMIDT, H.,
1261 HEMMINKI, K., NILSSON, B. & HOULSTON, R. S. 2016. Genome-wide
1262 association study identifies multiple susceptibility loci for multiple
1263 myeloma. *Nat Commun,* 7**,** 12050.
1264 MORAN, I., AKERMAN, I., VAN DE BUNT, M., XIE, R., BENAZRA, M., NAMMO, T.,
1265 ARNES, L., NAKIC, N., GARCIA-HURTADO, J., RODRIGUEZ-SEGUI, S.,
1266 PASQUALI, L., SAUTY-COLACE, C., BEUCHER, A., SCHARFMANN, R., VAN
1267 ARENSBERGEN, J., JOHNSON, P. R., BERRY, A., LEE, C., HARKINS, T., GMYR,
1268 V., PATTOU, F., KERR-CONTE, J., PIEMONTI, L., BERNEY, T., HANLEY, N.,
1269 GLOYN, A. L., SUSSEL, L., LANGMAN, L., BRAYMAN, K. L., SANDER, M.,
1270 MCCARTHY, M. I., RAVASSARD, P. & FERRER, J. 2012. Human beta cell
1271 transcriptome analysis uncovers lncRNAs that are tissue-specific,
1272 dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell*
1273 *Metab,* 16**,** 435-48.
1274 MORRIS, A. P., VOIGHT, B. F., TESLOVICH, T. M., FERREIRA, T., SEGRE, A. V.,
1275 STEINTHORSDOTTIR, V., STRAWBRIDGE, R. J., KHAN, H., GRALLERT, H.,
1276 MAHAJAN, A., PROKOPENKO, I., KANG, H. M., DINA, C., ESKO, T., FRASER,
1277 R. M., KANONI, S., KUMAR, A., LAGOU, V., LANGENBERG, C., LUAN, J.,
1278 LINDGREN, C. M., MULLER-NURASYID, M., PECHLIVANIS, S., RAYNER, N.
1279 W., SCOTT, L. J., WILTSHIRE, S., YENGO, L., KINNUNEN, L., ROSSIN, E. J.,
1280 RAYCHAUDHURI, S., JOHNSON, A. D., DIMAS, A. S., LOOS, R. J., VEDANTAM,
1281 S., CHEN, H., FLOREZ, J. C., FOX, C., LIU, C. T., RYBIN, D., COUPER, D. J., KAO,
1282 W. H., LI, M., CORNELIS, M. C., KRAFT, P., SUN, Q., VAN DAM, R. M.,
1283 STRINGHAM, H. M., CHINES, P. S., FISCHER, K., FONTANILLAS, P.,
1284 HOLMEN, O. L., HUNT, S. E., JACKSON, A. U., KONG, A., LAWRENCE, R.,
1285 MEYER, J., PERRY, J. R., PLATOU, C. G., POTTER, S., REHNBERG, E.,
1286 ROBERTSON, N., SIVAPALARATNAM, S., STANCAKOVA, A., STIRRUPS, K.,
1287 THORLEIFSSON, G., TIKKANEN, E., WOOD, A. R., ALMGREN, P., ATALAY,
1288 M., BENEDIKTSSON, R., BONNYCASTLE, L. L., BURTT, N., CAREY, J.,
1289 CHARPENTIER, G., CRENSHAW, A. T., DONEY, A. S., DORKHAN, M.,
1290 EDKINS, S., EMILSSON, V., EURY, E., FORSEN, T., GERTOW, K., GIGANTE, B.,
1291 GRANT, G. B., GROVES, C. J., GUIDUCCI, C., HERDER, C., HREIDARSSON, A.
1292 B., HUI, J., JAMES, A., JONSSON, A., RATHMANN, W., KLOPP, N., KRAVIC, J.,
1293 KRJUTSKOV, K., LANGFORD, C., LEANDER, K., LINDHOLM, E., LOBBENS, S.,
1294 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.

