Germline *cis* variant determines epigenetic regulation of the anti-cancer drug metabolism gene dihydropyrimidine dehydrogenase (*DPYD*)

Ting Zhang [1], Alisa Ambrodji [2,3], Huixing Huang [1], Kelly J. Bouchonville [1], Amy S. Etheridge [4], Remington E. Schmidt [1], Brianna M. Bembenek [1], Zoey B. Temesgen [1], Zhiquan Wang [5], Federico Innocenti [4], Deborah Stroka [6], Robert B. Diasio [1], Carlo R. Largiadèr [2], and Steven M. Offer [1,7,8,*]

1 Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, USA.

2 Department of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, CH-3010 Bern, Switzerland.

3 Graduate School for Cellular and Biomedical Sciences, University of Bern, Freiestrasse 1, CH-3010 Bern, Switzerland.

4 Eshelman School of Pharmacy, Division of Pharmacotherapy and Experimental Therapeutics, University of North Carolina, Chapel Hill, NC 27599, USA.

5 Division of Hematology, Department of Medicine, Mayo Clinic, Rochester, MN 55905 USA.

6 Department of Visceral Surgery and Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland.

7 Department of Pathology, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA.

8 Lead contact.

* Correspondence: soffer@uiowa.edu; offer.steven1@mayo.edu

ABSTRACT

Enhancers are critical for regulating tissue-specific gene expression, and genetic variants within enhancer regions have been suggested to contribute to various cancer-related processes, including therapeutic resistance. However, the precise mechanisms remain elusive. Using a well-defined drug-gene pair, we identified an enhancer region for dihydropyrimidine dehydrogenase (DPD, *DPYD* gene) expression that is relevant to the metabolism of the anti-cancer drug 5-fluorouracil (5-FU). Using reporter systems, CRISPR genome edited cell models, and human liver specimens, we demonstrated *in vitro* and *vivo* that genotype status for the common germline variant (rs4294451; 27% global minor allele frequency) located within this novel enhancer controls *DPYD* transcription and alters resistance to 5-FU. The variant genotype increases recruitment of the transcription factor CEBPB to the enhancer and alters the level of direct interactions between the enhancer and *DPYD* promoter. Our data provide insight into the regulatory mechanisms controlling sensitivity and resistance to 5-FU.

KEYWORDS

Epigenetics, Gene Regulation, 5-Fluorouracil, Chemotherapy, CEBPB, Chemotherapy, Biomarkers, Drug Resistance.

52 **INTRODUCTION**

53 Therapeutic resistance has been reported for nearly all anti-cancer drugs [1], and as
54 many as 90% of mortalities in cancer can be linked to drug resistance [2]. Enhancer-
55 mediated regulation of gene expression has been increasingly implicated in multiple
56 cancer-related processes, including resistance, response, and toxicity to cancer
57 therapies [3, 4]. However, the specific molecular mechanisms through which epigenetic
58 changes drive these processes remain mostly elusive.

59

60 Enhancers regulate gene expression by recruiting transcription factors (TFs) and
61 subsequently interacting with the promoter region of a target gene to drive expression.
62 The activity of enhancers can be affected by localized epigenetic state and genetic
63 variations that affect TF binding. More than 90% of disease-associated genetic variants
64 identified in genome-wide association studies (GWAS) lie in the non-coding portion of
65 the genome [5, 6]. Many of these GWAS variants map to putative enhancers [7].
66 Regardless, translating enhancer GWAS variants into disease mechanisms remains a
67 largely unmet challenge [8].

68

69 In previous studies, we identified a *cis* expression quantitative trait locus (eQTL) for the
70 chemotherapeutic metabolism gene dihydropyrimidine dehydrogenase enzyme (DPD,
71 encoded by the *DPYD* gene) [9]. DPD is the initial and rate-defining enzyme for
72 conversion of the commonly used chemotherapeutic 5-fluoruracil (5-FU) into inactive
73 metabolites. Hepatic DPD eliminates approximately 85% of circulating 5-FU within

74 minutes of administration [10, 11], and deficiency of DPD is associated with severe
75 (clinical grade ≥3) toxicity to 5-FU [12]. Clinical and preclinical studies have identified
76 missense and splicing variants of *DPYD* that consistently correlate with reduced DPD
77 function and increased 5-FU toxicity risk [12-18]. However, these variants explain <20%
78 of reported cases of grade ≥3 5-FU toxicities [12, 19], and the mechanisms contributing
79 to 5-FU resistance, and therefore means to overcome resistance, remain elusive.
80
81 In the present study, we used human liver tissues and cellular models to identify and
82 characterize a novel *cis* enhancer element capable of modulating *DPYD* expression.
83 We additionally provide mechanistic data demonstrating that transcription-factor driven
84 expression of *DPYD* is dependent on allele status for a common germline genetic
85 variant located within this enhancer, suggesting that the variant could be a valuable
86 biomarker of 5-FU toxicity risk. Furthermore, the genotype-dependent regulation of DPD
87 expression offers a potential mechanism for overcoming resistance in patients carrying
88 the risk variant.
89
90
91 **RESULTS**
92
93 **Identification of *DPYD* enhancer regions**
94 Candidate *cis* enhancer regions for *DPYD* (NM_000110.4) were identified using
95 GeneHancer [20] and Ensembl Regulatory Build [21] data mapped to GRCh37/hg19.

96 We defined the upstream boundary of the target region using enhancer data supported
97 by multiple methods and gene-enhancer links supported by multiple methods as
98 implemented within GeneHancer (i.e., putative enhancer regions with “double-elite”
99 status; **Figure 1A**). We then mapped annotated transcription factor binding sites in the
100 target region using ChIP-seq data from the Ensembl Regulatory Build as an additional
101 layer of evidence (**Figure 1A**). In all, three candidate proximal upstream enhancer
102 regions were identified, which will be referred to herein as E9, E16, and E20 based on
103 the relevant location upstream of the *DPYD* transcription start site (**Figure 1A**).
104
105 CRISPR-interference (CRISPRi) and -activation (CRISPRa) were used to determine
106 which of these regions were capable of regulating *DPYD* expression. For CRISPRi, we
107 generated cell lines that stably expressed the Krüppel-associated box (KRAB) domain
108 of Kox1 fused to nuclease-deficient Cas9 protein (dCas9-KRAB; [22]). For CRISPRa,
109 we created cell lines that stably expressed dCas9 fused to the histone acetyltransferase
110 E1A binding protein P300 (dCas9-P300; [23]). Six guide RNAs (gRNAs) were used to
111 target dCas9 fusion proteins to each region of interest, and quantitative reverse
112 transcription PCR (qRT-PCR) was used to monitor changes in *DPYD* expression
113 following transfection.
114
115 Targeting of dCas9-KRAB (CRISPRi) to the E9 region significantly decreased *DPYD*
116 expression relative to control for four of six gRNAs (**Figure 1B**). Targeting of dCas9-
117 KRAB to E16 resulted in a significant decrease in *DPYD* expression for one of six

118 gRNAs (**Figure 1C**), and no significant reductions in *DPYD* expression were noted
119 following transfection of gRNAs for E20 in dCas9-expressing cells; however, two gRNAs
120 increased *DPYD* expression (**Figure 1D**). For reciprocal CRISPRa experiments,
121 targeting dCas9-P300 to E9 significantly increased *DPYD* expression for four of six
122 gRNAs tested (**Figure 1E**), whereas gRNAs specific to E16 and E20 did not elicit any
123 significant changes in *DPYD* expression (**Figures 1F** and **1G**). Consistent with the
124 results in HepG2 cells, CRISPRi and CRISPRa targeted to the E9 region significantly
125 altered *DPYD* expression in HCT116 cells (**Figures S1A** and **S1D**), whereas no
126 significant changes were noted when targeted to E16 or E20 (**Figures S1B**, **S1C**, **S1E**,
127 and **S1F**).
128
129 To confirm that the expected epigenetic changes were induced at the E9 region, we
130 performed chromatin immunoprecipitation (ChIP) coupled with quantitative PCR (ChIP-
131 qPCR) in dCas9-KRAB and -P300 expressing cells transfected with E9 gRNAs. As
132 expected, targeting dCas9-KRAB to E9 resulted in a localized increase in H3K9me3
133 compared to controls, indicating a shift from active to inactive chromatin at E9 (**Figure**
134 **2A**; the primer nearest E9 is denoted by the gray box). Targeting of dCas-P300 to E9
135 caused changes consistent with epigenetic activation of the region, including a localized
136 gain of H3K27ac-marked chromatin (**Figure 2B**). Similar changes in localized histone
137 modifications were noted when dCas9-KRAB and dCas9-P300 HCT116 cells were
138 directed to E9 (**Figure 2C–D**). Collectively, these results indicate that the E9 region
139 might act as a *cis* regulatory element modulating *DPYD* expression.

