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

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

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

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

Abstract

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

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 (Scott et al., 2017, Mahajan et al., 2014, Morris et al., 2012, Voight et al.,
46 2010), the majority of which affect insulin secretion and beta cell function
47 (Dimas et al., 2014, Wood et al., 2017). However, most GWAS signals map to the
48 non-coding genome and identification of the molecular mechanisms through
49 which non-coding variants exert their effect has proven challenging. Several
50 studies have demonstrated that T2D-associated variants map disproportionately
51 to regulatory elements, particularly those which influence RNA expression and
52 cellular function of human pancreatic islets. (Parker et al., 2013, Pasquali et al.,
53 2014, van de Bunt et al., 2015, Olsson et al., 2014, Dayeh et al., 2014, Volkov et
54 al., 2017, Varshney et 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).
69
70 Most importantly, in part due to historical challenges in accessing human islet
71 material or authentic human cellular models, reference annotations of the islet

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

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

2. Results:

2.1 Characterising the DNA methylation landscape of human pancreatic islets

To characterise the human islet methylome and characterise the role of DNA methylation with respect to T2D genetic risk, we performed WGBS (mean coverage 13X) in human pancreatic islet DNA samples isolated from 10 non-diabetic cadaveric donors of European descent. Methylation levels across the genome were highly correlated across individual donors (mean CpG methylation Spearman's rho across 10 individual WGBS donors=0.71, 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 ascertained from non-diabetic donors of European descent (5
113 overlapping with those from whom WGBS data were generated). As with the
114 WGBS data, methylation levels were highly correlated across individuals (mean
115 CpG 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 for 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 with high precision and
125 coverage ~480k CpG sites (approximately 2% of all sites in the genome),
126 selected primarily because they are located near gene promoters and CpG-island
127 regions. The focus of the 450k array on these regions, which tend to be less
128 variable in terms of methylation, explains the high 450k array correlation levels
129 between donors. In addition, this selective design results in marked differences
130 in the distributions of genome-wide methylation values between WGBS and the
131 450k array. Whilst the WGBS data revealed the expected pattern of widespread
132 high methylation levels with hypomethylation (<50%) restricted to 11.2%
133 (2.6M/23.3M CpG sites) of the genome, the array disproportionately
134 interrogated those hypomethylated sites (218k [46%] of all 450k CpG probes)
135 (Kolmogorov–Smirnov (KS) test for difference, D=0.40, $P<2.2x10^{-16}$) (Figure 1B).
136 These differences in methylation distribution were also evident within specific
137 islet regulatory elements from previously defined standard chromatin state

138 maps (Parker et al., 2013) (Figure 1C, Supplementary Figure 1C-D). We found
139 significant (FDR<0.05) differences between the methylation levels of CpG sites
140 accessed on the array, and those interrogated by WGBS, across most islet
141 chromatin states: the largest differences were observed for weak promoters
142 (median WGBS=0.71 vs. median 450k=0.11, KS test D=0.51, $P<2.2x10^{-16}$,) and
143 weak enhancers (WGBS=0.87 vs. 450k median=0.76, D=0.39, $P<2.2x10^{-16}$,
144 Supplementary Figure 1D).
145
146 In terms of coverage, most chromatin states, apart from promoters, were poorly
147 represented by CpG sites directly interrogated by the array: for example the
148 array assayed only ~2.9% of CpG sites in strong enhancer states (2.7-3.8%
149 depending on strong enhancer subtype, Figure 1C). Although methylation levels
150 were previously reported to be highly correlated across short (0.1-2kb) genomic
151 distances (Zhang et al., 2015, Bell et al., 2011, Eckhardt et al., 2006, Guo et al.,
152 2017), the observed significant differences in the methylation distribution
153 (Figure 1C, Supplementary Figure 1D) across chromatin states including weak
154 promoter (median size 600bp) and enhancer subtypes (median size ranges from
155 200-1200bp) indicate that these correlative effects are not strong enough to
156 counterbalance the low coverage of the 450k array. These findings are consistent
157 with 450k array content being focused towards CpG-dense hypomethylated and
158 permissive promoter regions. This highlights the limited capacity of the array to
159 comprehensively interrogate the global DNA methylome, in particular at distal
160 regulatory states such as enhancers.
161
162 To understand the value of these data to reveal molecular mechanisms at GWAS
163 loci, where we and others had shown enrichment for islet enhancer states
164 (Pasquali et al., 2014, Gaulton et al., 2015a, Parker et al., 2013), we were
165 interested to see how the selective coverage of the array might impact on its
166 ability to interrogate methylation in GWAS-identified regions. We used the
167 largest currently available T2D DIAGRAM GWAS data set (involving 26.7k cases
168 and 132.5k controls of predominantly European origin, see dataset section for
169 details) to identify the "credible sets" of variants at each locus which collectively

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

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

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

203 **2.2 Integration islet methylation and other epigenomic annotations**

204

205 Studies in a variety of other tissues have shown that hypomethylation is a strong
206 indicator of regulatory function (Stadler et al., 2011). More specifically,
207 continuous stretches of CpG-poor Low-Methylated Regions (LMRs, with
208 methylation ranging from 10-50% and containing fewer than 30 CpG sites)
209 denote potential distal regulatory elements such as enhancers, while stretches of
210 CpG-rich UnMethylated Regions (UMRs, containing more than 30 CpG sites) are
211 more likely to represent proximal regulatory elements including promoters
212 (Burger et al., 2013). We detected 37.1k LMRs, 13.6k UMRs (Figure 2A) and
213 10.7k Partially Methylated Domains (PMDs, Methods and Supplementary Figure
214 2A-B). PMDs represent large regions of unordered methylation states associated
215 with DNA sequence features (Gaidatzis et al., 2014). As anticipated, we found
216 significant enrichment of LMRs with weak and strong enhancer states as defined
217 by islet chromatin state maps derived from existing ChIP-seq data (Parker et al.,
218 2013) (69.2% of islet LMRs overlapped islet strong and weak enhancer states,
219 $\log_2$FE=2.2-2.9, Bonferroni P<0.05, (Figure 2B, Supplementary Figure 1C).
220 Similarly, UMRs were enriched for islet active promoter chromatin states (90.8%
221 of UMRs overlapped islet active promoters, $\log_2$FE=3.9, FDR <0.05, Figure 2B).

222

223 To further characterise these hypomethylation domains, we overlapped
224 information from analyses of islet cis-expression QTLs (eQTLs) (van de Bunt et
225 al., 2015) and islet ChIP-seq transcription factor binding sites (TFBS) (Pasquali
226 et al., 2014). We observed marked enrichment for eQTLs (LMR $\log_2$FE=1.1, UMR
227 $\log_2$FE=2.7, Bonferroni P<0.05) and TFBS (LMR $\log_2$FE=4.1-4.6; UMR $\log_2$FE=2.4-
228 3.9, Bonferroni P<0.05, Figure 2B). These observations confirm that islet LMRs
229 and UMRs correspond to important tissue-specific regulatory regions,
230 overlapping cis-regulatory annotations known to be enriched for T2D GWAS
231 signals (Pasquali et al., 2014, Gaulton et al., 2015b).

232

233 We also considered the relationship between LMR and UMR regions defined in
234 our non-diabetic islet WGBS, and a complementary set of methylation-based
235 annotations previously derived from WGBS of islets from 6 T2D and 8 control

236 individuals (Volkov et al., 2017). In that study, comparisons between diabetic
237 and non-diabetic islets had been used to define a set of 25,820 "disease
238 differentially methylated regions" (dDMRs, min absolute methylation difference
239 5% and P<0.02). We found only limited overlap between these dDMRs and the
240 UMRs and LMRs from our data: of the 25,820 dDMRs, 2.2% overlapped LMRs
241 and 2.4% UMRs. This overlap was slightly greater than expected by chance
242 (Bonferroni P<0.05, LMR $\log_2$FE=1.0 and promoter-like UMRs $\log_2$FE=1.1, Figure
243 2B) but more modest than seen for the other regulatory annotations. Similarly,
244 we also observed that dDMRs showed more modest ($\log_2$FE=0.4-1.0), but still
245 significant (Bonferroni P<0.05) levels of enrichment with respect to all other
246 islet regulatory annotations (Figure 2B). The modest enrichment of dDMRs
247 indicates that only a fraction of these regions correspond to islet genomic
248 regulatory sites. Given that T2D risk variants preferentially map in islet
249 regulatory sites, the corollary is that most dDMRs are unlikely to directly
250 contribute to the mediation of genetic T2D risk.
251
252 **2.3 Refining islet enhancer function using methylation and open chromatin**
253 **data**
254
255 To further characterise the regulatory potential of hypomethylated regions,
256 including LMRs and UMRs, we combined the islet WGBS methylation data with
257 chromatin accessibility data generated from ATAC-seq assays of 17 human islet
258 samples (from non-diabetic donors of European descent; mean read count after
259 filtering =130M, Supplementary Figure 2C). We identified a total of 141k open
260 chromatin regions based on read depth, peak width and signal-to-noise ratio
261 (see Methods). These regions of islet open chromatin showed substantial overlap
262 (78%) with equivalent regions described in a recent study of two human islets
263 (Varshney et al., 2017) ($\log_2$FE=2.8 compared to random sites, not shown). In
264 addition, our islet ATAC-Seq sites demonstrated substantial overlap with LMRs:
265 53% of LMRs overlapped 16% of all ATAC-seq peaks (LMR $\log_2$FE=3.8 compared
266 to randomised sites, Figure 2B). Almost all UMRs (98%) were contained within
267 regions overlapping (13% of) ATAC-seq peaks (UMR $\log_2$FE=3.4 compared to
268 randomised sites, Figure 2B).

269

270 To fully leverage information across multiple overlapping islet epigenome
271 assays, we generated augmented chromatin state maps, using chromHMM (Ernst
272 and Kellis, 2012). These maps combined the WGBS methylation and ATAC-Seq
273 open chromatin data with previously generated ChIP-seq marks (Figure 3A,
274 Supplementary Figure 3A). For these analyses, we initially used a single
275 definition for hypomethylated regions (methylation<60%) that captured both
276 UMRs and LMRs (see Methods).
277
278 This augmented and larger set of 15 islet chromatin states retained the broad
279 classification of regulatory elements that included promoters (positive for
280 H3K4me3), transcribed and genic regions (H3K36me3), strong enhancers
281 (H3K4me1; H3K27ac), weak enhancers (H3K4me1), insulators (CTCF) and
282 repressed elements (H3K27me) (Figure 3A). The addition of islet methylation
283 and open chromatin data expanded existing chromatin state definitions to
284 provide new subclasses, particularly amongst enhancer elements. Here, we
285 observed two subclasses of strong enhancers and three of weak enhancers
286 (Figure 3A). We denote the strong enhancer subtypes as "open" (n=32k genome-
287 wide), characterised by open chromatin and hypomethylation, and "closed"
288 (n=110k) with closed chromatin and hypermethylation (Figure 3A). The three
289 weak enhancer states we denote as "open" (n=38k: open chromatin,
290 hypomethylation), "lowly-methylated" (n=78k; closed chromatin,
291 hypomethylation) and "closed" (n=206k: closed chromatin, hypermethylation).
292 No equivalent class of "lowly-methylated" strong enhancers was observed in the
293 15-state model. When comparing these chromatin states to those identified
294 using only ChIP-seq marks ((Parker et al., 2013), Supplementary Figure 1C,), the
295 two strong enhancer subclasses we identified subdivided the "strong enhancer
296 1" state as described by Parker (defined by H3K27ac and H3K4me1). Additional
297 comparison to "stretch" enhancer clusters (Parker et al., 2013), showed that
298 there was considerable overlap between the "open" strong and weak enhancer
299 states we identify here and previously-described "stretch" enhancer states
300 (16.1k out of 23k stretch enhancer overlapped 32k out of 70.1k "open"
301 enhancers). Even so, most (55%) "open" enhancer states, and in particular

302 “open weak enhancers” (70%), were not captured within “stretch” enhancer
303 intervals, and we regard these as distinct islet enhancer subclasses.
304
305 To understand the relationship of these various state definitions to genetic
306 variants influencing T2D risk, we applied the hierarchical modelling approach
307 FGWAS to the same sets of large-scale GWAS data for T2D (from DIAGRAM (Scott
308 et al., 2017)) and FG (ENGAGE (Horikoshi et al., 2015)) described in section 2.1.
309 FGWAS allowed us to combine GWAS and genomic data to determine the
310 genome-wide enrichment within islet regulatory features for variants associated
311 with T2D risk. These enrichment priors were then used to generate credible
312 variant sets that are informed by both GWAS and genomic data, as described in
313 section 2.4.
314
315 In single-feature analyses, we found significant enrichment (lower limit of
316 Confidence Interval (CI)>0) limited to four enhancer states (open weak
317 enhancers, both types of strong enhancer and H3K36me3 marked genic
318 enhancers) (Figure 3B, Table 1). To take into account protein-coding variant and
319 conserved sequence effects, we also included CoDing exon Sequence (CDS)
320 (Carlson and Maintainer, 2015) and CONServed sequence (CONS) (Lindblad-Toh
321 et al., 2011) as additional annotations which were previously found to be
322 strongly enriched for T2D GWAS signal (Finucane et al., 2015).We observed
323 significant enrichment for CDS and CONS sequence in the single state results
324 (Figure 3B, Table 1). FGWAS multi-feature analyses for T2D, incorporating all
325 annotations positive in single-element analyses, retained both subclasses of
326 strong enhancer, the subclass of open weak enhancers, genic enhancers and CDS
327 in the joint model (Figure 3C and Methods). Conserved sequence annotations
328 were not retained in the joint model.
329
330 We observed markedly different levels of enrichment for T2D association
331 between and within open and closed enhancer states (Figure 3B-3C, Table 1).
332 Using these augmented chromatin state maps, we demonstrated clear
333 enrichment for T2D association for the subset of “open” weak enhancers (12% of
334 all weak enhancer sites) with no evidence of enrichment in the remaining

