CRISPR Screening Uncovers a Long-Range Enhancer for *ONECUT1* in Pancreatic Differentiation and Links a Diabetes Risk Variant

Samuel Joseph Kaplan[1,2], Wilfred Wong[1,3], Jielin Yan[2,4], Julian Pulecio[2], Hyein S. Cho[2], Qianzi Li[1,3], Jiahui Zhao[1], Jayanti Leslie-Iyer[5], Jonathan Kazakov[5], Dylan Murphy[1], Renhe Luo[2,4], Kushal K. Dey[3], Effie Apostolou[6], Christina S. Leslie[3], Danwei Huangfu[2,*]

1. Weill Cornell Graduate School of Medical Sciences, Weill Cornell Medical College, New York, NY 10065, USA.
2. Developmental Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA.
3. Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA.
4. Louis V. Gerstner Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA.
5. Intern, Developmental Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA.
6. Meyer Cancer Center, Division of Neuro-Oncology, Department of Neurology, Sandra and Edward Meyer Cancer Center, New York-Presbyterian Hospital/Weill Cornell Medicine, New York, NY 10065, USA.

* Correspondence: huangfud@mskcc.org

Abstract

Functional enhancer annotation is a valuable first step for understanding tissue-specific transcriptional regulation and prioritizing disease-associated non-coding variants for investigation. However, unbiased enhancer discovery in physiologically relevant contexts remains a major challenge. To discover regulatory elements pertinent to diabetes, we conducted a CRISPR interference (CRISPRi) screen in the human pluripotent stem cell (hPSC) pancreatic differentiation system. Among the enhancers uncovered, we focused on a long-range enhancer ~664 kb from the *ONECUT1* promoter, as coding mutations in *ONECUT1* cause pancreatic hypoplasia and neonatal diabetes. Homozygous enhancer deletion in hPSCs was associated with a near-complete loss of *ONECUT1* gene expression and compromised pancreatic differentiation. This enhancer contains a confidently fine-mapped type 2 diabetes (T2D) associated variant (rs528350911) which disrupts a GATA motif. Introduction of the risk variant into hPSCs revealed substantially reduced binding of key pancreatic transcription factors (GATA4, GATA6 and FOXA2) on the edited allele, accompanied by a slight reduction of *ONECUT1* transcription, supporting a causal role for this risk variant in metabolic disease. This work expands our knowledge about transcriptional regulation in pancreatic development through the characterization of a long-range enhancer and highlights the utility of enhancer discovery in disease-relevant settings for understanding monogenic and complex disease.

Main

A spectrum of genetic causality often influences the onset and varying severity of symptoms for a disease[1]. This concept is exemplified in metabolic conditions like diabetes, where a single nucleotide alteration in a protein coding region can cause severe neonatal diabetes mellitus (NDM), while the aggregation of polygenic effects is thought to confer susceptibility to type 2 diabetes (T2D)[2,3]. However, most genetic variants have moderate effects and are in unannotated non-coding genomic regions, with the affected genes typically unknown[3,4]. Therefore, there is a pressing need to understand the functional impact of disease-associated variants uncovered by genome-wide association studies (GWAS) and manage the burgeoning set of variants of unknown significance identified through the widespread adoption of whole genome sequencing.

It is hypothesized that many genetic variants affect enhancers[5], and alterations in their sequences can affect gene expression, contributing to both Mendelian and complex disease traits. A clear illustration of this impact is shown by a mutation in a *SHH* enhancer, ~1 Mb from the gene promoter, causing limb malformation[6]. This represents one of a handful of clinical genetics examples where enhancer mutations have a pronounced pathological impact[7]. While the specific functional consequences of individual sequence alterations in common complex diseases are generally unknown, a significant proportion of disease-associated genomic variations identified through GWAS are located within putative enhancers[5]. Together, these findings have contributed to the inception of the enhanceropathy concept[7]. However, diagnosis of suspected enhanceropathies remains difficult due to the lack of functional annotation of most enhancers. While putative enhancers can be identified based on genomic features like chromatin accessibility, achieving scalable functional characterization in tissue-relevant settings poses a

60 considerable challenge. One obstacle is the variable distance that enhancers can have from their
61 target genes. In our recent interrogation of developmental enhancers, we found that 41% of the
62 enhancer identified were located >100 kb away from the target gene promoter[8], yet practical
63 considerations often lead many functional enhancer screens to confine their search to regions
64 within 100 kb of the target gene promoter[9]. Consequently, the overall prevalence of long-
65 distance enhancer-promoter regulations remains unclear despite a growing set of long-range
66 enhancers characterized through genetic studies[10].
67
68 Human pluripotent stem cell (hPSC) differentiation represents a powerful system for disease
69 modeling. The integration of genetic perturbation and genomic approaches has proven
70 instrumental for investigating the cascade of gene and enhancer activation during development,
71 holding promise for understanding long-term health implications. We and others have shown
72 neonatal diabetes (NDM) gene knockout in hPSCs recapitulates key aspects of human genetic
73 conditions[11]. In addition, the hPSC platform has been used to study the impact of clinically
74 observed recessive mutations in a *PTF1A* enhancer, shedding light on its role in pancreatic
75 agenesis and NDM[12,13]. However, many cases of NDM lack a known genetic cause even with
76 whole exome sequencing, underscoring the importance of identifying developmental
77 enhancers[14]. Significant progress in biochemical characterization has narrowed the search space
78 for enhancers, but many putative enhancers have not been functionally interrogated.
79 Furthermore, the extensive epigenetic rewiring during development further swells the number of
80 potential enhancers, adding complexity to the characterization task[15,16]. This challenge is
81 becoming more tractable with the use of CRISPR-Cas9 and dCas9-based CRISPR interference
82 (CRISPRi) tools, enabling the simultaneous perturbation of genes or enhancers en masse[9]. We

83 and others have applied CRISPR-Cas9 in hPSCs to uncover genes important for gastrulation and
84 pancreatic development[17-22]. Recent CRISPRi screens have further identified enhancers involved
85 in hPSC endoderm and cardiac differentiation[8,23,24]. These advances motivated us to leverage our
86 epigenetic characterization of the hPSC pancreatic differentiation system[25] to uncover enhancers
87 with potential roles in diabetes.
88
89 The hPSC pancreatic differentiation system directs cells progressively through the definitive
90 endoderm (DE), gut tube (GT), and posterior foregut (PFG, also known as PP1) stages that
91 model embryonic development in humans[26]. Building on our recent success in using a cell
92 identity gene as a readout for enhancer discovery in cell state transitions[8], we utilized the
93 developmental initiation of diabetes-associated gene *PDX1* expression at the PFG stage as a
94 readout for the onset of pancreatic differentiation[11,17]. We then selected candidate transcription
95 factors (TFs) likely to influence PDX1+ identity commitment and hypothesized that repressing
96 enhancers of these TFs would affect PDX1 expression. Our subsequent screen indeed uncovered
97 enhancers of multiple TF genes that affected PDX1 activation during pancreatic specification.
98 We focused on a distal enhancer of *ONECUT1*, located ~664 kb from the gene promoter. This
99 choice was motivated by the presence of a high-confidence fine-mapped T2D-associated single
100 nucleotide polymorphism (SNP) within this enhancer[27] and clinical genetics findings indicating
101 that coding mutations in *ONECUT1* cause NDM[28,29]. We show that *ONECUT1* enhancer deletion
102 in hPSCs resulted in an almost complete absence of *ONECUT1* transcript, protein, along with a
103 reduced proportion of cells expressing PDX1 during pancreatic differentiation. Enhancer
104 deletion had locus-wide effects within the topologically associating domain (TAD), with reduced
105 anti-sense transcription and decreased H3K27ac levels. Finally, we generated hPSC lines

106 harboring the T2D variant (rs528350911) and found the risk variant significantly impaired the
107 binding of key pancreatic transcription factors GATA4, GATA6, and FOXA2, but had a smaller
108 effect on *ONECUT1* transcription compared to enhancer knockout. Our study not only identified
109 an essential enhancer of *ONECUT1* in pancreatic development but also developed and reduced to
110 practice a framework for prioritizing disease-associated variants for investigation.
111

112 ## Results

113 **CRISPRi enhancer screen design**

114 To discover enhancers involved in pancreatic development and diabetes, we conducted a
115 CRISPRi screen in hPSC pancreatic differentiation, using PDX1 expression as a readout. PDX1
116 is a key marker for the initiation of pancreatic differentiation and is indispensable for mammalian
117 pancreatic development[30,31]. Homozygous and heterozygous mutations in *PDX1* cause NDM and
118 maturity onset diabetes of the young (MODY), respectively[32,33]. Additionally, genomic variants
119 in the *PDX1* locus are associated with elevated T2D risk[34]. To identify putative enhancer regions
120 for study, we first selected 23 candidate TFs (including PDX1) based on their likely impact on
121 PDX1+ induction and diabetes by integrating findings from murine experiments[30,31,35-53], clinical
122 genetics, and our previous CRISPR-Cas9 screens in hPSC pancreatic differentiation (Fig. 1B).
123 11 of these TFs have known roles in human metabolic disorders or NDM[29,32,54-59]. 14 TFs,
124 including GATA6, RFX6, and ONECUT1, have also been shown to affect pancreatic
125 differentiation and PDX1 expression in hPSC disease models or CRISPR-Cas9 coding screens
126 for regulators of DE and PFG specification[17,18,28,60,61].
127

128 TADs have been proposed to play a key role in enhancer function[62], so we conducted Hi-C
129 assays on PFG stage cells and used TAD information to delineate a ~2-6 Mb linear window
130 around each TF for interrogation (Sup. Table 1). Within each window, we collated all accessible
131 regions identified based on ATAC-seq peaks across the DE, GT, and PFG stages for
132 interrogation (Fig. 1C, Sup. Table 1)[25]. These accessible regions also included 758 annotated
133 promoters, facilitating the benchmarking of CRISPRi screen results against our previous
134 genome-scale CRISPR-Cas9 coding screen conducted in the same differentiation context[17].
135 Thus, we designed a library with 37,184 gRNAs targeting 6,443 regions (avg. length of 300 bp),
136 and 1100 safe-targeting[63] and 463 non-targeting gRNAs[64], and cloned the sequences into a
137 lentiviral backbone (Sup. Table 1).
138

139 **CRISPRi screen identifies enhancers and genes necessary for human pancreatic**
140 **differentiation**

141 We conducted a pooled CRISPRi screen in our hPSC pancreatic differentiation system[17], with
142 two independent infections and differentiations referred to as Screen 1 and 2 (Fig. 1A). For each
143 screen, we infected inducible dCas9-KRAB H1 hPSCs[65] with the gRNA library and induced
144 dCas9-KRAB expression. Following differentiation, we enriched for PDX1+ and PDX1- cells at
145 the PFG stage through fluorescence-activated cell sorting (FACS) (Fig. 1D, Sup. Fig. 1A).
146 Subsequently, next-generation sequencing determined gRNA abundance within the sorted
147 populations, and MAGeCK analysis[66] led us to prioritize 36 top hits causing a PDX1 decrease
148 (Fig. 1E, Sup. Fig 1B, Sup. Table 1). Both promoter (19) and non-promoter (17) hits have greater
149 sequence conservation compared to all regions investigated, as shown by the average
150 conservation scores calculated from PhyloP100[67] (Sup. Fig. 1C). Furthermore, we examined all

151 interrogated promoter regions and found the gene hits to be largely concordant with our previous
152 Cas9 coding screen conducted in a similar pancreatic differentiation context[17] (Fig. 1F). These
153 findings support the utility of CRISPRi repression for the discovery of both coding and non-
154 coding regulators of hPSC differentiation.
155
156 We next focused on non-promoter hits and examined H3K27ac, a histone mark associated with
157 active enhancers[68]. Non-promoter hit regions did not have H3K27ac peaks in undifferentiated
158 hPSCs, but gained H3K27ac at various stages during pancreatic differentiation (Fig. 1G,H),
159 indicating stage-specific enhancer activity. We further reasoned that each enhancer hit is likely to
160 correspond to a promoter hit for the regulated gene. Indeed, almost all discovered enhancers
161 were linked to a promoter hit within the linear neighborhood, resulting in a total of 15 enhancer-
162 gene pairs (Fig. 1I, Sup. Fig. 1D). This approach supports the specificity of the enhancer-gene
163 regulations. In addition, we observed that the enhancer-promoter distances ranged from 2 to 664
164 kb, with three enhancers found beyond 100 kb from the transcription start site (TSS) (Fig. 1J).
165 Thus, unbiased screening in hPSC pancreatic differentiation has uncovered many previously
166 undiscovered enhancers, expanding our understanding of gene regulation during human pancreas
167 development.
168
169 ***ONECUT1e-664kb* deletion causes allele-specific loss of *ONECUT1* expression**
170 The CRISPRi screen identified the *ONECUT1* promoter as well as a site ~664 kb from the
171 *ONECUT1* TSS, denoted as *ONECUT1e-664kb* henceforth. We observed that at the PFG stage,
172 the *ONECUT1* promoter showed significantly increased Hi-C contact with a region
173 encompassing *ONECUT1e-664kb* compared to the DE stage (Fig. 2A, 2B). This observation

174 supports 3D enhancer-promoter interaction and suggests substantial enhancer rewiring during
175 differentiation. We chose *ONECUT1e-664kb* for further investigation because *ONECUT1*
176 mutations cause pancreatic agenesis in humans[28,29] and *ONECUT1* is necessary for efficient
177 hPSC differentiation to pancreatic progenitors[28]. Consistent with the clinical findings, *ONECUT1*
178 null mice had a reduced number of endodermal cells expressing PDX1 at pancreatic budding[53]
179 accompanied by compromised pancreatic organogenesis[69-71]. To validate the regulatory impact
180 of *ONECUT1e-664kb* on *ONECUT1*, we used CRISPR-Cas9 deletion as an orthogonal method
181 to complement our dCas9-KRAB based repression approach. Using paired gRNAs flanking a
182 ~750 bp region, we generated two heterozygous and one homozygous clonal enhancer deletion
183 lines (Fig. 2C). *ONECUT1* transcription is activated during the GT to PFG transition (Sup. Fig.
184 2A), so we differentiated the homozygous and heterozygous enhancer knockout (eKO) lines to
185 the PFG stage and assessed PDX1 and ONECUT1 protein levels via flow cytometry (Fig. 2D).
186 Homozygous eKO led to a near-complete depletion of ONECUT1+ cells (Fig. 2E). In addition,
187 the proportion of cells activating PDX1 expression was significantly lower in homozygous eKO
188 cells, showing that the loss of ONECUT1 impaired pancreatic differentiation (Fig. 2F). These
189 findings were further confirmed through immunofluorescence staining for PDX1 and ONECUT1
190 (Sup. Fig. 2B). While heterozygous lines did not exhibit a significantly lower percentage of
191 PDX1+ or ONECUT1+ cells, there was a significant decrease in the mean fluorescence intensity
192 signal of ONECUT1 among the PDX1+ cells in heterozygous cells compared to wild-type (WT)
193 cells (Fig. 2G). This suggested that the heterozygous cells had lower ONECUT1 expression
194 compared to WT cells.
195

196 To directly assess the *cis* effect of heterozygous eKO on *ONECUT1* transcription, we pursued a
197 more quantitative approach and designed an allele-specific droplet digital PCR (ddPCR) assay
198 for the *ONECUT1* transcript. Given the substantial linear genomic distance between
199 *ONECUT1e-664kb* and the *ONECUT1* gene, we first called and phased all SNPs in the
200 *ONECUT1* locus after conducting targeted long-read Nanopore sequencing with adaptive
201 sampling (Sup. Fig. 2D). This approach allowed us to identify a SNP in the 3'UTR of *ONECUT1*
202 for allele-specific ddPCR (Fig. 2H). We further identified a genotyping SNP close to
203 *ONECUT1e-664kb* for phasing the deleted enhancer in the heterozygous lines via PCR (Fig. 2I,
204 Sup. Fig. 2D). Both heterozygous lines had enhancer deletions on the same allele (designated as
205 "Allele 2") (Fig. 2J). ddPCR assays performed on RNA extracted from cells differentiated to the
206 PFG stage showed that homozygous eKO lines lost >95% of transcription from both alleles (Fig.
207 2K). Heterozygous lines exhibited a specific loss of transcription from Allele 2, resulting in a
208 significantly skewed ratio of transcripts from Allele 2 over Allele 1. This skewed ratio was not
209 observed in control ddPCR assays conducted on genomic DNA (gDNA) (Sup. Fig. 2C). These
210 findings demonstrate that the transcriptional impact of *ONECUT1e-664kb* deletion on
211 *ONECUT1* expression is due to direct in cis regulation, rather than indirect effect through the
212 downregulation of the ONECUT1 TF or other mechanisms. Notably, one of the heterozygous
213 eKO lines (Het. 1) had the enhancer sequence inverted on Allele 1, and *ONECUT1* transcription
214 from Allele 1 was not significantly affected by this inversion. Together, these results establish
215 *ONECUT1e-664kb* as a critical *cis*-regulatory element for *ONECUT1* activation in pancreatic
216 development.
217
218 ***ONECUT1* enhancer deletion causes transcriptional and epigenomic changes in the locus**