140

141 **Identification of rs4294451 as a putative regulatory SNP within E9**

142 In an expression quantitative trait loci (eQTL) genome-wide association study (GWAS)
143 conducted in human liver specimens, Etheridge *et al*. previously identified three
144 independent haplotype blocks that significantly associated with altered *DPYD*
145 expression [9]. One of the blocks identified in that study spanned the E9 region,
146 prompting us to investigate if genetic variants in E9 could potentially perturb *DPYD*
147 regulation. Based on the biological role of enhancers, we postulated that a causal
148 functional single nucleotide polymorphism (SNP) could interfere with transcription factor
149 binding. Within the Ensembl Regulatory Build [21], we identified a region of
150 approximately 150 bp showing evidence for binding by multiple transcription factors in
151 HepG2 cells, but not in other cell types tested (**Figure S2**). Within this region, we
152 identified the variant rs4294451, which is a tag SNP for the previously identified eQTL
153 block [9], prompting us to hypothesize that this variant might be the causal eQTL SNP.

154

155 To assess the contribution of rs4294451 alleles to *DPYD* regulation, we first performed
156 reporter assays by cloning the *DPYD* promoter and E9 enhancer region containing
157 either the reference or variant allele for rs4294451 (T or A, respectively) into a luciferase
158 reporter vector (**Figure 3A**). Reporter vectors containing the T-allele yielded significantly
159 higher luciferase activity compared to vectors containing only the *DPYD* promoter
160 (p=0.00054), whereas those containing the A allele within E9 showed a more modest
161 increase (p=0.070, **Figure 3B**). Directly comparing E9-containing reporters, vectors

162 containing the T allele showed significantly higher activity than those with the A allele
163 (p=0.046, **Figure 3B**), suggesting that the variant could impact *DPYD* expression.
164
165 **Impact of rs4294451 genotype on *DPYD* expression and regulation**
166 To directly characterize the impact of rs4294451 alleles on *DPYD* expression at the
167 endogenous locus, we used CRISPR-mediated genome editing to create matched
168 isogenic HCT116 knock-in cell models for each genotype. Consistent with the results
169 from the reporter assay (**Figure 3B**), cells homozygous for the A allele had significantly
170 lower *DPYD* expression compared to cells that were homozygous for the T allele
171 (p=0.0011, **Figure 3C**). Heterozygous rs4294451 A/T cells displayed intermediate
172 expression compared to cells with homozygous A/A and T/T genotypes (p=0.00013 and
173 p=0.028, respectively; **Figure 3C**). At the chromatin level, cells homozygous for the A
174 allele showed reductions in H3K27ac and accumulation of H3K9me at E9 compared to
175 cells that were homozygous or heterozygous for the T allele (**Figure 3D–3E**). This
176 finding is consistent with the A allele being associated with a less active epigenetic state
177 at E9. Cells carrying a heterozygous genotype for rs4294451 displayed an intermediate
178 epigenetic state, which is consistent with one epigenetically active and one inactive
179 copy of E9 (**Figure 3D–3E**). Similar epigenetic differences between rs4294451
180 genotypes were noted at E9 in liver specimens obtained from human donors (**Figure**
181 **S3A–S3B**).
182
183 **Allele-specific expression of *DPYD* transcripts**

184 To precisely determine the impact of rs4294451 alleles on *DPYD* expression *in vivo*, we
185 measured allele-specific expression using reverse transcriptase digital droplet PCR
186 (RT-ddPCR) using donor liver tissues. Because rs4294451 falls outside of the coding
187 region, we used coding region variants as proxies to measure strand-specific
188 expression. These coding variants consisted of *DPYD*-c.85T>C (rs1801265) and
189 c.496A>G (rs2297595), both of which are in linkage disequilibrium (LD) with the
190 rs4294451 T allele (c.85C: D′=0.92, R^2=0.80; c.496G: D′= 0.67, R^2= 0.21). Imperfect LD
191 between rs4294451 and proxies allowed us to use samples that are heterozygous for
192 the coding region SNP, but homozygous for rs429441 genotype, as an additional level
193 of control. In samples measured using the c.85 genotype, the fractional abundance of
194 the C-allele ranged from 0.542 to 0.602 in samples heterozygous for rs4294451 (**Figure**
195 **3F**). Expressed in terms of relative expression, the C-allele was expressed at levels 18–
196 51% higher than the T-allele (i.e., 0.542/0.458=1.18 and 0.602/0.398=1.51). In contrast,
197 the sample that was homozygous for the rs4294451 T allele showed allelic expression
198 at the c.85 locus that was indistinguishable from the DNA control (**Figure 3F**), indicating
199 equal expression of both transcripts. In measurements using c.496, a significant allelic-
200 imbalance of 0.565 (CI: 0.546–0.584) in favor of the variant c.496 G-allele compared to
201 the DNA control (CI: 0.488–0.511) was observed (**Figure 3G**). This again indicated
202 higher expression of the linked rs4294451 T allele. Notably, liver specimens that were
203 homozygous at the rs4294451 locus again showed c.496 allelic expression that was
204 indistinguishable from the DNA control (**Figure 3G**). These data indicate that rs4294451
205 significantly affects the expression of *DPYD* in the liver.

206

207 **Rs4294451 genotype affects interactions between the *DPYD* promoter and E9**

208 We next investigated the impact of rs4294451 genotype on intra-chromatin interactions
209 between the E9 region and the *DPYD* promoter using quantitative analysis of chromatin
210 conformation capture (3C-qPCR) in knock-in cells. The schematic in **Figure 4A** shows
211 the location of primers used for 3C-qPCR relative to the *DPYD* transcription start site
212 (TSS) and E9 region, as well as the location of *Hin*dIII digestion sites. When using
213 primers anchored at the *DPYD* promoter, the E9 region showed a significantly stronger
214 interaction with the promoter in cells carrying the T allele compared to those
215 homozygous for the A allele ($p=8.9 \times 10^{-4}$ comparing TT to AA; ANOVA $p=3.0 \times 10^{-4}$
216 across all three genotypes; **Figure 4C**). With primers anchored to the E9 region, cells
217 homozygous for the rs4294451 T allele showed stronger promoter interaction than
218 those carrying the A allele ($p=4.9 \times 10^{-4}$ comparing TT to AA; ANOVA $p=1.2 \times 10^{-4}$
219 across all three genotypes; **Figure 4D**).

220

221 **Rs4294451 genotype confers differential sensitivity to 5-FU**

222 Having demonstrated that the rs4294451 genotype impacts *DPYD* expression, we next
223 evaluated the impact on 5-FU sensitivity using real-time cellular analysis (RTCA). RTCA
224 provides a measure of the number of viable cells present in a culture. We have
225 previously demonstrated the utility of this technology for measuring differences in drug-
226 sensitivity to 5-FU and that those differences directly correlate with DPD enzyme
227 function within cells [14, 17]. Consistent with lower expression of *DPYD*, knock-in cells

228 homozygous for the rs4294451 A allele were significantly more sensitive to 5-FU than
229 cells homozygous for the T allele (IC_{50} concentrations were 9.1 and 95.0 μM 5-FU,
230 respectively; $p<0.0001$; **Figure 4E**). Heterozygous T/A cells showed an intermediate
231 IC_{50} value of 29.9 μM 5-FU (**Figure 4E**).
232
233 **Transcription factor binding within E9 is affected by rs4294451 status**
234 Based on ENCODE data, rs4294451 is localized to a region that previously has been
235 shown to be bound by multiple transcription factors (**Figure S2**). To determine if known
236 transcription factor binding sites could be affected by allele status at rs4294451, we
237 computed binding scores for JUND, MYBL2, HNF4A, CEBPB, FOXA1, and FOXA2 for
238 sequences comprising the E9 region containing the T and A alleles at rs4294451.
239 CEBPB, FOXA1, and FOXA2 showed differential predicted binding scores between the
240 rs4294451 A- and T-containing query sequences (data not shown). For CEBPB, the
241 differential binding site had a higher score than other sites in the queried region that did
242 not overlap with rs4294451, suggesting that the SNP could affect CEBPB binding in the
243 region. While predicted binding in the region varied for FOXA1 and FOXA2 depending
244 on rs4294451 genotype, other binding sites that did not overlap with the variant site
245 showed stronger predicted binding, indicating that the SNP was likely not affecting the
246 critical binding site in the region. Binding scores did not differ by genotype for JUND or
247 HNF4A, and no binding site above the threshold score was detected in the region for
248 MYBL2.
249