335 subclasses ("closed" and "lowly-methylated") (Figure 3B and Table 1). This
336 concentration of enrichment amongst a relatively small subset of the weak
337 enhancers was consistent with the lack of enrichment across all weak enhancers
338 defined solely on the basis of H3K4me1 signal ((Parker et al., 2013), single state
339 $\log_2$FE=0.9, CI=-2.5 to 2.0, Table 1, Supplementary Figure 1C). We also saw
340 differences in enrichment signal between open and closed strong enhancers,
341 with the most marked enrichment amongst open strong enhancers (22% of the
342 total, Figure 3B-C, Table 1). This effect was particularly obvious in the joint-
343 analysis (open strong enhancer joint FE=4.1, CI=3.3 to 4.8 vs. closed strong
344 enhancer joint $\log_2$FE=2.4, CI=0.5 to 3.3, Figure 3C).
345
346 Hypomethylation and open chromatin are highly correlated, but the observed
347 difference in T2D enrichment between the weak enhancer states (particularly
348 between "lowly-methylated" and "open" which differ markedly with respect to
349 chromatin status) points to a primary role for open chromatin. To test this
350 further, we regenerated chromatin state maps using different subsets of the data
351 (ChIP-only, with optional addition of methylation and/or open chromatin
352 information, see Methods and Supplementary Figure 3A-3B). These analyses
353 confirmed that the T2D GWAS enrichment signal was predominantly driven by
354 the distribution of islet open chromatin (Supplementary Figure 3C).
355
356 We further evaluated the role of subclasses of DNA methylation regulatory
357 region with respect to T2D GWAS enrichment. We divided hypomethylated
358 (<60% methylated) sequence into enhancer-like LMRs (6.5% of all
359 hypomethylated sequence), promoter-like UMRs (7.5% of hypomethylated
360 sequence), as well as PMDs (61% of hypomethylated sequence). The remaining
361 25% of hypomethylated sequence did not fit any category. LMRs were
362 significantly (CI>0) enriched ($\log_2$FE=3.2, CI=2.3 to 3.9) for T2D association
363 signals consistent with their co-localisation with distal regulatory elements, and
364 displayed modestly increased enrichment compared to enhancer states derived
365 from ChIP-seq alone (Figure 3D, Supplementary Table 1). In contrast, no
366 significant enrichment was found for human islet (promoter-like) UMRs
367 ($\log_2$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

368 no evidence that recently-described regions of T2D-associated differential
369 methylation (dDMRs: derived from comparison of WGBS data from islets of
370 diabetic and non-diabetic individuals) were enriched for genome-wide T2D
371 association signals (log_2FE=-24.6, CI=-44.6 to 3.7) (Figure 3D, Supplementary
372 Table 1).
373
374 Finally, since the hypomethylation signal for T2D enrichment was concentrated
375 in LMRs (Figure 3D, Supplementary Table 1), we reran a FGWAS joint-analysis
376 combining open chromatin peaks, LMRs and ChIP-only states using a nested
377 model (Figure 3E, Supplementary Figure 3D-E, see Methods). This confirmed
378 that the improvement in enrichment was mainly driven by open chromatin but
379 showed that LMRs also contributed significantly and independently to the
380 enrichment (Figure 3E, Supplementary Table 2).
381
382 FGWAS analysis for FG corroborated the observations from T2D analysis.
383 Despite reduced power of the FG GWAS data due to a lower number of
384 significantly associated FG GWAS loci, both single feature and joint-model
385 analyses of human islet epigenome data found significant enrichment in strong
386 enhancer states with the strongest enrichment in enhancers with open
387 chromatin and hypomethylation (Supplementary Figure 4A-B and Table 1). In
388 addition, evaluation of the relative contributions of ATAC-seq open chromatin
389 and DNA methylation to FG GWAS enrichment across both single-feature
390 (Supplementary Figure 4C-D) and joint-model analysis (Supplementary Figure
391 4E-F and Supplementary Table 3) indicated that open chromatin was primarily
392 responsible for the enhanced enrichment.
393
394 Overall, these analyses demonstrate that the addition of open chromatin and
395 DNA methylation data to ChIP-seq marks enhances the resolution of regulatory
396 annotation for human islets. In particular, it defines subsets of weak and strong
397 enhancers that differ markedly with respect to the impact of genetic variation on
398 T2D risk. Although DNA accessibility and hypomethylation status are strongly
399 correlated and provide broadly similar enrichments, the effects of the former
400 predominate. In line with the dominance of open chromatin status for T2D GWAS

401 enrichment, we observed that T2D risk in relation to methylation status is
402 primarily invested in hypomethylated LMRs (i.e. enhancers) rather than UMRs,
403 dDMRs or PMDs.
404
405 **2.4 Augmented chromatin maps and open chromatin allelic imbalance**
406 **refine likely causal variants at *ADCY5* , *CDC123,* and *KLHDC5***
407
408 We next deployed the insights from the global FGWAS enrichment analyses to
409 define the molecular mechanisms at individual T2D susceptibility loci, refining
410 T2D causal variant localisation using the combination of genetic data (from fine-
411 mapping) and the genome-wide patterns of epigenomic enrichment.
412
413 Specifically, we applied FGWAS to the T2D DIAGRAM GWAS data (Scott et al.,
414 2017) under the joint model (Figure 3C) derived from the augmented chromatin
415 state maps. We divided the genome into 2327 segments (average size 5004 SNPs
416 or 1.2Mb) and identified 52 segments significantly associated with T2D genome-
417 wide (segmental FGWAS PPA >=0.9 or single variant GWAS $P< 5x10^{-8}$, see
418 Methods for details). These corresponded to 49 known T2D associated regions
419 representing that subset of the ~120 known T2D GWAS loci which passed those
420 significance/filtering criteria in this European-only dataset. We then calculated
421 reweighted PPAs for each variant within each segment and generated
422 reweighted 99% credible sets. (Of note, in line with traditional GWAS
423 nomenclature, locus names were defined based on proximity between the lead
424 variant and the closest gene and does not, of itself, indicate any causal role for
425 the gene in T2D susceptibility).
426
427 Consistent with the increased T2D GWAS enrichment of states including open
428 chromatin and DNA methylation information, we found that analyses using
429 enrichments from the augmented chromatin state model (combining ChIP-seq,
430 ATAC-seq and WGBS data) were associated with smaller 99% credible sets
431 (median of 17 SNPs) than those derived from FGWAS enrichment derived from
432 ChIP-seq data alone (median 23). In parallel, the PPA for the best variant per
433 locus increased (median 0.39 vs 0.31). Individual T2D GWAS locus results are

434 shown in Figure 4A-B. We also expanded the FGWAS PPA analysis to investigate
435 open chromatin and DNA methylation effects on fine-mapping and found that the
436 reduction in 99% credible set size and increase in maximum variant PPA was
437 driven predominantly by open chromatin (Supplementary Figure 5,
438 Supplementary Table 4). This demonstrates that the inclusion of open chromatin
439 maps helps to improve prioritisation of causal variants at many T2D GWAS loci.
440
441 A subset of T2D GWAS signals are known to influence T2D risk through a
442 primary effect on insulin secretion, whilst others act primarily through insulin
443 resistance. We used previous categorisations of T2D GWAS loci based on the
444 patterns of association with quantitative measurements of metabolic function
445 and anthropometry (Wood et al., 2017, Dimas et al., 2014), to define a set of
446 15/48 loci most clearly associated with deficient insulin secretion (and therefore
447 most likely to involve islet dysfunction). At 11 of these 15 loci, we found that islet
448 "open strong enhancer" states, and to a lesser extent "open weak enhancer" and
449 "closed strong enhancer", captured more than 60% of the PPA (median 92%,
450 Figure 4C). Variants in these islet enhancer subclasses also captured at least 95%
451 of the PPA at 4 T2D GWAS loci that could not be classified according to
452 physiological association data but which have been previously implicated in
453 human islet genome or functional regulation based on islet eQTL (van de Bunt et
454 al., 2015) or mQTL (Olsson et al., 2014) data (Figure 4C, genes highlighted in
455 bold). In contrast, at 3/6 of the insulin resistance and all but 5 unclassified loci,
456 the PPA was mostly (>50%) attributable to other non-islet enhancer states
457 (across all insulin resistance and unclassified loci, DNA not overlapping islet
458 enhancers and defined as "Other" capture a median PPA of 64%). Thus, islet
459 regulatory annotations are particularly useful for fine-mapping T2D GWAS loci
460 that affect insulin secretion and beta-cell function.
461
462 To obtain additional evidence to support the localisation of causal variants, we
463 tested for allelic imbalance in ATAC-seq open chromatin data. We selected 54
464 variants within 33 T2D-associated GWAS segments for testing of allelic
465 imbalance on the basis of (a) a reweighted variant PPA >=10% and (b) overlap
466 with an enriched regulatory state within the FGWAS T2D joint-model (Figure 4D,

467 Supplementary Table 5). Of these, 20 variants (at 16 loci) had sufficient numbers
468 of heterozygous samples (>2) and ATAC_Seq read depth (depth>9 and at least 5
469 reads for each allele). After correcting for mapping bias using WASP, we
470 observed the strongest evidence for allelic imbalance (FDR< 0.05) at 3 out of the
471 20 variants (rs11257655 near *CDC123* and *CAMK1D*, rs10842991 near *KLHDC5*
472 and rs11708067 at *ADCY5*) (Table 2). All three overlapped refined islet open
473 strong or open weak enhancer regions characterised by open chromatin and
474 hypomethylation.
475
476 Variant rs11257655 accounts for 95% of the reweighted PPA (compared to a
477 PPA of 20% from genetic data alone) at the *CDC123/CAMK1D* locus, overlaps an
478 "open strong enhancer" region (Figure 5A) and the risk allele correlates with
479 increased chromatin accessibility. The same variant is in high LD (r2=0.82) with
480 the lead variant for a cis-eQTL for *CAMK1D* in islets (van de Bunt et al., 2015). In
481 experimental assays (Fogarty et al., 2014), the T2D-risk allele has been shown to
482 be associated with increased *CAMK1D* gene expression and enhanced binding of
483 the FOXA1 and FOXA2 transcription factors. These data all point to rs11257655
484 as the causal variant at this locus.
485
486 At *KLHDC5*, no clear causal variant emerged based on genetic fine-mapping data
487 alone as the credible set contained 23 variants in high mutual LD (r^2>0.8, top
488 variant PPA<5%, Figure 5B). Of these, variants rs10771372 (genetic fine-
489 mapping PPA= 5%), rs10842992 (genetic fine mapping PPA=5%) and
490 rs10842991 (genetic fine-mapping PPA=3%) overlapped "open strong
491 enhancer" regions (Figure 5B), such that their reweighted PPAs rose to 21%
492 (rs10771372), 21% (rs10842992) and 13% (rs10842991), respectively. We
493 observed allelic imbalance only at rs10842991 with the T2D-risk C allele
494 showing greater chromatin accessibility (binomial $P=4.1x10^{-3}$, Table 2). This
495 variant further overlapped a predicted TFBS motif for PAX6 as determined by
496 the software tool FIMO (Grant et al., 2011): the T2D-risk allele was predicted to
497 enhance PAX6 transcription factor binding consistent with the allelic effects on
498 increasing chromatin accessibility (Supplementary Figure 6A). This strong
499 enhancer region is almost exclusively found in islets, with strong enhancer

500 H3K27ac states overlapping rs10842991 in only two non-islet (heart and
501 smooth muscle) Epigenome Roadmap tissues (out of 99 tissues with 18-state
502 chromatin state information, Figure 5B)(Roadmap Epigenomics et al., 2015).
503 Islet eQTL data (Varshney et al., 2017) also links rs10842991 and close proxy
504 SNPs (including rs7960190) to islet transcription with the risk allele increasing
505 *KLHDC5* expression. These data prioritise rs10842991 as the likely causal
506 variant at the *KLHDC5* T2D GWAS locus, and indicate a likely molecular
507 mechanism involving modified PAX6 transcription factor binding and an impact
508 on *KLHDC5* expression and islet function.
509
510 The third example of allelic imbalance mapped to the *ADCY5* locus. Fine-mapping
511 based solely on genetic data could not prioritise a distinct causal variant due to
512 multiple variants in high LD (range for top 5 variants=12-26%, Figure 5C).
513 However, reweighting of variants based on epigenomic annotation clearly
514 prioritised variant rs11708067: this SNP overlapped an "open weak enhancer"
515 and captured most of the reweighted PPA (PPA=92%). Allelic imbalance analysis
516 also showed that the T2D-risk A allele was associated with decreased chromatin
517 accessibility (binomial $P=1.2x10^{-6}$, Table 2). The same lead variant maps to an
518 islet cis-eQTL and methylation QTL (Figure 5C, Supplementary Figure 6B) at
519 which the T2D-risk allele is associated with reduced *ADCY5* expression and
520 increased *ADCY5* gene body DNA methylation.
521
522 To further understand the role of the rs11708067 variant, we performed ATAC-
523 seq and Next Generation Capture-C, in the glucose-responsive human beta-cell
524 line EndoC-βH1 (n=3). We targeted the *ADCY5* promoter to define distal regions
525 interacting with the promoter, and confirmed physical contact with the
526 hypomethylated open chromatin enhancer region harbouring rs11708067
527 (Figure 5C, Supplementary Figure 6C). To resolve the significance of the
528 interaction between the restriction fragment encompassing rs11708067 and the
529 *ADCY5* promoter, we used the programme peakC(de Wit and Geeven, 2017)
530 (*https://github.com/deWitLab/peakC)* to evaluate the interactions of 12
531 fragments covering the lead SNP rs11708067 and 15 SNPs in high LD (r2 > 0.8)
532 across a region of 47kb. After adjusting for multiple testing using FDR correction,

533 only two fragments yielded a significant normalised read number over
534 background. This included the open-chromatin overlapping fragment containing
535 rs11708067 and another fragment harbouring rs2877716, rs6798189,
536 rs56371916 (Supplementary Figure 6D). These SNPs fall into a region that did
537 not show evidence of open chromatin.
538
539 These findings support rs11708067 as the likely causal variant affecting islet
540 accessible chromatin (in line with another recent study (Roman et al., 2017)),
541 and link the open and hypomethylated enhancer element in which it sits to
542 regulation of *ADCY5* expression in islets.
543
544 **3. Discussion**
545
546 A key challenge in the quest to describe the molecular mechanisms through
547 which GWAS signals influence traits of interest, involves the identification of the
548 causal variants responsible and, given that most lie in non-coding sequence, the
549 characterisation of the regulatory elements which they perturb. This underpins
550 efforts to define the effector genes through which these variants operate and to
551 reconstruct the biological networks that are central to disease pathogenesis.
552
553 Genetic and physiological studies have highlighted the singular importance of
554 pancreatic islet dysfunction in type 2 diabetes, but epigenomic characterisation
555 of this tissue has been limited in large-scale community projects such as ENCODE
556 and GTex. The present study seeks to address this deficit by describing, in
557 unprecedented detail, genome-wide patterns of methylation and chromatin
558 accessibility in human islet material. We have combined these data with existing
559 islet epigenomic marks to generate a refined regulatory map which, based on the
560 evidence of improved enrichment for T2D association signals, offers more
561 granular annotation of functional impact.
562
563 Our data show that, for DNA methylation, the signal of T2D predisposition is
564 primarily associated with enhancer-like LMRs rather than other categories of
565 methylation elements including UMRs, dDMRs or PMDs. We highlight the strong