219 The large impact of *ONECUT1e-664kb* deletion on *ONECUT1* transcription led us to investigate
220 its potential effects on transcription at the broader genomic locus. We performed TruSeq
221 Stranded Total (Ribodepletion) RNA-seq on WT, heterozygous, and homozygous eKO cells
222 differentiated to the GT and PFG stages. Consistent with our ddPCR results, homozygous eKO
223 cells had an approximately 30-fold decrease in *ONECUT1* reads at the PFG stage compared to
224 reads in WT cells. Some enhancers may serve as "hubs" that simultaneously regulate the
225 expression of multiple genes in a locus[72]. While we did not detect any significant transcription
226 changes of genes in the adjacent TADs, we did observe a statistically significant decrease in
227 *WDR72* expression in homozygous *ONECUT1e-664kb* cells (Fig. 3A). Notably, *WDR72* is the
228 only other protein-coding gene in the same TAD as *ONECUT1*. However, ddPCR assays did not
229 show a statistically significant difference in *WDR72* transcripts, despite a noticeable trend
230 towards a decrease (Sup. Fig. 3A), which may be attributed to the low expression of *WDR72* in
231 the PFG stage. In summary, we conclude that *ONECUT1e-664kb* primarily regulates *ONECUT1*
232 transcription during pancreatic differentiation.
233
234 We further investigated the impact of *ONECUT1e-664kb* deletion on *ONECUT1* promoter
235 antisense (PAS) RNA transcription. Transcription-associated PAS RNAs have been extensively
236 documented[73,74], and antisense RNAs can be quantified in strand-specific RNA-seq datasets[75].
237 Indeed, we observed an adjacent non-coding region antisense to *ONECUT1* that had lower PAS
238 transcription in homozygous eKO cells compared to WT cells (Fig. 3B). Quantifying antisense
239 transcripts in this region across three differentiation replicates showed that during the GT-PFG
240 transition, *ONECUT1* transcriptional activation in WT cells was associated with a significant
241 increase in *ONECUT1* PAS transcripts. These transcripts were significantly reduced in both

242 heterozygous and homozygous eKO lines at the PFG stage (Fig. 3C). In contrast, sense
243 transcripts across the same non-coding region showed no change during the GT-PFG transition
244 and were unperturbed upon enhancer deletion (Sup. Fig. 3B). Thus, enhancer deletion has
245 transcriptional effects beyond the decrease of the *ONECUT1* transcript, illustrating an additional
246 transcriptional phenotype of enhancers loss.
247
248 H3K27ac is an epigenetic mark associated with active enhancers and transcription. A previous
249 study documented that *PTF1A* enhancer deletion caused H3K27ac loss at the *PTF1A* promoter as
250 well as adjacent regions in the genomic locus[12]. However, given that *PTF1A* encodes a TF, it is
251 challenging to distinguish *trans* and *cis* effects of enhancer deletion. We reasoned that analyzing
252 allele-specific H3K27ac ChIP-seq reads could aid in examining *cis*-consequences of enhancer
253 deletion, as a *cis*-effect would manifest as an imbalance in allele-specific reads within
254 heterozygous eKO cells. We performed H3K27ac ChIP-seq on PFG cells differentiated from
255 WT, heterozygous, and homozygous eKO lines. First, we compared WT and homozygous eKO
256 cells. The decrease in H3K27ac in the *ONECUT1* TAD was the most pronounced of all regions
257 on chromosome 15 (Sup. Fig. 3C), strongly suggesting cis effects. Among the 16 regions with
258 the most significant H3K27ac reduction (adjusted p-value < 1E-15) (Fig. 3D), 10 contained
259 heterozygous SNPs. These SNPs allowed us to determine changes occurring *in cis* with enhancer
260 deletion by calculating the allele-specific read ratio in both WT and heterozygous eKO cells
261 (Sup. Fig. 3D). As controls, we identified regions with heterozygous SNPs but did not meet the
262 significance threshold. For all top differential peaks, H3K27ac was decreased in cis to the allele
263 with enhancer deletion in heterozygous eKO cells, with some positions losing nearly all
264 H3K27ac (Fig. 3E,3F), whereas control regions showed a near 1:1 allele ratio in both WT and

265 heterozygous eKO cells. This demonstrates that *ONECUT1e-664kb* controls the activity of
266 multiple enhancers within the locus, arguing for direct crosstalk among the cis-regulatory
267 elements. In addition, the reduction of PAS transcripts and H3K27ac signals in *cis* with enhancer
268 deletion also illustrate the impact of enhancer dysregulation on genome function beyond gene
269 transcription.
270
271 **T2D-associated SNP decreases pancreatic TF recruitment to *ONECUT1e-664kb***
272 To explore the potential impact of *ONECUT1e-664kb* in clinical genetics, we examined SNP
273 metabolic disease-association statistics in the *ONECUT1* locus using data from a recent T2D
274 GWAS meta-analysis[27] and other studies[76,77]. We observed one distinct T2D disease association
275 signal around the *ONECUT1* gene with a largely unresolved 99% credible set spanning ~100kb
276 genomic region (hg38, chr15:52777944-52873484). We identified a second independent T2D
277 disease-causal signal 664 kb away from *ONECUT1* and close to *WDR72*, that is driven by a
278 single variant, rs528350911 (G/C, reverse strand, minor allele frequency = 0.0068)[27] (Fig 4A).
279 This variant is directly within *ONECUT1e-664kb* and is highly confidently fine-mapped with the
280 posterior probability of association (PPA) close to 1 (0.994)[27]. To predict if rs528350911 would
281 affect enhancer activation in pancreatic development, we trained a ChromBPNet[78,79] model on
282 PFG-stage ATAC-seq data[25]. Then, we computed per nucleotide DeepLift[80] attribution scores
283 across the region for both the reference and alternate sequences of rs528350911. The decreased
284 attribution for an entire GATA motif suggests that the G in the GATA motif is crucial for the
285 model's genomic accessibility prediction, and the alternate sequence associated with T2D risk
286 leads to an attribution loss at the site (Fig. 4B).
287

288 GATA4 is known for its important role in pancreatic differentiation[60] and is linked to neonatal
289 diabetes[81]. At the PFG stage, GATA4, along with key pancreatic TFs GATA6 and FOXA2,
290 binds to the *ONECUT1e-664kb* enhancer region (Sup. Fig. 4A)[25,60]. This motivated us to
291 investigate the molecular impact of rs528350911. We engineered three heterozygous hPSC lines
292 with the risk variant: two carried the variant on Allele 2, and one on Allele 1 (Fig. 4C). While all
293 lines differentiated to the PFG stage had a similar proportion of ONECUT1 and PDX1
294 expressing cells (Fig. 4D, Sup. Fig. 4B), we observed a significant reduction in ONECUT1 MFI
295 for two of the three lines (Fig. 4E). Allele-specific ddPCR quantification of *ONECUT1*
296 transcription (Fig. 2J) confirmed that the same two lines had a small but significant allele-
297 specific reduction in *ONECUT1* transcription (Fig. 4F).
298
299 To investigate the impact of rs528350911 on TF binding at *ONECUT1-e664kb* following
300 chromatin immuno-precipitation (ChIP), we developed an allele-specific ddPCR assay targeting
301 the reference and reference and alternate (T2D associated) nucleotides (Fig. 4G). By utilizing the
302 heterozygous lines, this allele-specific assay supports an internally controlled experimental
303 system that mitigates technical variabilities arising from *in vitro* differentiation, ChIP, and PCR.
304 We selected two hPSC lines that contained the risk variant on different alleles; one line had a
305 significant impact on *ONECUT1* expression, while the other did not show a significant effect as
306 mentioned above. After differentiating the lines to the PFG stage, we performed ChIP for
307 GATA4, GATA6, and FOXA2. The precipitated DNA was then used as the template in the
308 ddPCR assay, revealing a ~4-fold reduction in GATA4 binding on the allele containing the
309 disease-associated SNP in both lines (Fig. 4H). We also observed significantly decreased

310 GATA6 and FOXA2 binding. Thus, despite a mild effect on *ONECUT1* transcription, the disease
311 SNP substantially reduces TF binding to *ONECUT1e-664kb*.
312

313 ## Discussion

314 Pancreatic developmental enhancers have been found in mice by integrating pre-existing
315 knowledge of important pancreatic genes with annotations such as evolutionary conservation and
316 DNA accessibility. Using this approach, several murine *PDX1* enhancers were identified, and
317 knockout mice had a spectrum of pancreatic phenotypes depending on the specific region that
318 was removed[82,83]. A similar approach also identified a putative *GATA4* enhancer, which was
319 subsequently verified by a transgenic reporter assay in endoderm and endoderm-derived tissues
320 including the pancreas and duodenum[84]. These studies set the stage for integration of
321 biochemical epigenetic marks into the enhancer discovery framework, which led to the discovery
322 of a *PTF1a* enhanceropathy responsible for pancreatic agenesis and monogenic diabetes in
323 humans[12,13]. The studies also highlight the value of developmental enhancer discovery for
324 elucidating mechanisms of organogenesis and disease diagnosis. Building upon these findings,
325 our study expanded the scope of enhancer discovery by targeting developmentally accessible
326 chromatin regions for large-scale functional interrogation using dCas9-KRAB. Our study not
327 only identified the human *PDX1* and *GATA4* enhancers corresponding to the aforementioned
328 mouse enhancers (Sup. Table 1), but also uncovered additional enhancers that had not been
329 previously described. These newly discovered can serve as a diagnostic resource not only for the
330 large number of monogenic diabetes patients with no known genetic cause[14] but also for
331 comprehending variants identified through population-level association studies.
332

333 The identification of *ONECUT1e-664kb* adds to a select group of well-documented long-range
334 regulatory sequences that activate transcription over genomic intervals exceeding 500 kb[85-93].
335 These enhancers offer a valuable model for further investigation into long-range transcriptional
336 regulation and its potential implications in human disease. Our findings underscore the value of
337 unbiased discovery, and we speculate that gene regulation by long-range enhancers is more
338 common than appreciated. From a technical standpoint, the substantial genomic distance between
339 *ONECUT1e-664kb* and its target gene presented a challenge for definitively demonstrating the
340 cis impact of enhancer deletion. In addition to conducting Hi-C to support 3D enhancer-promoter
341 interaction, we phased the variants in the *ONECUT1* locus and employed high-sensitivity
342 variant-specific ddPCR assays in heterozygous lines. As a result, we uncovered two mechanisms
343 beyond transcriptional control by which regulatory regions could influence disease traits: loss of
344 promoter antisense transcription and H3K27ac in a genomic locus. We also demonstrate a near-
345 complete, allele-specific loss of *ONECUT1* transcription resulting from the deletion of
346 *ONECUT1e-664kb*. Considering results from *ONECUT1* knockout studies in mice[53] and hPSC
347 pancreatic differentiation[28], we infer ONECUT1 protein loss to be the primary driver of the
348 pancreatic differentiation phenotype in our study. Our findings are also in concordance with the
349 reduced number of PDX1 expressing cells during pancreatic budding in *ONECUT1* null mice
350 that contributes to pancreatic hypoplasia[53]. We predict that loss of *ONECUT1e-664kb* in an
351 individual would phenocopy the severe pancreatic hypoplasia observed in cases of *ONECUT1*
352 protein-coding mutations in clinical settings[28,29].
353
354 In previous GWAS, the variant rs528350911 is assigned to the closest gene – *WDR72* [27].
355 Considering the strong effect of *ONECUT1e-664kb* deletion on *ONECUT1* expression and the

356 known association of *ONECUT1* with diabetes, this variant most probably confers disease
357 susceptibility through its effect on ONECUT1 expression. Our work and demonstrates how the
358 integration of functional screening, precise genome editing, deep learning models, and 3D
359 genomic information can improve the functional pairing of variants with genes. The introduction
360 of the disease variant into hPSCs did not have a strong effect on *ONECUT1* transcription, but its
361 robust effect on TF binding to *ONECUT1e-664kb* suggests that assays other than gene
362 expression could be valuable for characterizing the molecular consequences of variants in
363 endogenous loci. We speculate that the rs528350911 risk variant may represent a latent disease
364 vulnerability, which could be exposed by factors not fully modeled in our current differentiation
365 system such as age and environment[94,95,]. Given the considerable distance between *ONECUT1e-*
366 *664kb* and the *ONECUT1* gene, as well as the limited impact of the rs528350911 risk variant on
367 *ONECUT1* transcription, it is unlikely that this variant would have been prioritized for
368 investigation in a high-throughput variant modeling screen that relies on prime or base editing
369 technologies. Higher assay sensitivity is likely necessary to detect the subtle effects of a point
370 variant[96] like rs528350911 compared to the more pronounced effects of enhancer repression with
371 dCas9-KRAB[97]. Our work demonstrates that integrating *in vitro* functional enhancer discovery
372 with population-level genomics provides a rational approach to efficiently prioritize variants for
373 deeper investigation. This framework holds promise for understanding the exponentially growing
374 set of disease-associated variants discovered through population-wide GWAS and whole genome
375 sequencing efforts.
376

Acknowledgments

We thank Dr. Thomas Vierbuchen for valuable advice. We thank Hanuman Kale and Emily DeBitetto for assisting with experiments not included in the manuscript and acknowledge the assistance from the following Memorial Sloan Kettering Cancer Center (MSKCC) Cores: Integrated Genomics Operation, Flow Cytometry, Gene Editing & Screening, Molecular Cytogenetics, and Stem Cell Research. We thank Dr. Ralph Garippa, and Sanjoy Mehta for assistance with CRISPR library generation and sequencing, Cassidy Cobbs for providing advice regarding the next-generation sequencing experiments, Dr. Rui Gardner for assistance to design the sorting strategies for the CRISPR-based screenings, Dr. Stephanie Chrysanthou for conducting nanopore sequencing with adaptive sampling, and Dr. Andrea Farina for designing and executing the ddPCR assays.

Author Contributions

Conceptualization: SJK, DH

Methodology: SJK, JY, WW, JP, HSC, RL, DM, KD, EA, CSL, DH

Investigation: SJK, JY, WW, JP, HSC, JLI, JK, JZ, QL, DM, RL

Visualization: SJK, WW, QL

Funding acquisition: DH

Project administration: DH

Supervision: SJK, DH

Writing (original draft): SJK, DH

Writing (review & editing): all authors

Funding

399 This study was funded in part by the National Institutes of Health grant U01DK128852 (C.S.,
400 D.H., E.A.), National Institutes of Health grant U01HG012051 (D.H.), National Institutes of
401 Health grant R01DK096239 (D.H.), Starr Tri-I Stem Cell Initiative #2019-001 (D.H. and E.A.),
402 NIH/NCI MSKCC Cancer Center Support Grant P30CA008748, and a National Institutes of
403 Health T32 training grant T32GM008539 (S.J.K.).

404 **Competing interests**

405 Authors declare that they have no competing interests.

406 **STAR Methods**

407 KEY RESOURCES TABLE

408 RESOURCE AVAILABILITY

409 - Lead contact

410 - Materials availability

411 - Data and code availability

412 EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

413 - Primary Cells

414 METHOD DETAILS

415 - hPSC culture

416 - hPSC directed pancreatic differentiation

417 - gRNA library assembly

418 - Lentiviral gRNA libraries

419 - CRISPR dCas9-KRAB screens

420 - gRNA sequencing and screen analysis

421 - Generation of *ONECUT1e-664kb* clonal knockout hPSC lines

422 - Generation of clonal hPSC lines with disease-associated variant

423 - *ONECUT1* locus SNP phasing

424 - *ONECUT1e-664kb* heterozygous knockout line phasing

425 - *ONECUT1e-664kb* heterozygous rs528350911 knock-in line phasing

426 - Immunofluorescence staining

427 - Flow cytometry analysis

428 - Droplet digital PCR (ddPCR) and analysis

429 - Hi-C and analysis

430 - ChIP and ChIP-seq assays

431 - Enhancer knockout line H3K27ac ChIP-seq analysis

432 - chromBPNet variant modeling

433 QUANTIFICATION AND STATISTICAL ANALYSIS

434

435 RESOURCE AVAILABILITY

436 Lead Contact

437 Further information and requests for resources and reagents should be directed to Danwei

438 Huangfu (huangfud@mskcc.org).

439 Materials availability

440 All the hPSC cell lines and plasmids generated in this study are available upon reasonable

441 request.

442 Data and Code availability

443 All the sequencing data generated in this study is available at GEO under accession codes

444 GSE261390, GSE261391, GSE261392, GSE263174. The Hi-C data from the PFG cells are

445 available in the 4DN Nucleome Data Portal (https://data.4dnucleome.org) under accession

446 number 4DNFIQBPRZZD. Previously published data that were re-analyzed here are available at

447 GEO under code GSE114102, and the 4DN Nucleome Data Portal, accession numbers:

448 4DNFIKCEWMIA, 4DNFI7RGXYFY, 4DNFIRLG5UWL. Source data are provided with this

449 paper. All other data supporting the findings of this study are available from the leading author

450 upon reasonable request. Publicly available software and packages were used throughout this

451 study according to each developer's instructions.