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.,

1393 VADLAMUDI, S., CANNON, M. E., DIDION, J. P., HENSLEY, J., KIRILUSHA, A.,
1394 PROGRAM, N. C. S., BONNYCASTLE, L. L., TAYLOR, D. L., WATANABE, R.,
1395 MOHLKE, K. L., BOEHNKE, M., COLLINS, F. S., PARKER, S. C. & STITZEL, M.
1396 L. 2017. Genetic regulatory signatures underlying islet gene expression
1397 and type 2 diabetes. *Proc Natl Acad Sci U S A,* 114**,** 2301-2306.
1398 VENTHAM, N. T., KENNEDY, N. A., ADAMS, A. T., KALLA, R., HEATH, S., O'LEARY,
1399 K. R., DRUMMOND, H., CONSORTIUM, I. B., CONSORTIUM, I. C., WILSON, D.
1400 C., GUT, I. G., NIMMO, E. R. & SATSANGI, J. 2016. Integrative epigenome-
1401 wide analysis demonstrates that DNA methylation may mediate genetic
1402 risk in inflammatory bowel disease. *Nat Commun,* 7**,** 13507.
1403 VOIGHT, B. F., SCOTT, L. J., STEINTHORSDOTTIR, V., MORRIS, A. P., DINA, C.,
1404 WELCH, R. P., ZEGGINI, E., HUTH, C., AULCHENKO, Y. S., THORLEIFSSON,
1405 G., MCCULLOCH, L. J., FERREIRA, T., GRALLERT, H., AMIN, N., WU, G.,
1406 WILLER, C. J., RAYCHAUDHURI, S., MCCARROLL, S. A., LANGENBERG, C.,
1407 HOFMANN, O. M., DUPUIS, J., QI, L., SEGRE, A. V., VAN HOEK, M.,
1408 NAVARRO, P., ARDLIE, K., BALKAU, B., BENEDIKTSSON, R., BENNETT, A. J.,
1409 BLAGIEVA, R., BOERWINKLE, E., BONNYCASTLE, L. L., BENGTSSON
1410 BOSTROM, K., BRAVENBOER, B., BUMPSTEAD, S., BURTT, N. P.,
1411 CHARPENTIER, G., CHINES, P. S., CORNELIS, M., COUPER, D. J.,
1412 CRAWFORD, G., DONEY, A. S., ELLIOTT, K. S., ELLIOTT, A. L., ERDOS, M. R.,
1413 FOX, C. S., FRANKLIN, C. S., GANSER, M., GIEGER, C., GRARUP, N., GREEN,
1414 T., GRIFFIN, S., GROVES, C. J., GUIDUCCI, C., HADJADJ, S., HASSANALI, N.,
1415 HERDER, C., ISOMAA, B., JACKSON, A. U., JOHNSON, P. R., JORGENSEN, T.,
1416 KAO, W. H., KLOPP, N., KONG, A., KRAFT, P., KUUSISTO, J., LAURITZEN, T.,
1417 LI, M., LIEVERSE, A., LINDGREN, C. M., LYSSENKO, V., MARRE, M.,
1418 MEITINGER, T., MIDTHJELL, K., MORKEN, M. A., NARISU, N., NILSSON, P.,
1419 OWEN, K. R., PAYNE, F., PERRY, J. R., PETERSEN, A. K., PLATOU, C.,
1420 PROENCA, C., PROKOPENKO, I., RATHMANN, W., RAYNER, N. W.,
1421 ROBERTSON, N. R., ROCHELEAU, G., RODEN, M., SAMPSON, M. J., SAXENA,
1422 R., SHIELDS, B. M., SHRADER, P., SIGURDSSON, G., SPARSO, T.,
1423 STRASSBURGER, K., STRINGHAM, H. M., SUN, Q., SWIFT, A. J., THORAND,
1424 B., et al. 2010. Twelve type 2 diabetes susceptibility loci identified
1425 through large-scale association analysis. *Nat Genet,* 42**,** 579-89.