566 correlation between islet methylation status and chromatin accessibility but
567 demonstrate that open chromatin predominantly contributes to defining the
568 regulatory impact associated with genetic T2D risk. Finally, we demonstrate
569 how these enhanced epigenomic annotations, when analysed in concert with
570 genetic fine-mapping data and information from allelic imbalance in chromatin
571 accessibility allow us to home in on likely causal variants at T2D association
572 signals such as those near *ADCY5, CDC123* and *KLHDC5*.
573
574 While previous studies had explored the candidacy of selected variants at the
575 *CDC123* (Fogarty et al., 2014) and *ADCY5* (Olsson et al., 2014, Hodson et al., 2014,
576 van de Bunt et al., 2015) loci with respect to islet regulation and T2D
577 predisposition, our integrative analysis of T2D GWAS and epigenetic data has
578 enabled a detailed and comprehensive analysis that considers the regulatory
579 impact of all variants at these loci across multiple islet samples. Our analysis
580 implicates the rs11257655 and rs11708067 variants as the most likely causal
581 variants at the *CDC123* and *ADCY5* loci respectively and highlights their
582 relationship to islet enhancer activity. The findings at *ADCY5* are supported by a
583 recent paper that found allelic imbalance in H3K27 acetylation involving the
584 rs11708067 variant in a single human islet sample, and which observed that
585 deletion of the relevant enhancer element led to reduction in both *ADCY5* gene
586 expression and insulin secretion (Roman et al., 2017).
587
588 At the *KLHDC5* locus, local LD frustrated efforts to define the causal variant using
589 genetic data alone, but the integration of genetic and epigenetic data pinpointed
590 rs10842991 as the likely culprit based on its impact on chromatin accessibility in
591 an open enhancer region. Evidence that this variant co-localises with an islet cis-
592 eQTL signal points to *KLHDC5* as the likely downstream target (Varshney et al.,
593 2017). Overall, our integrative approach provides useful insights into the
594 functional mechanisms through which T2D GWAS signals operate. Our findings
595 mirror those from other studies, which have, in various ways, and for other
596 complex traits, combined diverse epigenomic annotations to explore the basis of
597 genetic risk (Wang et al., 2016).
598

599 The whole genome methylation data generated in the present study also allowed
600 us to evaluate the likely contribution of previously identified T2D-associated
601 dDMRs (Volkov et al., 2017) with respect to T2D predisposition. These dDMRs,
602 defined on the basis of observed differences in methylation between islets
603 recovered from diabetic and non-diabetic donors, cover a substantial part of the
604 genome, but we were able to show that only a small minority of these overlap
605 functional islet regulatory regions. As a consequence, dDMR regions as a whole
606 had no significant enrichment for T2D association signals. This suggests that
607 most of the dDMR signal involves stochastic effects and/or the secondary
608 consequences on methylation of the diabetic state. However, we cannot exclude
609 that some of the dDMR signals are causal contributors to the diabetic phenotype
610 either because they reflect environmental rather than genetic predisposition, or
611 because they accelerate further perturbation of islet dysfunction as diabetes
612 develops.
613
614 Although we provide highly detailed functional fine-mapping of T2D genetic
615 variants to uncover causal variants, the FGWAS approach applied in this study is
616 limited in its ability to determine the effect of multiple variants at individual loci.
617 Specifically, FGWAS relies on the assumption of a single causal variant within
618 each region, which may not necessarily be true for all loci. This assumption could
619 be violated where there are multiple independent signals at a given locus, or
620 where there are multiple (small effect size) variants on a single risk haplotype
621 which jointly impact the phenotype. Analysis methods that combine functional
622 fine-mapping with conditional analysis and consider LD and haplotype patterns
623 are likely to provide a more complete overview of the causal interactions at T2D
624 GWAS loci.
625
626 In addition, while the present study characterises islet epigenome status and
627 variability in chromatin accessibility in substantially larger numbers of islet
628 samples than those previously reported (Gaulton et al., 2015b, Parker et al.,
629 2013, Pasquali et al., 2014, Varshney et al., 2017), the number of islet
630 preparations for which these data were available was still limited. As a result,
631 our power to detect allelic imbalance in chromatin accessibility was restricted to

632 sites with common variants and relatively large effects. We anticipate that
633 expansion of these sample numbers will extend our capacity to detect such allelic
634 imbalance, and offer more granular insights into the relationships between
635 genetic variation and methylation status. A further limitation is that the genomic
636 data we analysed was generated only from islet samples from non-diabetic
637 donors. Whilst causal inference is possible through the integration of basal
638 epigenomic annotations with genetic data, addition of epigenomic data from
639 islets recovered from diabetic donors has the potential to add a further
640 dimension to such analyses, and to unravel what are likely to be complex causal
641 relationships between genetic variants, epigenomic phenotypes and disease
642 states (Gutierrez-Arcelus et al., 2013). Finally, future work should also focus on
643 experimental validation of likely causal variants and mechanisms e.g. differential
644 binding of the TF PAX6 could be tested at the *KLHDC5* rs10842991 variant
645 through electrophoretic mobility shift assays. Our ongoing research efforts are
646 now concentrated on improving the fine-mapping analysis and expanding these
647 genomic enrichment analyses in larger numbers of human islet samples from
648 healthy and diabetic islets. By coupling the integration of these data with
649 empirical functional studies, we expect to provide an increasingly complete
650 description of the causal interactions between DNA methylation, chromatin
651 state, RNA expression and T2D susceptibility.
652
653 **4. Methods**
654
655 **4.1 Human Pancreatic islet samples**
656
657 **4.1.2 WGBS and 450k array human pancreatic islet sample collection**
658 Human islets were retrieved from deceased Caucasian non-diabetic donors from
659 the Oxford DRWF Human Islet Isolation Facility (n=34) and at the Alberta
660 Diabetes Institute in Edmonton in Canada (n=10). For the analysis only samples
661 with a purity >70% were used as determined by dithizone labeling. The Human
662 Research Ethics Board at the University of Alberta (Pro00001754), the
663 University of Oxford's Oxford Tropical Research Ethics Committee (OxTREC
664 Reference: 2–15), or the Oxfordshire Regional Ethics Committee B (REC

665 reference: 09/H0605/2) approved the studies. All organ donors provided
666 informed consent for use of pancreatic tissue in research.
667
668 For all WGBS (n=10) and a subset of 450k array samples (n=18) human
669 pancreatic islet DNA was extracted from 100,000-150,000 islet cells using Trizol-
670 (Ambion, UK or Sigma Aldrich, Canada) as described previously (van de Bunt et
671 al., 2015). For the remaining 23 samples islet DNA was extracted using the
672 ReliaPrep gDNA Tissue Miniprep system (Promega). Extracted DNA was stored
673 at -80°C before further use.
674
675 **4.1.2 ATAC-seq human pancreatic islet sample collection**
676 Human pancreatic islets preparations (n=18) were retrieved from 17 deceased
677 non-diabetic donors of European descent from the Oxford DRWF Human Islet
678 Isolation Facility and stored for 1-3 days in CMRL or UW media. The latter were
679 reactivated in CMRL for 1h before processing them further. Approximately
680 50,000 islet cells per sample were hand-picked and immediately processed for
681 ATAC-seq as described previously (Buenrostro et al., 2013), however, an
682 additional round of purification was performed using Agencourt AMPure XP
683 magnetic beads (Beckman Coulter).
684
685
686 **4.2. WGBS data generation**
687
688 **4.2.1 Bisulphite conversion**
689 400ng of DNA per human islet samples (n=10) were sent as part of a
690 collaborative effort to the Blizard Institute, Queen Mary University, London, UK
691 and bisulphite- converted using the Ovation Ultralow Methyl-Seq DR Multiplex
692 System 1-8 (Nugen) and purified using Agencourt AMPure beads (Beckman
693 Coulter) as described previously (Lowe et al., 2013).
694
695 **4.2.2 Library generation and processing of reads**
696 The libraries were sequenced by members of the High-Throughput Genomics
697 group at the Wellcome Trust Centre for Human Genetics, University of Oxford,

698 Oxford, UK. Samples were sequenced as multiplex libraries across 3 HiSeq2000
699 lanes with 100bp paired-end read length (including a PhIX spike-in of 5%) to
700 obtain high-coverage read data. The obtained reads were trimmed using a
701 customized python3 script (10bp at the start and 15bp at the end) and aligned to
702 hg19 using the software Bismark (settings: L,0,-0.6, version 0.12.5,
703 RRID:SCR_005604)(Krueger and Andrews, 2011). Specifically, paired-end
704 alignment of trimmed reads was performed and unmapped reads from read 1
705 were realigned using Bismark and merged with the paired-end alignment using
706 samtools (Li et al., 2009) (version 0.1.19, RRID:SCR_002105) in order to increase
707 mapping efficiency. Coverage for the merged paired-end and realigned HiSeq
708 read alignments was estimated for the human mappable genome (NCBI hg19 2.8
709 billion base pairs excluding gaps and unmappable and blacklisted regions
710 according to UCSC and Encode(EncodeProjectConsortium, 2012)) using bedtools
711 (version v2.21.0) (Quinlan, 2014).
712
713 **4.2.3 WGBS DNA methylation quantification and prediction of**
714 **hypomethylated regulatory regions**
715 CpG methylation levels were determined for each sample by calculating the ratio
716 of unmodified C (methylated) and bisulphite converted T (unmethylated) alleles
717 using BiFAST (first described here (Lowe et al., 2013)). High-pass pooled WGBS
718 data was generated by adding methylated and unmethylated read counts across
719 individual low-pass samples to then estimate the average beta methylation
720 levels.
721
722 Regulatory regions were identified using the R package methylseek
723 (RRID:SCR_006513) (Burger et al., 2013). After removing PMDs, which represent
724 highly heterogenous methylation states determined by DNA sequence features
725 (Gaidatzis et al., 2014), LMRs (<30 CpGs) and UMRs (>30 CpGs) were predicted
726 in hypomethylated regions (<50%) at an FDR of 0.05. The methylation level and
727 FDR parameter was inferred from the data as suggested by the methylseek
728 workflow (Burger et al., 2013).
729
730 **4.3 450k DNA methylation array data generation**

731 In total, 41 samples were processed for the Illumina Infinium
732 HumanMethylation450 BeadChip (Illumina, San Diego, CA). Of these 18 samples
733 were bisulphite-converted and processed as part of a collaboration at the UCL
734 Cancer Institute, University College London, London, UK while the remaining 23
735 samples were processed in OCDEM, University of Oxford, Oxford, UK. The DNA
736 was bisulphite converted using the EZ DNA Methylation™ Kit (© Zymogen
737 Research Corp, Irvine, CA) and hybridised to the Illumina 450k array and
738 scanned with iScan (Illumina) according to the manufacturer's protocol.
739
740 The resulting data was analysed using the Package minfi (RRID:SCR_012830)
741 (Aryee et al., 2014) and custom R scripts ((RCoreTeam), R version 3.0.2,
742 RRID:SCR_001905). Specifically, CpG sites with a detection P-value >0.01 were
743 removed from the analysis and samples with >5% of CpG sites failing this
744 threshold (n=9) were also removed from the analysis.
745
746 Following separate quantile normalisation of signal intensities derived from
747 methylated and unmethylated Type I probes and Type II probes, methylation
748 levels (ß) were estimated, based on the intensities of the methylated (M) and
749 unmethylated (U) signal in the following way: $\beta = M/(M+U+100)$. To correct for
750 batch effects the ComBat function implemented in the sva (Johnson et al., 2007,
751 Leek et al.) package was used (Supplementary Figure 7).
752
753 **4.4 ATAC-seq data generation**
754
755 **4.4.1 Sequencing of ATAC-seq reads**
756 ATAC-seq libraries were sequenced at the High-Throughput Genomics group
757 which is part of the Wellcome Trust Centre for Human Genetics, University of
758 Oxford, Oxford, UK. Samples were sequenced as 4-6plex libraries across 1-3
759 Hiseq2500 lanes with 50bp paired-end read length.
760
761 **4.4.2 Processing of ATAC-seq reads**
762 Raw FASTQ reads were processed with an in-house pipeline (first described in
763 (Hay et al., 2016) and on the website

764 http://userweb.molbiol.ox.ac.uk/public/telenius/PipeSite.html). Specifically,
765 library and sequencing quality was checked with FASTQC (RRID:SCR_014583)
766 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and reads were
767 mapped to the human genome (hg19) via bowtie (Langmead et al., 2009)
768 (version 1.1.0, RRID:SCR_005476) with default settings but -m 2, and maxins
769 2000 which allows mapping of reads with a maximum number of 2 alignments
770 and a maximum insert size of 2000bp. For reads that could not be aligned the
771 first time, adapters were removed with Trim Galore at the 3 prime end
772 (RRID:SCR_011847, settings -length 10, -qualFilter 20,
773 http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to enhance
774 the chance of mapping. The resulting trimmed reads were then mapped again
775 with bowtie. Any remaining unmapped and trimmed reads were processed with
776 FLASH (Magoc and Salzberg, 2011) (version 1.2.8, RRID:SCR_005531, settings -m
777 9 -x 0.125) which combines overlapping read pairs and reconstructs read pairs
778 without overlap. These are then realigned a third time using bowtie. PCR
779 duplicates are then removed from the mapped bam files using samtools rmdup
780 function (Li et al., 2009). Additionally, all reads overlapping any of the
781 "unmappable" UCSC Duke blacklisted hg19 regions (EncodeProjectConsortium,
782 2012) are also removed from the final bam file.
783 Open chromatin peaks were called through the aforementioned in-house
784 pipeline by applying sample-specific read depth and width parameters, which
785 were chosen based on the signal to noise ratio of a given sample.
786
787 **4.5 ChIP-seq data and identification of chromatin states**
788
789 **4.5.1 Processing of available ChIP-seq data**
790 Human islet ChIP-seq histone mark and TFBS data were obtained from various
791 sources: H3K4me1, *CTCF* and H3K27ac (Pasquali et al., 2014), H3K36me3 and
792 H3K4me3 (Moran et al., 2012) and H3K27me3 (Roadmap Epigenomics et al.,
793 2015). Available raw fastq files were aligned to hg19 using bowtie1 (version
794 1.1.1) with modified default settings (-m 1 which removes reads with more than
795 1 valid alignment and -n 1 which defines the maximum number of mismatches in
796 seed) and PCR duplicates were removed from the aligned bam files with Picard