452 # EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

453 ## Primary cells

454 The parental hESC line H1 (NIHhESC-10-0043) is available from WiCell under a material
455 transfer agreement.

456

457 # METHOD DETAILS

458 ## hPSC culture

459 Experiments were performed with H1 human embryonic stem cells (NIHhESC-10-0043), which
460 were regularly confirmed to be mycoplasma-free by the Memorial Sloan Kettering Cancer
461 Center (MSKCC) Antibody & Bioresource Core Facility. All experiments were conducted per
462 NIH guidelines and approved by the Tri-SCI Embryonic Stem Cell Research Oversight
463 (ESCRO) Committee. hPSCs were maintained in Essential 8 (E8) medium (Thermo Fisher
464 Scientific, A1517001) on vitronectin (Thermo Fisher Scientific, A14700) pre-coated plates at
465 37°C with 5% CO_2. The Rho-associated protein kinase (ROCK) inhibitor Y-276325 (5 μM;
466 Selleck Chemicals, S1049) was added to the E8 medium immediately after passaging or thawing
467 of hPSCs for 24 hours.

468 ## hPSC directed pancreatic differentiation

469 hPSCs were maintained in E8 medium for 2 days to reach ~80% confluence. Cells were washed
470 with PBS and differentiated to DE, GT, and PFG stages following previously described
471 protocols[17,98]. hPSCs were rinsed with PBS and first differentiated into DE using S1/2 medium
472 supplemented with 100 ng/ml activin A (Bon Opus Biosciences) for 3 days and CHIR99021
473 (Stemgent, 04-0004-10) for 2 days (first day, 5 μM; second day, 0.5 μM). DE cells were rinsed
474 with PBS and then exposed to S1/2 medium supplemented with 50 ng/ml KGF (FGF7)

475 (PeproTech, 100-19) and 0.25 mM vitamin C (VitC) (Sigma-Aldrich, A4544) for 2 days to reach
476 GT stage. For PFG stage, cells were switched to S3/4 medium supplemented with 50 ng/ml
477 FGF7, 0.25 mM VitC and 1 μM retinoic acid (RA) (Sigma-Aldrich, R2625) for 2 days to reach
478 PFG stage.

479 gRNA library assembly

480 23 transcription factors expressed in pancreatic differentiation were selected based on known
481 clinical roles in metabolic disease and demonstrated pancreatic functional requirement from
482 CRISPR screens[29,54-57]. Around each transcription factor, loci within hg19 were defined to
483 include at least the TAD containing the gene, with TADs called on PFG-stage Hi-C data as
484 previously done[65]. To select putative enhancers for interrogation in these loci, Genrich (0.6.1)[99]
485 was used to call peaks at an FDR-adjusted p-value of 0.01 on previously published DE, GT, and
486 PFG stage ATAC-seq data (Sup. Table 1)[25]. For each region in the loci, and some additional
487 regions, CHOPCHOPv3 was used to design gRNAs for each region[100]. For regions <750 bp, the
488 top 5 ranked gRNAs were taken, and for regions >750 bp, 1 gRNA/150bp of sequence was
489 selected for interrogation. Finally, 1100 safe harbor[63], 463 non-targeting gRNAs[64] were included
490 in the library for a total of 38,747 gRNAs. Oligos were synthesized, amplified, and restriction
491 cloned into lentiGuide-puro[101] (Addgene; 52963) backbone by the MSKCC Gene Editing &
492 Screening Core Facility. Cloned plasmid libraries were PCR amplified to incorporate adapters
493 for NGS. Samples were purified and sequenced using Illumina HiSeq 2500 platform. FASTQ
494 files were clipped by position and reads were mapped back to the reference library file to show
495 relative abundance of reads per gRNA. Reads within each sample were normalized to total
496 number of mapped reads and library size. The overall representation of the libraries was charted

497 over a one-log fold change to evaluate if any gRNA was over- or under-represented in the final
498 library.

499 Lentiviral gRNA libraries

500 gRNA lentiviral library generation was performed as previously described[17]. In brief, a total of
501 9.45 µg of the library plasmid combined with 6.75 µg lentiviral packaging vector psPAX2 and
502 1.36 µg vesicular stomatitis virus G (VSV-G) envelope expressing plasmid pMD2.G (Addgene
503 plasmids 12260 and 12259) were transfected with the JetPRIME (VMR; 89137972) reagent into
504 10E6 293T cells in a 10cm plate to produce the lentiviral particles. Fresh medium was changed
505 24h after transfection and viral supernatant was collected at 48 and 72h after transfection, spun at
506 1000 rpm for 5 minutes, 0.45um filtered, and stored at −80°C.

507 CRISPR dCas9-KRAB screens

508 Two independent screens were performed following the same procedure. H1 hPSC (i)dCas9-
509 KRAB cells were used[65]. The lentiviral library was transduced into 12E6 hPSCs distributed in 3
510 x 10 cm plates following the conditions (MOI: 4, protamine sulfate: 6 µg/ml, Ri: 10 µM). After
511 one day of recovery, doxycycline (DOX) was added and cells were kept under puromycin
512 selection for 2 days. After two days, 30E6 selected cells were seeded into 6-well plates (5E5 per
513 well) in VTN-coated plates, grown for 48 hours in E8 medium. After washing with Phosphate
514 Buffered Saline (PBS) w/o Ca2+ & Mg2+, cells were differentiated to PFG cells, with
515 doxycycline added throughout the nine-days differentiation process. After, cells were stained
516 with LIVE/DEAD reagent, fixed, and stained with the PDX1 antibody, and then a fluorescent
517 secondary antibody. They were then sorted by the MSKCC cytometry facility based on 2ndary
518 antibody fluorescence using FACSAria sorters (BD Biosciences). Positive and negative cells
519 were sorted aiming for a minimum 1000X representation (number of cells sorted*MOI)/(unique

520 gRNAs in library) per condition. For replicate 1, approximately 10E6 PDX1- and 10E6 PDX1+
521 cells were sorted. For replicate 2, approximately 20E6 PDX1- and 20E6 PDX1+ cells were
522 sorted. Cells were pelleted and kept at -80°C for downstream gRNA sequencing.

523 ## gRNA sequencing and screen analysis

524 gRNA enrichment sequencing was performed by MSKCC Gene Editing & Screening Core
525 Facility as previously described, but with modifications to enable DNA extraction from fixed
526 cells[17]. Cells were decrosslinked and treated with proteinase K (0.1 mg/ml) overnight followed
527 by genomic DNA extraction with QIAGEN Blood & Cell Culture DNA Maxi Kit (QIAGEN;
528 13362) and quantified by Qubit (Thermo-Scientific) following the manufacturer's guidelines. A
529 quantity of gDNA covering the representation of gRNAs described above was amplified with
530 oligos containing the Illumina adapters and multiplexing barcodes by PCR. Amplicons were
531 quantified by Qubit and Bioanalyzer (Agilent) and sequenced on the Illumina HiSeq 2500
532 platform. The gRNA library sequences were used to align the sequencing reads and the counts
533 for each gRNA were determined with MAGeCK. Screens were analyzed with MAGeCK RRA
534 (0.5.9.5), and the two screen replicates were analyzed independently (Sup. Table 1). To calculate
535 a single value for the magnitude of effect, a ratio was taken between the positive and negative
536 scores and log-transformed. The top 36 hits associated with a PDX1 decrease had a MAGeCK
537 neg.FDR<0.1 in replicate 2 and caused a PDX1 decrease in both replicates (Sup. Table 1). Hits
538 overlapped with annotated promoter regions were designated as "Promoter Hits," with the rest
539 designated as "Non-Promoter Hits." Hits overlapped with annotated promoter regions were
540 designated as "Promoter Hits," with the rest designated as "Non-Promoter Hits." Promoter
541 overlap was defined as overlap with an interrogated region ± 1.5kb of Gencode (v18) TSS sites.
542 For Figure 1F, if there were multiple overlapped regions, the most significant hit was used for

543 plotting. To link enhancers to the putatively regulated genes, the nearest linearly adjacent
544 "Promoter Hit" was inferred to be the regulated gene of the "Non-Promoter Hit." To determine
545 hit and non-hit overlap with H3K27ac ChIP-seq peaks during differentiation, Genrich (0.6.1)[99]
546 was used to call peaks at an FDR-adjusted p-value of 0.01 on previously published ES, DE, GT,
547 and PFG stage H3K27ac ChIP-seq data[25]. Conservation scores of regions were determined by
548 calculating the average of per nucleotide conservation across each region from PhyloP100 per-
549 nucleotide conservation scores[67].

550 Generation of *ONECUT1e-664kb* clonal knockout hPSC lines

551 H iCas9 hPSC lines[60] were used to generate three enhancer deletion lines carrying ONECUT1e-
552 664kb deletions. Two gRNAs targeting the flanking regions of the enhancer were designed, and
553 knockouts generated as previously described but with some modifications[8]. gRNAs and tracer
554 RNA were ordered from IDT (Alt-R CRISPR-Cas9 crRNA and 1072532) and added at a 15 nM
555 final concentration. In brief, gRNA/tracer RNA and Lipofectamine RNAiMAX (Thermo Fisher
556 Scientific, 13778030) were diluted separately in Opti-MEM (Invitrogen, 31985070), mixed
557 together and incubated for 15 min at room temperature, and added dropwise to freshly seeded
558 iCas9 hPSCs in a 24-well plate. Doxycycline (2 µg/ml) was added the day before transfection,
559 the day of transfection and 1 day after transfection to induce the Cas9 expression. Three days
560 after transfection, hPSCs were dissociated into single cells and ~1-3E3 cells were plated into one
561 100-mm tissue culture dish for colony formation. After ~10 days of expansion, single colonies
562 were picked. Genomic DNA from crude cell lysate was used for PCR genotyping (primers:
563 ONECUT1e_KOgenot_F, ONECUT1e_KOgenot_R). Heterozygous lines contained a band
564 corresponding to the WT band and a smaller band corresponding the knockout band, while
565 homozygous lines only had smaller bands. Further genotyping was done through TOPO cloning

566 of the genotyping amplicons, which determined the exact genotypes of both alleles for all lines
567 (Sup. Table 2). gRNA target sequences and primers used for PCR and sequencing are listed in
568 Supplementary Table 3.

569 Generation of clonal hPSC lines with disease-associated variant

570 A process similar to enhancer deletion line generation was used to generate heterozygous knock-
571 in SNP lines carrying the rs528350911 variant. During transfection, a single gRNA proximal to
572 the variant site was used, and an ssDNA HDR template containing the disease-associated variant
573 was added to the mixture at the Lipofectamine transfection step (Sup. Table 3) at a concentration
574 of 20 nM. Genomic DNA from crude cell lysate was used for PCR (primers:
575 ONECUT1e_KOgenot_F, ONECUT1e_KOgenot_R). All picked clones were sequenced, and
576 heterozygous lines picked by the presence of a double peak on the spectrogram that
577 corresponded to the insertion of the disease-associated variant. For lines containing the
578 introduced variant, no indels or other changes from the reference genome were detected in the
579 amplicon.

580 *ONECUT1* locus SNP phasing

581 To phase all SNPs in the *ONECUT1* locus, Nanopore sequencing with adaptive sampling of the
582 region chr15:52,707,645-53,586,740 (and controls regions) was conducted on an H1 iCas9 line
583 with the rs528350911 heterozygous variant (line N11). DNA was extracted using Monarch®
584 HMW DNA Extraction Kit for Cells & Blood (NEB, T3050), according to the manufacturer's
585 instructions. DNA quantity was measured using a QuantiT (Thermo Fisher Scientific), purity on
586 a NanoDrop (Thermo Fisher Scientific) and fragment-size distribution on a TapeStation
587 (Agilent). Prior to ONT library preparations, DNA was sheared to ~15-20 kb fragment size using
588 Covaris g-TUBE. Sequencing libraries were prepared from ~1–2ug of DNA, using native library

589 preparation kit, SQK-LSK114, according to the manufacturer's instructions. Each library was
590 loaded onto a PromethION flow cell, R10.4.1, and sequenced on an ONT PromethION P24
591 device. The sample was run for 72 h, with 1 nuclease flush and reload performed during the run
592 to maximize sequencing yield.
593 Reads were aligned with minimap2 (2.26)[102], with samtools[103] coverage estimating ~138X
594 coverage across the region enriched with adaptive sampling. WhatsHap (2.2)[104] was used to
595 phase the variants in the locus as well as assemble a haplotype containing the *ONECUT1* gene,
596 enhancer, and surrounding regions (Sup. Fig. 2D). WhatsHap was also used to phase-tag reads
597 for visualization in Sup. Fig. D and E. From these results, a heterozygous SNP was identified in
598 the 3'UTR of the *ONECUT1* transcript, as well as the enhancer-proximal Genotyping SNP.

599 ## *ONECUT1e-664kb* heterozygous knockout line phasing

600 To assess which allele contained the knocked-out enhancer in the heterozygous enhancer
601 deletion lines (Het. 1, Het. 2), a forward primer was designed centromeric to the Genotyping
602 SNP and a reverse primer targeting the enhancer sequence that would be present on the non-
603 deletion allele. Since line designated as Het. 1 contained an inversion of the WT enhancer
604 sequence, two different reverse primers were needed. Thus, for Het. 1, and amplicon was
605 generated with the primer pair ONECUT1e_GenotSNP_F,
606 ONECUT1e_GenotSNP_R_WT_inv_e, and for Het.2, the pair ONECUT1e_GenotSNP_F,
607 ONECUT1e_GenotSNP_R_WT_e. The amplicon was sent for sanger sequencing with the
608 ONECUT1e_GenotSNP_F primer used as the sequencing primer. The Genotyping SNP resulting
609 from this sequencing was used to determine the allele containing the deletion.

610 ## *ONECUT1e-664kb* heterozygous rs528350911 knock-in line phasing

611 Amplicons were generated with the genotyping primers (ONECUT1e_KOgenot_F,
612 ONECUT1e_KOgenot_R) and sanger sequenced with both primers. To phase the variant, we
613 conducted a PCR amplifying an ~8.1kb fragment containing both the T2D-associated variant and
614 the Genotyping SNP (primers: ONECUT1e_GenotSNP_F , ONECUT1e_GenotSNP_R_WT_e).
615 The amplicon was sent to Plasmidsaurus for linear fragment sequencing. Reads were aligned
616 with minimap2 (2.26), and variants were phased with longshot (0.4.5)[105].

617 Immunofluorescence staining

618 In vitro pancreatic differentiated cells were fixed in 4% paraformaldehyde (Thermo Fisher
619 Scientific, 50980495) for 10 min at room temperature. After washing with PBST (PBS with
620 0.1% Triton X-100) three times, cells were blocked in 5% donkey serum in PBST buffer for
621 30 min at room temperature. Primary and second antibodies were diluted in the blocking solution
622 (Sup. Table 4). Cells were incubated with primary antibodies overnight at 4°C, followed by 1 hr
623 staining for secondary antibodies at room temperature. The cells were stained with 4′,6-
624 diamidino-2-phenylindole (DAPI) for ~15 min at room temperature. Images were taken using a
625 confocal laser scanning platform (Leica TCS SP5).

626 Flow cytometry analysis

627 Cells were dissociated using TrypLE Select and resuspended in FACS buffer (5% FBS in PBS).
628 Live-Dead Fixable Violet Dead cell stain (Invitrogen, L34955) was used to discriminate dead
629 cells from live cells. Live/dead staining was performed per manufacturer's instructions in FACS
630 buffer, followed by fixation and intracellular staining with a FOXP3 staining buffer set
631 (eBioscience, 00-5523-00) following the manufacturer's instructions. Permeabilization/fixation
632 was performed at room temperature for 1 hr. Antibody staining was performed in
633 permeabilization buffer. Antibodies for this study are listed in Supplementary Table 8. Cells

634 were then analyzed using BD LSRFortessa. Flow cytometry analysis and figures were generated
635 using FlowJo v.10.

636 RNA-seq and analysis

637 RNA was extracted from differentiated cells with the Zymo Quick-RNA MiniPrep kit (R1055).
638 TruSeq stranded total (Ribodepletion) RNA-seq was performed on three independent
639 differentiations by the MSKCC Integrated Genomics Operation (IGO) Core. Briefly, STAR
640 alignment was performed followed by featureCounts[106] quantification (Sup. Table 6) and
641 normalization with DESeq2 median of ratios prior to visualization in Fig. 3A, Fig. 3C, and Sup.
642 Fig. 3B. Differential gene expression was performed and adjusted p-values computed with
643 DESeq2. Strand-specific analysis of promoter antisense transcription analysis was performed
644 with featureCounts strand specific quantification of a region telomeric to *ONECUT1*
645 (chr15:52805588-53391232). For promoter antisense transcript visualization, samtools was used
646 to extract sense and antisense alignments from .bam files from which bigwig coverage files were
647 generated.

648 Droplet digital PCR (ddPCR) and analysis

649 Assays specific for the detection of SNPs (Sup. Table 5) were designed (using Primer3Plus, if
650 needed) and ordered through Bio-Rad. Cycling conditions were tested to ensure optimal
651 annealing/extension temperature as well as optimal separation of positive from empty droplets.
652 Optimization was done with a known positive control. A constant amount of RNA or DNA was
653 used within each experimental reaction. Reactions were partitioned into a median of ~17,500
654 droplets per well using the QX200 droplet generator. Plates were read and analyzed with the
655 QuantaSoft software to assess the number of positive droplets positive.