1426 VOLKOV, P., BACOS, K., OFORI, J. K., ESGUERRA, J. L., ELIASSON, L., RONN, T. &
1427 LING, C. 2017. Whole-Genome Bisulfite Sequencing of Human Pancreatic
1428 Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes
1429 Pathogenesis. *Diabetes,* 66**,** 1074-1085.
1430 WANG, X., TUCKER, N. R., RIZKI, G., MILLS, R., KRIJGER, P. H., DE WIT, E.,
1431 SUBRAMANIAN, V., BARTELL, E., NGUYEN, X. X., YE, J., LEYTON-MANGE, J.,
1432 DOLMATOVA, E. V., VAN DER HARST, P., DE LAAT, W., ELLINOR, P. T.,
1433 NEWTON-CHEH, C., MILAN, D. J., KELLIS, M. & BOYER, L. A. 2016.
1434 Discovery and validation of sub-threshold genome-wide association study
1435 loci using epigenomic signatures. *Elife,* 5.
1436 WICKHAM, H. 2009. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag*
1437 *New York*.
1438 WOOD, A. R., JONSSON, A., JACKSON, A. U., WANG, N., VAN LEEWEN, N., PALMER,
1439 N. D., KOBES, S., DEELEN, J., BOQUETE-VILARINO, L., PAANANEN, J.,
1440 STANCAKOVA, A., BOOMSMA, D. I., DE GEUS, E. J., EEKHOFF, E. M.,
1441 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.

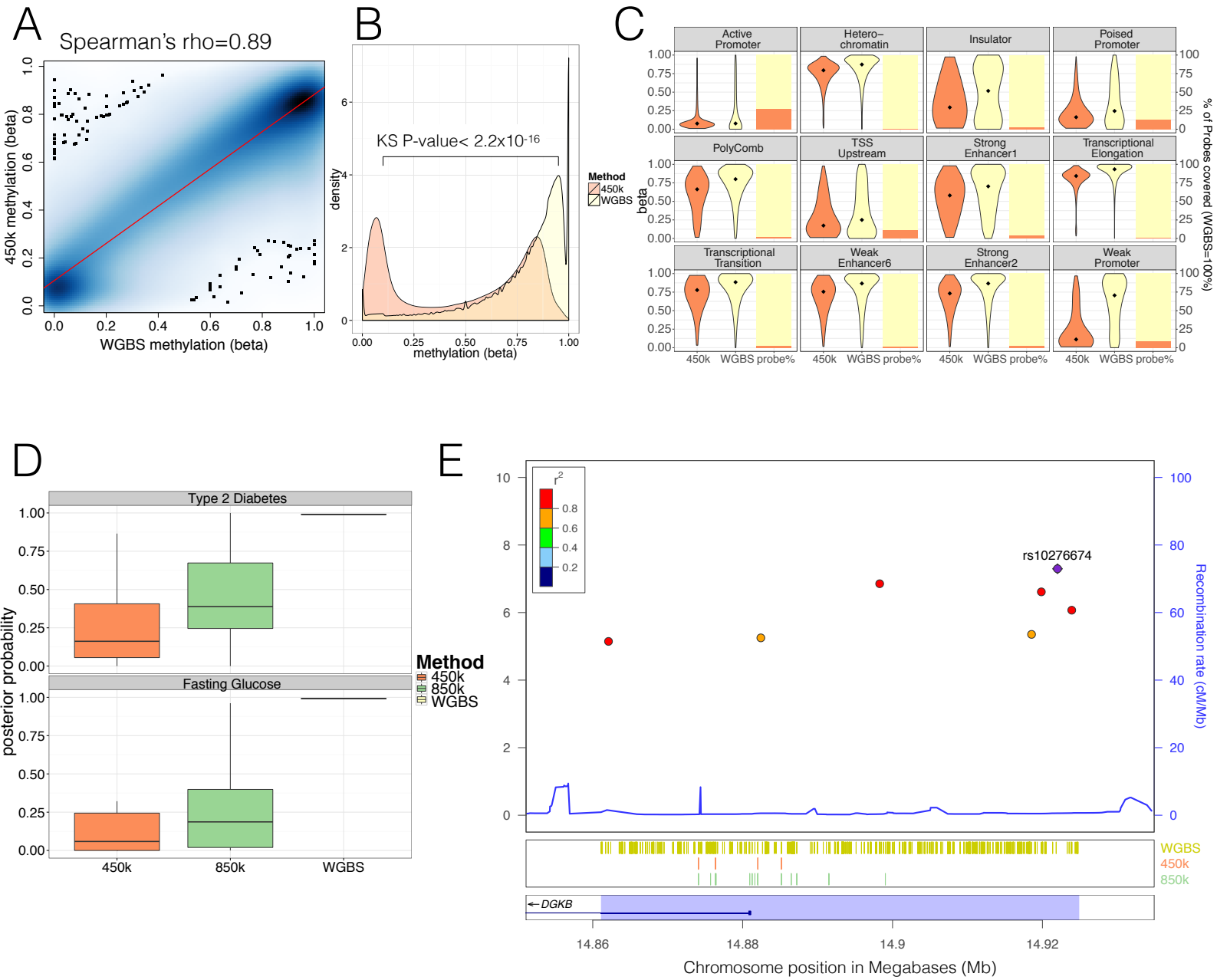

FIGURE 1. Comparison of human pancreatic islet WGBS and 450k methylation data across the genome. A) Smooth Scatter plot shows Spearman's rho correlation between the 450k array (x-axis) and WGBS (y-axis) at overlapping sites. Darker colour indicates higher density of sites. B) Comparison of the 450k array (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).