797 tools (RRID:SCR_006525, v1.119, http://broadinstitute.github.io/picard/). The
798 resulting reads were converted into bed format using the bedtools bamToBed
799 function (Quinlan, 2014) (RRID:SCR_006646, version v2.21.0) and extended by
800 200bp towards the 3' end to account for fragment size.
801
802 **4.5.2 Identification of chromatin states using chromHMM**
803
804 Binarised 200bp density maps from the bed files of the 6 ChIP-seq marks were
805 created using a Poisson distribution implemented in the BinaryBed function of
806 the ChromHMM software as described in (Ernst and Kellis, 2012, Ernst et al.,
807 2011). From these epigenomic density maps, 11 ChIP-only chromatin states
808 were derived using a multivariate Hidden Markov Model implemented in the
809 Learnmodel function (standard settings, h19 genome) of the software
810 ChromHMM (Ernst and Kellis, 2012).
811
812 To generate additional sets of chromatin states based on ChIP-seq, ATAC-seq and
813 DNA methylation data, ATAC-seq open chromatin and DNA methylation status
814 were binarised. Specifically, ATAC-seq peaks (presence/absence) and whole-
815 genome CpG methylation status (hypermethylation/hypomethylation based on a
816 threshold of 60% methylation) were binarised across 200bp windows of the
817 genome.
818
819 These binarised 200bp ChIP-seq, ATAC-seq and DNA methylation maps were
820 combined and used to generate 3 sets of chromatin states derived from ChIP and
821 DNA methylation data (ChIP+Meth), ChIP and ATAC-seq data (ChIP+ATAC) or
822 ChIP, ATAC-seq and DNA methylation data (ChIP+ATAC+Meth) using the
823 Learnmodel ChromHMM function (Figure 3A and Supplementary Figure 3A-B) .
824 As suggested by (Ernst et al., 2011), after evaluating models with up to 20
825 chromatin states, a 15 state model was chosen based on the resolution provided
826 by identified states
827
828 **4.6 *ADCY5* Capture C analysis and ATAC-seq in EndoC-ßH1**

829 Next-generation Capture-C was performed in order to map physical chromatin
830 interactions with the *ADCY5* promoter in EndoC-ßH1 (RRID:CVCL_L909) cell
831 lines (n = 3) (see protocol in Methods in (Davies et al., 2016)).
832 In brief, chromatin conformation capture (3C) libraries were generated by
833 formaldehyde fixation prior to DpnII restriction enzyme digestion and
834 subsequent DNA ligation. Following cross-link reversal, DNA extraction and
835 sonication, sequencing adapters were added to sonicated fragments (~200bp).
836 Library fragments were subjected to a double capture through hybridisation
837 with a biotinylated oligonucleotide probes encompassing the *ADCY5* promoter
838 and enriched using streptavidin bead pulldown. PCR amplified fragments were
839 then sequenced through paired-end sequencing (Illumina Next-Seq). An *in silico*
840 restriction enzyme digestion was performed on the set of reconstructed
841 fragments (from paired-end sequenced reads) using the DpnII2E.pl script
842 (Davies, 2015)(https://github.com/Hughes-Genome-Group/captureC).
843 Uncaptured reads and PCR duplicates were removed prior to mapping to the
844 human genome (hg19) with Bowtie (Langmead et al., 2009)(v 1.1.0). Removal of
845 PCR duplicates and classification of fragments as 'capture' (i.e. including the
846 *ADCY5* promoter) or 'reporter' (outside the capture fragment on exclusion
847 region) was performed with the CCanalyser2.pl wrapper (Davies,
848 2015)(https://github.com/Hughes-Genome-Group/captureC). Unique mapped
849 interactions were normalized to the total number of *cis* interactions (i.e. same
850 chromosome) per 100,000 interactions. Significant chromatin interactions were
851 determined from a rank-sum test implemented in the program peakC (de Wit
852 and Geeven, 2017)(https://github.com/deWitLab/peakC). Specifically, we
853 evaluated interactions involving all SNPs in high LD (r2 > 0.8) with the lead
854 rs11708067. The lead variant (rs11708067) was in high LD with 15 SNPs
855 (mapping to 12 DpnII fragments) that spanned a region of 47kb. We applied the
856 Benjamini-Hochberg correction to control the false discovery rate for the set of
857 p-values corresponding to each restriction fragment within the 47kb region at
858 the ADCY5 locus.
859

860 In addition, ATAC-seq was performed in 50,000 cells of EndoC-ßH1 cell lines
861 (n=3) and the data was analysed in the same way as described above for human
862 islet samples.
863
864 Endo-βH1 cells were obtained from Endocells and have been previously
865 authenticated (Ravassard et al., 2011). In addition, the cell line was tested and
866 found negative for mycoplasma contamination.
867
868 **4.7 Overlaying generated epigenomic datasets generated here with other**
869 **genomic regulatory regions**
870
871 CpG sites and/or hypomethylated regulatory regions identified from the WGBS
872 and/or 450k array data were overlapped with existing islet chromatin state
873 maps (Parker et al., 2013), islet transcription factor binding sites (*FOXA2, MAFB,*
874 *NKX2.2, NKX6.1, PDX1*), T2D-associated islet dDMRs (Dayeh et al., 2014) and
875 eQTLs (van de Bunt et al., 2015). Similarly, ATAC-seq open chromatin peaks
876 generated here were overlapped with publicly available ATAC-seq peaks
877 (Varshney et al., 2017).
878
879 In addition, we also obtained the 850k array manifest file to determine overlap
880 of 850k array CpG sites with GWAS credible set regions
881 (https://support.illumina.com/downloads/infinium-methylationepic-v1-0-
882 product-files.html).
883
884 **4.8 Genetic datasets used in this study**
885 Credible sets from the DIAGRAM (Scott et al., 2017)(involving 26.7k cases and
886 132.5k controls of predominantly European origin, imputed to the 1000G March
887 2012 reference panel) and ENGAGE (Horikoshi et al., 2015)(including 46.7k
888 individuals, imputed to the 1000G March 2012 reference panel) consortium
889 were used to compare the ability of the 450k, 850k and WGBS methylation array
890 to interrogate T2D and FG GWAS regions.
891

892 The DIAGRAM and ENGAGE GWAS SNP summary level data was used for the
893 FGWAS analysis to determine enrichment of regulatory annotations in T2D and
894 FG GWAS signal.
895
896 Furthermore, data from (Wood et al., 2017) and (Dimas et al., 2014) were used
897 to categories T2D GWAS loci into physiological groups of insulin secretion,
898 insulin resistance or unclassified loci.
899
900 **4.9 Statistical and computational analysis**
901
902 **4.9.1 Enrichment analysis of identified regulatory annotations in other**
903 **genomic annotations**
904
905 Enrichment of hypomethylated regulatory regions (LMRs and UMRs, result
906 section 2.2.) and ATAC-seq open chromatin peaks (result section 2.3) in the
907 aforementioned genomic annotations (method section 4.6) was determined
908 through 100,000 random permutations. P-values and fold enrichment was
909 determined by comparing the true overlap results to the permuted overlap
910 results. The resulting P-values were multiple testing corrected using Bonferroni
911 correction (an adjusted P-value < 0.05 was considered significant).
912
913 **4.9.2 FGWAS enrichment analysis**
914 FGWAS (Pickrell, 2014) (version 0.3.6) applied a hierarchical model that
915 determined shared properties of loci affecting a trait. The FGWAS model used
916 SNP-based GWAS summary level data and divided the genome into windows
917 (setting "k"=5000 which represents the number of SNPs per window), which are
918 larger than the expected LD patterns in the population. The model assumed that
919 each window either contained a single SNP that affected the trait or that there
920 was no SNP in the window that influenced the trait. The model estimated the
921 prior probability of a window to contain an association and the conditional prior
922 probability that a SNP within the window was the causal variant. These prior
923 probabilities were variable, dependent on regional annotations and estimated

924 based on enrichment patterns of annotations across the genome using a Bayes
925 approach.
926
927
928
929 **4.9.2.1 FGWAS Single state analysis**
930 FGWAS was used with standard settings to determine enrichment of individual
931 islet chromatin states, LMRs, UMRs, PMDS and ATAC-seq open chromatin peaks,
932 CDS and CONS sequence in DIAGRAM (setting "cc" was applied for use with T2D-
933 case-control GWAS data) and ENGAGE GWAS SNP summary level data.
934
935 For each individual annotation, the model provided maximum likelihood
936 enrichment parameters and annotations were considered as significantly
937 enriched if the parameter estimate and 95% CI was above zero.
938
939 **4.9.2.2 FGWAS Joint model analysis**
940 To determine the maximum likelihood model the following approach suggested
941 by (Pickrell, 2014) was used for each set of chromatin states (ChIP-only,
942 ChIP+ATAC, ChIP+Meth and ChIP+ATAC+Meth), separately. In addition, CDS and
943 CONS sequenced were used as well for each set of chromatin states in the joint
944 analysis. Firstly, a model was fitted for each annotation individually to identify
945 all annotations that were significantly enriched with the trait. Secondly, the
946 annotation with the highest increase (and enrichment) in the maximum log-
947 likelihood was added to the model and the analysis was repeated with the
948 additional annotation. Thirdly, annotations were added as long as they increase
949 the maximum log-likelihood of the newly tested model. Fourthly, a 10-fold cross-
950 validation approach was used after determining a penalty parameter based on
951 the maximum likelihood of a penalised log-likelihood function to avoid
952 overfitting. Fifthly, each annotation was dropped from the model and the
953 maximum cross-validation likelihood was evaluated. If a reduced model has a
954 higher cross-validation maximum likelihood, additional annotations are dropped
955 until the model cannot be further improved. This model was described as the
956 best fitted model and used for the remaining analysis. The maximum likelihood

957 enrichment parameters and 95% CI for each annotation of the best model were
958 reported (independent of significance).
959
960 **4.9.2.3 Comparing FGWAS enrichment parameter across chromatin states**
961 Initially, similar enhancer chromatin states derived from the 4 different
962 ChromHMM analyses (ChIP-only, ChIP+ATAC, ChIP+Meth, ChIP+ATAC+Meth)
963 were compared. Similarity was determined based on shared histone chromatin
964 marks according to the chromHMM emission parameters. Further comparisons
965 between the ChIP-only and ChIP+ATAC+Meth model were performed based on
966 the reweighted FGWAS maximum variant PPA and the number of reweighted
967 99% credible set variants per T2D locus (for details regarding FGWAS PPA see
968 next section).
969
970 However, considering that the chromatin states were derived from distinct sets
971 of annotations across different analyses of ChromHMM, a direct comparison was
972 not fully possible. Hence, a nested model approach was used to further dissect
973 the contribution of open chromatin and DNA methylation to the enrichment.
974 Specifically, an FGWAS analysis was performed that combined the ChIP-only
975 chromHMM states with raw LMRs (representing DNA methylation) and ATAC-
976 seq peaks (representing open chromatin). After determining the best maximum-
977 likelihood cross-validation model (combining ChIP-only, ATAC-seq and LMR
978 states) a nested model and log-likelihood ratio test were used to determine the
979 contribution of each annotation to the model (Supplementary Figure 3D).
980
981 **4.9.3 Reweighting of variant PPA and testing of allelic imbalance**
982 The enrichment priors derived from the FGWAS maximum likelihood model
983 were used as a basis for evaluating both the significance and functional impact of
984 associated variants in GWAS regions; allowing variants that map into
985 annotations that show global enrichment to be afforded extra weight.
986
987
988 Specifically, variants at significant GWAS regions with a high FGWAS PPA (PPA
989 >=10%) and overlapping open enhancer states were prioritised for further

990 follow-up. Genome-wide significance of loci was determined based on P-values
991 ($P < 5x10^{-8}$) or a regional FGWAS PPA >= 90% (representing the sum of the PPAs
992 of all SNPs in a given region). The latter threshold is based on a recommendation
993 from (Pickrell, 2014) who observed that a regional PPA of 90% or above can be
994 used to identify sub-threshold GWAS loci.
995
996 Of the prioritised variants, only variants with at least 2 heterozygous samples
997 and ATAC_Seq read depth of at least 9 reads (minimum 5 reads for each allele)
998 were tested for allelic imbalance.
999
1000 To avoid read-mapping and reference allele bias the software WASP (van de
1001 Geijn et al., 2015)(Version 0.2261) was used to remove reads associated with
1002 mapping bias. In short, reads of the unfiltered bam file that overlapped the
1003 variant of interest were identified. For each read overlapping an SNP, the
1004 genotype of that SNP was changed to the alternative allele and the read was
1005 remapped using bwa (Li and Durbin, 2009) (version 0.5.8c). Any read that failed
1006 to realign in the same position in the genome was discarded. Ultimately, PCR
1007 duplicates were filtered using the WASP "rmdup_pe.py" script, which removed
1008 duplicated reads randomly (independent of the mapping score) to avoid any
1009 bias.
1010
1011 Allelic imbalance was determined using a binomial test as implemented in R.
1012
1013 **4.9.4 Identification of TFBS at SNPs that display allelic imbalance**
1014
1015 The tool "Fimo"(Grant et al., 2011) implemented in the "meme" software
1016 package (RRID:SCR_001783) was applied to identify TF motifs that significantly
1017 (FDR <0.05) matched the sequence overlapping a SNP variant showing allelic
1018 imbalance (20bp up and downstream).
1019
1020 **4.9.5 Overlap of regulatory regions**
1021 Overlap between genomic regulatory regions was performed using bedtools
1022 intersectBed function (Quinlan, 2014) (version 2.21.0). Summary statistics