250 Based on the above, we sought to determine if CEBPB could regulate *DPYD*
251 expression through the E9 region and if regulation was affected by rs4294451
252 genotype. We first used reporter assays in conjunction with CEBPB. CEBPB expression
253 significantly increased reporter activity relative to GFP control for luciferase vectors
254 containing the *DPYD* promoter, indicating that the promoter contains CEBPB
255 recognition sites ($p=9.1x10^{-4}$; **Figure 5A**). In cells overexpressing CEBPB, plasmids
256 containing the promoter and E9 region resulted in higher luciferase activity compared to
257 those containing only the *DPYD* promoter, regardless of the presence of the A or T
258 allele (p=0.0036 and p=0.0017, respectively), suggesting that rs4294451 genotype does
259 not completely disrupt CEBPB regulation through E9 (**Figure 5A**). Comparing results for
260 the A and T alleles at rs4294451 in the presence of overexpressed CEBPB indicates
261 that the T allele is likely more responsive to CEBPB (**Figure 5A**; p=0.036).
262
263 To directly examine CEBPB binding to the E9 region and to determine if binding affinity
264 differs between the rs4294451 A and T alleles, we performed ChIP with CEBPB
265 antibodies in isogenic knock-in HCT116 cells for the rs4294451 A/A, A/T, and T/T
266 genotypes. Cells homozygous for the reference T allele showed significantly higher
267 CEBPB occupancy at both E9 and the *DPYD* promoter than A/T or A/A cells (**Figure**
268 **5B–5C**). These results suggest that rs4294451 genotype determined the binding
269 potential for CEBPB at the E9 enhancer region, which, in turn, affects CEBPB-driven
270 expression of *DPYD* from the promoter.
271

272 **Upregulation of *DPYD* expression by CEBPB is dependent on the rs4294451-T**
273 **allele**
274 To further characterize the role of CEBPB in regulating *DPYD*, we disrupted CEBPB
275 expression using two independent shRNAs (denoted as “sh1” and “sh2”) in HCT116
276 knock-in cells carrying rs4294451 A/A, A/T, and T/T genotypes. Knockdown of CEBPB
277 was confirmed at the protein (**Figure 6A**) and mRNA (**Figures 6B–6D**) levels.
278 Knockdown of CEBPB significantly reduced *DPYD* expression in rs4294451 T/A and
279 T/T cells, but not in rs4294451 A/A cells (**Figures 6E–6F**). CEBPB knockdown also
280 reduced occupancy of the transcription factor at both E9 and the *DPYD* TSS in
281 rs4294451 A/T and T/T cells, but not A/A cells, and the level of CEBPB occupancy at
282 both regions under CEBPB knockdown conditions is similar to that in control shRNA-
283 treated A/A cells (**Figures S4A–S4B**).
284
285 We next determined the extent to which CEBPB contributed to 5-FU chemoresistance in
286 cells with the T allele of rs4294451 (**Figure 4E**). In rs4294451 A/A cells, knockdown of
287 CEBPB did not affect 5-FU sensitivity (**Figure 6H**). In contrast, CEBPB knockdown
288 significantly reduced the IC_{50} for 5-FU in rs4294451 T/T cells ($p<0.0001$; **Figure 6I**).
289
290 **The rs4294451-T allele is enriched in individuals of African ancestry**
291 Data from the Genome Aggregation Database (gnomAD [24]) were used to estimate the
292 frequency of the rs4294451-T allele in global populations (**Figure S5A**). The highest
293 minor allele frequency (MAF) was noted for African/African American individuals (40%

294 MAF), where the MAF was lowest in East Asian individuals (7%). For comparison,
295 individuals of European (Non-Finnish) ancestry, the population with the highest number
296 of individuals reported in gnomAD, had a MAF of 23%. Within Latino-Admixed American
297 gnomAD subjects, similar differences in MAFs were noted in local ancestry-informed
298 frequency data (**Figure S5B**).
299
300
301 **DISCUSSION**
302
303 The antitumor efficacy and risk of severe adverse events associated with 5-FU are
304 determined by the overall bioavailability of the drug in plasma. As the critical
305 determinant of 5-FU pharmacokinetics, liver DPD expression is pivotal to both the risk of
306 severe adverse events and therapeutic resistance in 5-FU chemotherapy at opposite
307 ends of the exposure spectrum. This is underscored by the narrow therapeutic window
308 for 5-FU, with toxicity and efficacy occurring at partially overlapping drug exposure
309 levels [25]. Deleterious germline coding-region *DPYD* variants have been linked to
310 severe 5-FU toxicity [12, 13]; however, these variants are responsible for only a small
311 fraction of severe adverse events in 5-FU use and are unlikely to contribute to drug
312 resistance [26]. Elevated expression of DPD in tumor cells is known to confer 5-FU
313 resistance [27, 28], and upregulation of hepatic DPD expression has been shown to
314 reduce drug exposures and promote the development of 5-FU–resistant tumors [29].
315 However, the mechanisms regulating DPD expression are not well characterized, nor is

316 it understood how the regulatory processes can be altered to support the development
317 of 5-FU resistance.
318
319 In the current study, we identified a novel *cis* enhancer region for *DPYD* that is located
320 approximately 9 kb upstream from the gene's transcription start site. We demonstrated
321 that CEBPB is a critical transcription factor for *DPYD* that binds to this enhancer region,
322 termed E9, promoting enhancer–promoter interactions and increasing *DPYD*
323 expression. We also showed that the allele status of the germline variant rs4294451,
324 located within the E9 region, can affect CEBPB-driven *DPYD* expression and
325 sensitivity/resistance to 5-FU, making it a strong candidate biomarker for 5-FU toxicity
326 risk and potentially tumor resistance to 5-FU–based cancer therapy. These findings are
327 consistent with the recent identification of a haplotype block linked to the rs4294451 T
328 allele that was significantly associated with elevated *DPYD* expression in human liver
329 tissues [9]. In the present manuscript, we show that the rs4294451 T allele is enriched
330 for active chromatin marks in both human liver (**Figure S3**) and in cellular knock-in
331 models (**Figure 3D–E**). Furthermore, using an innovative digital droplet RT-qPCR–
332 based approach, we demonstrate that the rs4294451 T allele is associated with
333 elevated expression of the *cis DPYD* transcript in human liver (**Figure 3F–G**).
334
335 Allele frequency data retrieved from gnomAD suggest that a majority (65-70%,
336 estimated from allele frequencies) of individuals of African ancestry carry the
337 rs4294451-T allele (**Figure S5**), whereas only about 35% of individuals of European

338 ancestry are predicted to be carriers of the T allele. African American patients have
339 worse overall survival in colorectal cancer compared to white patients, owing to
340 biological and non-biological factors. While differential access to healthcare, treatment
341 bias, and socioeconomic factors have been shown to contribute to the poorer prognosis
342 [30], other unrecognized factors also contribute to this difference [31, 32]. Our data
343 support a hypothesis that higher systemic 5-FU catabolism to due elevated liver DPD
344 expression in carriers of the rs4294451-T allele results in lower exposure to active anti-
345 tumor metabolites of 5-FU. Additionally, we demonstrate that colorectal cancer cell lines
346 likely retain the enhancer functions associated with the rs4294451-T allele, suggesting
347 that tumor cells carrying this variant could more readily inactive 5-FU via increased DPD
348 expression. The higher likelihood of carrying the T-allele in individuals of African
349 ancestry would therefore place them at greater risk. Additional studies will be needed to
350 investigate the degree to which rs4294451-T contributes to survival and progression in
351 colorectal cancer and in other solid cancers frequently treated with 5-FU.
352
353 The transcription factor CEBPB is a member of the CCAAT Enhancer Binding Protein
354 family, a group of transcription factors that contain basic leucine zipper (bZIP) domains
355 and is highly expressed in liver [33, 34]. Multiple CEBPB isoforms have been detected,
356 with some acting as transcriptional activators and others as inhibitors. The data
357 presented herein indicate that active isoforms of CEBPB are up-regulating *DPYD*
358 through binding to the E9 enhancer region. Inhibitory isoforms of CEBPB have also
359 been shown to be important for certain physiological processes including tumorigenesis