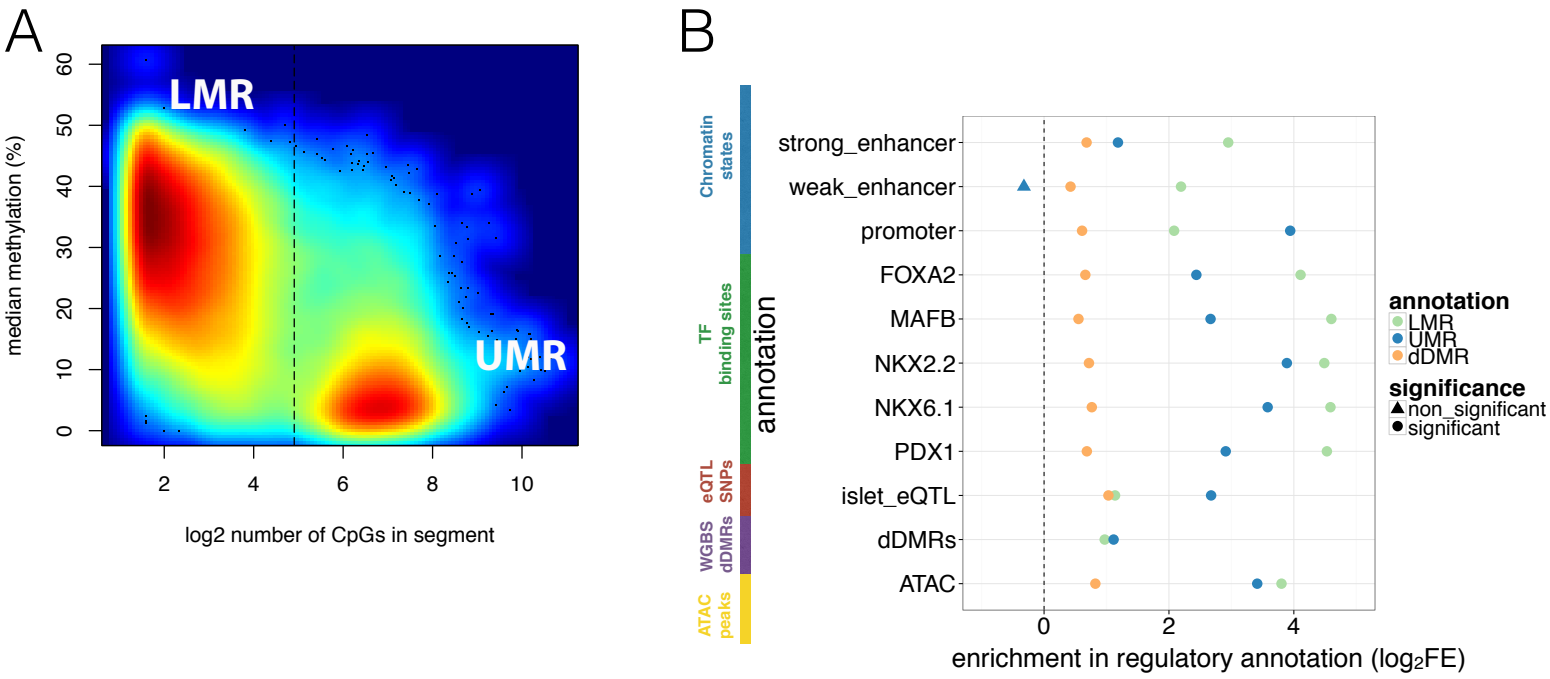

FIGURE 2. Overlap of WGBS hypomethylation and ATAC-seq open chromatin peaks with regulatory annotation . A) Methylation levels in percent (y-axis) and $\log_2$ CpG density (x-axis) of UMR and LMR regulatory regions with the dashed line indicating the CpG-number (30 CpGs) that distinguishes LMRs and UMRs. 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. The shape indicates significance (a dot corresponds to an FDR <0.05 while a triangle indicates FDR> 0.05).