1023 across 200bp windows were determined using bedtools mapBed function.
1024 Random permutations of regulatory regions were performed by applying the
1025 bedtools shuffleBed function.
1026
1027 **4.9.6 Statistical analysis**
1028 All statistical analysis (unless otherwise stated) was performed using R (version
1029 3.0.2) including Spearman's correlation analysis to compare the 450k and WGBS
1030 array, the KS-test to compare 450k and WGBS DNA methylation distributions,
1031 the binomial test to evaluate allelic imbalance and principal component analysis
1032 to identify batch effects in the 450k data. Significance is defined as $P<0.05$ unless
1033 otherwise stated.
1034
1035 **4.9.7 Visualisation and figure generation**
1036 All figures unless otherwise stated were generated using R (version 3.0.2)
1037 and/or ggplot2(Wickham, 2009). Figure 1E was generated using locuszoom
1038 (Pruim et al., 2010). Chromatin state CHiP-seq enrichment maps (Figure 3A,
1039 Supplementary Figure 3A-B) were generated using chromHMM (Ernst and Kellis,
1040 2012). The genome-browser views (Figure 5) were generated using the UCSC
1041 genome browser tool (Kent et al., 2002).
1042
1043 **5. Acknowledgements**
1044 We thank the High-Throughput Genomics Group at the Wellcome Trust Centre
1045 for Human Genetics (funded by Wellcome Trust grant reference 090532) for the
1046 generation of the Sequencing data.
1047 MT was supported by a Wellcome Trust Doctoral Studentship**.** MvdB was
1048 supported by a Novo Nordisk postdoctoral fellowship run in partnership with
1049 the University of Oxford. JMT is supported by Wellcome Trust Strategic Award.
1050 VN is funded by the European Union Horizon 2020 Programme (T2DSYSTEMS).
1051 SB was supported by a Royal Society Wolfson Research Merit Award
1052 (WM100023). ALG is a Wellcome Trust Senior Fellow in Basic Biomedical
1053 Science (095101/Z/10/Z and 200837/Z/16/Z) and MIM is a Wellcome Trust
1054 Senior Investigator. The research was supported by the National Institute for
1055 Health Research (NIHR) Oxford Biomedical Research Centre (BRC). This work

1056 was also supported by EU (HEALTH-F4-2007-201413), Wellcome Trust
1057 (090367, 090532, 106130, 098381 and 099673/Z/12/Z) and NIH (U01-
1058 DK105535, U01-DK085545, R01-DK098032 and R01-MH090941) grants. The
1059 views expressed are those of the author(s) and not necessarily those of the NHS,
1060 the NIHR or the Department of Health.

1061 **6. Competing interests**

1062 The authors do not have any competing interest.

1063 **7. References**
1064

1065 ARYEE, M. J., JAFFE, A. E., CORRADA-BRAVO, H., LADD-ACOSTA, C., FEINBERG, A.
1066 P., HANSEN, K. D. & IRIZARRY, R. A. 2014. Minfi: a flexible and
1067 comprehensive Bioconductor package for the analysis of Infinium DNA
1068 methylation microarrays. *Bioinformatics,* 30**,** 1363-9.
1069 BELL, J. T., PAI, A. A., PICKRELL, J. K., GAFFNEY, D. J., PIQUE-REGI, R., DEGNER, J.
1070 F., GILAD, Y. & PRITCHARD, J. K. 2011. DNA methylation patterns
1071 associate with genetic and gene expression variation in HapMap cell lines.
1072 *Genome Biol,* 12**,** R10.
1073 BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J.
1074 2013. Transposition of native chromatin for fast and sensitive epigenomic
1075 profiling of open chromatin, DNA-binding proteins and nucleosome
1076 position. *Nat Methods,* 10**,** 1213-8.
1077 BURGER, L., GAIDATZIS, D., SCHUBELER, D. & STADLER, M. B. 2013.
1078 Identification of active regulatory regions from DNA methylation data.
1079 *Nucleic Acids Res,* 41**,** e155.
1080 CARLSON, M. & MAINTAINER, B. 2015. TxDb.Hsapiens.UCSC.hg19.knownGene:
1081 Annotation package for TxDb object(s). *R package version 3.2.2*.
1082 DAVIES, J. O., TELENIUS, J. M., MCGOWAN, S. J., ROBERTS, N. A., TAYLOR, S.,
1083 HIGGS, D. R. & HUGHES, J. R. 2016. Multiplexed analysis of chromosome
1084 conformation at vastly improved sensitivity. *Nat Methods,* 13**,** 74-80.
1085 DAVIES, J. O. J. 2015. captureC. *Github.* https://github.com/Hughes-Genome-
1086 Group/captureC . *aae8046*
1087 DAYEH, T., VOLKOV, P., SALO, S., HALL, E., NILSSON, E., OLSSON, A. H.,
1088 KIRKPATRICK, C. L., WOLLHEIM, C. B., ELIASSON, L., RONN, T., BACOS, K.
1089 & LING, C. 2014. Genome-wide DNA methylation analysis of human
1090 pancreatic islets from type 2 diabetic and non-diabetic donors identifies
1091 candidate genes that influence insulin secretion. *PLoS Genet,* 10**,**
1092 e1004160.
1093 DE WIT, E. & GEEVEN, G. 2017. peakC. *Github.*
1094 https://github.com/deWitLab/peakC . *a2ac8c1*.
1095 DIMAS, A. S., LAGOU, V., BARKER, A., KNOWLES, J. W., MAGI, R., HIVERT, M. F.,
1096 BENAZZO, A., RYBIN, D., JACKSON, A. U., STRINGHAM, H. M., SONG, C.,
1097 FISCHER-ROSINSKY, A., BOESGAARD, T. W., GRARUP, N., ABBASI, F. A.,
1098 ASSIMES, T. L., HAO, K., YANG, X., LECOEUR, C., BARROSO, I.,
1099 BONNYCASTLE, L. L., BOTTCHER, Y., BUMPSTEAD, S., CHINES, P. S.,
1100 ERDOS, M. R., GRAESSLER, J., KOVACS, P., MORKEN, M. A., NARISU, N.,

PAYNE, F., STANCAKOVA, A., SWIFT, A. J., TONJES, A., BORNSTEIN, S. R., CAUCHI, S., FROGUEL, P., MEYRE, D., SCHWARZ, P. E., HARING, H. U., SMITH, U., BOEHNKE, M., BERGMAN, R. N., COLLINS, F. S., MOHLKE, K. L., TUOMILEHTO, J., QUERTEMOUS, T., LIND, L., HANSEN, T., PEDERSEN, O., WALKER, M., PFEIFFER, A. F., SPRANGER, J., STUMVOLL, M., MEIGS, J. B., WAREHAM, N. J., KUUSISTO, J., LAAKSO, M., LANGENBERG, C., DUPUIS, J., WATANABE, R. M., FLOREZ, J. C., INGELSSON, E., MCCARTHY, M. I., PROKOPENKO, I. & INVESTIGATORS, M. 2014. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes,* 63**,** 2158-71.

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

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

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

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

ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R., COYNE, M., KU, M., DURHAM, T., KELLIS, M. & BERNSTEIN, B. E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473**,** 43-9.

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

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

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

GAULTON, K. J., FERREIRA, T., LEE, Y., RAIMONDO, A., MAGI, R., RESCHEN, M. E., MAHAJAN, A., LOCKE, A., RAYNER, N. W., ROBERTSON, N., SCOTT, R. A., PROKOPENKO, I., SCOTT, L. J., GREEN, T., SPARSO, T., THUILLIER, D., YENGO, L., GRALLERT, H., WAHL, S., FRANBERG, M., STRAWBRIDGE, R. J.,

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

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

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

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

1198 GUO, S., DIEP, D., PLONGTHONGKUM, N., FUNG, H. L., ZHANG, K. & ZHANG, K.
1199 2017. Identification of methylation haplotype blocks aids in
1200 deconvolution of heterogeneous tissue samples and tumor tissue-of-
1201 origin mapping from plasma DNA. *Nat Genet,* 49**,** 635-642.
1202 GUTIERREZ-ARCELUS, M., LAPPALAINEN, T., MONTGOMERY, S. B., BUIL, A.,
1203 ONGEN, H., YUROVSKY, A., BRYOIS, J., GIGER, T., ROMANO, L., PLANCHON,
1204 A., FALCONNET, E., BIELSER, D., GAGNEBIN, M., PADIOLEAU, I., BOREL, C.,
1205 LETOURNEAU, A., MAKRYTHANASIS, P., GUIPPONI, M., GEHRIG, C.,
1206 ANTONARAKIS, S. E. & DERMITZAKIS, E. T. 2013. Passive and active DNA
1207 methylation and the interplay with genetic variation in gene regulation.
1208 *Elife,* 2**,** e00523.
1209 HANNON, E., SPIERS, H., VIANA, J., PIDSLEY, R., BURRAGE, J., MURPHY, T. M.,
1210 TROAKES, C., TURECKI, G., O'DONOVAN, M. C., SCHALKWYK, L. C., BRAY,
1211 N. J. & MILL, J. 2016. Methylation QTLs in the developing brain and their
1212 enrichment in schizophrenia risk loci. *Nat Neurosci,* 19**,** 48-54.
1213 HAY, D., HUGHES, J. R., BABBS, C., DAVIES, J. O. J., GRAHAM, B. J., HANSSEN, L.,
1214 KASSOUF, M. T., MARIEKE OUDELAAR, A. M., SHARPE, J. A., SUCIU, M. C.,
1215 TELENIUS, J., WILLIAMS, R., RODE, C., LI, P. S., PENNACCHIO, L. A.,
1216 SLOANE-STANLEY, J. A., AYYUB, H., BUTLER, S., SAUKA-SPENGLER, T.,
1217 GIBBONS, R. J., SMITH, A. J. H., WOOD, W. G. & HIGGS, D. R. 2016. Genetic
1218 dissection of the alpha-globin super-enhancer in vivo. *Nat Genet,* 48**,** 895-
1219 903.
1220 HODSON, D. J., MITCHELL, R. K., MARSELLI, L., PULLEN, T. J., GIMENO BRIAS, S.,
1221 SEMPLICI, F., EVERETT, K. L., COOPER, D. M., BUGLIANI, M., MARCHETTI,
1222 P., LAVALLARD, V., BOSCO, D., PIEMONTI, L., JOHNSON, P. R., HUGHES, S.
1223 J., LI, D., LI, W. H., SHAPIRO, A. M. & RUTTER, G. A. 2014. ADCY5 couples
1224 glucose to insulin secretion in human islets. *Diabetes,* 63**,** 3009-21.
1225 HORIKOSHI, M., MGI, R., VAN DE BUNT, M., SURAKKA, I., SARIN, A. P., MAHAJAN,
1226 A., MARULLO, L., THORLEIFSSON, G., HGG, S., HOTTENGA, J. J.,
1227 LADENVALL, C., RIED, J. S., WINKLER, T. W., WILLEMS, S. M.,
1228 PERVJAKOVA, N., ESKO, T., BEEKMAN, M., NELSON, C. P., WILLENBORG,
1229 C., WILTSHIRE, S., FERREIRA, T., FERNANDEZ, J., GAULTON, K. J.,
1230 STEINTHORSDOTTIR, V., HAMSTEN, A., MAGNUSSON, P. K., WILLEMSEN,
1231 G., MILANESCHI, Y., ROBERTSON, N. R., GROVES, C. J., BENNETT, A. J.,
1232 LEHTIMKI, T., VIIKARI, J. S., RUNG, J., LYSSENKO, V., PEROLA, M., HEID, I.
1233 M., HERDER, C., GRALLERT, H., MULLER-NURASYID, M., RODEN, M.,
1234 HYPPONEN, E., ISAACS, A., VAN LEEUWEN, E. M., KARSSEN, L. C.,
1235 MIHAILOV, E., HOUWING-DUISTERMAAT, J. J., DE CRAEN, A. J., DEELEN, J.,
1236 HAVULINNA, A. S., BLADES, M., HENGSTENBERG, C., ERDMANN, J.,
1237 SCHUNKERT, H., KAPRIO, J., TOBIN, M. D., SAMANI, N. J., LIND, L.,
1238 SALOMAA, V., LINDGREN, C. M., SLAGBOOM, P. E., METSPALU, A., VAN
1239 DUIJN, C. M., ERIKSSON, J. G., PETERS, A., GIEGER, C., JULA, A., GROOP, L.,
1240 RAITAKARI, O. T., POWER, C., PENNINX, B. W., DE GEUS, E., SMIT, J. H.,
1241 BOOMSMA, D. I., PEDERSEN, N. L., INGELSSON, E., THORSTEINSDOTTIR,
1242 U., STEFANSSON, K., RIPATTI, S., PROKOPENKO, I., MCCARTHY, M. I.,
1243 MORRIS, A. P. & CONSORTIUM, E. 2015. Discovery and Fine-Mapping of
1244 Glycaemic and Obesity-Related Trait Loci Using High-Density Imputation.
1245 *PLoS Genet,* 11**,** e1005230.