656 Procedure for RNA: after PicoGreen quantification of RNA, droplet generation was performed
657 on a QX200 ddPCR system (Bio-Rad catalog # 1864001) using cDNA generated from 0.1-2 ng
658 total RNA with the One-Step RT-ddPCR Advanced Kit for Probes (Bio-Rad catalog # 1864021)
659 according to the manufacturer's protocol with reverse transcription at 42°C and
660 annealing/extension at 60°C.
661 Procedure for DNA: after PicoGreen quantification of DNA, 0.25-9 ng gDNA or ChIP-ed DNA
662 was combined with locus-specific primers; FAM- and HEX-labeled probes; Hae III (DNA
663 extracted form hPSC SNP lines), Hind III (ChIP-ed DNA), or Mse I (DNA extracted
664 *ONECUT1e-664kb* hPSC knockout lines); and digital PCR Supermix for probes (no dUTP).
665 Emulsified PCRs were run on a 96-well thermal cycler using cycling conditions identified during
666 the optimization step (95°C 10'; 40 cycles of 94°C 30' and 53°C (rs528350911 assays only) or
667 60°C 1'; 98°C 10'; 4°C hold).

668 ## Hi-C and analysis

669 Hi-C was performed as previously described with cells differentiated to the PFG stage[8]. HiC-Pro
670 (3.1.0)[107] was used to align the individual Hi-C replicates to GRCh38, GCA_000001405.15,
671 with alternative contigs removed, and then final merged libraries were obtained by combining all
672 *.allValidPairs files where duplicates across biological replicates are maintained. DE libraries
673 consisted of three biological replicates (two HUES8, one H1 with accessions
674 4DNFIKCEWMIA.hic, 4DNFI7RGXYFY.hic, and 4DNFIRLG5UWL.hic respectively), while
675 the PFG libraries consisted of two biological replicates differentiated to the PFG stage (merged
676 accession 4DNFIQBPRZZD.hic). For visualization, Hi-C data was visualized at 25 KB
677 resolution.

678 Virtual 4C counts for the regions of interest were extracted using utilities from plotGardener
679 (1.6.4)[108] and plotted as a one-dimensional signal. Hi-C matrix counts for the regions of interest
680 were also extracted using plotGardener and plotted as a rectangular region. Observed counts are
681 library normalized prior to plotting so that the total number of reads between panels is
682 comparable. TADs that were previously identified were also overlaid on the data. Counts within
683 the vicinity of the *ONECUT1* promoter, all contacts whose start or end bins sit within a 1.2Mb
684 window around the *ONECUT1* promoter were extracted with strawr (0.0.91) and passed to
685 DESeq2 (1.40.1)[109]. The Wald test was used to assess significance of all differential contacts
686 within that window, and significant hits anchored on the *ONECUT1* containing bin are reported
687 at an FDR of 0.1.
688 For calling topologically associating domains, merged Hi-C files were first ICE normalized
689 through HiTC (1.34.0)[110] and then topologically associating domains were identified with Hi-
690 CDC+ (1.5.2)[111] using its interface to TopDom at 50 kb. Window size for scanning was set at 5.

691 ## ChIP and ChIP-seq assays

692 ChIP was performed as previously described[18]. Briefly, for each sample, around 30E6 cells were
693 crosslinked with 1% formaldehyde for 10 minutes at 37°C and quenched with 0.125 M glycine.
694 Fixed cells were collected in cold PBS buffer and washed with the same. Then the cells were
695 snap frozen and saved at -80°C. On the day of immunoprecipitation, cell pellet was thawed on
696 ice, resuspended in 700uL SDS buffer (1% SDS, 10 mM EDTA, 50 mM Tris–HCl, pH 8), and
697 incubated for 10min on ice. Sonication was performed on a Branson Sonifier 150 set at 30%
698 amplitude for 5.5min (10s on/off pulsing). Supernatant was pre-cleared with Dynabeads Protein
699 G (#10004D) and then incubated overnight with antibodies (10 ug for transcription factors, 5 ug
700 for H3K27ac, Sup. Table 4) at 4°C.). On the next day, the ChIP samples were incubated with

701 Dynabeads Protein G for 6 hours at 4°C. Then the beads were pelleted and sequentially washed
702 twice with low salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl, pH 8, 150
703 mM NaCl), then high salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl, pH 8,
704 500 mM NaCl), and finally TE buffer (10 mM Tris–HCl, pH 8, 1 mM EDTA). The DNA was
705 eluted from the beads by incubating in elution buffer (1% SDS, 0.1 M NaHCO3) at 65°C for
706 15min and decrosslinked with 190 mM NaCl at 65°C overnight. A total of 5 µl 0.5 M EDTA, 10
707 µl 1 M Tris–HCl, pH 6.5, and 1 µl Proteinase K (20 mg/ml) were added to 260 uL of the de-
708 crosslinked product and incubated for 1 hr at 45°C. DNA was isolated by using QIAquick PCR
709 purification kit (Qiagen, 28104; QIAGEN). H3K27ac ChIP samples were submitted to MSKCC
710 Integrated Genomics Operation core for NGS library preparation and sequencing. GATA4,
711 GATA6, and FOXA2 ChIP samples were submitted for ddPCR.

712 H3K27ac ChIP-seq Analysis

713 Sequenced data were aligned to the hg38 reference genome using bowtie2 (2.5.1)[112]. Peak calling
714 was performed using MACS2 (2.2.9.1)[113] with the respective ChIP input as the control and the
715 default P value cutoff and default extension size. Irreproducible discovery rate (IDR)[114] with a
716 cutoff of 0.01 was used to filter pick up reproducible peaks, where at least 2 pairwise
717 comparisons of the three replicates passed the cutoff. Peaks were filtered by ENCODE exclusion
718 list version 2 for hg38[115]. The quantification and differential signal intensity of the samples were
719 calculated in DESeq2[109]. The integer counts, log(fold-change), and adjusted P values are
720 provided in Supplementary table 7). The replicates were combined and the signal track was
721 generated from MACS2. We used MACS2 subtract to generate the difference of the signal track.
722 All signal tracks were visualized using Integrative Genomics Viewer (IGV)[116].

723 For allele-specific analysis, GATK ASEReadCounter was used to quantify reads at heterozygous
724 positions phased by long-read sequencing and ratios were calculated for heterozygous positions
725 (score filters: PHASE_QUAL >= 140, and QUAL >= 200) with more than 100 reads total across
726 all WT and Het. 2 replicates. For figure 4F, the "Phased Region" is the haplotype block
727 containing *ONECUT1* called by WhatsHap (chr15:50462878-54550826) and indicated in Sup.
728 Fig. 2D.

729 ## chromBPNet variant modeling

730 A ChromBPNet[78,79] model was trained to predict the ATAC-seq profile using bulk ATAC-seq
731 data from the PFG stage[25]. We utilized default hyperparameters with an input size of 2114 base
732 pairs and an output size of 1000 base pairs for training both the bias model and the ChromBPNet
733 model. For the bias model training, we selected regions that do not overlap with any ATAC-seq
734 peaks in the four stages of pancreas differentiation[25], applying a bias threshold factor of 0.3 to
735 filter out high-count regions in the sets. This resulted in a training set of 54957 regions. For
736 ChromBPNet training, we used ATAC-seq peaks only from the PFG stage with IDR values
737 greater than 830, resulting in a training set of 52541 regions. During training for both models, we
738 held out peaks from chromosomes 15 and 13 for validation. For each specific region of interest,
739 we used the 2114 base pairs around the summit center as input and obtained the DeepLift
740 attribution score using the trained ChromBPNet model. Specifically, we employed the original
741 attribution pipeline from the ChromBPNet GitHub repository, which involves computing the
742 DeepLift score over 20 reference inputs shuffled from the original sequence. To find the GATA
743 specific motifs, we ran FIMO[117] on our region of interest using the positional weight matrix
744 (PWMs) of all GATA variants from the CIS-BP2.00 dataset[118], and filtered out the motif hits
745 with p-values < 0.05.

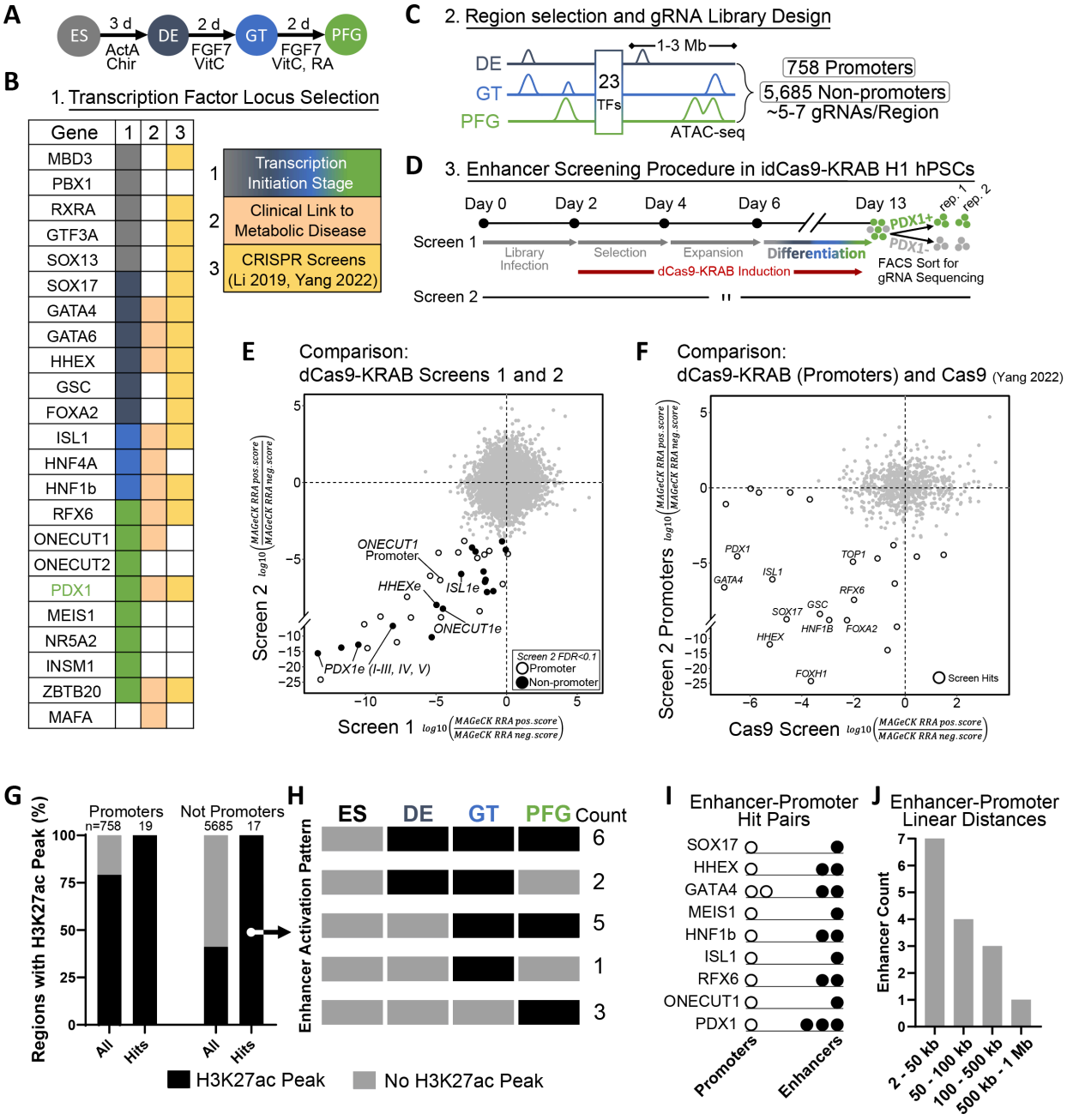
746

747 ## QUANTIFICATION AND STATISTICAL ANALYSIS

748 All datapoints refer to biological repeats. No statistical method was used to predetermine sample
749 sizes. The investigators were not blinded to allocation during experiments and outcome
750 assessment. No data were excluded from the analyses unless the differentiation experiment itself
751 failed. The number of biological and technical replicates are reported in the legend of each
752 figure. Flow cytometry analysis were derived from at least three independent experiments unless
753 specified in the legends. For ChIP-seq and bulk RNA-seq, quantification and statistics were
754 derived from at least three independent experiments. CRISPR dCas9-KRAB screening was
755 performed twice. Quantification is shown as the mean ± s.d. All the statistical analysis methods
756 are indicated in the figure legends and methods.

757

Fig. 1


A
ES
3 d
ActA
Chir
DE
2 d
FGF7
VitC
GT
2 d
FGF7
VitC, RA
PFG
B
1. Transcription Factor Locus Selection
Gene
1
2
3
MBD3
PBX1
RXRA
GTF3A
SOX13
SOX17
GATA4
GATA6
HHEX
GSC
FOXA2
ISL1
HNF4A
HNF1b
RFX6
ONECUT1
ONECUT2
PDX1
MEIS1
NR5A2
INSM1
ZBTB20
MAFA
1 Transcription Initiation Stage
2 Clinical Link to Metabolic Disease
3 CRISPR Screens (Li 2019, Yang 2022)
C
2. Region selection and gRNA Library Design
DE
GT
PFG
23 TFs
1-3 Mb
ATAC-seq
758 Promoters
5,685 Non-promoters
~5-7 gRNAs/Region
D
3. Enhancer Screening Procedure in idCas9-KRAB H1 hPSCs
Day 0
Day 2
Day 4
Day 6
Day 13
Screen 1
Library Infection
Selection
Expansion
Differentiation
dCas9-KRAB Induction
PDX1+
PDX1-
rep. 1
rep. 2
FACS Sort for gRNA Sequencing
Screen 2
E
Comparison:
dCas9-KRAB Screens 1 and 2
Screen 2 log10(MAGeCK RRA pos.score / MAGeCK RRA neg.score)
Screen 1 log10(MAGeCK RRA pos.score / MAGeCK RRA neg.score)
ONECUT1 Promoter
HHEXe
ISL1e
ONECUT1e
PDX1e (I-III, IV, V)
Screen 2 FDR<0.1
Promoter
Non-promoter
F
Comparison:
dCas9-KRAB (Promoters) and Cas9 (Yang 2022)
Screen 2 Promoters log10(MAGeCK RRA pos.score / MAGeCK RRA neg.score)
Cas9 Screen log10(MAGeCK RRA pos.score / MAGeCK RRA neg.score)
PDX1
TOP1
ISL1
GATA4
RFX6
SOX17
GSC
HNF1B
FOXA2
HHEX
FOXH1
Screen Hits
G
Regions with H3K27ac Peak (%)
Promoters
n=758
19
Not Promoters
5685
17
All
Hits
All
Hits
H3K27ac Peak
No H3K27ac Peak
H
ES
DE
GT
PFG
Count
Enhancer Activation Pattern
6
2
5
1
3
I
Enhancer-Promoter Hit Pairs
SOX17
HHEX
GATA4
MEIS1
HNF1b
ISL1
RFX6
ONECUT1
PDX1
Promoters
Enhancers
J
Enhancer-Promoter Linear Distances
Enhancer Count
2 - 50 kb
50 - 100 kb
100 - 500 kb
500 kb - 1 Mb

760 **Fig. 1: CRISPRi repression screen to discover pancreatic differentiation enhancers**

761 A) hPSC stepwise pancreatic differentiation protocol schematic. ActA, activin A; CHIR,
762 CHIR99021; RA, retinoic acid; VitC, vitamin C. B) Gene selection rationale for enhancer
763 discovery screen.
764 C) Putative enhancer region selection and gRNA design schematic.
765 D) Screening procedure schematic.
766 E) Results from two screen replicates, each point represents a genomic region. 36 regions
767 with a FDR < 0.1 in replicate 2 and have a decrease in PDX1 in both replicates.
768 F) Results from screen replicate 2 compared to whole genome PDX1 expression screen.
769 Genes showing enrichment in both screens are labeled.
770 G) Region overlap with H3K27ac peaks at any stage during differentiation (ES, DE, GT,
771 PFG).
772 H) Differentiation stages at which non-promoter screen hits overlap with H3K27ac peaks,
773 and the number of hits with each activation pattern.
774 I) Enhancer-gene pair assignment.
775 J) Enhancer-gene pair linear distances.
776