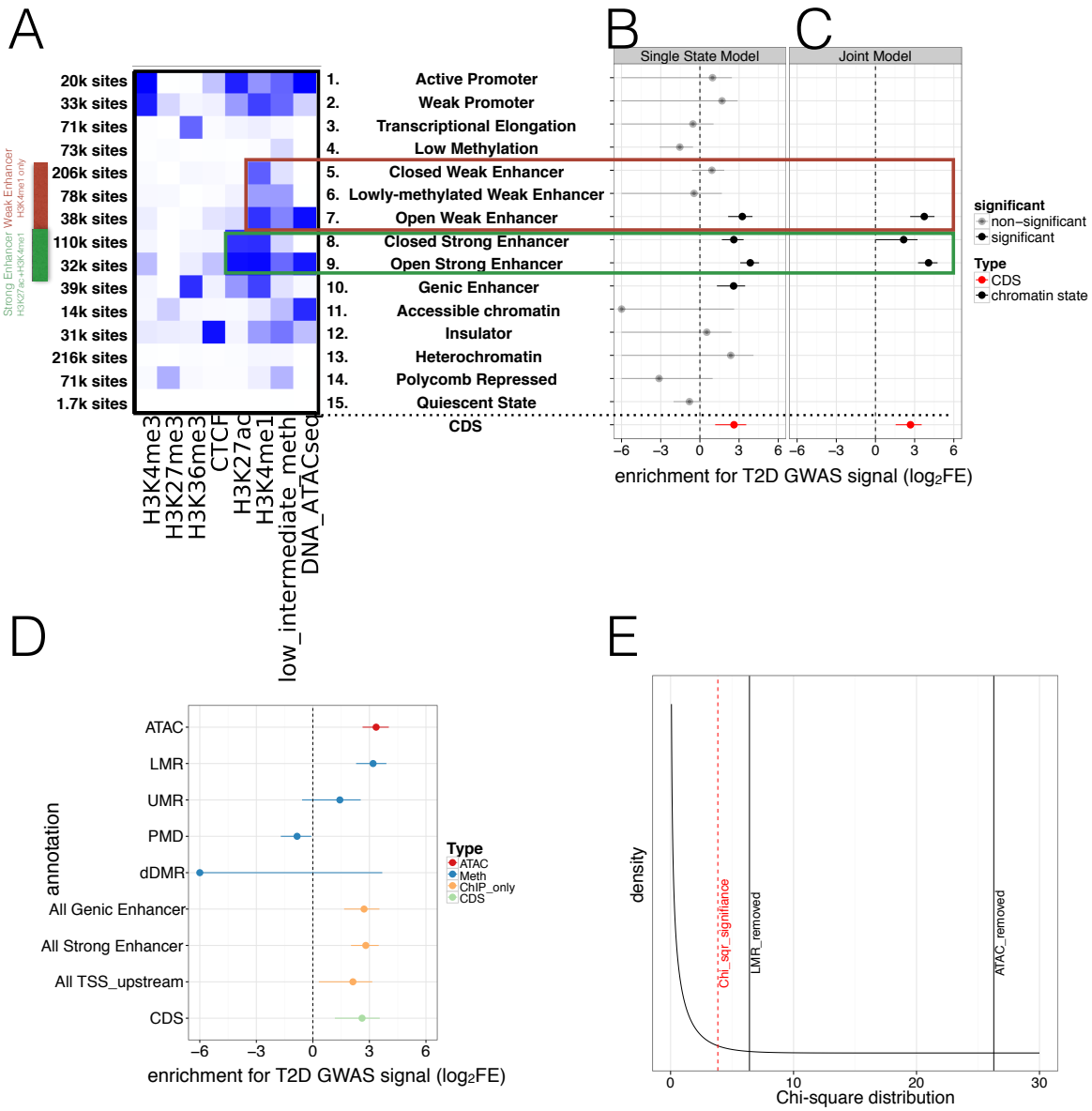

FIGURE 3. Integration of islet epigenetic data to refine chromatin regulatory states and enrichment of these states in T2D GWAS data. A) 15 chromatin states (y-axis) were derived from ChIP histone marks, DNA methylation and ATAC-seq open chromatin annotations (x-axis) using chromHMM. For each state the 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). 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, CDS is shown to also include the effect of protein-coding regions. Significantly enriched annotations are shown in black or red (CDS). 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. 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 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.