1246 JOHNSON, W. E., LI, C. & RABINOVIC, A. 2007. Adjusting batch effects in
1247 microarray expression data using empirical Bayes methods. *Biostatistics,*
1248 8**,** 118-27.
1249 KATO, N., LOH, M., TAKEUCHI, F., VERWEIJ, N., WANG, X., ZHANG, W., KELLY, T.
1250 N., SALEHEEN, D., LEHNE, B., MATEO LEACH, I., DRONG, A. W., ABBOTT, J.,
1251 WAHL, S., TAN, S. T., SCOTT, W. R., CAMPANELLA, G., CHADEAU-HYAM, M.,
1252 AFZAL, U., AHLUWALIA, T. S., BONDER, M. J., CHEN, P., DEHGHAN, A.,
1253 EDWARDS, T. L., ESKO, T., GO, M. J., HARRIS, S. E., HARTIALA, J., KASELA,
1254 S., KASTURIRATNE, A., KHOR, C. C., KLEBER, M. E., LI, H., MOK, Z. Y.,
1255 NAKATOCHI, M., SAPARI, N. S., SAXENA, R., STEWART, A. F., STOLK, L.,
1256 TABARA, Y., TEH, A. L., WU, Y., WU, J. Y., ZHANG, Y., AITS, I., DA SILVA
1257 COUTO ALVES, A., DAS, S., DORAJOO, R., HOPEWELL, J. C., KIM, Y. K.,
1258 KOIVULA, R. W., LUAN, J., LYYTIKAINEN, L. P., NGUYEN, Q. N., PEREIRA, M.
1259 A., POSTMUS, I., RAITAKARI, O. T., SCANNELL BRYAN, M., SCOTT, R. A.,
1260 SORICE, R., TRAGANTE, V., TRAGLIA, M., WHITE, J., YAMAMOTO, K.,
1261 ZHANG, Y., ADAIR, L. S., AHMED, A., AKIYAMA, K., ASIF, R., AUNG, T.,
1262 BARROSO, I., BJONNES, A., BRAUN, T. R., CAI, H., CHANG, L. C., CHEN, C. H.,
1263 CHENG, C. Y., CHONG, Y. S., COLLINS, R., COURTNEY, R., DAVIES, G.,
1264 DELGADO, G., DO, L. D., DOEVENDANS, P. A., GANSEVOORT, R. T., GAO, Y.
1265 T., GRAMMER, T. B., GRARUP, N., GREWAL, J., GU, D., WANDER, G. S.,
1266 HARTIKAINEN, A. L., HAZEN, S. L., HE, J., HENG, C. K., HIXSON, J. E.,
1267 HOFMAN, A., HSU, C., HUANG, W., HUSEMOEN, L. L., HWANG, J. Y., et al.
1268 2015. Trans-ancestry genome-wide association study identifies 12
1269 genetic loci influencing blood pressure and implicates a role for DNA
1270 methylation. *Nat Genet,* 47**,** 1282-93.
1271 KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER,
1272 A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC.
1273 *Genome Res,* 12**,** 996-1006.
1274 KRUEGER, F. & ANDREWS, S. R. 2011. Bismark: a flexible aligner and methylation
1275 caller for Bisulfite-Seq applications. *Bioinformatics,* 27**,** 1571-2.
1276 LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and
1277 memory-efficient alignment of short DNA sequences to the human
1278 genome. *Genome Biol,* 10**,** R25.
1279 LEEK, J. T., JOHNSON, W. E., PARKER, H. S., JAFFE, A. E. & STOREY, J. D. sva:
1280 Surrogate Variable Analysis. *R package version 3.8.0.*
1281 LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-
1282 Wheeler transform. *Bioinformatics,* 25**,** 1754-60.
1283 LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH,
1284 G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S.
1285 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,*
1286 25**,** 2078-9.
1287 LINDBLAD-TOH, K., GARBER, M., ZUK, O., LIN, M. F., PARKER, B. J., WASHIETL, S.,
1288 KHERADPOUR, P., ERNST, J., JORDAN, G., MAUCELI, E., WARD, L. D., LOWE,
1289 C. B., HOLLOWAY, A. K., CLAMP, M., GNERRE, S., ALFOLDI, J., BEAL, K.,
1290 CHANG, J., CLAWSON, H., CUFF, J., DI PALMA, F., FITZGERALD, S., FLICEK,
1291 P., GUTTMAN, M., HUBISZ, M. J., JAFFE, D. B., JUNGREIS, I., KENT, W. J.,
1292 KOSTKA, D., LARA, M., MARTINS, A. L., MASSINGHAM, T., MOLTKE, I.,
1293 RANEY, B. J., RASMUSSEN, M. D., ROBINSON, J., STARK, A., VILELLA, A. J.,
1294 WEN, J., XIE, X., ZODY, M. C., BROAD INSTITUTE SEQUENCING, P., WHOLE

GENOME ASSEMBLY, T., BALDWIN, J., BLOOM, T., CHIN, C. W., HEIMAN, D., NICOL, R., NUSBAUM, C., YOUNG, S., WILKINSON, J., WORLEY, K. C., KOVAR, C. L., MUZNY, D. M., GIBBS, R. A., BAYLOR COLLEGE OF MEDICINE HUMAN GENOME SEQUENCING CENTER SEQUENCING, T., CREE, A., DIHN, H. H., FOWLER, G., JHANGIANI, S., JOSHI, V., LEE, S., LEWIS, L. R., NAZARETH, L. V., OKWUONU, G., SANTIBANEZ, J., WARREN, W. C., MARDIS, E. R., WEINSTOCK, G. M., WILSON, R. K., GENOME INSTITUTE AT WASHINGTON, U., DELEHAUNTY, K., DOOLING, D., FRONIK, C., FULTON, L., FULTON, B., GRAVES, T., MINX, P., SODERGREN, E., BIRNEY, E., MARGULIES, E. H., HERRERO, J., GREEN, E. D., HAUSSLER, D., SIEPEL, A., GOLDMAN, N., POLLARD, K. S., PEDERSEN, J. S., LANDER, E. S. & KELLIS, M. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature,* 478**,** 476-82.

LOWE, R., GEMMA, C., BEYAN, H., HAWA, M. I., BAZEOS, A., LESLIE, R. D., MONTPETIT, A., RAKYAN, V. K. & RAMAGOPALAN, S. V. 2013. Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics,* 8**,** 445-54.

MAGOC, T. & SALZBERG, S. L. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics,* 27**,** 2957-63.

MAHAJAN, A., GO, M. J., ZHANG, W., BELOW, J. E., GAULTON, K. J., FERREIRA, T., HORIKOSHI, M., JOHNSON, A. D., NG, M. C., PROKOPENKO, I., SALEHEEN, D., WANG, X., ZEGGINI, E., ABECASIS, G. R., ADAIR, L. S., ALMGREN, P., ATALAY, M., AUNG, T., BALDASSARRE, D., BALKAU, B., BAO, Y., BARNETT, A. H., BARROSO, I., BASIT, A., BEEN, L. F., BEILBY, J., BELL, G. I., BENEDIKTSSON, R., BERGMAN, R. N., BOEHM, B. O., BOERWINKLE, E., BONNYCASTLE, L. L., BURTT, N., CAI, Q., CAMPBELL, H., CAREY, J., CAUCHI, S., CAULFIELD, M., CHAN, J. C., CHANG, L. C., CHANG, T. J., CHANG, Y. C., CHARPENTIER, G., CHEN, C. H., CHEN, H., CHEN, Y. T., CHIA, K. S., CHIDAMBARAM, M., CHINES, P. S., CHO, N. H., CHO, Y. M., CHUANG, L. M., COLLINS, F. S., CORNELIS, M. C., COUPER, D. J., CRENSHAW, A. T., VAN DAM, R. M., DANESH, J., DAS, D., DE FAIRE, U., DEDOUSSIS, G., DELOUKAS, P., DIMAS, A. S., DINA, C., DONEY, A. S., DONNELLY, P. J., DORKHAN, M., VAN DUIJN, C., DUPUIS, J., EDKINS, S., ELLIOTT, P., EMILSSON, V., ERBEL, R., ERIKSSON, J. G., ESCOBEDO, J., ESKO, T., EURY, E., FLOREZ, J. C., FONTANILLAS, P., FOROUHI, N. G., FORSEN, T., FOX, C., FRASER, R. M., FRAYLING, T. M., FROGUEL, P., FROSSARD, P., GAO, Y., GERTOW, K., GIEGER, C., GIGANTE, B., GRALLERT, H., GRANT, G. B., GRROP, L. C., GROVES, C. J., GRUNDBERG, E., GUIDUCCI, C., HAMSTEN, A., HAN, B. G., HARA, K., HASSANALI, N., et al. 2014. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet,* 46**,** 234-44.

MALLER, J. B., MCVEAN, G., BYRNES, J., VUKCEVIC, D., PALIN, K., SU, Z., HOWSON, J. M., AUTON, A., MYERS, S., MORRIS, A., PIRINEN, M., BROWN, M. A., BURTON, P. R., CAULFIELD, M. J., COMPSTON, A., FARRALL, M., HALL, A. S., HATTERSLEY, A. T., HILL, A. V., MATHEW, C. G., PEMBREY, M., SATSANGI, J., STRATTON, M. R., WORTHINGTON, J., CRADDOCK, N., HURLES, M., OUWEHAND, W., PARKES, M., RAHMAN, N., DUNCANSON, A., TODD, J. A., KWIATKOWSKI, D. P., SAMANI, N. J., GOUGH, S. C., MCCARTHY, M. I.,

DELOUKAS, P. & DONNELLY, P. 2012. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet,* 44**,** 1294-301.

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

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

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

insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet,* 44**,** 981-90.

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

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

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

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

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

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

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

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

ROADMAP EPIGENOMICS, C., KUNDAJE, A., MEULEMAN, W., ERNST, J., BILENKY, M., YEN, A., HERAVI-MOUSSAVI, A., KHERADPOUR, P., ZHANG, Z., WANG, J., ZILLER, M. J., AMIN, V., WHITAKER, J. W., SCHULTZ, M. D., WARD, L. D., SARKAR, A., QUON, G., SANDSTROM, R. S., EATON, M. L., WU, Y. C., PFENNING, A. R., WANG, X., CLAUSSNITZER, M., LIU, Y., COARFA, C., HARRIS, R. A., SHORESH, N., EPSTEIN, C. B., GJONESKA, E., LEUNG, D., XIE, W., HAWKINS, R. D., LISTER, R., HONG, C., GASCARD, P., MUNGALL, A. J., MOORE, R., CHUAH, E., TAM, A., CANFIELD, T. K., HANSEN, R. S., KAUL, R., SABO, P. J., BANSAL, M. S., CARLES, A., DIXON, J. R., FARH, K. H., FEIZI, S., KARLIC, R., KIM, A. R., KULKARNI, A., LI, D., LOWDON, R., ELLIOTT, G., MERCER, T. R., NEPH, S. J., ONUCHIC, V., POLAK, P., RAJAGOPAL, N., RAY, P., SALLARI, R. C., SIEBENTHALL, K. T., SINNOTT-ARMSTRONG, N. A., STEVENS, M., THURMAN, R. E., WU, J., ZHANG, B., ZHOU, X., BEAUDET, A.

1440 E., BOYER, L. A., DE JAGER, P. L., FARNHAM, P. J., FISHER, S. J., HAUSSLER,
1441 D., JONES, S. J., LI, W., MARRA, M. A., MCMANUS, M. T., SUNYAEV, S.,
1442 THOMSON, J. A., TLSTY, T. D., TSAI, L. H., WANG, W., WATERLAND, R. A.,
1443 ZHANG, M. Q., CHADWICK, L. H., BERNSTEIN, B. E., COSTELLO, J. F.,
1444 ECKER, J. R., HIRST, M., MEISSNER, A., MILOSAVLJEVIC, A., REN, B.,
1445 STAMATOYANNOPOULOS, J. A., WANG, T. & KELLIS, M. 2015. Integrative
1446 analysis of 111 reference human epigenomes. *Nature,* 518**,** 317-30.
1447 ROMAN, T. S., CANNON, M. E., VADLAMUDI, S., BUCHKOVICH, M. L., WOLFORD, B.
1448 N., WELCH, R. P., MORKEN, M. A., KWON, G. J., VARSHNEY, A., KURSAWE,
1449 R., WU, Y., JACKSON, A. U., SEQUENCING PROGRAM, N. C., ERDOS, M. R.,
1450 KUUSISTO, J., LAAKSO, M., SCOTT, L. J., BOEHNKE, M., COLLINS, F. S.,
1451 PARKER, S. C. J., STITZEL, M. L. & MOHLKE, K. L. 2017. A Type 2 Diabetes-
1452 Associated Functional Regulatory Variant in a Pancreatic Islet Enhancer
1453 at the Adcy5 Locus. *Diabetes*.
1454 SCOTT, R. A., SCOTT, L. J., MAGI, R., MARULLO, L., GAULTON, K. J., KAAKINEN, M.,
1455 PERVJAKOVA, N., PERS, T. H., JOHNSON, A. D., EICHER, J. D., JACKSON, A.
1456 U., FERREIRA, T., LEE, Y., MA, C., STEINTHORSDOTTIR, V.,
1457 THORLEIFSSON, G., QI, L., VAN ZUYDAM, N. R., MAHAJAN, A., CHEN, H.,
1458 ALMGREN, P., VOIGHT, B. F., GRALLERT, H., MULLER-NURASYID, M.,
1459 RIED, J. S., RAYNER, W. N., ROBERTSON, N., KARSSEN, L. C., VAN
1460 LEEUWEN, E. M., WILLEMS, S. M., FUCHSBERGER, C., KWAN, P.,
1461 TESLOVICH, T. M., CHANDA, P., LI, M., LU, Y., DINA, C., THUILLIER, D.,
1462 YENGO, L., JIANG, L., SPARSO, T., KESTLER, H. A., CHHEDA, H., EISELE, L.,
1463 GUSTAFSSON, S., FRANBERG, M., STRAWBRIDGE, R. J., BENEDIKTSSON,
1464 R., HREIDARSSON, A. B., KONG, A., SIGURETHSSON, G., KERRISON, N. D.,
1465 LUAN, J., LIANG, L., MEITINGER, T., RODEN, M., THORAND, B., ESKO, T.,
1466 MIHAILOV, E., FOX, C., LIU, C. T., RYBIN, D., ISOMAA, B., LYSSENKO, V.,
1467 TUOMI, T., COUPER, D. J., PANKOW, J. S., GRARUP, N., HAVE, C. T.,
1468 JORGENSEN, M. E., JORGENSEN, T., LINNEBERG, A., CORNELIS, M. C., VAN
1469 DAM, R. M., HUNTER, D. J., KRAFT, P., SUN, Q., EDKINS, S., OWEN, K. R.,
1470 PERRY, J. R., WOOD, A. R., ZEGGINI, E., TAJES-FERNANDES, J., ABECASIS, G.
1471 R., BONNYCASTLE, L. L., CHINES, P. S., STRINGHAM, H. M., KOISTINEN, H.
1472 A., KINNUNEN, L., SENNBLAD, B., MUHLEISEN, T. W., NOTHEN, M. M.,
1473 PECHLIVANIS, S., BALDASSARRE, D., GERTOW, K., HUMPHRIES, S. E.,
1474 TREMOLI, E., KLOPP, N., MEYER, J., STEINBACH, G., et al. 2017. An
1475 Expanded Genome-Wide Association Study of Type 2 Diabetes in
1476 Europeans. *Diabetes*.
1477 STADLER, M. B., MURR, R., BURGER, L., IVANEK, R., LIENERT, F., SCHOLER, A.,
1478 VAN NIMWEGEN, E., WIRBELAUER, C., OAKELEY, E. J., GAIDATZIS, D.,
1479 TIWARI, V. K. & SCHUBELER, D. 2011. DNA-binding factors shape the
1480 mouse methylome at distal regulatory regions. *Nature,* 480**,** 490-5.
1481 VAN DE BUNT, M., MANNING FOX, J. E., DAI, X., BARRETT, A., GREY, C., LI, L.,
1482 BENNETT, A. J., JOHNSON, P. R., RAJOTTE, R. V., GAULTON, K. J.,
1483 DERMITZAKIS, E. T., MACDONALD, P. E., MCCARTHY, M. I. & GLOYN, A. L.
1484 2015. Transcript Expression Data from Human Islets Links Regulatory
1485 Signals from Genome-Wide Association Studies for Type 2 Diabetes and
1486 Glycemic Traits to Their Downstream Effectors. *PLoS Genet,* 11**,**
1487 e1005694.