360 and liver regeneration [35, 36]. Our over-expression studies used the full-length active
361 isoform. Therefore, we cannot rule out a role for the inactive isoform participating in the
362 regulation of *DPYD* expression. In addition to acting as a homodimer, CEBPB can also
363 heterodimerize with other CEBP family proteins and interact with other transcription
364 factors, including P300/CBP, CREB, NFKB, AP1, and NFAT, to co-regulate gene
365 expression [37, 38]. Additional studies are underway to characterize the role of
366 additional regulatory factors within the *DPYD* enhancer region identified in this
367 manuscript.
368
369 *In silico* analyses suggested that the rs4294451 A allele created a stronger binding site
370 for CEBPB within the E9 enhancer region. However, our data demonstrate that the T
371 allele is associated with higher reporter activity (**Figure 3B**), higher DPD expression in
372 both knock-in cells (**Figure 3C**) and human liver tissues (**Figure 3F–G**), and localized
373 epigenetic activation (**Figure 3D–E**). Enrichment of CEBPB to the E9 region was
374 likewise shown to be higher in cells with the rs4294451 T allele (**Figure 5**), and higher
375 levels of interaction were also noted between the E9 region with the T allele and the
376 *DPYD* promoter (**Figure 4**). These findings are also consistent with previously reported
377 eQTL results for the haplotype block linked to rs4294451, where the T allele was
378 associated with higher levels of *DPYD* expression in human liver specimens [9], and our
379 observation that the T allele is associated with cellular resistance to 5-FU (**Figure 4E**).
380 The binding of CEBPB to specific motifs has recently been shown to be cell-type
381 specific and to rely on non-consensus binding motifs with binding strengths that can be

382 modulated by the sequence and structure of surrounding DNA regions [39, 40],
383 providing a possible explanation for the discrepancy between predicted and observed
384 results.
385
386 To our knowledge, this is the first report directly linking CEBPB to 5-FU metabolism and
387 the first mechanistic data demonstrating *cis* effects of a regulatory variant on *DPYD*
388 expression. The role of CEBPB in modulating 5-FU resistance and toxicity is not without
389 precedent. For example, mir-191 is abnormally expressed in several cancers and has
390 been associated with both 5-FU resistance and the regulation of CEBPB expression
391 [41]; however, the CEBPB–*DPYD* regulatory axis has not previously been recognized.
392 Furthermore, CEBPB signaling has been shown to be activated in colorectal cancer
393 cells following treatment with 5-FU [42], suggesting that CEBPB-mediated activation of
394 *DPYD* expression might represent a dynamic response to therapy.
395
396 While the contributions of regulatory variants to 5-FU metabolism have not been widely
397 studied to date, previous studies have explored DPD regulation. Our laboratories
398 previously characterized a *trans*-acting regulatory variant for DPD located within the
399 microRNA mir-27a that was subsequently shown to further increase 5-FU toxicity risk in
400 individuals that carried deleterious nonsynonymous *DPYD* variants [43, 44]. The variant
401 at rs4294451 is in LD with the *DPYD* variants c.85T>C and c.496G>A, which served as
402 coding region proxies for allele-specific expression in our present study (**Figure 3F–G**).
403 The haplotypes defined by these two coding-region variants together with a third variant

404 (c.1129-5923C>G/rs75017182) previously associated with varied levels of systemic
405 DPD activity [45]. A subsequent retrospective analysis indicated that these haplotype
406 differences translate to differential risk of severe 5-FU toxicity [46]; however, additional
407 studies are needed that are powered to evaluate more than the most common
408 haplotypes. Taken together, these results suggest that the differential regulatory effects
409 of rs4294451 alleles could further impact overall DPD activity and, by extension,
410 modulate the risk of severe 5-FU–related toxicity conferred by coding or splice-variants
411 in *DPYD*. Further work is also needed to identify interactions between the linked
412 variants that impact DPD enzyme activity and the regulatory variant to define systemic
413 DPD function and 5-FU toxicity risk.
414
415
416 **METHODS**
417
418 **Cells**
419 HEK293T, HCT116, and HepG2 cell lines were obtained from the American Tissue
420 Collection Center (ATCC, Manassas, VA). All cell lines were cultured in Dulbecco's
421 modified Eagle's medium (DMEM; Corning, Corning, NY) supplemented with 10% fetal
422 bovine serum (FBS; MilliporeSigma, Billerica, MA), 2 mM L-glutamate (Corning), and 1x
423 penicillin/streptomycin solution (MilliporeSigma). Cells were maintained at 37°C with 5%
424 CO_2. To support cell attachment and expansion, HepG2 cells were grown on plates
425 coated with 5% Matrigel (Corning).

426

427 **Liver tissues**

428 Human liver tissues used for ChIP analyses were processed through Dr. Mary Relling's
429 laboratory at St. Jude Children's Research Hospital, part of the Pharmacogenetics of
430 Anticancer Agents Research (PAAR) Group and were provided by the Liver Tissue Cell
431 Distribution System funded by NIH Contract #N01-DK-7-0004/HHSN267200700004C
432 and by the Cooperative Human Tissue Network. The acquisition and use of specimens
433 for this manuscript was conducted with the approval of the University of North Carolina
434 at Chapel Hill IRB (study number 10-2253), which has designated the use of these
435 livers for the current analyses as nonhuman subject research and the need for direct
436 consent for use in this study was waived. Human liver tissues used for allele-specific
437 expression were obtained from the University Clinic of Visceral Surgery and Medicine,
438 Inselspital, Bern, Switzerland. Specimens were from donated material from patients who
439 had undergone liver surgery at Inselspital and signed a written consent form for remnant
440 tissues to be used in research (KEK-BE:2016-02202). Patients with impaired liver
441 function due to cirrhosis or other conditions were excluded from analyses.

442

443 **Quantitative RT-PCR (RT-qPCR)**

444 Total RNA was extracted from cells using the Direct-zol RNA Kit (Zymo Research,
445 Tustin, CA) following the manufacturer’s protocol. Reverse transcription into cDNA was
446 performed using the Transcriptor Reverse Transcriptase kit (Roche, Indianapolis, IN)
447 and random hexamer primers (Roche) according to the manufacturer’s instructions.

448 Quantitative PCR (qPCR) was performed on a LightCycler 480 System (Roche) using
449 LightCycler 480 SYBR Green I Master Mix reagents (Roche). Primers used for RT-
450 qPCR are listed in **Supplementary Table S1**. RNA expression was normalized to the
451 reference gene L32, and relative gene expression was calculated using the $2^{-\Delta\Delta CT}$
452 method. For all analyses, three independent experiments were performed; for each
453 experiment, gene expression was assessed in triplicate.
454
455 **Western blot analysis**
456 Whole-cell lysates were extracted using the RIPA lysis buffer system (Santa Cruz
457 Biotechnology, Dallas, TX), separated by SDS-PAGE, and transferred to PVDF
458 membrane (MilliporeSigma). Blots were blocked using Odyssey Blocking Buffer (LI-
459 COR Biosciences, Lincoln, NE) and incubated with primary antibody at 4°C overnight.
460 Primary antibodies consisted of anti-CEBPB (PA5-27244; 1:1,000 dilution; Thermo
461 Fisher Scientific, Waltham, MA), anti-alpha Tubulin (ab4074; 1:7,500 dilution; Abcam,
462 Waltham, MA), and anti-cas9 (sc-517386; 1:1,000 dilution; Santa Cruz Biotechnology,
463 Dallas, TX). Membranes were washed and incubated with anti-mouse and anti-rabbit
464 secondary antibodies (#926-32212 and #926-68073; both 1:5,000 dilution; LI-COR
465 Biosciences) for 1 hour at room temperature. Blots were imaged using the Odyssey
466 infrared imaging system (LI-COR Biosciences).
467
468 **Plasmids**

469 The gRNA expression empty vector lentiGuide-Puro was a gift from Feng Zhang
470 (Addgene_52963). pCMV-FLAG LAP2 was a gift from Joan Massague
471 (Addgene_15738). The oligonucleotides targeting enhancer regions (E9, E16, and E20)
472 were designed using GuideScan [47], hybridized, phosphorylated, and cloned into
473 lentiGuide-Puro via *Bsm*BI sites. The luciferase expression vector pGL4.10 was
474 purchased from Promega (Madison, WI). Lentivirus vectors expressing shRNAs
475 targeting CEBPB were obtained from the University of Minnesota Genomics Center
476 (Sh1: TRCN0000007440 and Sh2: TRCN0000007442).
477
478 **CRISPR inactivation (CRISPRi) and activation (CRISPRa)**
479 HepG2 cell lines that overexpress dCas9-KRAB and dCas9-P300 were generated by
480 lentiviral transduction using lenti-EF1a-dCas9-KRAB-Puro and pLV-dCas9-p300-P2A-
481 PuroR, respectively. To generate lentiviral particles, HEK293T cells were co-
482 transfected with lenti-EF1a-dCas9-KRAB-Puro plasmid (a gift from Kristen Brennand;
483 Addgene_99372) or pLV-dCas9-p300-P2A-PuroR plasmid (a gift from Charles
484 Gersbach; Addgene_83889), psPAX2 (a gift from Didier Trono; Addgene_12260), and
485 pMD2.G (a gift from Didier Trono; Addgene_12259) using TransIT-Lenti Transfection
486 Reagent (Mirus Bio, Madison, WI). A 3:1 ratio of transfection reagent to total plasmid
487 was used for all transfections. For all transfections, medium was changed 14 hours after
488 transfection, and viral supernatants were collected 34 hours later. Supernatants were
489 filtered using 0.45 μm PVDF filters (MilliporeSigma) to remove debris/cells and used
490 directly for transductions. For transductions, HCT116 cells or HepG2 cells were seeded