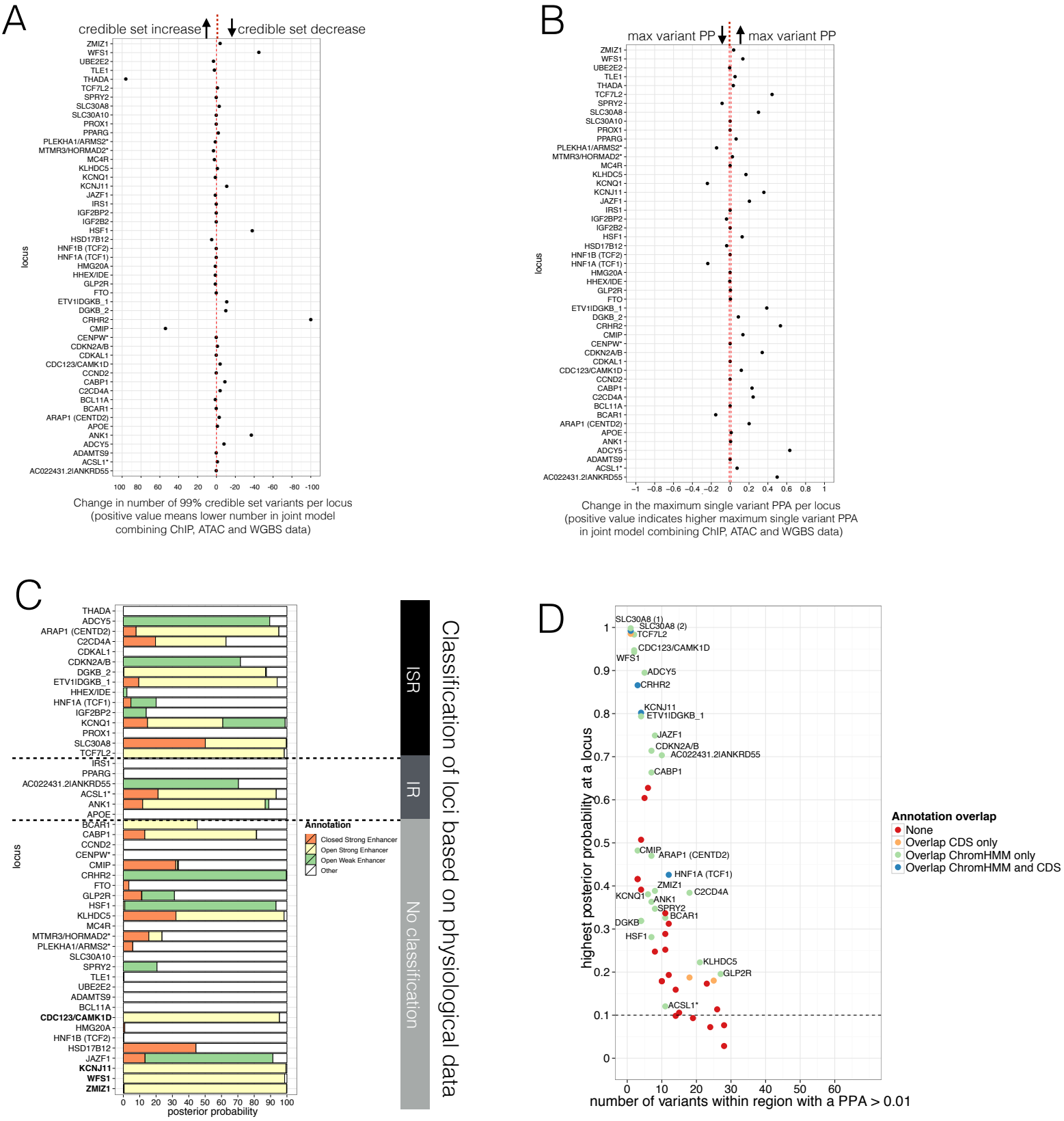

FIGURE 4. Evaluating Posterior Probabilities (PP) derived from the FGWAS maximum likelihood model at significant T2D GWAS loci. (A) Per locus the difference in the number of 99% credible set variants between ChIP+ATAC+Meth and ChIP-only model is shown (positive values indicate a reduction in the number 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.

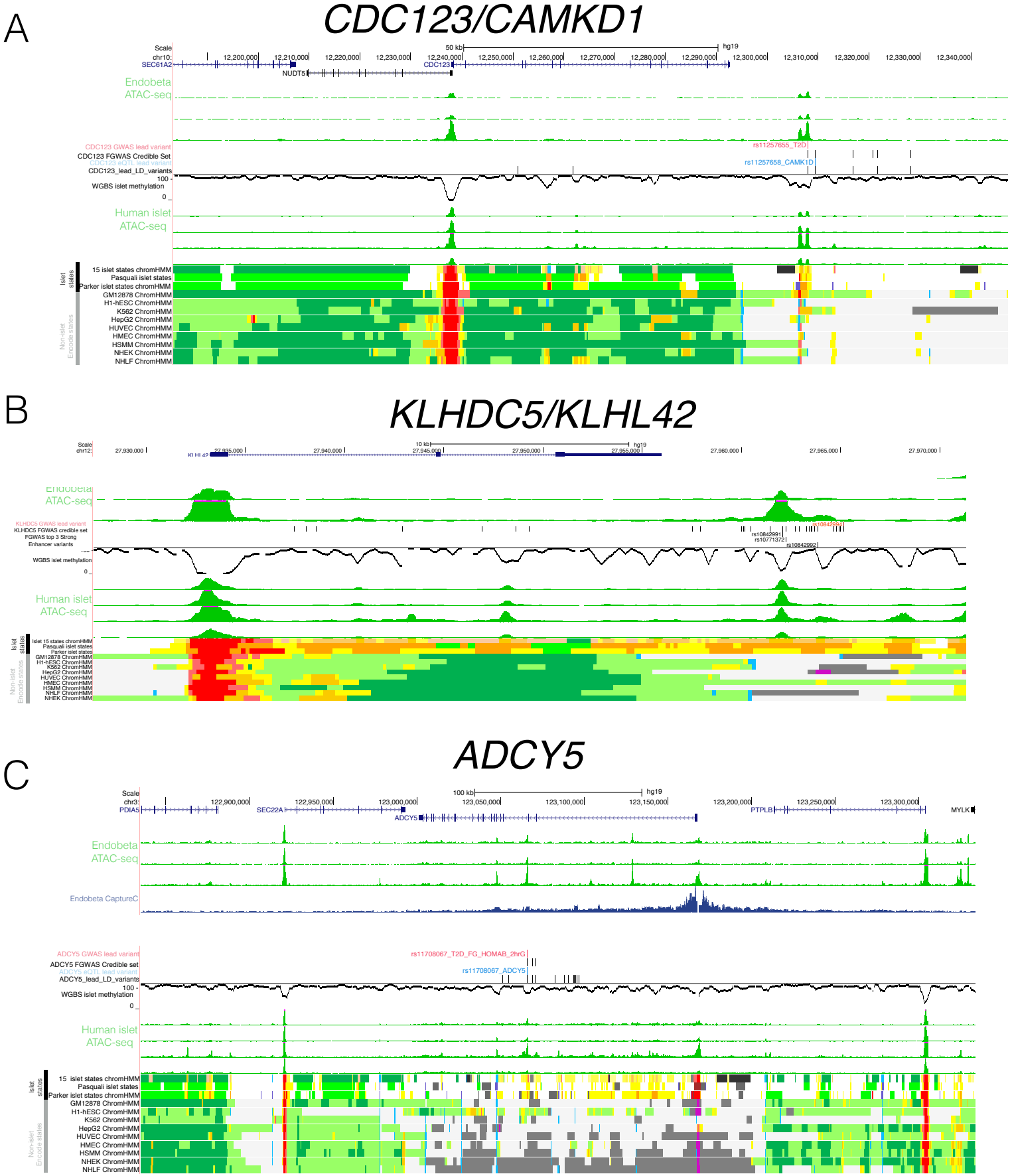

FIGURE 5. Epigenome Landscape of selected loci with allelic imbalance. For each locus A) *CDC123*, B) *KLHDC5* and C) *ADCY5* the following information is shown: 3 ATAC-seq Endoß tracks (green, top), variant level information (depending on the region GWAS lead SNP red, credible set black, eQTL blue and 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).