1488 VAN DE GEIJN, B., MCVICKER, G., GILAD, Y. & PRITCHARD, J. K. 2015. WASP:
1489 allele-specific software for robust molecular quantitative trait locus
1490 discovery. *Nat Methods,* 12**,** 1061-3.
1491 VARSHNEY, A., SCOTT, L. J., WELCH, R. P., ERDOS, M. R., CHINES, P. S., NARISU, N.,
1492 ALBANUS, R. D., ORCHARD, P., WOLFORD, B. N., KURSAWE, R.,
1493 VADLAMUDI, S., CANNON, M. E., DIDION, J. P., HENSLEY, J., KIRILUSHA, A.,
1494 PROGRAM, N. C. S., BONNYCASTLE, L. L., TAYLOR, D. L., WATANABE, R.,
1495 MOHLKE, K. L., BOEHNKE, M., COLLINS, F. S., PARKER, S. C. & STITZEL, M.
1496 L. 2017. Genetic regulatory signatures underlying islet gene expression
1497 and type 2 diabetes. *Proc Natl Acad Sci U S A,* 114**,** 2301-2306.
1498 VENTHAM, N. T., KENNEDY, N. A., ADAMS, A. T., KALLA, R., HEATH, S., O'LEARY,
1499 K. R., DRUMMOND, H., CONSORTIUM, I. B., CONSORTIUM, I. C., WILSON, D.
1500 C., GUT, I. G., NIMMO, E. R. & SATSANGI, J. 2016. Integrative epigenome-
1501 wide analysis demonstrates that DNA methylation may mediate genetic
1502 risk in inflammatory bowel disease. *Nat Commun,* 7**,** 13507.
1503 VOIGHT, B. F., SCOTT, L. J., STEINTHORSDOTTIR, V., MORRIS, A. P., DINA, C.,
1504 WELCH, R. P., ZEGGINI, E., HUTH, C., AULCHENKO, Y. S., THORLEIFSSON,
1505 G., MCCULLOCH, L. J., FERREIRA, T., GRALLERT, H., AMIN, N., WU, G.,
1506 WILLER, C. J., RAYCHAUDHURI, S., MCCARROLL, S. A., LANGENBERG, C.,
1507 HOFMANN, O. M., DUPUIS, J., QI, L., SEGRE, A. V., VAN HOEK, M.,
1508 NAVARRO, P., ARDLIE, K., BALKAU, B., BENEDIKTSSON, R., BENNETT, A. J.,
1509 BLAGIEVA, R., BOERWINKLE, E., BONNYCASTLE, L. L., BENGTSSON
1510 BOSTROM, K., BRAVENBOER, B., BUMPSTEAD, S., BURTT, N. P.,
1511 CHARPENTIER, G., CHINES, P. S., CORNELIS, M., COUPER, D. J.,
1512 CRAWFORD, G., DONEY, A. S., ELLIOTT, K. S., ELLIOTT, A. L., ERDOS, M. R.,
1513 FOX, C. S., FRANKLIN, C. S., GANSER, M., GIEGER, C., GRARUP, N., GREEN,
1514 T., GRIFFIN, S., GROVES, C. J., GUIDUCCI, C., HADJADJ, S., HASSANALI, N.,
1515 HERDER, C., ISOMAA, B., JACKSON, A. U., JOHNSON, P. R., JORGENSEN, T.,
1516 KAO, W. H., KLOPP, N., KONG, A., KRAFT, P., KUUSISTO, J., LAURITZEN, T.,
1517 LI, M., LIEVERSE, A., LINDGREN, C. M., LYSSENKO, V., MARRE, M.,
1518 MEITINGER, T., MIDTHJELL, K., MORKEN, M. A., NARISU, N., NILSSON, P.,
1519 OWEN, K. R., PAYNE, F., PERRY, J. R., PETERSEN, A. K., PLATOU, C.,
1520 PROENCA, C., PROKOPENKO, I., RATHMANN, W., RAYNER, N. W.,
1521 ROBERTSON, N. R., ROCHELEAU, G., RODEN, M., SAMPSON, M. J., SAXENA,
1522 R., SHIELDS, B. M., SHRADER, P., SIGURDSSON, G., SPARSO, T.,
1523 STRASSBURGER, K., STRINGHAM, H. M., SUN, Q., SWIFT, A. J., THORAND,
1524 B., et al. 2010. Twelve type 2 diabetes susceptibility loci identified
1525 through large-scale association analysis. *Nat Genet,* 42**,** 579-89.
1526 VOLKOV, P., BACOS, K., OFORI, J. K., ESGUERRA, J. L., ELIASSON, L., RONN, T. &
1527 LING, C. 2017. Whole-Genome Bisulfite Sequencing of Human Pancreatic
1528 Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes
1529 Pathogenesis. *Diabetes,* 66**,** 1074-1085.
1530 WANG, X., TUCKER, N. R., RIZKI, G., MILLS, R., KRIJGER, P. H., DE WIT, E.,
1531 SUBRAMANIAN, V., BARTELL, E., NGUYEN, X. X., YE, J., LEYTON-MANGE, J.,
1532 DOLMATOVA, E. V., VAN DER HARST, P., DE LAAT, W., ELLINOR, P. T.,
1533 NEWTON-CHEH, C., MILAN, D. J., KELLIS, M. & BOYER, L. A. 2016.
1534 Discovery and validation of sub-threshold genome-wide association study
1535 loci using epigenomic signatures. *Elife,* 5.

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

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

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

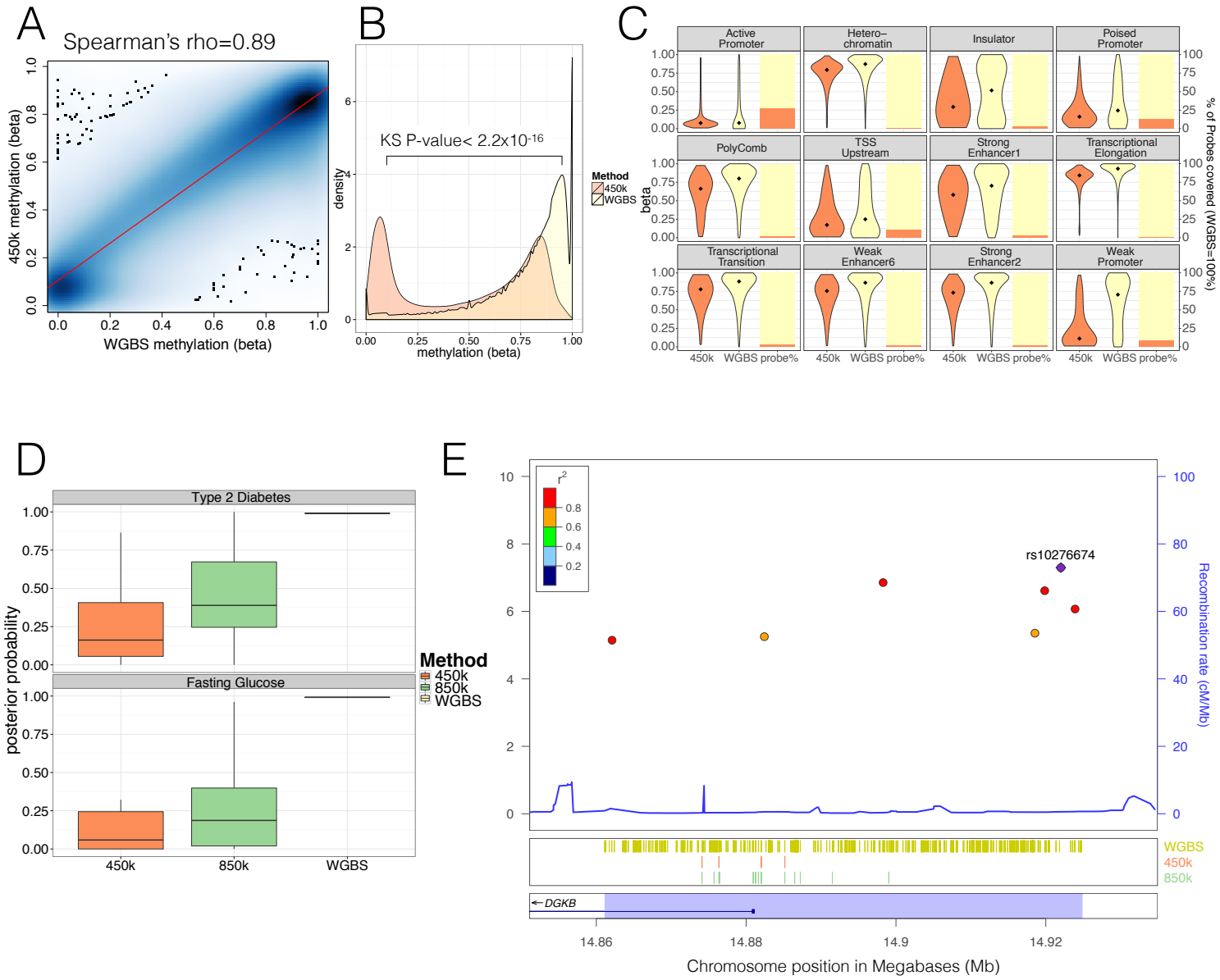

FIGURE 1. Comparison of human pancreatic islet WGBS and 450k methylation data across the genome. A) Smooth Scatter plot shows Spearman's rho correlation between the 450k array (x-axis) and WGBS (y-axis) at overlapping sites. Darker colour indicates higher density of sites. B) Comparison of the 450k array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

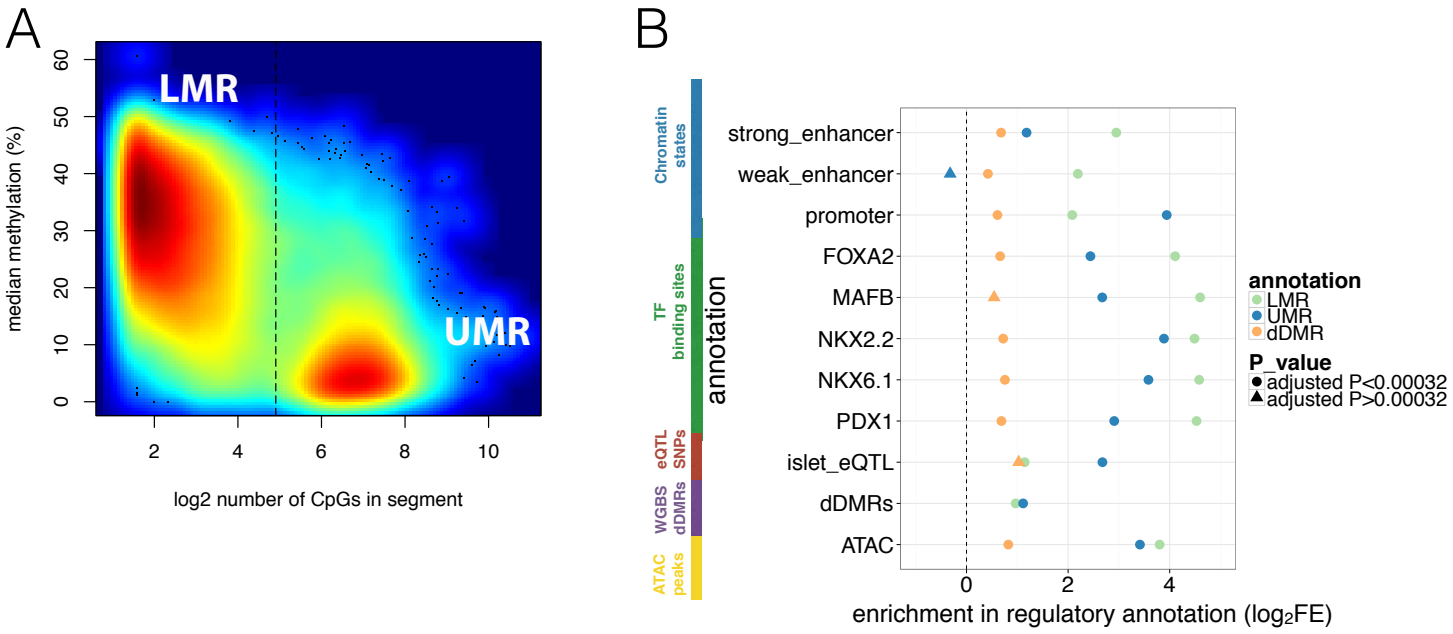

FIGURE 2. Overlap of WGBS hypomethylation and ATAC-seq open chromatin peaks with regulatory annotation . A) Methylation levels in percent (y-axis) and $\log_2$ CpG density (x-axis) of UMR and LMR regulatory regions with the dashed line indicating the CpG-number (30 CpGs) that distinguishes LMRs and UMRs.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