491 at a density of 4×10^5 cells per well in 6-well plates and incubated with 500 µL virus-
492 containing supernatant, 12.5 µg/mL polybrene (MilliporeSigma), and 1.5 mL fresh
493 DMEM culture medium. Medium was changed after 24 hours. Cells were treated with 1
494 µg/mL puromycin to initiate selection for transduced cells 48 hours after transduction.
495 Expression of dCas9-KRAB and dCas9-P300 was confirmed by western blotting. Guide
496 RNAs (gRNAs) for each target region were identified and designed using GuideScan
497 [47]. Oligonucleotides corresponding to each gRNA were obtained from IDT (Coralville,
498 IA), hybridized, phosphorylated using T4 polynucleotide kinase (New England Biolabs,
499 Ipswich, MA), and ligated into digested BsmBI-digested (New England Biolabs)
500 lentiGuide-Puro vector (a gift from Feng Zhang; Addgene_52963). Cell lines stably
501 expressing dCas9-KRAB or dCas9-P300 were transfected with plasmids encoding
502 gRNAs or lentiGuide empty vector using TransIT-X2 (Mirus Bio). RNA was extracted 2
503 days after transfection, and *DPYD* expression was measured by RT-qPCR.
504

505 **Chromatin immunoprecipitation coupled with quantitative PCR (ChIP-qPCR)**

506 ChIP assays were performed using the ChIP-IT Express Enzymatic Kit (Active Motif,
507 Carlsbad, CA) following manufacturer's directions. One million cells were harvested,
508 washed, and crossed linked using 1% formaldehyde (Thermo Fisher Scientific) in
509 serum-free medium for 10 minutes followed by quenching with 125 mM glycine for 5
510 minutes at room temperature. The chromatin was digested using enzymatic shearing
511 cocktail provided by the kit to an average size of 200–1000 bp. Two percent of the
512 sheared chromatin was retained as input control. Approximately 25 µg sheared

513 chromatin was incubated with 2 µg H3K27ac antibody (ab4729; Abcam), 2 µg
514 H3K9me3 antibody (ab8898; Abcam), 2 µg CEBPB antibody (PA5-27244; Thermo
515 Fisher Scientific), or 2 µg control normal Rabbit IgG antibody (antibody 2729; Cell
516 Signaling Technology, Danvers, MA), in the presence of protein G magnetic beads,
517 ChIP buffer, and protease inhibitor cocktail (all Active Motif) for 4 hours at 4°C. Magnetic
518 beads were washed, chromatin was eluted, cross-linking was reversed, and proteinase
519 K treatment performed using reagents provided in the kit following manufacturer's
520 directions. DNA was purified using the QIAquick PCR Purification Kit (Qiagen,
521 Germantown, MD). Purified DNA was used for subsequent qPCR reactions using SYBR
522 Green I Master Mix on a LightCycler 480 System (Roche). Enrichment was calculated
523 using the following formula: (a) % ChIP = $2^{(\text{Input Ct - ChIP Ct})}$ * (dilution factor) (100); (b) %
524 IgG = $2^{(\text{Input Ct - IgG Ct})}$ * (dilution factor) (100); (c) Fold Enrichment = % ChIP ÷ % IgG.
525 Primers used for ChIP-qPCR are listed in **Supplementary Table S1**.
526

527 **Luciferase reporter assays**

528 The *DPYD* promoter region, consisting of the 1154 bp of genomic DNA directly
529 upstream of the *DPYD* TSS, was amplified by PCR, digested with EcoRV and HindIII
530 (New England Biolabs), and cloned into compatible sites on the pGL4.10 vector
531 (Promega). The 1392-bp region comprising the E9 region was PCR amplified from
532 genomic DNA and cloned upstream of the *DPYD* promoter using KpnI and SacI sites
533 (New England Biolabs). A vector containing the A allele of rs4294451 within E9 was
534 confirmed by Sanger sequencing. The vector containing the rs4294451 T allele was

535 generated by site-directed mutagenesis and confirmed by sequencing. All primers used
536 in vector construction are listed in **Supplementary Table S1**. For reporter assays, 10^5
537 HEK293T cells were seeded into 24-well plates and co-transfected with pGL4.10-based
538 plasmids and pRL-SV40 Renilla luciferase plasmid (Promega). After 48 hours,
539 luciferase activity was measured using the Dual-Glo Luciferase Assay (Promega)
540 following manufacturer's recommendations on a Synergy HTX Multimode Plate Reader
541 (Agilent Technologies, Santa Clara, CA).
542
543 **Knock-in cell lines for rs4294451 genotypes**
544 Knock-in cell lines were generated using CRISPR/Cas9 gene editing. Homology-
545 directed repair (HDR) donor templates and target-specific Alt-R crRNA were designed
546 using the Alt-R HDR Design Tool (IDT). Equimolar amounts of crRNA (IDT) and
547 common Alt-R tracrRNA (IDT) were annealed to form the gRNA duplex. RNP
548 complexes were formed by combining gRNA with Alt-R S.p. Cas9 Nuclease V3 (IDT) to
549 a final Cas9:gRNA ratio of 4:4.8. RNA complex and Alt-R HDR Donor Oligos were
550 transfected into HCT116 cells by electroporation (Lonza Nucleofector 96-well Shuttle
551 System; Lonza, Bend, OR) using parameters provided by the manufacturer. Seventy-
552 two hours after transfection, serial dilutions were performed to obtain single-cell clones.
553 Clones were expanded, genomic DNA was isolated, and rs42944551 genotype was
554 determined by rhAmp Genotyping (assay ID: Hs.GT.rs4294451.A.1; IDT) using rhAmp
555 Genotyping Master Mix and universal probe Reporter Mix (both IDT).
556

557 **Allele-specific gene expression**

558 Allele-specific expression of *DPYD* was measured using reverse transcriptase droplet
559 digital PCR (RT-ddPCR) by targeting variants in the coding region of *DPYD* (c.85T>C
560 and c.496A>G). Linkage disequilibrium (LD) between variants was calculated using
561 LDpair implemented within LDlink [48]. DNA was extracted from donor liver tissues
562 using the QIAamp DNA Mini Kit (Qiagen). For RNA extraction, tissues were lysed in
563 QIAzol (Qiagen), and RNA was extracted using the miRneasy Kit (Qiagen) with on-
564 column DNA digestion using RNase-free DNase (Qiagen). RNA quality was assessed
565 using an Agilent 2100 Bioanalyzer running 2100 Expert Software v.B.02.10 using
566 Agilent RNA 6000 Nano kits.

567

568 DNA samples were genotyped for c.85, c.496, and rs4294451 loci using TaqMan SNP
569 Genotyping assays (Applied Biosystems, Waltham, MA). Tissues that were
570 heterozygous for at least one of the coding region SNPs (i.e., c.85T>C and/or
571 c.496A>G) were suitable for allele-specific expression analysis because expression
572 from both alleles could be discriminated using the coding-region markers in mRNA.
573 Allele-specific expression was measured using RT-ddPCR with the One-Step RT-
574 ddPCR Advanced Kit for Probes (Qiagen) on a QX200 ddPCR Droplet Reader (BioRad,
575 Hercules, CA). Fractional abundance was calculated using QuantaSoft software
576 (BioRad). Poisson distributions were determined using Quantasoft and were used to
577 define 95% confidence intervals. To address possible biases associated with differing

578 probe efficacies caused by differential probe binding affinities or specificities, fractional
579 abundances are also reported relative to those measured in DNA.
580
581 **Quantitative analysis of chromatin conformation capture (3C-qPCR)**
582 Quantitative analysis of chromatin conformation capture (3C-qPCR) was performed as
583 described by Hagère *et al*. [49] with minor modifications. Cells were trypsinized using
584 0.25% w/v trypsin-EDTA (Thermo Fisher Scientific) and resuspended in 1% FBS
585 (MilliporeSigma) in DMEM (Corning) for counting by flow cytometry using a NovoCyte
586 3000 RYB system (Agilent Technologies). Cells (10^7) were pelleted by centrifugation at
587 300 × *g* at 22°C for 5 minutes. Supernatant was discarded, and cell pellets were
588 resuspended in 500 μL of 10% FBS in DMEM. Single-cell suspensions were obtained
589 by filtration through a 40-μm cell strainer (Corning). Crosslinking was performed by
590 adding 9.5 mL of 1% formaldehyde in 10% FBS in PBS per 10^7 cells. Reactions were
591 incubated at 22°C while tumbling for 10 minutes. Reactions were transferred to ice and
592 crosslinking was quenched by the addition of ice-cold glycine (MilliporeSigma) to
593 achieve a final concentration of 0.125M. Samples were centrifuged at 300 × *g* at 4°C for
594 5 minutes. Supernatant was removed and discarded. Crosslinked cell pellets were lysed
595 in 5 mL cold lysis buffer (10 mM Tris-HCl, pH7.5; 10 mM NaCl; 5 mM MgCl; 0.1 mM
596 EGTA; 1x Roche Complete protease inhibitor cocktail). Lysis was allowed to proceed for
597 10 minutes on ice with intermittent gentle pipetting to obtain homogeneous suspensions
598 of nuclei. Nuclei were pelleted at 500 × *g* at 4°C for 5 minutes and resuspended in 0.5
599 mL of 1.2x NEBuffer r2.1 (New England Biolabs) containing 3% sodium dodecyl sulfate