Fig. 2

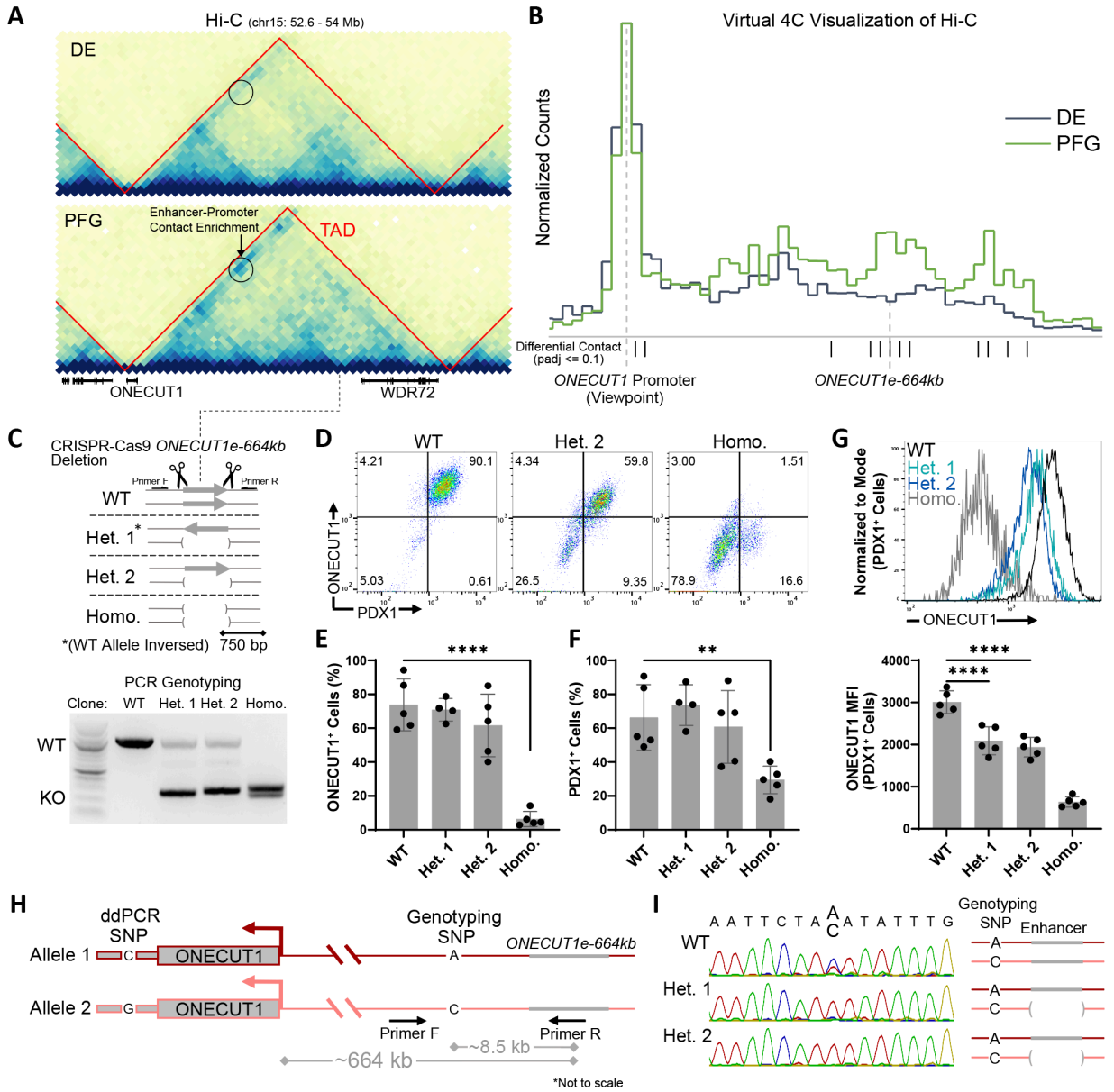

A
Hi-C (chr15: 52.6 - 54 Mb)
DE
PFG
Enhancer-Promoter Contact Enrichment
TAD
ONECUT1
WDR72
B
Virtual 4C Visualization of Hi-C
Normalized Counts
DE
PFG
Differential Contact (padj <= 0.1)
ONECUT1 Promoter (Viewpoint)
ONECUT1e-664kb
C
CRISPR-Cas9 ONECUT1e-664kb Deletion
Primer F
Primer R
WT
Het. 1*
Het. 2
Homo.
*(WT Allele Inversed)
750 bp
PCR Genotyping
Clone: WT Het. 1 Het. 2 Homo.
WT
KO
D
WT
Het. 2
Homo.
ONECUT1
PDX1
4.21
90.1
5.03
0.61
4.34
59.8
26.5
9.35
3.00
1.51
78.9
16.6
G
Normalized to Mode (PDX1+ Cells)
WT
Het. 1
Het. 2
Homo.
ONECUT1
E
ONECUT1+ Cells (%)
****
WT
Het. 1
Het. 2
Homo.
F
PDX1+ Cells (%)
**
ONECUT1 MFI (PDX1+ Cells)
****
****
H
ddPCR SNP
Genotyping SNP
Allele 1
Allele 2
ONECUT1
ONECUT1e-664kb
Primer F
Primer R
~664 kb
~8.5 kb
*Not to scale
I
Genotyping
SNP
Enhancer
WT
Het. 1
Het. 2
J
ddPCR Assay
Allele 1
Allele 2
ONECUT1
K
Allele 1
Allele 2
Allele 2 / Allele 1
ONECUT1 Allele 1 RNA Concentration (copies/uL)
ONECUT1 Allele 2 RNA Concentration (copies/uL)
ONECUT1 Transcript Ratio
ns
ns
*
*
*
*
****
****
WT
Het. 1
Het. 2
Homo.

Fig. 2: *ONECUT1e-664kb* deletion and pancreatic differentiation characterization

A) Hi-C data at the definitive endoderm and posterior foregut stage differentiated hPSCs are shown around the *ONECUT1* locus. Contact between *ONECUT1e-664kb* and *ONECUT1* promoter region are circled to show the increase in contact frequency between the two loci. Red lines demarcate TAD boundaries identified at 50 kb.

B) 1D slices of the 2D contact map were obtained with the *ONECUT1* promoter set as the bin of origin. Hi-C data was library scaled and shows the changes observed in (A). Significantly different bins were determined by DESeq2.

C) H1 hPSC enhancer deletion line schematic and genotyping PCR.

D) Representative flow cytometry plots of WT and enhancer deletion cells differentiated to the PFG stage.

E) Proportion of cells achieving a ONECUT1+ identity. Each dot represents one independent experiment (n = 5 independent experiments) and data are presented as the mean ± s.d. One-way analysis of variance (ANOVA) followed by Dunnett multiple comparisons test versus WT control.

F) Proportion of cells achieving a PDX1+ identity. Each dot represents one independent experiment (n = 5 independent experiments) and data are presented as the mean ± s.d. One-way analysis of variance (ANOVA) followed by Dunnett multiple comparisons test versus WT control.

G) Representative histograms of ONECUT1 MFI from flow cytometry plots of unedited and enhancer knockout cells differentiated to the PFG stage. Quantification of all replicates of ONECUT1 MFI of PDX1+ cells.

H) Schematic for identifying ddPCR and Genotyping SNPs.

I) Spectrogram of genotyping SNP sequencing and schematic demonstrating allele that contained the deletion in heterozygous lines.

J) ddPCR assay design schematic.

K) Allele-specific *ONECUT1* transcript quantification with ddPCR on RNA extracted from PFG stage cells. Each allele is shown in a separate plot, with the expression ratio shown in the last plot. Each symbol represents one independent differentiation (n = 3 independent experiments), average of two technical replicates per differentiation. Data are presented as the mean ± s.d. One-way analysis of variance (ANOVA) followed by Dunnett multiple comparisons test versus WT control.

Fig. 3

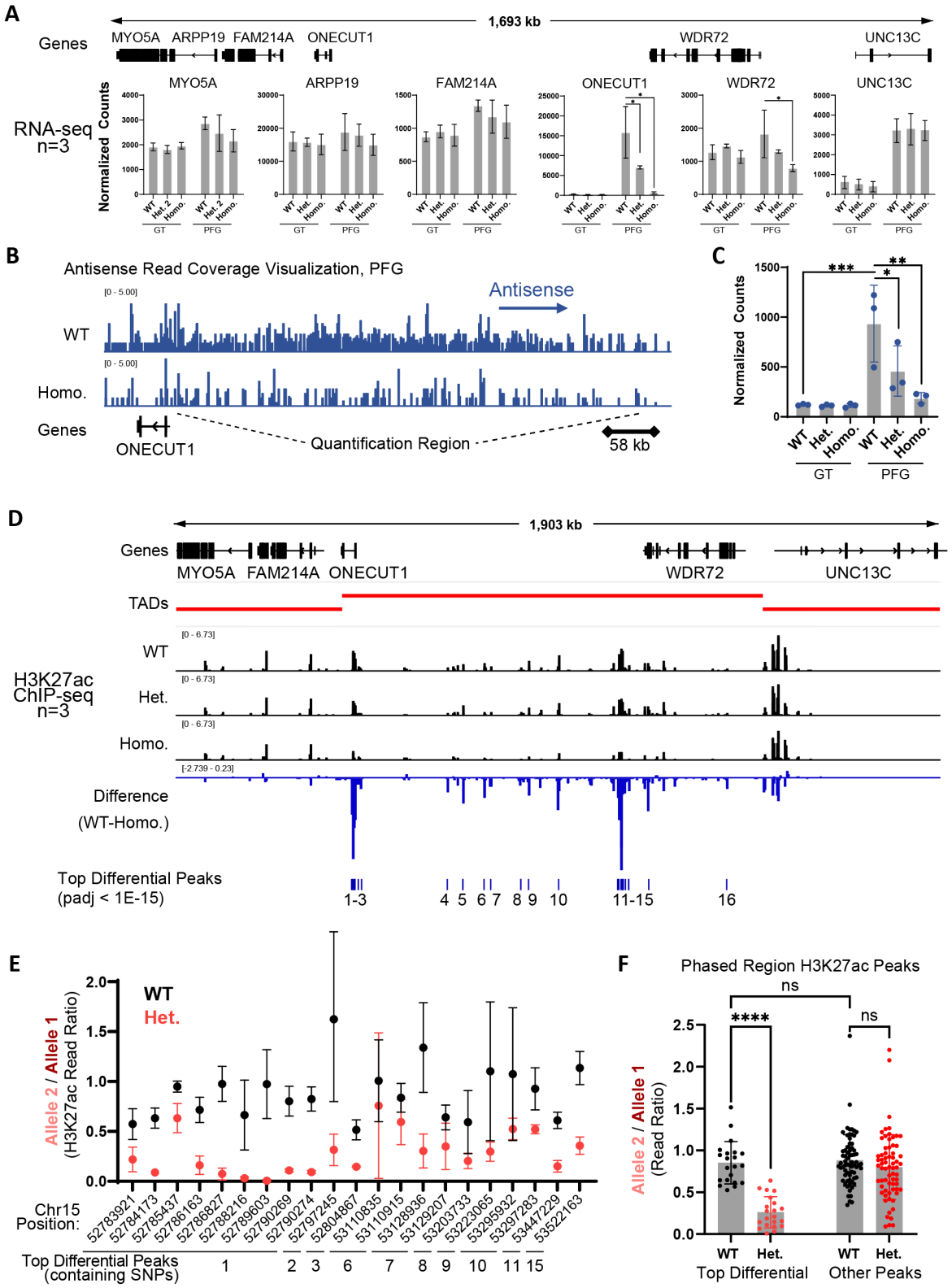

A
1,693 kb
Genes
MYO5A ARPP19 FAM214A ONECUT1
WDR72
UNC13C
RNA-seq
n=3
Normalized Counts
MYO5A
ARPP19
FAM214A
ONECUT1
WDR72
UNC13C
WT
Het. 2
Homo.
GT
PFG
B
Antisense Read Coverage Visualization, PFG
Antisense
[0 - 5.00]
WT
Homo.
Genes
ONECUT1
Quantification Region
58 kb
C
Normalized Counts
***
**
*
WT
Het.
Homo.
GT
PFG
D
1,903 kb
Genes
MYO5A FAM214A ONECUT1
WDR72
UNC13C
TADs
H3K27ac
ChIP-seq
n=3
WT
Het.
Homo.
[0 - 6.73]
[-2.739 - 0.23]
Difference
(WT-Homo.)
Top Differential Peaks
(padj < 1E-15)
1-3
4 5 6 7 8 9 10
11-15
16
E
WT
Het.
Allele 2 / Allele 1
(H3K27ac Read Ratio)
Chr15
Position:
52783921
52784173
52785437
52786163
52786827
52788216
52789603
52790269
52790274
52797245
52804867
53110835
53110915
53128936
53129207
53203733
53223065
53295932
53297283
53447229
53522163
Top Differential Peaks
(containing SNPs)
1 2 3 6 7 8 9 10 11 15
F
Phased Region H3K27ac Peaks
ns
****
ns
Allele 2 / Allele 1
(Read Ratio)
WT
Het.
Top Differential
Other Peaks

Fig. 3: Transcriptional and epigenetic consequences of *ONECUT1e-664kb* knockout

A) RNA-seq normalized read counts for genes in *ONECUT1* locus at GT and PFG stages from WT, heterozygous (het. 2), and homozygous enhancer knockout cells. * indicates padj<0.005 computed by DESeq2.

B) Visualization of *ONECUT1* promoter antisense RNA read coverage from strand-specific RNA-seq of PFG-stage WT and homozygous enhancer knockout cells (representative differentiation replicate).

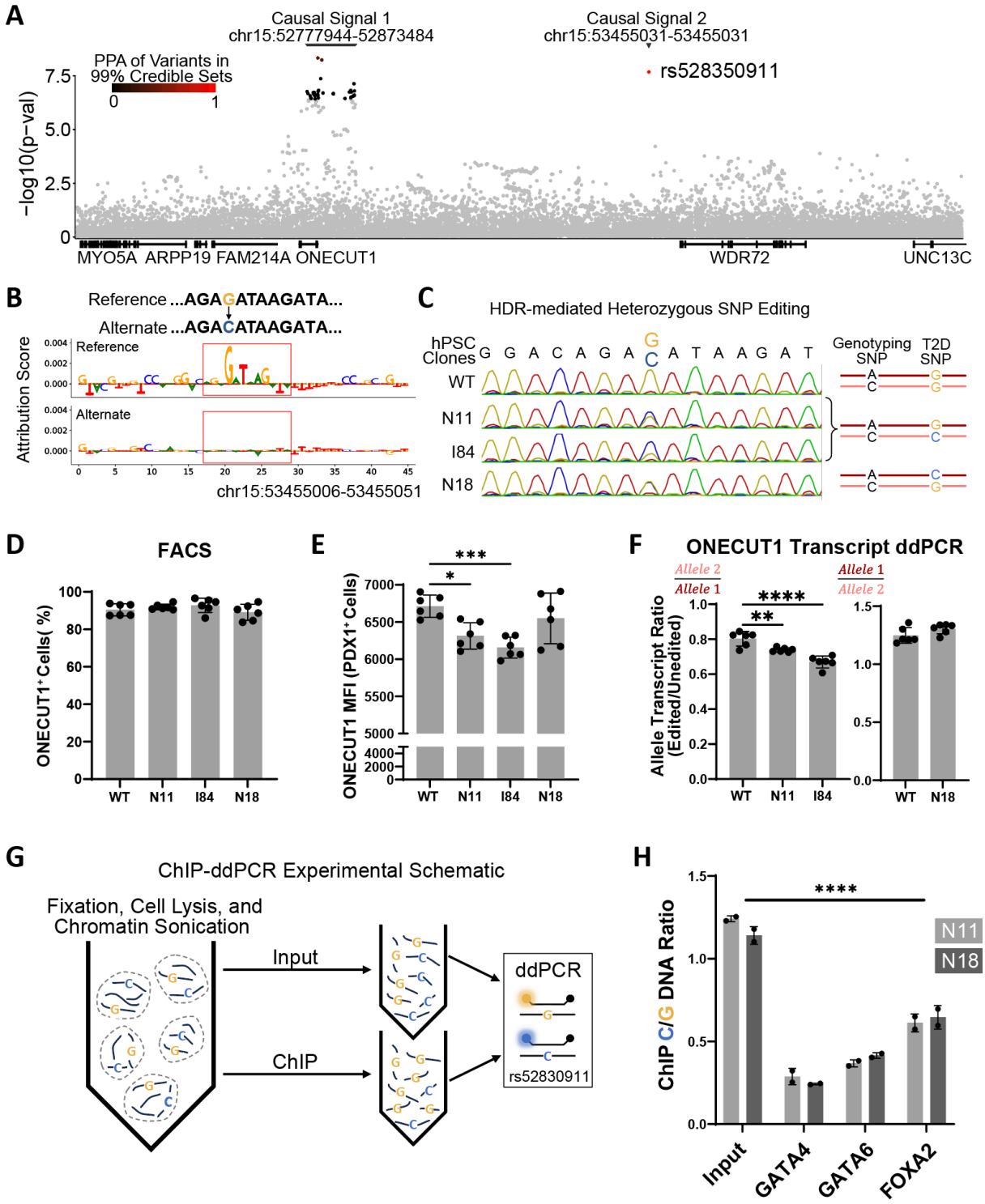
C) Quantification of (B) from three differentiations. One-way analysis of variance (ANOVA) followed by Dunnett multiple comparisons test versus WT control.

D) H3K27ac ChIP-seq at PFG stage of WT, heterozygous, and homozygous enhancer knockout cells. TADs called from PFG stage Hi-C data. Difference calculated via MACS2 subtract. Differential peaks between WT and homozygous enhancer knockout cells quantified with DESeq2. *ONECUT1e-664kb* deletion region is within peak #12.

E) H3K27ac is decreased in an allele-specific manner within significantly affected H3K27ac peaks.

F) Quantification of average H3K27ac read ratios at all heterozygous SNP positions in phased locus in top significantly different peaks (WT vs Homo, padj<1E-15) and other peaks. Each point represents a heterozygous SNP position, average of three replicates. Ordinary one-way ANOVA with Šidák's multiple comparisons test, with a single pooled variance.