	Chromatin States	Total number of states	T2D log2FE (CI)	FG log2FE (CI)
Augmented chromatin state maps	**1. Active Promoter**	20k	1 (-20 to 2.4)	2.6 (-0.5 to 4.3)
	2. Weak Promoter	33k	1.7 (-20 to 2.9)	1.5 (-20 to 4.1)
	3. Transcriptional Elongation	71k	-0.5 (-20 to 1)	NA
	4. Low Methylation	73k	-1.5 (-3.1 to -0.5)	-2.2 (-5.2 to -0.5)
	5. Closed Weak Enhancer	206k	0.9 (-0.6 to 1.9)	**2.1 (0.2 to 3.4)**
	6. Lowly-methylated Weak Enhancer	78k	-0.5 (-20 to 1.7)	-29.8 (-49.8 to 1.8)
	7. Open Weak Enhancer	38k	**3.2 (2.2 to 4)**	**3.6 (0.8 to 5.1)**
	8. Closed Strong Enhancer	110k	**2.6 (1.7 to 3.3)**	**4 (2.6 to 5.3)**
	9. Open Strong Enhancer	32k	**3.9 (3.1 to 4.5)**	**5.2 (3.7 to 6.5)**
	10. Genic Enhancer	39k	**2.6 (1.3 to 3.5)**	2.9 (-20 to 4.7)
	11. Accessible chromatin	14k	-26.9 (-46.9 to 2.6)	-25.4 (-45.4 to 3.9)
	12. Insulator	31k	0.5 (-20 to 2.4)	1.5 (-20 to 4.1)
	13. Heterochromatin	216k	2.4 (-20 to 4.1)	2 (-1.3 to 4.5)
	14. Polycomb Repressed	71k	-3.1 (-20 to 1)	-30.5 (-50.5 to 1.7)
	15. Quiescent State	1.7k	-0.8 (-2 to 0.1)	-28.2 (-48.2 to 0)
CDS	**CDS**	NA	**2.6 (1.2 to 3.5)**	2.7 (-0.2 to 4.3)
Parker states	**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 (log_2FE) 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	0.89	0.30 (39 A VS 93 G alleles)	2.9E-06	risk allele A closed
rs11257655	CDC123	4.0E-08	0.95	0.39 (279 C VS 436 T alleles)	4.7E-09	risk allele T open
rs10842991	KLHDC5	7.3E-07	0.13	0.67 (79 C VS 39 T alleles)	2.9E-04	risk C allele open

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