600 (SDS, MilliporeSigma). Samples were incubated at 37°C for 1 hour with shaking.
601 Following incubation, Triton X-100 (MilliporeSigma) was added to a final concentration
602 of 2%. Reactions were incubated at 37°C for 1 hour with shaking. To digest DNA, 600 U
603 of HindIII (New England Biolabs) was added, and reactions were incubated for 16 hours
604 at 37°C with shaking. To inactivate digestion, SDS was added to a final concentration of
605 1.6%, and reactions were incubated at 65°C for 25 minutes with shaking. Excess SDS
606 was sequestered by the addition of Triton X-100 to a final concentration of 1%. Samples
607 were divided into two aliquots, one for ligation and the other for non-ligation control.
608 Ligation was performed on a 7-fold dilution of HindIII-digested chromatin using 100 units
609 of Quick T4 DNA ligase (New England Biolabs) at 16°C for 16 hours, followed by 1 hour
610 at 22°C. Proteinase K (300 μg; New England Biolabs) was added to ligation mixtures
611 and non-ligated controls, and samples were incubated at 65°C for 16 hours to reverse
612 crosslinking. DNA was subsequently purified by adding 300 μg RNase A (Thermo-
613 Fisher) followed by a 30-minute incubation at 37°C and subsequent phenol-chloroform
614 extraction as described [49]. DNA pellets were washed with 70% ethanol and dissolved
615 in 10 mM Tris pH 7.5.
616
617 For qPCR, directional primers were designed within each fragment as depicted in
618 **Figure 4A**. Two anchors, one localized to the fragment containing the *DPYD* promoter
619 and the other to the E9 region, were selected and paired with primers designed to
620 "walk" across the length of the surrounding region. Reactions were carried out in 10 μL
621 reaction volumes, consisting of 5 pmol each primer and 1 μL of a 1:50 dilution of each

622 3C sample. Amplification was performed using LightCycler 480 SYBR Green I Master
623 Mix (Roche) on a LightCycler 480 thermocycler (Roche). PCR conditions were 95°C for
624 10 minutes, and 40 cycles of 95°C for 15 seconds, 65°C for 1 minute, and 72°C for 15
625 seconds. Enrichment was calculated as $2^{-(cp\ (ligated\ DNA) - cp\ (non\text{-}ligated\ DNA))}$. Enrichment with
626 the fragment containing the nearest primer was used as the control for interaction
627 frequency for further normalization between replicate experiments. Specifically, when
628 primer 5 was used as the anchor, data were normalized to the average of values
629 obtained from using primers 5 and 4. When primer 7 was used as the anchor, data were
630 normalized to the average of values from primers 7 and 8. All experiments were
631 performed in triplicate. Primer sequences and positions relative to the *DPYD*
632 transcription start site (TSS) for qPCR are listed in **Supplementary Table S2**.
633

634 **Cellular sensitivity to 5-FU**

635 Cell viability was monitored using the xCELLigence MP Real-Time Cell Analysis (RTCA)
636 system (Agilent) as previously described [14]. Background impedance values for RTCA
637 E-View plates (Agilent) were obtained using complete DMEM prior to plating cells. Cells
638 were seeded at a density of 5,000 cells per well and incubated for 20 hours, at which
639 time medium was removed and replaced with medium containing serial dilutions of 5-FU
640 ranging from 1.25 μM to 100 μM. Impedance values (expressed as cell index, CI, units)
641 were recorded every 15 minutes over the course of the experiment to monitor
642 proliferation. Results represent the average of three independent cultures. To account
643 for minute differences in plating and potential cell loss during drug addition, relative CI

644 was calculated as the CI measured 48 hours after 5-FU divided by the CI recorded
645 immediately after treatment. Four parameter logistic non-linear regression analysis was
646 used to determine IC_{50} concentrations (GraphPad Prism version 9, GraphPad Software,
647 San Diego, CA).
648
649 **CEBPB knockdown cells**
650 Lentiviral particles for CEBPB knockdown were generated by transfecting shRNA
651 plasmids (TRCN0000007440 and TRCN0000007442) or scramble shRNA control (a gift
652 from David Sabatini; Addgene_1864 [50]), psPAX2 (a gift from Didier Trono;
653 Addgene_12260), and pMD2.G (a gift from Didier Trono; Addgene_12259) into
654 HEK293T cells using TransIT-Lenti (Mirus Bio, Madison, WI). A 3:1 ratio of transfection
655 reagent to total plasmid was used for all transfections. For all transfections, medium
656 was changed 14 hours after transfection, and viral supernatants were collected 34
657 hours later. Supernatants were filtered using 0.45 μm PVDF filters (Millipore) to remove
658 debris/cells and used directly for transductions. For transductions, target cells were
659 seeded at a density of 4×10^5 cells per well in 6-well plates and incubated with 500 µL
660 virus-containing supernatant, 12.5 µg/mL polybrene (MilliporeSigma), and 1.5 mL fresh
661 DMEM culture medium.
662
663 **Analysis of transcription factor binding motifs**
664 Potential transcription factor binding sites within the E9 region were determined using
665 TFBSTools [51]. DNA sequence corresponding to the 101-nucleotide region centered

666 on rs4294451 was retrieved from genome build GRCh38.p13 (NC_000001.11). A
667 second DNA string was created to mimic the sequence corresponding to the rs4294451
668 A allele. Position frequency matrices for each transcription factor were retrieved from
669 JASPAR CORE 2022 [52] and converted to log-scale position weight matrices using the
670 toPWM method implemented in TFBSTools. JASPAR includes binding site information
671 from multiple sources. ENCODE data [53] were available for CEBPB (matrix ID
672 MA04661) and JUND (MA0491.1). In absence of ENCODE data, alternatives including
673 PAZAR [54] (HNF4A, MA0114.2; FOXA1, MA0148.3), REMAP [55] (FOXA2,
674 MA0047.3), and data from an individual publication [56] (MYBL2, MA0777.1) were used.
675 EP300 is a cofactor that does not recognize a specific DNA motif on its own; instead, it
676 interacts with various DNA-binding factors to modify chromatin and facilitate activation
677 of target genes. As such, P300 does not have a DNA-binding motif and was not
678 included in analysis. Nucleotide sequences were scanned using the patterns presented
679 in the position weight matrices to identify putative transcription factor binding sites.
680 Forward and reverse strands were searched, and the 70th percentile between the
681 minimum and maximum possible value for a matrix was used as the minimum threshold
682 score. Empirical p-values for each score were calculated by an exact method using the
683 TFMPvalue R package. R version 4.2.2 was used for analyses
684
685 **Statistical analyses**
686 Significance was defined as $p<0.05$ unless otherwise noted in the text. Pairwise
687 comparisons were performed using unpaired two-tailed Student's t tests calculated

688 using GraphPad Prism version 9. One-way ANOVA statistics were calculated using
689 GraphPad Prism. Summary statistics pertaining to allele-specific expression were
690 calculated using Quantasoft software as described in the Allele-specific gene
691 expression section. Transcription factor binding predictions and associated analyses
692 were performed in R version 4.2.2 as described above.
693
694 **Availability of data and materials**
695 All data generated or analyzed during this study are included in this published article
696 and its supplementary information files.
697

Acknowledgements

This work was supported by the National Cancer Institute of the National Institutes of Health under award number R01CA251065 (SMO, P.I.). Allele-specific expression experiments were supported by the Swiss National Science Foundation under award number 320030_212583 (CRL, P.I.). The funding bodies did not contribute to the design of the study, the collection, analysis, and interpretation of data, or in writing the manuscript.

705 **REFERENCES**

706

707 1. Ramos, A., S. Sadeghi, and H. Tabatabaeian, *Battling Chemoresistance in*
708 *Cancer: Root Causes and Strategies to Uproot Them.* Int J Mol Sci, 2021.
709 **22**(17).