835 **Fig. 4:**



836

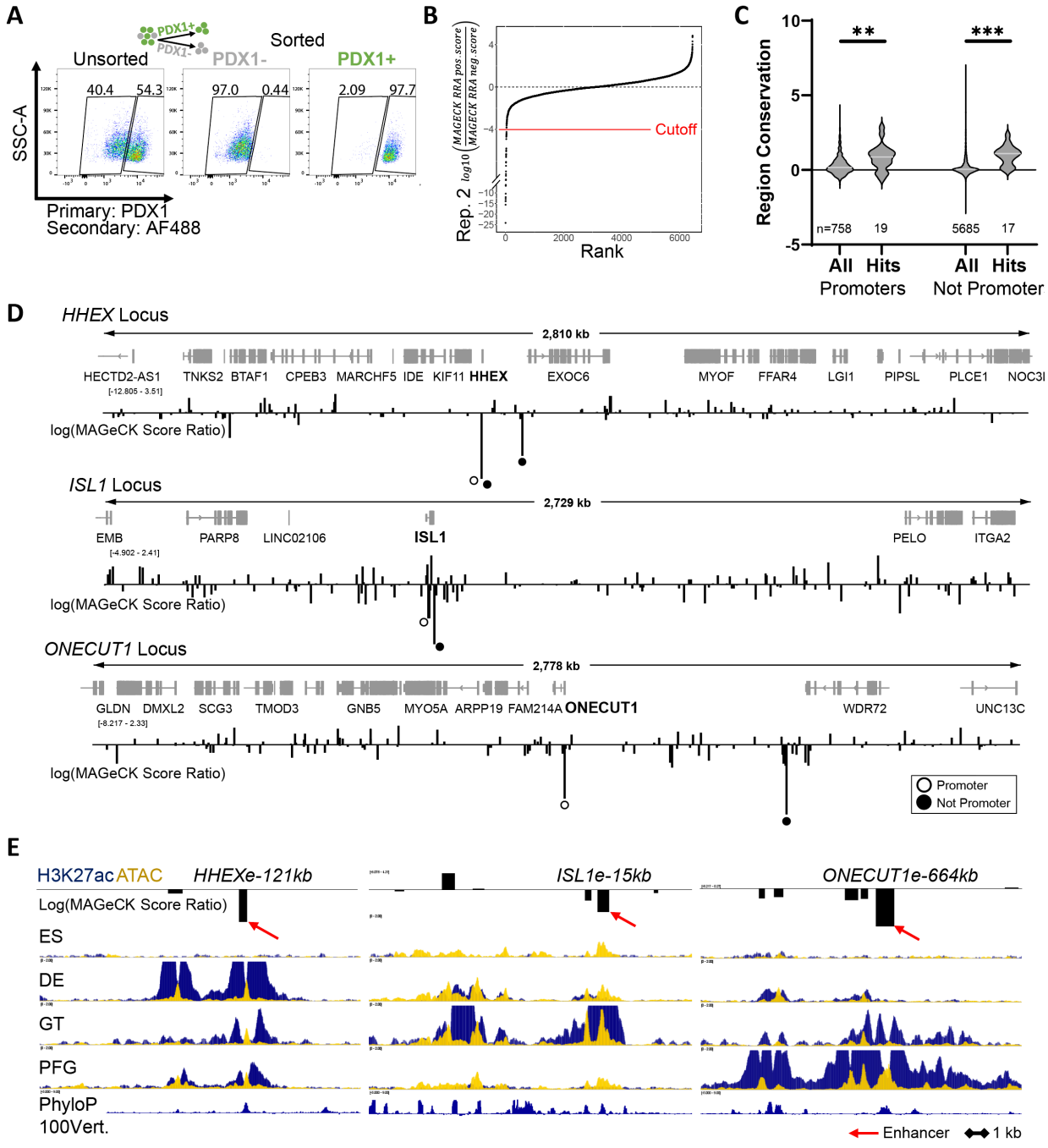
837

838 **Fig. 4: T2D-associated SNP hPSC modeling and pancreatic differentiation characterization**
839 A) LocusZoom plot of variants from Mahajan et al (2018b) T2D GWAS meta-analysis
840 unadjusted for BMI. Variants within credible sets are colored by the PPA score.
841 B) Decreased DeepLift attribution score at GATA4 motif upon introduction of variant.
842 Red box represents a GATA4/6 binding motif hit by FIMO ($p < 0.001$).
843 C) Spectrogram of disease-associated SNP sequencing and schematic demonstrating
844 allele that contains the introduced SNP.
845 D) Proportion of ONECUT1+ cells at PFG stage. Each symbol represents one
846 independent experiment (n = 6 independent experiments) and data are presented as the
847 mean ± s.d. One-way analysis of variance (ANOVA) followed by Dunnett multiple
848 comparisons test versus WT control.
849 E) ONECUT1 MFI of PDX1+ cells. Each symbol represents one independent experiment
850 (n = 6 independent experiments) and data are presented as the mean ± s.d. One-way
851 analysis of variance (ANOVA) followed by Dunnett multiple comparisons test versus
852 WT control.
853 F) *ONECUT1* transcript ratio of edited allele divided by unedited allele. Each symbol
854 represents one independent differentiation (n = 6 independent experiments), average of
855 two technical replicates per differentiation. Data are presented as the mean ± s.d. One-
856 way analysis of variance (ANOVA) followed by Dunnett multiple comparisons test
857 versus WT control.
858 G) ChIP-ddPCR schematic.
859 H) ChIPed DNA ratio.Each symbol represents one independent differentiation and ChIP
860 experiment, average of two technical replicates per differentiation. Two-way analysis of
861 variance (ANOVA) followed by Tukey's multiple comparisons test versus WT control.
862

Supplemental Information

Supplementary Figures:

Sup. Fig. 1



868 **Sup. Fig. 1, Related to Figure 1: Supplemental results and analysis for CRISPRi screens**

869 A) Representative flow cytometry plot pre- and post-sort PDX1 sorting.

870 B) Rank plot of replicate 2 MAGeCK scores with FDR < 0.1

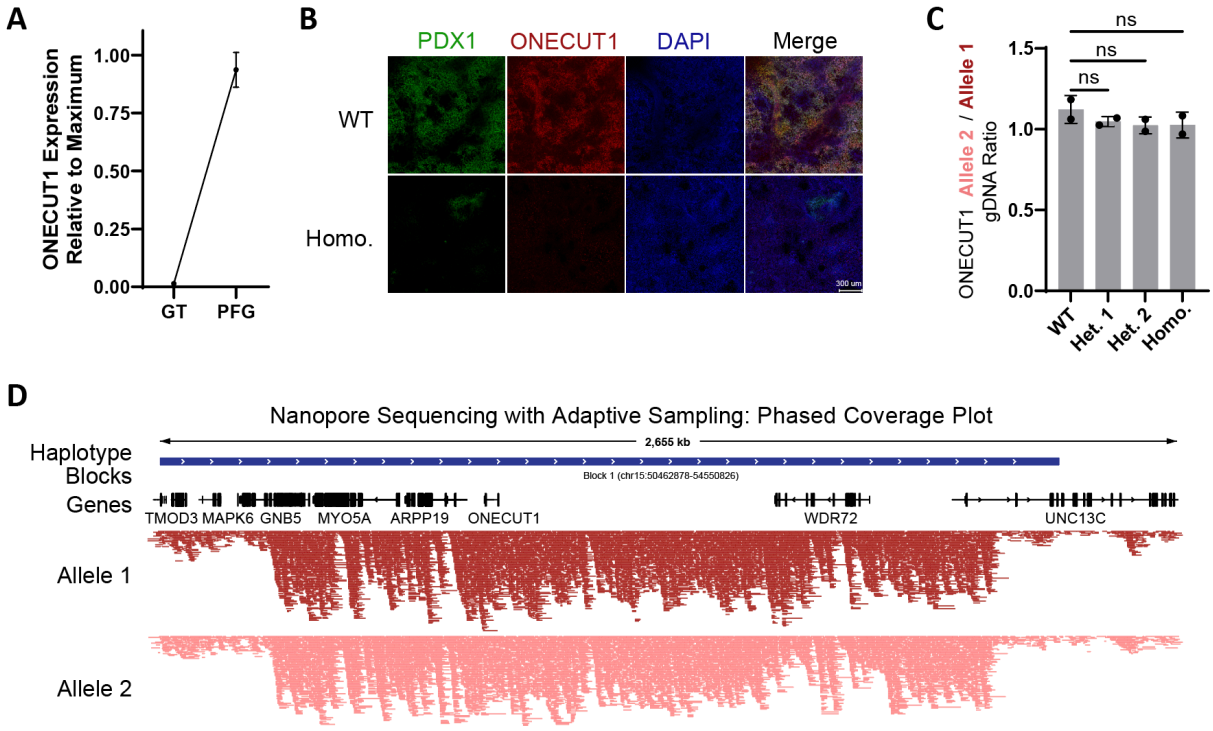
871 C) Calculated conservation scores for each region within indicated group. Ordinary one-
872 way ANOVA with Šidák's multiple comparisons test, with a single pooled variance.

873 D) Representative IGV plots for enhancer-gene pair rationale. Each bar represents a
874 MAGeCK score from replicate 2. A negative bar indicates a region caused net decrease
875 in the proportion of PDX1+ cells, with a larger bar indicating a more significant change
876 in PDX1 enrichment. Hollow black points represent promoter region hits, filled black
877 points represent non-promoter region hits.

878 E) H3K27ac ChIP-seq and ATAC-seq at representative enhancer loci.

879

880 **Sup. Fig. 2**

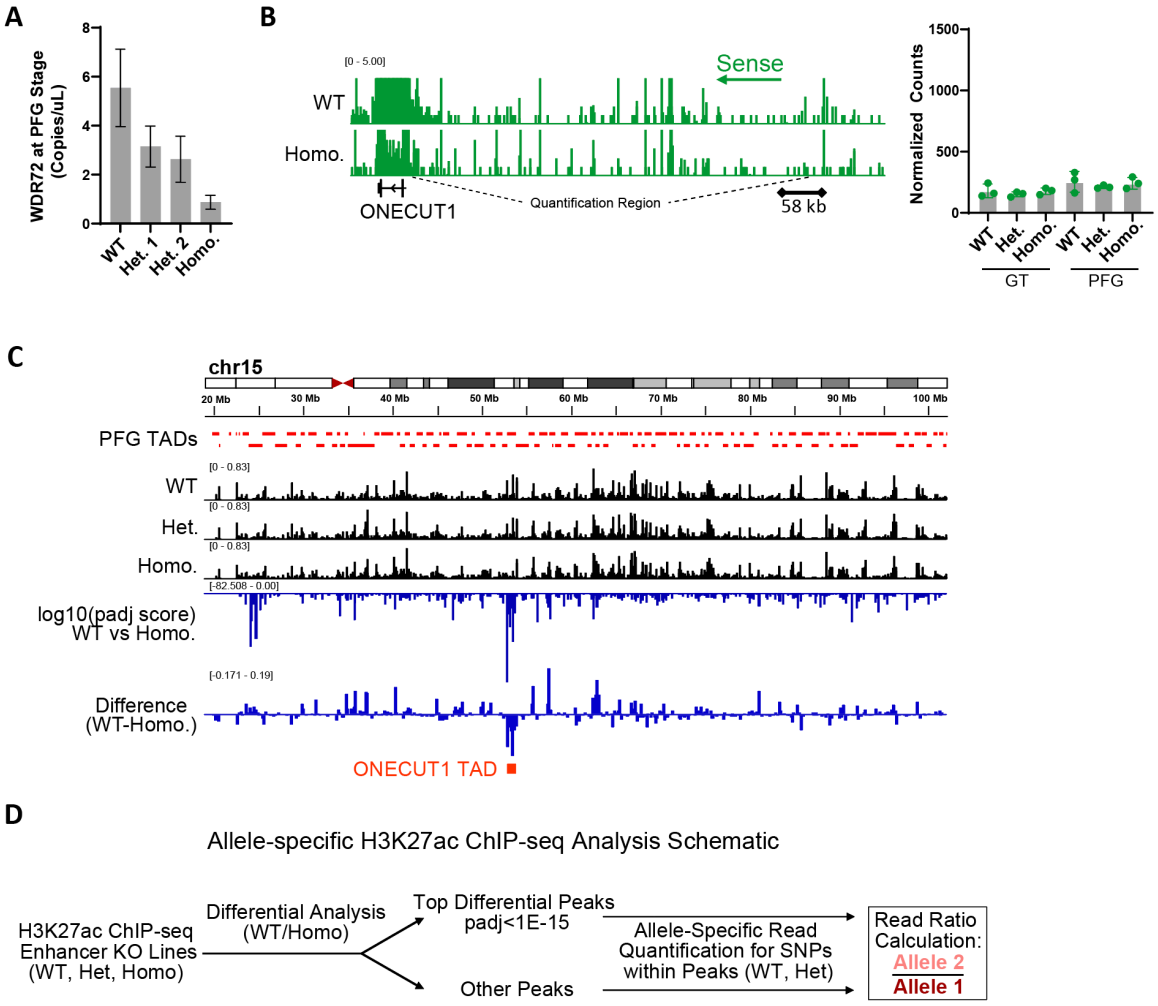


881

882 **Sup. Fig. 2, Related to Figure 2: *ONECUT1e-664kb* deletion supplemental results and**
883 **analysis**

884 A) *ONECUT1* transcript ddPCR on RNA from WT cells at GT and PFG stages.
885 B) PFG-stage immunofluorescence for PDX1 and ONECUT1 of WT and homozygous
886 enhancer knockout cells.
887 C) *ONECUT1* transcript ddPCR on DNA from WT and enhancer knockout cells. One-
888 way analysis of variance (ANOVA) followed by Dunnett multiple comparisons test
889 versus WT control.
890 D) Plot of phased reads and haplotype blocks from nanopore sequencing of H1 hPSCs
891 with adaptive sampling.
892

893 **Sup. Fig. 3**

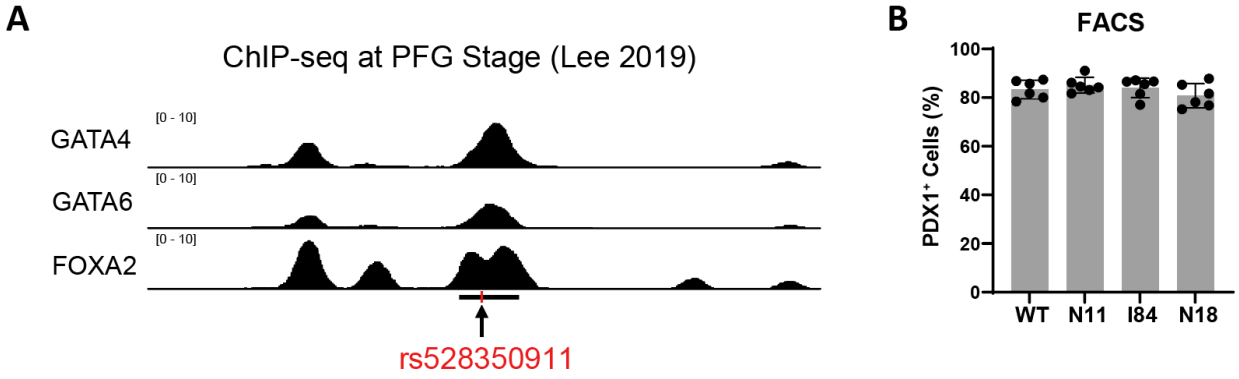


894

895 **Sup. Fig. 3, Related to Figure 3: Supplemental RNA-seq and H3K27ac ChIP-seq results**
896 **and analysis**

897 A) Non allele-specific *WDR72* ddPCR results of WT and enhancer knockout cells
898 differentiated to the PFG stage.
899 B) Visualization of sense read coverage in region telomeric to *ONECUT1* from strand-
900 specific RNA-seq of PFG-stage WT and homozygous enhancer knockout cells
901 (representative differentiation replicate). Quantification from three differentiations. One-
902 way analysis of variance (ANOVA) followed by Dunnett multiple comparisons test
903 versus WT control.
904 C) Full chromosome view of H3K27ac in WT and enhancer knockout cells showing the
905 *ONECUT1* TAD has the most significantly affected regions, as well as the regions with
906 the greatest magnitude loss of H3K27ac.
907 D) Allele-specific H3K27ac ChIP-seq analysis schematic.

908 **Sup. Fig. 4**



909

910 **Sup. Fig. 4, Related to Figure 4: Supplemental results for T2D-associated SNP hPSC**
911 **modeling**

912 A) Visualization of ChIP-seq at *ONECUT1-e664kb* locus conducted at PFG stage. Black
913 bar indicates *ONECUT1e-664kb*.
914 B) Bar plots of proportion of cells achieving a PDX1+ identity. Each symbol represents
915 one independent experiment (n = 6 independent experiments) and data are presented as
916 the mean ± s.d. One-way analysis of variance (ANOVA) followed by Dunnett multiple
917 comparisons test versus WT control.
918

919 **Supplementary Tables:**

920 **Table S1:** Enhancer screen region selection and analysis results. Sheet 1: Transcription factor
921 loci selected for CRISPRi interrogation. Sheet 2: ATAC-seq accessible regions selected for
922 interrogation. Sheet 3: Read abundance quantification of each gRNA. Sheet 4: MAGeCK-RRA
923 region-level analysis of two screens. Sheet 5: Top screen hits, FDR<0.1 in screen 2.

924 **Table S2:** Engineered hPSC line allele genotypes.

925 **Table S3:** List of gRNA target sequences and primer sequences for PCR genotyping.