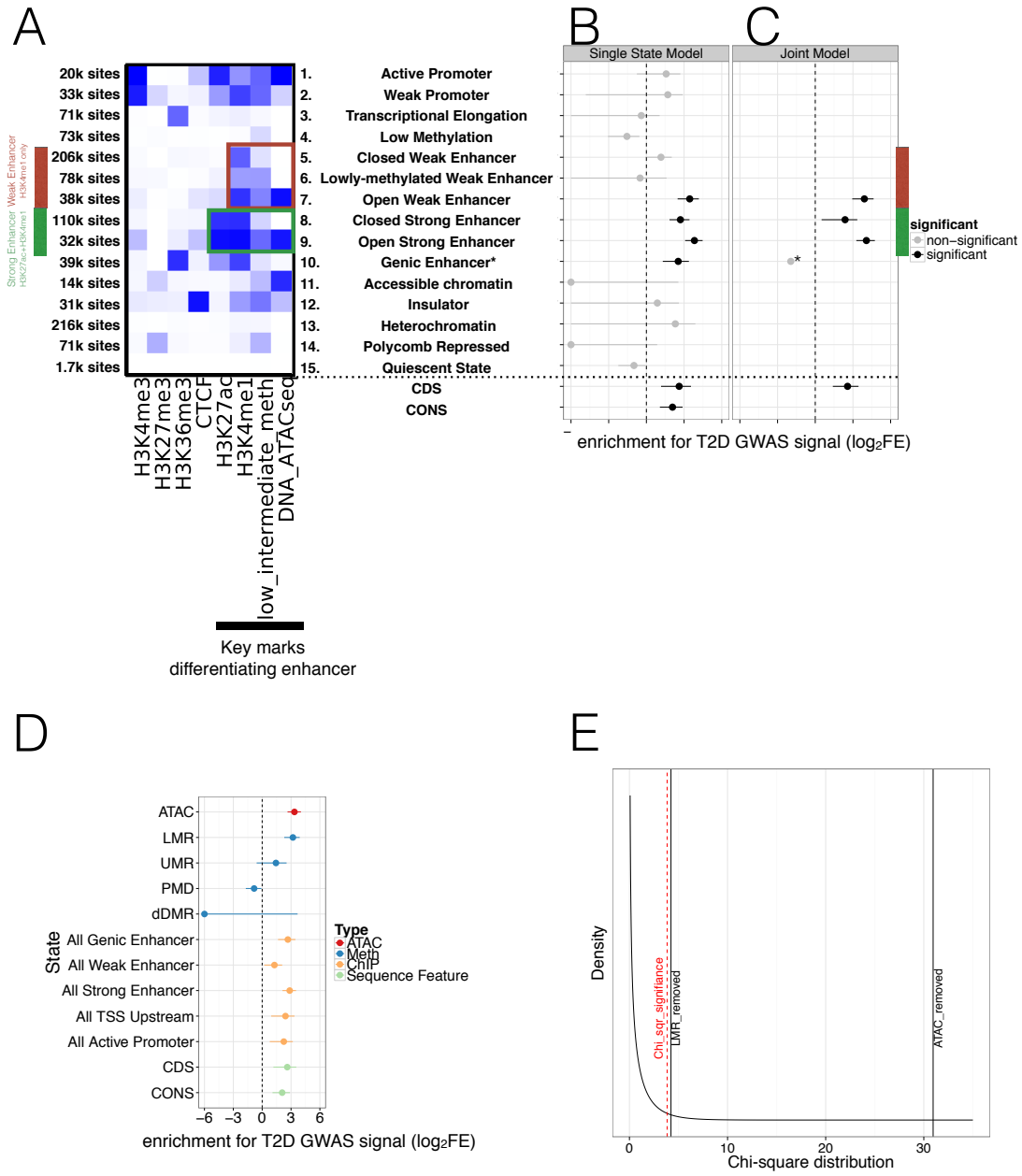

FIGURE 3. Integration of islet epigenetic data to refine chromatin regulatory states and enrichment of these states in T2D GWAS data. A) 15 chromatin states (y-axis) were derived from ChIP histone marks, DNA methylation and ATAC-seq open chromatin annotations (x-axis) using chromHMM. For each state the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

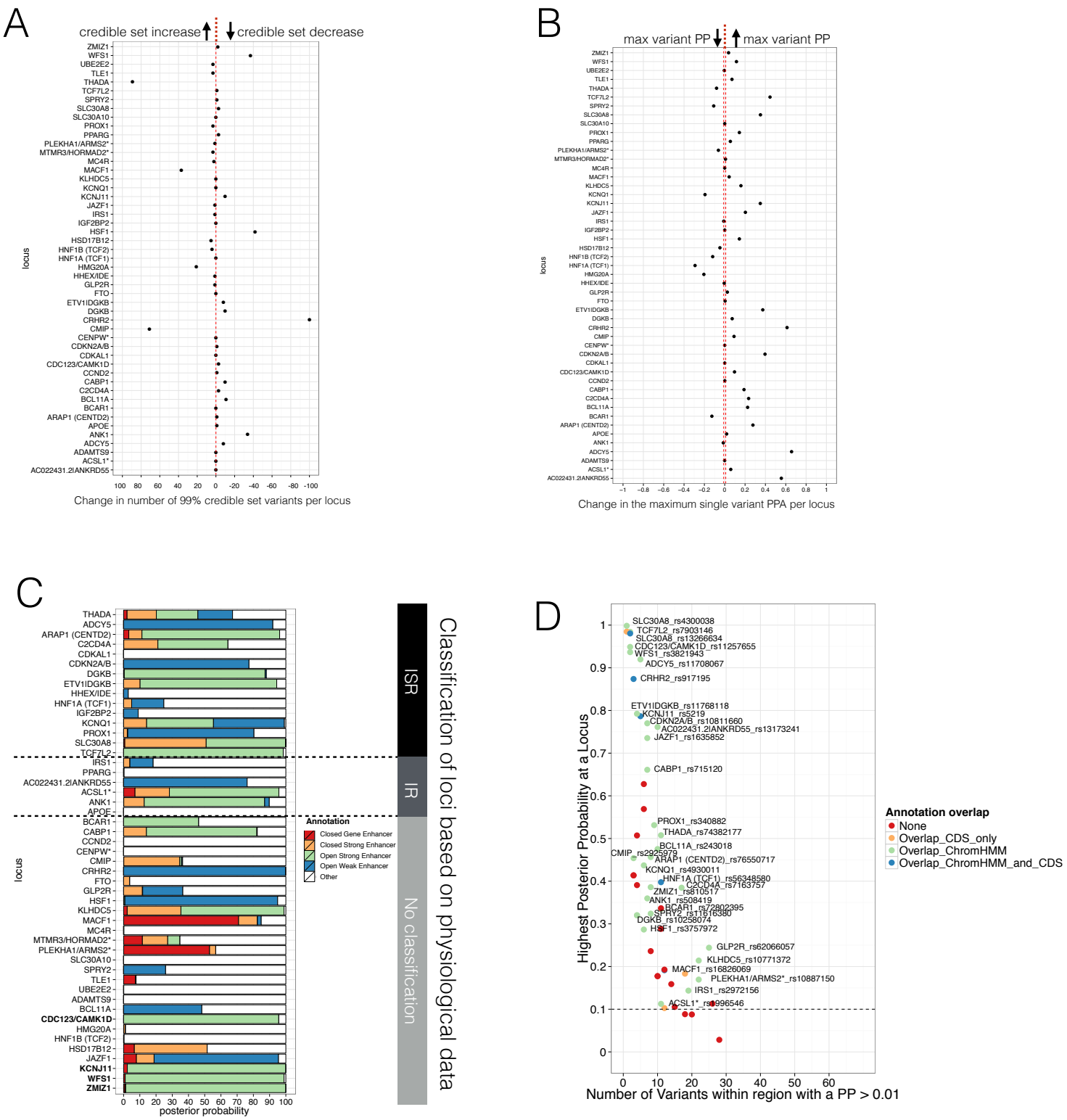

FIGURE 4. Evaluating Posterior Probabilities (PP) derived from the FGWAS maximum likelihood model at significant T2D GWAS loci. (A) Per locus the difference in the number of 99% credible set variants between ChIP+ATAC+Meth and ChIP-only model is shown (positive values indicate a reduction in the number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

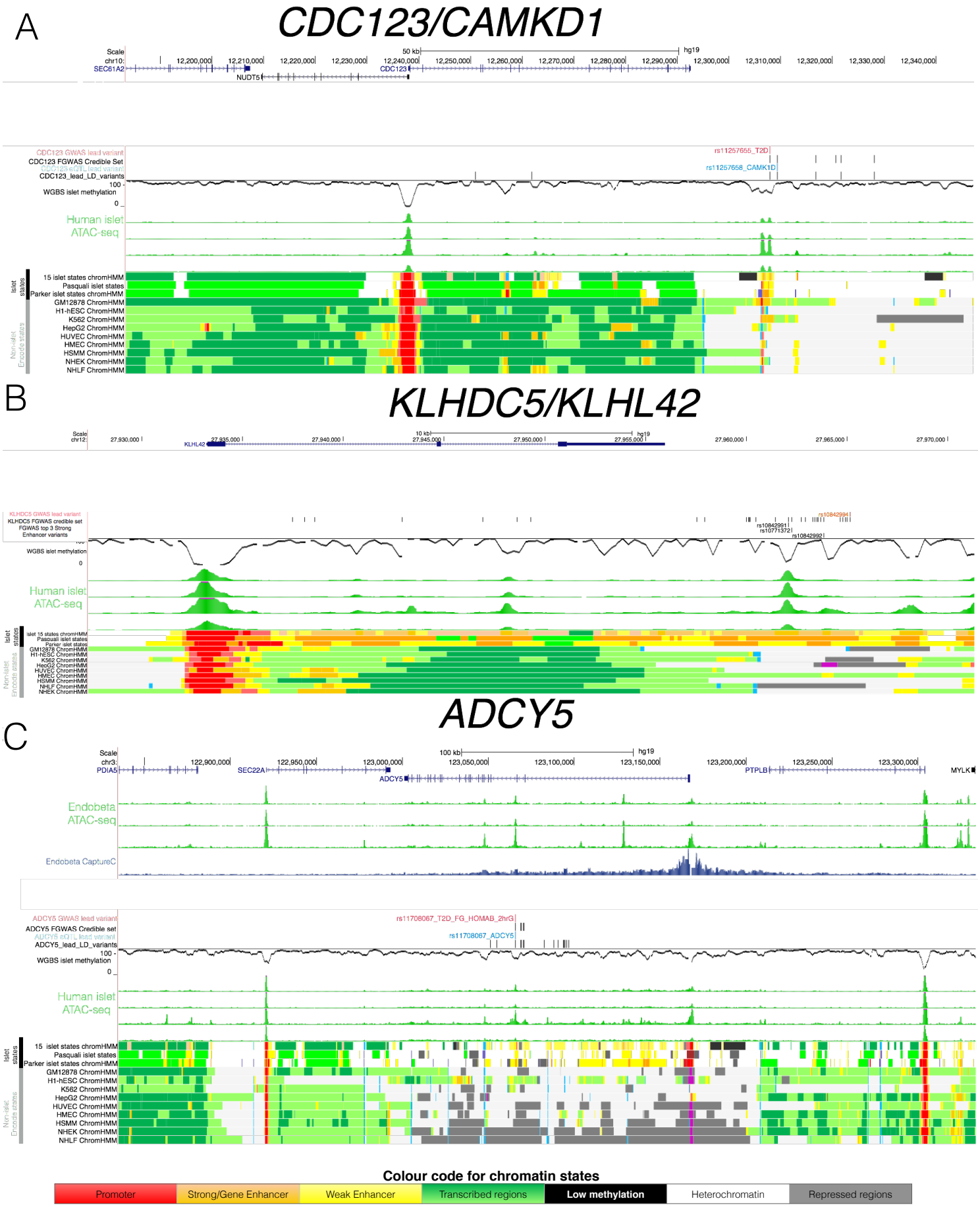

FIGURE 5. Epigenome Landscape of selected loci with allelic imbalance. For each locus A) *CDC123*, B) *KLHDC5* and C) *ADCY5* the following information is shown: 3 ATAC-seq Endoß tracks (green, top), variant level information (depending on the region GWAS lead SNP red, credible set black, eQTL blue and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

	Chromatin States	Total number of states	T2D log2FE (CI)	FG log2FE (CI)
Augmented chromatin state maps	**1. Active Promoter**	20k	1.6 (-0.8 to 2.7)	2.7 (0 to 4.1)
	2. Weak Promoter	33k	1.7 (-4.8 to 2.9)	2.7 (-0.1 to 4.2)
	3. Transcriptional Elongation	71k	-0.4 (-20 to 1.1)	-26.1 (-46.1 to 1)
	4. Low Methylation	73k	-1.5 (-3.1 to -0.6)	-1.7 (-4.2 to -0.3)
	5. Closed Weak Enhancer	206k	1.2 (-0.1 to 2)	1.7 (0 to 2.9)
	6. Lowly-methylated Weak Enhancer	78k	-0.5 (-20 to 1.6)	-26.7 (-46.7 to 1.6)
	7. Open Weak Enhancer	38k	**3.4 (2.5 to 4.2)**	3.1 (-0.6 to 4.6)
	8. Closed Strong Enhancer	110k	**2.7 (1.8 to 3.4)**	**3.3 (2 to 4.4)**
	9. Open Strong Enhancer	32k	**3.8 (3.1 to 4.5)**	**4.3 (2.8 to 5.5)**
	10. Genic Enhancer	39k	**2.5 (1.3 to 3.4)**	**2.9 (0.8 to 4.3)**
	11. Accessible chromatin	14k	-25.2 (-45.2 to 2.5)	-28.4 (-48.4 to 3.7)
	12. Insulator	31k	0.9 (-20 to 2.6)	-0.6 (-20 to 3.6)
	13. Heterochromatin	216k	2.3 (-20 to 3.9)	1.8 (-1.5 to 4)
	14. Polycomb Repressed	71k	-25.5 (-45.5 to 0.9)	-33.2 (-53.2 to 1.5)
	15. Quiescent State	1.7k	-1 (-2.2 to -0.1)	-28.6 (-48.6 to -0.6)
CDS/CONS	**CDS**	NA	**2.6 (1.2 to 3.5)**	2.7 (-0.2 to 4.3)
	CONS	NA	**2.1 (1.1 to 2.9)**	**1.9 (0.2 to 3.2)**
Parker 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	0.92	0.29 (38 A VS 94 G alleles)	1.2E-06	risk allele A closed
rs11257655	CDC123	4.0E-08	0.95	0.39 (278 C VS 435 T alleles)	4.5E-09	risk allele T open
rs10842991	KLHDC5	7.3E-07	0.13	0.64 (75 C VS 43 T alleles)	4.1E-03	risk C allele open

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