710 2. Mansoori, B., A. Mohammadi, S. Davudian, S. Shirjang, and B. Baradaran, *The*
711 *Different Mechanisms of Cancer Drug Resistance: A Brief Review.* Adv Pharm
712 Bull, 2017. **7**(3): p. 339-348.

713 3. Koutsi, M.A., M. Pouliou, L. Champezou, G. Vatsellas, A.I. Giannopoulou, C.
714 Piperi, and M. Agelopoulos, *Typical Enhancers, Super-Enhancers, and Cancers.*
715 Cancers (Basel), 2022. **14**(18).

716 4. Lauschke, V.M., Y. Zhou, and M. Ingelman-Sundberg, *Novel genetic and*
717 *epigenetic factors of importance for inter-individual differences in drug*
718 *disposition, response and toxicity.* Pharmacol Ther, 2019. **197**: p. 122-152.

719 5. Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter,
720 M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E.
721 Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle,
722 A.S. Whittemore, M. Boehnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines,
723 T.F. Mackay, S.A. McCarroll, and P.M. Visscher, *Finding the missing heritability*
724 *of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.

725 6. Maurano, M.T., R. Humbert, E. Rynes, R.E. Thurman, E. Haugen, H. Wang, A.P.
726 Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin,
727 S. Stehling-Sun, A.K. Johnson, T.K. Canfield, E. Giste, M. Diegel, D. Bates, R.S.
728 Hansen, S. Neph, P.J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C.
729 Cotsapas, N. Sotoodehnia, I. Glass, S.R. Sunyaev, R. Kaul, and J.A.
730 Stamatoyannopoulos, *Systematic localization of common disease-associated*
731 *variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.

732 7. Boix, C.A., B.T. James, Y.P. Park, W. Meuleman, and M. Kellis, *Regulatory*
733 *genomic circuitry of human disease loci by integrative epigenomics.* Nature,
734 2021. **590**(7845): p. 300-307.

735 8. Claringbould, A. and J.B. Zaugg, *Enhancers in disease: molecular basis and*
736 *emerging treatment strategies.* Trends Mol Med, 2021. **27**(11): p. 1060-1073.

737 9. Etheridge, A.S., P.J. Gallins, D. Jima, K.A. Broadaway, M.J. Ratain, E. Schuetz,
738 E. Schadt, A. Schroder, C. Molony, Y. Zhou, K.L. Mohlke, F.A. Wright, and F.
739 Innocenti, *A New Liver Expression Quantitative Trait Locus Map From 1,183*
740 *Individuals Provides Evidence for Novel Expression Quantitative Trait Loci of*

741 *Drug Response, Metabolic, and Sex-Biased Phenotypes.* Clin Pharmacol Ther,
742 2019.

743 10. Sommadossi, J.P., D.A. Gewirtz, R.B. Diasio, C. Aubert, J.P. Cano, and I.D.
744 Goldman, *Rapid catabolism of 5-fluorouracil in freshly isolated rat hepatocytes as*
745 *analyzed by high performance liquid chromatography.* J Biol Chem, 1982.
746 **257**(14): p. 8171-6.

747 11. Heggie, G.D., J.P. Sommadossi, D.S. Cross, W.J. Huster, and R.B. Diasio,
748 *Clinical pharmacokinetics of 5-fluorouracil and its metabolites in plasma, urine,*
749 *and bile.* Cancer Res, 1987. **47**(8): p. 2203-6.

750 12. Amstutz, U., L.M. Henricks, S.M. Offer, J. Barbarino, J.H.M. Schellens, J.J.
751 Swen, T.E. Klein, H.L. McLeod, K.E. Caudle, R.B. Diasio, and M. Schwab,
752 *Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for*
753 *Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017*
754 *Update.* Clin Pharmacol Ther, 2018. **103**(2): p. 210-216.

755 13. Lee, A.M., Q. Shi, E. Pavey, S.R. Alberts, D.J. Sargent, F.A. Sinicrope, J.L.
756 Berenberg, R.M. Goldberg, and R.B. Diasio, *DPYD variants as predictors of 5-*
757 *fluorouracil toxicity in adjuvant colon cancer treatment (NCCTG N0147).* J Natl
758 Cancer Inst, 2014. **106**(12).

759 14. Shrestha, S., C. Zhang, C.R. Jerde, Q. Nie, H. Li, S.M. Offer, and R.B. Diasio,
760 *Gene-Specific Variant Classifier (DPYD-Varifier) to Identify Deleterious Alleles of*
761 *Dihydropyrimidine Dehydrogenase.* Clin Pharmacol Ther, 2018. **104**(4): p. 709-
762 718.

763 15. Offer, S.M., C.C. Fossum, N.J. Wegner, A.J. Stuflesser, G.L. Butterfield, and
764 R.B. Diasio, *Comparative functional analysis of DPYD variants of potential*
765 *clinical relevance to dihydropyrimidine dehydrogenase activity.* Cancer Res,
766 2014. **74**(9): p. 2545-54.

767 16. Offer, S.M., A.M. Lee, L.K. Mattison, C. Fossum, N.J. Wegner, and R.B. Diasio,
768 *A DPYD variant (Y186C) in individuals of african ancestry is associated with*
769 *reduced DPD enzyme activity.* Clin Pharmacol Ther, 2013. **94**(1): p. 158-66.

770 17. Offer, S.M., N.J. Wegner, C. Fossum, K. Wang, and R.B. Diasio, *Phenotypic*
771 *profiling of DPYD variations relevant to 5-fluorouracil sensitivity using real-time*
772 *cellular analysis and in vitro measurement of enzyme activity.* Cancer Res, 2013.
773 **73**(6): p. 1958-68.

774 18. Nie, Q., S. Shrestha, E.E. Tapper, C.S. Trogstad-Isaacson, K.J. Bouchonville,
775 A.M. Lee, R. Wu, C.R. Jerde, Z. Wang, P.A. Kubica, S.M. Offer, and R.B. Diasio,
776 *Quantitative Contribution of rs75017182 to Dihydropyrimidine Dehydrogenase*

mRNA Splicing and Enzyme Activity. Clin Pharmacol Ther, 2017. **102**(4): p. 662-670.

19. Meulendijks, D., L.M. Henricks, G.S. Sonke, M.J. Deenen, T.K. Froehlich, U. Amstutz, C.R. Largiader, B.A. Jennings, A.M. Marinaki, J.D. Sanderson, Z. Kleibl, P. Kleiblova, M. Schwab, U.M. Zanger, C. Palles, I. Tomlinson, E. Gross, A.B. van Kuilenburg, C.J. Punt, M. Koopman, J.H. Beijnen, A. Cats, and J.H. Schellens, *Clinical relevance of DPYD variants c.1679T>G, c.1236G>A/HapB3, and c.1601G>A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data.* Lancet Oncol, 2015. **16**(16): p. 1639-50.

20. Fishilevich, S., R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, N. Rosen, A. Kohn, M. Twik, M. Safran, D. Lancet, and D. Cohen, *GeneHancer: genome-wide integration of enhancers and target genes in GeneCards.* Database (Oxford), 2017. **2017**.

21. Zerbino, D.R., S.P. Wilder, N. Johnson, T. Juettemann, and P.R. Flicek, *The ensembl regulatory build.* Genome Biol, 2015. **16**: p. 56.

22. Yeo, N.C., A. Chavez, A. Lance-Byrne, Y. Chan, D. Menn, D. Milanova, C.C. Kuo, X. Guo, S. Sharma, A. Tung, R.J. Cecchi, M. Tuttle, S. Pradhan, E.T. Lim, N. Davidsohn, M.R. Ebrahimkhani, J.J. Collins, N.E. Lewis, S. Kiani, and G.M. Church, *An enhanced CRISPR repressor for targeted mammalian gene regulation.* Nat Methods, 2018. **15**(8): p. 611-616.

23. Hilton, I.B., A.M. D'Ippolito, C.M. Vockley, P.I. Thakore, G.E. Crawford, T.E. Reddy, and C.A. Gersbach, *Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers.* Nat Biotechnol, 2015. **33**(5): p. 510-7.

24. Chen, S., L.C. Francioli, J.K. Goodrich, R.L. Collins, M. Kanai, Q. Wang, J. Alföldi, N.A. Watts, C. Vittal, L.D. Gauthier, T. Poterba, M.W. Wilson, Y. Tarasova, W. Phu, M.T. Yohannes, Z. Koenig, Y. Farjoun, E. Banks, S. Donnelly, S. Gabriel, N. Gupta, S. Ferriera, C. Tolonen, S. Novod, L. Bergelson, D. Roazen, V. Ruano-Rubio, M. Covarrubias, C. Llanwarne, N. Petrillo, G. Wade, T. Jeandet, R. Munshi, K. Tibbetts, g.P. Consortium, A. O'Donnell-Luria, M. Solomonson, C. Seed, A.R. Martin, M.E. Talkowski, H.L. Rehm, M.J. Daly, G. Tiao, B.M. Neale, D.G. MacArthur, and K.J. Karczewski, *A genome-wide mutational constraint map quantified from variation in 76,156 human genomes.* bioRxiv, 2022: p. 2022.03.20.485034.