926 **Table S4:** Antibody target, company or lab that supplied antibody, catalog number, application,
927 and quantity or dilution used are indicated. For ChIP experiments, the amount used for cells from
928 around half of a 15cm plate (roughly 40-50 million cells) is indicated.

929 **Table S5:** ddPCR assay details.

930 **Table S6:** RNA-seq conducted on WT, heterozygous (het. 2) and homozygous *ONECUT1e-*
931 *664kb* enhancer knockout cells. Sheet 1: DESeq2 normalized gene counts. Sheet 2: DESeq2
932 differential analysis between WT and heterozygous enhancer knockout cells at PFG stage. Sheet
933 3: DESeq2 differential analysis between WT and homozygous enhancer knockout cells at PFG
934 stage.

935 **Table S7:** H3K27ac ChIP-seq conducted on PFG-stage WT, heterozygous (het. 2), and
936 homozygous *ONECUT1e-664kb* enhancer knockout cells. Sheet 1: Raw integer counts of
937 H3K27ac reads. Sheet 2: DESeq2 calculated differential statistics between pairwise comparisons
938 among WT, heterozygous, and homozygous enhancer knockout cells. Sheet 3: Phased read
939 counts at indicated position, read ratios for each replicate, average read ratios for each genotype.
940 Top_Differential_Peak indicates whether the position falls within a top differential peak (WT vs
941 Homo, padj>1E-15).

942 # References

943 1. Katsanis, N. (2016). The continuum of causality in human genetic disorders. Genome Biology *17*.
944 10.1186/s13059-016-1107-9.
945 2. Lemelman, M.B., Letourneau, L., and Greeley, S.A.W. (2018). Neonatal Diabetes Mellitus: An
946 Update on Diagnosis and Management. Clinics in Perinatology.
947 3. Croucha, D.J.M., and Bodmer, W.F. (2020). Polygenic inheritance, GWAS, polygenic risk scores,
948 and the search for functional variants. Proceedings of the National Academy of Sciences of the
949 United States of America.
950 4. Zhang, F., and Lupski, J.R. (2015). Non-coding genetic variants in human disease. Human
951 Molecular Genetics.
952 5. Edwards, S.L., Beesley, J., French, J.D., and Dunning, M. (2013). Beyond GWASs: Illuminating the
953 dark road from association to function. American Journal of Human Genetics.
954 6. Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill,
955 R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing
956 limb and fin and is associated with preaxial polydactyly. Human Molecular Genetics.
957 10.1093/hmg/ddg180.
958 7. Claringbould, A., and Zaugg, J.B. (2021). Enhancers in disease: molecular basis and emerging
959 treatment strategies. Trends in Molecular Medicine.
960 8. Luo, R., Yan, J., Oh, J.W., Xi, W., Shigaki, D., Wong, W., Cho, H.S., Murphy, D., Cutler, R., Rosen,
961 B.P., et al. (2023). Dynamic network-guided CRISPRi screen identifies CTCF-loop-constrained
962 nonlinear enhancer gene regulatory activity during cell state transitions. Nature Genetics *55*.
963 10.1038/s41588-023-01450-7.
964 9. Shukla, A., and Huangfu, D. (2018). Decoding the noncoding genome via large-scale CRISPR
965 screens. Current Opinion in Genetics and Development.
966 10. Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer–promoter contacts in gene
967 expression control. Nature Reviews Genetics.
968 11. Burgos, J.I., Vallier, L., and Rodríguez-Seguí, S.A. (2021). Monogenic Diabetes Modeling: In Vitro
969 Pancreatic Differentiation From Human Pluripotent Stem Cells Gains Momentum. Frontiers in
970 Endocrinology.
971 12. Miguel-Escalada, I., Maestro, M.Á., Balboa, D., Elek, A., Bernal, A., Bernardo, E., Grau, V., García-
972 Hurtado, J., Sebé-Pedrós, A., and Ferrer, J. (2022). Pancreas agenesis mutations disrupt a lead
973 enhancer controlling a developmental enhancer cluster. Developmental Cell *57*.
974 10.1016/j.devcel.2022.07.014.
975 13. Weedon, M.N., Cebola, I., Patch, A.-M., Flanagan, S.E., Franco, E.D., Caswell, R., Rodríguez-Seguí,
976 S.A., Shaw-Smith, C., Cho, C.H.H., Allen, H.L., et al. (2013). Recessive mutations in a distal PTF1A
977 enhancer cause isolated pancreatic agenesis. Nature Genetics 2013 46:1 *46*, 61-64.
978 10.1038/ng.2826.
979 14. Franco, E.D., Flanagan, S.E., Houghton, J.A.L., Allen, H.L., MacKay, D.J.G., Temple, I.K., Ellard, S.,
980 and Hattersley, A.T. (2015). The effect of early, comprehensive genomic testing on clinical care
981 in neonatal diabetes: An international cohort study. The Lancet *386*. 10.1016/S0140-
982 6736(15)60098-8.
983 15. Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of
984 validated and target-linked human enhancers. Nature Reviews Genetics.
985 16. Yan, J., and Huangfu, D. (2022). Epigenome rewiring in human pluripotent stem cells. Trends in
986 Cell Biology.
987 17. Yang, D., Cho, H., Tayyebi, Z., Shukla, A., Luo, R., Dixon, G., Ursu, V., Stransky, S., Tremmel, D.M.,
988 Sackett, S.D., et al. (2022). CRISPR screening uncovers a central requirement for HHEX in

989 pancreatic lineage commitment and plasticity restriction. Nature Cell Biology *24*.
990 10.1038/s41556-022-00946-4.
991 18. Li, Q.V., Dixon, G., Verma, N., Rosen, B.P., Gordillo, M., Luo, R., Xu, C., Wang, Q., Soh, C.L., Yang,
992 D., et al. (2019). Genome-scale screens identify JNK–JUN signaling as a barrier for pluripotency
993 exit and endoderm differentiation. Nature Genetics *51*. 10.1038/s41588-019-0408-9.
994 19. Dixon, G., Pan, H., Yang, D., Rosen, B.P., Jashari, T., Verma, N., Pulecio, J., Caspi, I., Lee, K.,
995 Stransky, S., et al. (2021). QSER1 protects DNA methylation valleys from de novo methylation.
996 Science *372*. 10.1126/science.abd0875.
997 20. Xu, X., Du, Y., Ma, L., Zhang, S., Shi, L., Chen, Z., Zhou, Z., Hui, Y., Liu, Y., Fang, Y., et al. (2021).
998 Mapping germ-layer specification preventing genes in hPSCs via genome-scale CRISPR screening.
999 iScience *24*. 10.1016/j.isci.2020.101926.
1000 21. Yilmaz, A., Braverman-Gross, C., Bialer-Tsypin, A., Peretz, M., and Benvenisty, N. (2020).
1001 Mapping Gene Circuits Essential for Germ Layer Differentiation via Loss-of-Function Screens in
1002 Haploid Human Embryonic Stem Cells. Cell Stem Cell *27*. 10.1016/j.stem.2020.06.023.
1003 22. Naxerova, K., Stefano, B.D., Makofske, J.L., Watson, E.V., de Kort, M.A., Martin, T.D., Dezfulian,
1004 M., Ricken, D., Wooten, E.C., Kuroda, M.I., et al. (2021). Integrated loss- And gain-of-function
1005 screens define a core network governing human embryonic stem cell behavior. Genes and
1006 Development *35*. 10.1101/GAD.349048.121.
1007 23. Armendariz, D.A., Goetsch, S.C., Sundarrajan, A., Sivakumar, S., Wang, Y., Xie, S., Munshi, N.V.,
1008 and Hon, G.C. (2023). CHD-associated enhancers shape human cardiomyocyte lineage
1009 commitment. eLife *12*. 10.7554/eLife.86206.
1010 24. Haswell, J.R., Mattioli, K., Gerhardinger, C., Maass, P.G., Foster, D.J., Peinado, P., Wang, X.,
1011 Medina, P.P., Rinn, J.L., and Slack, F.J. (2021). Genome-wide CRISPR interference screen
1012 identifies long non-coding RNA loci required for differentiation and pluripotency. PLoS ONE *16*.
1013 10.1371/journal.pone.0252848.
1014 25. Lee, K., Cho, H., Rickert, R.W., Li, Q.V., Pulecio, J., Leslie, C.S., and Huangfu, D. (2019). FOXA2 Is
1015 Required for Enhancer Priming during Pancreatic Differentiation. Cell Reports.
1016 10.1016/j.celrep.2019.06.034.
1017 26. Balboa, D., Iworima, D.G., and Kieffer, T.J. (2021). Human Pluripotent Stem Cells to Model Islet
1018 Defects in Diabetes. Frontiers in Endocrinology.
1019 27. Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J.,
1020 Steinthorsdottir, V., Scott, R.A., and Grarup, N. (2018). Fine-mapping type 2 diabetes loci to
1021 single-variant resolution using high-density imputation and islet-specific epigenome maps.
1022 Nature genetics *50*, 1505-1513.
1023 28. Philippi, A., Heller, S., Costa, I.G., Senée, V., Breunig, M., Li, Z., Kwon, G., Russell, R., Illing, A., Lin,
1024 Q., et al. (2021). Mutations and variants of ONECUT1 in diabetes. Nature Medicine *27*.
1025 10.1038/s41591-021-01502-7.
1026 29. Russ-Silsby, J., Patel, K.A., Laver, T.W., Hawkes, G., Johnson, M.B., Wakeling, M.N., Patil, P.P.,
1027 Hattersley, A.T., Flanagan, S.E., Weedon, M.N., and Franco, E.D. (2023). The Role of ONECUT1
1028 Variants in Monogenic and Type 2 Diabetes Mellitus. Diabetes *72*. 10.2337/db23-0498.
1029 30. Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L.M.,
1030 and Wright, C.V.E. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the
1031 rostral duodenum. Development *122*. 10.1242/dev.122.3.983.
1032 31. Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994). Insulin-promoter-factor 1 is required
1033 for pancreas development in mice. Nature *371*. 10.1038/371606a0.
1034 32. Stoffers, D.A., Zinkin, N.T., Stanojevic, V., Clarke, W.L., and Habener, J.F. (1997). Pancreatic
1035 agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence.
1036 Nature Genetics *15*. 10.1038/ng0197-106.

1037 33. Stoffers, D.A., Ferrer, J., Clarke, W.L., and Habener, J.F. (1997). Early-onset type-ll diabetes
1038 mellitus (MODY4) linked to IPF1. Nature genetics *17*, 138-139.
1039 34. Steinthorsdottir, V., Thorleifsson, G., Sulem, P., Helgason, H., Grarup, N., Sigurdsson, A.,
1040 Helgadottir, H.T., Johannsdottir, H., Magnusson, O.T., Gudjonsson, S.A., et al. (2014).
1041 Identification of low-frequency and rare sequence variants associated with elevated or reduced
1042 risk of type 2 diabetes. Nature Genetics *46*. 10.1038/ng.2882.
1043 35. Smith, S.B., Qu, H.-Q., Taleb, N., Kishimoto, N.Y., Scheel, D.W., Lu, Y., Patch, A.-M., Grabs, R.,
1044 Wang, J., and Lynn, F.C. (2010). Rfx6 directs islet formation and insulin production in mice and
1045 humans. Nature *463*, 775-780.
1046 36. Kropp, P.A., Zhu, X., and Gannon, M. (2019). Regulation of the Pancreatic Exocrine
1047 Differentiation Program and Morphogenesis by Onecut 1/Hnf6. CMGH *7*.
1048 10.1016/j.jcmgh.2019.02.004.
1049 37. Kim, S.K., Selleri, L., Lee, J.S., Zhang, A.Y., Gu, X., Jacobs, Y., and Cleary, M.L. (2002). Pbx1
1050 inactivation disrupts pancreas development and in Ipf1-deficient mice promotes diabetes
1051 mellitus. Nature Genetics *30*. 10.1038/ng860.
1052 38. Ahlgren, U., Pfaff, S.L., Jessell, T.M., Edlund, T., and Edlund, H. (1997). Independent requirement
1053 for ISL1 in formation of pancreatic mesenchyme and islet cells. Nature *385*. 10.1038/385257a0.
1054 39. Lee, C.S., Sund, N.J., Vatamaniuk, M.Z., Matschinsky, F.M., Stoffers, D.A., and Kaestner, K.H.
1055 (2002). Foxa2 controls Pdx1 gene expression in pancreatic β-cells in vivo. Diabetes *51*.
1056 10.2337/diabetes.51.8.2546.
1057 40. Barbera, J.P.M., Clements, M., Thomas, P., Rodriguez, T., Meloy, D., Kioussis, D., and
1058 Beddington, R.S.P. (2000). The homeobox gene Hex is required in definitive endodermal tissues
1059 for normal forebrain, liver and thyroid formation. Development *127*. 10.1242/dev.127.11.2433.
1060 41. Bort, R., Martinez-Barbera, J.P., Beddington, R.S.P., and Zaret, K.S. (2004). Hex homeobox gene-
1061 dependent tissue positioning is required for organogenesis of the ventral pancreas.
1062 Development *131*. 10.1242/dev.00965.
1063 42. Decker, K., Goldman, D.C., Grasch, C.L., and Sussel, L. (2006). Gata6 is an important regulator of
1064 mouse pancreas development. Developmental Biology *298*. 10.1016/j.ydbio.2006.06.046.
1065 43. Xuan, S., Borok, M.J., Decker, K.J., Battle, M.A., Duncan, S.A., Hale, M.A., Macdonald, R.J., and
1066 Sussel, L. (2012). Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic
1067 agenesis. Journal of Clinical Investigation *122*. 10.1172/JCI63352.
1068 44. Carrasco, M., Delgado, I., Soria, B., Martín, F., and Rojas, A. (2012). GATA4 and GATA6 control
1069 mouse pancreas organogenesis. Journal of Clinical Investigation *122*. 10.1172/JCI63240.
1070 45. Watt, A.J., Zhao, R., Li, J., and Duncan, S.A. (2007). Development of the mammalian liver and
1071 ventral pancreas is dependent on GATA4. BMC Developmental Biology *7*. 10.1186/1471-213X-7-
1072 37.
1073 46. Spence, J.R., Lange, A.W., Lin, S.C.J., Kaestner, K.H., Lowy, A.M., Kim, I., Whitsett, J.A., and Wells,
1074 J.M. (2009). Sox17 Regulates Organ Lineage Segregation of Ventral Foregut Progenitor Cells.
1075 Developmental Cell *17*. 10.1016/j.devcel.2009.05.012.
1076 47. Lorberbaum, D.S., Kishore, S., Rosselot, C., Sarbaugh, D., Brooks, E.P., Aragon, E., Xuan, S.,
1077 Simon, O., Ghosh, D., Mendelsohn, C., et al. (2020). Retinoic acid signaling within pancreatic
1078 endocrine progenitors regulates mouse and human β cell specification. Development
1079 (Cambridge) *147*. 10.1242/dev.189977.
1080 48. Haumaitre, C., Barbacci, E., Jenny, M., Ott, M.O., Gradwohl, G., and Cereghini, S. (2005). Lack of
1081 TCF2/vHNF1 in mice leads to pancreas agenesis. Proceedings of the National Academy of
1082 Sciences of the United States of America *102*. 10.1073/pnas.0405776102.
1083 49. Vanhorenbeeck, V., Jenny, M., Cornut, J.F., Gradwohl, G., Lemaigre, F.P., Rousseau, G.G., and
1084 Jacquemin, P. (2007). Role of the Onecut transcription factors in pancreas morphogenesis and in

1085 pancreatic and enteric endocrine differentiation. Developmental Biology *305*.
1086 10.1016/j.ydbio.2007.02.027.
1087 50. Nishimura, W., Kondo, T., Salameh, T., Khattabi, I.E., Dodge, R., Bonner-Weir, S., and Sharma, A.
1088 (2006). A switch from MafB to MafA expression accompanies differentiation to pancreatic β-
1089 cells. Developmental Biology *293*. 10.1016/j.ydbio.2006.02.028.
1090 51. Osipovich, A.B., Long, Q., Manduchi, E., Gangula, R., Hipkens, S.B., Schneider, J., Okubo, T.,
1091 Stoeckert, C.J., Takada, S., and Magnuson, M.A. (2014). Insm1 promotes endocrine cell
1092 differentiation by modulating the expression of a network of genes that includes Neurog3 and
1093 Ripply3. Development (Cambridge) *141*. 10.1242/dev.104810.
1094 52. Nissim, S., Weeks, O., Talbot, J.C., Hedgepeth, J.W., Wucherpfennig, J., Schatzman-Bone, S.,
1095 Swinburne, I., Cortes, M., Alexa, K., Megason, S., et al. (2016). Iterative use of nuclear receptor
1096 Nr5a2 regulates multiple stages of liver and pancreas development. Developmental Biology *418*.
1097 10.1016/j.ydbio.2016.07.019.
1098 53. Jacquemin, P., Lemaigre, F.P., and Rousseau, G.G. (2003). The Onecut transcription factor HNF-6
1099 (OC-1) is required for timely specification of the pancreas and acts upstream of Pdx-1 in the
1100 specification cascade. Developmental Biology *258*. 10.1016/S0012-1606(03)00115-5.
1101 54. PanelApp (2024). Genomics England PanelApp;https://panelapp.genomicsengland.co.uk (date
1102 accessed), Monogenic diabetes (Version 2.54).
1103 55. Horikoshi, M., Hara, K., Ito, C., Shojima, N., Nagai, R., Ueki, K., Froguel, P., and Kadowaki, T.
1104 (2007). Variations in the HHEX gene are associated with increased risk of type 2 diabetes in the
1105 Japanese population. Diabetologia *50*. 10.1007/s00125-007-0827-5.
1106 56. Shimomura, H., Sanke, T., Hanabusa, T., Tsunoda, K., Furuta, H., and Nanjo, K. (2000). Nonsense
1107 mutation of Islet-1 (Q310X) found in a type 2 diabetic patient with a strong family history.
1108 Diabetes *49*. 10.2337/diabetes.49.9.1597.
1109 57. Iacovazzo, D., Flanagan, S.E., Walker, E., Quezado, R., Barros, F.A.D.S., Caswell, R., Johnson, M.B.,
1110 Wakeling, M., Brändle, M., Guo, M., et al. (2018). MAFA missense mutation causes familial
1111 insulinomatosis and diabetes mellitus. Proceedings of the National Academy of Sciences of the
1112 United States of America *115*. 10.1073/pnas.1712262115.
1113 58. Ng, N.H.J., Jasmen, J.B., Lim, C.S., Lau, H.H., Krishnan, V.G., Kadiwala, J., Kulkarni, R.N., Ræder,
1114 H., Vallier, L., Hoon, S., and Teo, A.K.K. (2019). HNF4A Haploinsufficiency in MODY1 Abrogates
1115 Liver and Pancreas Differentiation from Patient-Derived Induced Pluripotent Stem Cells. iScience
1116 *16*. 10.1016/j.isci.2019.05.032.
1117 59. Haldorsen, I.S., Vesterhus, M., Ræder, H., Jensen, D.K., Søvik, O., Molven, A., and Njølstad, P.R.
1118 (2008). Lack of pancreatic body and tail in HNF1B mutation carriers. Diabetic Medicine *25*.
1119 10.1111/j.1464-5491.2008.02460.x.
1120 60. Shi, Z.D., Lee, K., Yang, D., Amin, S., Verma, N., Li, Q.V., Zhu, Z., Soh, C.L., Kumar, R., Evans, T., et
1121 al. (2017). Genome Editing in hPSCs Reveals GATA6 Haploinsufficiency and a Genetic Interaction
1122 with GATA4 in Human Pancreatic Development. Cell Stem Cell *20*. 10.1016/j.stem.2017.01.001.
1123 61. Trott, J., Alpagu, Y., Tan, E.K., Shboul, M., Dawood, Y., Elsy, M., Wollmann, H., Tano, V., Bonnard,
1124 C., Eng, S., et al. (2020). Mitchell-Riley syndrome iPSCs exhibit reduced pancreatic endoderm
1125 differentiation due to a mutation in RFX6. Development (Cambridge) *147*. 10.1242/dev.194878.
1126 62. Bolt, C.C., Lopez-Delisle, L., Hintermann, A., Mascrez, B., Rauseo, A., Andrey, G., and Duboule, D.
1127 (2022). Context-dependent enhancer function revealed by targeted inter-TAD relocation. Nature
1128 Communications *13*. 10.1038/s41467-022-31241-3.
1129 63. Morgens, D.W., Wainberg, M., Boyle, E.A., Ursu, O., Araya, C.L., Tsui, C.K., Haney, M.S., Hess,
1130 G.T., Han, K., Jeng, E.E., et al. (2017). Genome-scale measurement of off-target activity using
1131 Cas9 toxicity in high-throughput screens. Nature Communications *8*. 10.1038/ncomms15178.

1132 64. Sanson, K.R., Hanna, R.E., Hegde, M., Donovan, K.F., Strand, C., Sullender, M.E., Vaimberg, E.W.,
1133 Goodale, A., Root, D.E., Piccioni, F., and Doench, J.G. (2018). Optimized libraries for CRISPR-Cas9
1134 genetic screens with multiple modalities. Nature Communications *9*. 10.1038/s41467-018-
1135 07901-8.
1136 65. Pulecio, J., Tayyebi, Z., Liu, D., Wong, W., Luo, R., Damodaran, J.R., Kaplan, S., Cho, H., Yan, J.,
1137 Murphy, D., et al. (2023). Discovery of Competent Chromatin Regions in Human Embryonic Stem
1138 Cells. bioRxiv.
1139 66. Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu,
1140 X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale
1141 CRISPR/Cas9 knockout screens. Genome biology *15*, 1-12.
1142 67. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral
1143 substitution rates on mammalian phylogenies. Genome Research. 10.1101/gr.097857.109.
1144 68. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J.,
1145 Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from
1146 poised enhancers and predicts developmental state. Proceedings of the National Academy of
1147 Sciences of the United States of America. 10.1073/pnas.1016071107.
1148 69. Jacquemin, P., Durviaux, S.M., Jensen, J., Godfraind, C., Gradwohl, G., Guillemot, F., Madsen,
1149 O.D., Carmeliet, P., Dewerchin, M., Collen, D., et al. (2000). Transcription Factor Hepatocyte
1150 Nuclear Factor 6 Regulates Pancreatic Endocrine Cell Differentiation and Controls Expression of
1151 the Proendocrine Gene ngn3. Molecular and Cellular Biology *20*. 10.1128/mcb.20.12.4445-
1152 4454.2000.
1153 70. Pierreux, C.E., Poll, A.V., Kemp, C.R., Clotman, F., Maestro, M.A., Cordi, S., Ferrer, J., Leyns, L.,
1154 Rousseau, G.G., and Lemaigre, F.P. (2006). The Transcription Factor Hepatocyte Nuclear Factor-6
1155 Controls the Development of Pancreatic Ducts in the Mouse. Gastroenterology *130*.
1156 10.1053/j.gastro.2005.12.005.
1157 71. Zhang, H., Ables, E.T., Pope, C.F., Washington, M.K., Hipkens, S., Means, A.L., Path, G., Seufert, J.,
1158 Costa, R.H., Leiter, A.B., et al. (2009). Multiple, temporal-specific roles for HNF6 in pancreatic
1159 endocrine and ductal differentiation. Mechanisms of Development *126*.
1160 10.1016/j.mod.2009.09.006.
1161 72. Uyehara, C.M., and Apostolou, E. (2023). 3D enhancer-promoter interactions and multi-
1162 connected hubs: Organizational principles and functional roles. Cell Reports.
1163 73. Yang, F. (2022). Promoter antisense RNAs: beyond transcription by-products of active
1164 promoters. RNA biology *19*, 533-540.
1165 74. Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nature
1166 Structural and Molecular Biology.
1167 75. Mills, J.D., Kawahara, Y., and Janitz, M. (2013). Strand-Specific RNA-Seq Provides Greater
1168 Resolution of Transcriptome Profiling. Current Genomics *14*. 10.2174/1389202911314030003.
1169 76. Slieker, R.C., Donnelly, L.A., Fitipaldi, H., Bouland, G.A., Giordano, G.N., Åkerlund, M., Gerl, M.J.,
1170 Ahlqvist, E., Ali, A., Dragan, I., et al. (2021). Distinct Molecular Signatures of Clinical Clusters in
1171 People With Type 2 Diabetes: An IMI-RHAPSODY Study. Diabetes *70*. 10.2337/DB20-1281.
1172 77. Torres, J.M., Abdalla, M., Payne, A., Fernandez-Tajes, J., Thurner, M., Nylander, V., Gloyn, A.L.,
1173 Mahajan, A., and McCarthy, M.I. (2020). A Multi-omic Integrative Scheme Characterizes Tissues
1174 of Action at Loci Associated with Type 2 Diabetes. American Journal of Human Genetics *107*.
1175 10.1016/j.ajhg.2020.10.009.
1176 78. Brennan, K.J., Weilert, M., Krueger, S., Pampari, A., yun Liu, H., Yang, A.W.H., Morrison, J.A.,
1177 Hughes, T.R., Rushlow, C.A., Kundaje, A., and Zeitlinger, J. (2023). Chromatin accessibility in the
1178 Drosophila embryo is determined by transcription factor pioneering and enhancer activation.
1179 Developmental Cell *58*. 10.1016/j.devcel.2023.07.007.

1180 79. Pampari, A., Shcherbina, A., Nair, S., Schreiber, J., Patel, A., Wang, A., Kundu, S., Shrikumar, A.,
1181 and Kundaje, A. (2023). Bias factorized, base-resolution deep learning models of chromatin
1182 accessibility reveal cis-regulatory sequence syntax, transcription factor footprints and regulatory
1183 variants.
1184 80. Shrikumar, A., Greenside, P., and Kundaje, A. (2017). Learning important features through
1185 propagating activation differences.
1186 81. Shaw-Smith, C., Franco, E.D., Allen, H.L., Batlle, M., Flanagan, S.E., Borowiec, M., Taplin, C.E.,
1187 Velden, J.V.A.-V.D., Cruz-Rojo, J., Nanclares, G.P.D., et al. (2014). GATA4 mutations are a cause
1188 of neonatal and childhood-onset diabetes. Diabetes *63*. 10.2337/db14-0061.
1189 82. Yang, Y.P., Magnuson, M.A., Stein, R., and Wright, C.V.E. (2017). The mammal-specific Pdx1 area
1190 II enhancer has multiple essential functions in early endocrine cell specification and postnatal β-
1191 cell maturation. Development (Cambridge). 10.1242/dev.143123.
1192 83. Fujitani, Y., Fujitani, S., Boyer, D.F., Gannon, M., Kawaguchi, Y., Ray, M., Shiota, M., Stein, R.W.,
1193 Magnuson, M.A., and Wright, C.V.E. (2006). Targeted deletion of a cis-regulatory region reveals
1194 differential gene dosage requirements for Pdx1 in foregut organ differentiation and pancreas
1195 formation. Genes and Development. 10.1101/gad.1360106.
1196 84. Rojas, A., Schachterle, W., Xu, S.M., and Black, B.L. (2009). An endoderm-specific transcriptional
1197 enhancer from the mouse Gata4 gene requires GATA and homeodomain protein-binding sites
1198 for function in vivo. Developmental dynamics : an official publication of the American
1199 Association of Anatomists *238*. 10.1002/dvdy.22091.
1200 85. Kelly, M.R., Wisniewska, K., Regner, M.J., Lewis, M.W., Perreault, A.A., Davis, E.S., Phanstiel,
1201 D.H., Parker, J.S., and Franco, H.L. (2022). A multi-omic dissection of super-enhancer driven
1202 oncogenic gene expression programs in ovarian cancer. Nature Communications *13*.
1203 10.1038/s41467-022-31919-8.
1204 86. Long, H.K., Osterwalder, M., Welsh, I.C., Hansen, K., Davies, J.O.J., Liu, Y.E., Koska, M., Adams,
1205 A.T., Aho, R., Arora, N., et al. (2020). Loss of Extreme Long-Range Enhancers in Human Neural
1206 Crest Drives a Craniofacial Disorder. Cell Stem Cell *27*. 10.1016/j.stem.2020.09.001.
1207 87. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a long-
1208 range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation
1209 of the mouse limb. Development *132*. 10.1242/dev.01613.
1210 88. Uslu, V.V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N.A., Marioni, J.C., and Spitz, F. (2014).
1211 Long-range enhancers regulating Myc expression are required for normal facial morphogenesis.
1212 Nature Genetics *46*. 10.1038/ng.2971.
1213 89. Isoda, T., Moore, A.J., He, Z., Chandra, V., Aida, M., Denholtz, M., van Hamburg, J.P., Fisch, K.M.,
1214 Chang, A.N., Fahl, S.P., et al. (2017). Non-coding Transcription Instructs Chromatin Folding and
1215 Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. Cell *171*.
1216 10.1016/j.cell.2017.09.001.
1217 90. Osterwalder, M., Barozzi, I., Tissiéres, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A.,
1218 Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic
1219 robustness in mammalian development. Nature *554*. 10.1038/nature25461.
1220 91. Choi, H.I., Chai, J.C., Lee, Y.S., Jung, K.H., and Chai, Y.G. (2021). Targeting MYC-inducing
1221 enhancer-associated noncoding (MYC-IEANC) RNAs inhibits the proliferation of HCC cells.
1222 92. Luo, X., Liu, Y., Dang, D., Hu, T., Hou, Y., Meng, X., Zhang, F., Li, T., Wang, C., Li, M., et al. (2021).
1223 3D Genome of macaque fetal brain reveals evolutionary innovations during primate
1224 corticogenesis. Cell *184*. 10.1016/j.cell.2021.01.001.
1225 93. Gonen, N., Futtner, C.R., Wood, S., Garcia-Moreno, S.A., Salamone, I.M., Samson, S.C., Sekido,
1226 R., Poulat, F., Maatouk, D.M., and Lovell-Badge, R. (2018). Sex reversal following deletion of a
1227 single distal enhancer of Sox9. Science *360*. 10.1126/science.aas9408.

1228 94. Lascar, N., Brown, J., Pattison, H., Barnett, A.H., Bailey, C.J., and Bellary, S. (2018). Type 2
1229 diabetes in adolescents and young adults. The Lancet Diabetes and Endocrinology.
1230 95. Franks, P.W., Pearson, E., and Florez, J.C. (2013). Gene-environment and gene-treatment
1231 interactions in type 2 diabetes: Progress, pitfalls, and prospects. Diabetes Care.
1232 96. Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A.,
1233 Feeley, M.N., Sanson, K.R., Baidi, Y., et al. (2021). Massively parallel assessment of human
1234 variants with base editor screens. Cell *184*. 10.1016/j.cell.2021.01.012.
1235 97. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N.,
1236 Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-
1237 guided regulation of transcription in eukaryotes. Cell *154*, 442-442. 10.1016/j.cell.2013.06.044.
1238 98. Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N.,
1239 Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived
1240 in vitro from human pluripotent stem cells. Nature Biotechnology. 10.1038/nbt.3033.
1241 99. Gaspar, J.M. (2018). Genrich: detecting sites of genomic enrichment.
1242 100. Labun, K., Montague, T.G., Krause, M., Cleuren, Y.N.T., Tjeldnes, H., and Valen, E. (2019).
1243 CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids
1244 Research *47*. 10.1093/nar/gkz365.
1245 101. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for
1246 CRISPR screening. Nature Methods.
1247 102. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094-
1248 3100.
1249 103. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane,
1250 T., McCarthy, S.A., and Davies, R.M. (2021). Twelve years of SAMtools and BCFtools. GigaScience
1251 *10*. 10.1093/gigascience/giab008.
1252 104. Martin, M., Patterson, M., Garg, S., Fischer, S.O., Pisanti, N., Klau, G.W., Schöenhuth, A., and
1253 Marschall, T. (2016). WhatsHap: fast and accurate read-based phasing. bioRxiv.
1254 105. Edge, P., and Bansal, V. (2019). Longshot enables accurate variant calling in diploid genomes
1255 from single-molecule long read sequencing. Nature Communications *10*. 10.1038/s41467-019-
1256 12493-y.
1257 106. Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program
1258 for assigning sequence reads to genomic features. Bioinformatics *30*.
1259 10.1093/bioinformatics/btt656.
1260 107. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and
1261 Barillot, E. (2015). HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. Genome
1262 Biology *16*. 10.1186/s13059-015-0831-x.
1263 108. Kramer, N.E., Davis, E.S., Wenger, C.D., Deoudes, E.M., Parker, S.M., Love, M.I., and Phanstiel,
1264 D.H. (2022). Plotgardener: Cultivating precise multi-panel figures in R. Bioinformatics *38*.
1265 10.1093/bioinformatics/btac057.
1266 109. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
1267 dispersion for RNA-seq data with DESeq2. Genome Biology *15*. 10.1186/s13059-014-0550-8.
1268 110. Servant, N., Lajoie, B.R., Nora, E.P., Giorgetti, L., Chen, C.J., Heard, E., Dekker, J., and Barillot, E.
1269 (2012). HiTC: Exploration of high-throughput 'C' experiments. Bioinformatics *28*.
1270 10.1093/bioinformatics/bts521.
1271 111. Sahin, M., Wong, W., Zhan, Y., Deynze, K.V., Koche, R., and Leslie, C.S. (2021). HiC-DC+ enables
1272 systematic 3D interaction calls and differential analysis for Hi-C and HiChIP. Nature
1273 Communications *12*. 10.1038/s41467-021-23749-x.

1274 112. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient
1275 alignment of short DNA sequences to the human genome. Genome Biology *10*. 10.1186/gb-
1276 2009-10-3-r25.
1277 113. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers,
1278 R.M., Brown, M., and Li, W. (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology
1279 *9*, 1-9.
1280 114. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-
1281 throughput experiments. Annals of Applied Statistics *5*. 10.1214/11-AOAS466.
1282 115. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of
1283 Problematic Regions of the Genome. Scientific Reports *9*. 10.1038/s41598-019-45839-z.
1284 116. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and
1285 Mesirov, J.P. (2011). Integrative genomics viewer. Nature Biotechnology.
1286 117. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: Scanning for occurrences of a given motif.
1287 Bioinformatics *27*. 10.1093/bioinformatics/btr064.
1288 118. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi,
1289 H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic
1290 transcription factor sequence specificity. Cell *158*. 10.1016/j.cell.2014.08.009.

1291