25. Beumer, J.H., E. Chu, C. Allegra, Y. Tanigawara, G. Milano, R. Diasio, T.W. Kim, R.H. Mathijssen, L. Zhang, D. Arnold, K. Muneoka, N. Boku, and M. Joerger, *Therapeutic Drug Monitoring in Oncology: International Association of*

Therapeutic Drug Monitoring and Clinical Toxicology Recommendations for 5-Fluorouracil Therapy. Clin Pharmacol Ther, 2019. **105**(3): p. 598-613.

26. Henricks, L.M., C. Lunenburg, F.M. de Man, D. Meulendijks, G.W.J. Frederix, E. Kienhuis, G.J. Creemers, A. Baars, V.O. Dezentje, A.L.T. Imholz, F.J.F. Jeurissen, J.E.A. Portielje, R.L.H. Jansen, P. Hamberg, A.J. Ten Tije, H.J. Droogendijk, M. Koopman, P. Nieboer, M.H.W. van de Poel, C. Mandigers, H. Rosing, J.H. Beijnen, E.V. Werkhoven, A.B.P. van Kuilenburg, R.H.N. van Schaik, R.H.J. Mathijssen, J.J. Swen, H. Gelderblom, A. Cats, H.J. Guchelaar, and J.H.M. Schellens, *DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis.* Lancet Oncol, 2018. **19**(11): p. 1459-1467.

27. Jiang, W., Z. Lu, Y. He, and R.B. Diasio, *Dihydropyrimidine dehydrogenase activity in hepatocellular carcinoma: implication in 5-fluorouracil-based chemotherapy.* Clin Cancer Res, 1997. **3**(3): p. 395-9.

28. Kikuchi, O., S. Ohashi, Y. Nakai, S. Nakagawa, K. Matsuoka, T. Kobunai, T. Takechi, Y. Amanuma, M. Yoshioka, T. Ida, Y. Yamamoto, Y. Okuno, S. Miyamoto, H. Nakagawa, K. Matsubara, T. Chiba, and M. Muto, *Novel 5-fluorouracil-resistant human esophageal squamous cell carcinoma cells with dihydropyrimidine dehydrogenase overexpression.* Am J Cancer Res, 2015. **5**(8): p. 2431-40.

29. Li, L.H., H. Dong, F. Zhao, J. Tang, X. Chen, J. Ding, H.T. Men, W.X. Luo, Y. Du, J. Ge, B.X. Tan, D. Cao, and J.Y. Liu, *The upregulation of dihydropyrimidine dehydrogenase in liver is involved in acquired resistance to 5-fluorouracil.* Eur J Cancer, 2013. **49**(7): p. 1752-60.

30. Mayberry, R.M., R.J. Coates, H.A. Hill, L.A. Click, V.W. Chen, D.F. Austin, C.K. Redmond, C.M. Fenoglio-Preiser, C.P. Hunter, M.A. Haynes, and et al., *Determinants of black/white differences in colon cancer survival.* J Natl Cancer Inst, 1995. **87**(22): p. 1686-93.

31. Govindarajan, R., R.V. Shah, L.G. Erkman, and L.F. Hutchins, *Racial differences in the outcome of patients with colorectal carcinoma.* Cancer, 2003. **97**(2): p. 493-8.

32. Alexander, D.D., J. Waterbor, T. Hughes, E. Funkhouser, W. Grizzle, and U. Manne, *African-American and Caucasian disparities in colorectal cancer mortality and survival by data source: an epidemiologic review.* Cancer Biomark, 2007. **3**(6): p. 301-13.

33. Akira, S., H. Isshiki, T. Sugita, O. Tanabe, S. Kinoshita, Y. Nishio, T. Nakajima, T. Hirano, and T. Kishimoto, *A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family.* EMBO J, 1990. **9**(6): p. 1897-906.

853 34. Jakobsen, J.S., J. Waage, N. Rapin, H.C. Bisgaard, F.S. Larsen, and B.T. Porse,
854 *Temporal mapping of CEBPA and CEBPB binding during liver regeneration*
855 *reveals dynamic occupancy and specific regulatory codes for homeostatic and*
856 *cell cycle gene batteries.* Genome Res, 2013. **23**(4): p. 592-603.

857 35. Begay, V., J.J. Smink, C. Loddenkemper, K. Zimmermann, C. Rudolph, M.
858 Scheller, D. Steinemann, U. Leser, B. Schlegelberger, H. Stein, and A. Leutz,
859 *Deregulation of the endogenous C/EBPbeta LIP isoform predisposes to*
860 *tumorigenesis.* J Mol Med (Berl), 2015. **93**(1): p. 39-49.

861 36. Li, Y., E. Bevilacqua, C.B. Chiribau, M. Majumder, C. Wang, C.M. Croniger, M.D.
862 Snider, P.F. Johnson, and M. Hatzoglou, *Differential control of the*
863 *CCAAT/enhancer-binding protein beta (C/EBPbeta) products liver-enriched*
864 *transcriptional activating protein (LAP) and liver-enriched transcriptional inhibitory*
865 *protein (LIP) and the regulation of gene expression during the response to*
866 *endoplasmic reticulum stress.* J Biol Chem, 2008. **283**(33): p. 22443-56.

867 37. Miller, M., *Interactions of CCAAT/enhancer-binding protein beta with*
868 *transcriptional coregulators.* Postepy Biochem, 2016. **62**(3): p. 343-348.

869 38. Seo, J., D.D. Kocak, L.C. Bartelt, C.A. Williams, A. Barrera, C.A. Gersbach, and
870 T.E. Reddy, *AP-1 subunits converge promiscuously at enhancers to potentiate*
871 *transcription.* Genome Res, 2021. **31**(4): p. 538-550.

872 39. Cohen, D.M., H.W. Lim, K.J. Won, and D.J. Steger, *Shared nucleotide flanks*
873 *confer transcriptional competency to bZip core motifs.* Nucleic Acids Res, 2018.
874 **46**(16): p. 8371-8384.

875 40. Lountos, G.T., S. Cherry, J.E. Tropea, A. Wlodawer, and M. Miller, *Structural*
876 *basis for cell type specific DNA binding of C/EBPbeta: The case of cell cycle*
877 *inhibitor p15INK4b promoter.* J Struct Biol, 2022. **214**(4): p. 107918.

878 41. Zhang, X.F., K.K. Li, L. Gao, S.Z. Li, K. Chen, J.B. Zhang, D. Wang, R.F. Tu, J.X.
879 Zhang, K.X. Tao, G. Wang, and X.D. Zhang, *miR-191 promotes tumorigenesis of*
880 *human colorectal cancer through targeting C/EBPbeta.* Oncotarget, 2015. **6**(6): p.
881 4144-58.

882 42. Wang, D., L. Yang, W. Yu, Q. Wu, J. Lian, F. Li, S. Liu, A. Li, Z. He, J. Liu, Z.
883 Sun, W. Yuan, and Y. Zhang, *Colorectal cancer cell-derived CCL20 recruits*
884 *regulatory T cells to promote chemoresistance via FOXO1/CEBPB/NF-kappaB*
885 *signaling.* J Immunother Cancer, 2019. **7**(1): p. 215.

886 43. Offer, S.M., G.L. Butterfield, C.R. Jerde, C.C. Fossum, N.J. Wegner, and R.B.
887 Diasio, *microRNAs miR-27a and miR-27b directly regulate liver*
888 *dihydropyrimidine dehydrogenase expression through two conserved binding*
889 *sites.* Mol Cancer Ther, 2014. **13**(3): p. 742-51.

890 44. Amstutz, U., S.M. Offer, J. Sistonen, M. Joerger, R.B. Diasio, and C.R. Largiader,
891 *Polymorphisms in MIR27A Associated with Early-Onset Toxicity in*
892 *Fluoropyrimidine-Based Chemotherapy.* Clin Cancer Res, 2015. **21**(9): p. 2038-
893 44.

894 45. Hamzic, S., D. Scharer, S.M. Offer, D. Meulendijks, C. Nakas, R.B. Diasio, S.
895 Fontana, M. Wehrli, S. Schurch, U. Amstutz, and C.R. Largiader, *Haplotype*
896 *structure defines effects of common DPYD variants c.85T > C (rs1801265) and*
897 *c.496A > G (rs2297595) on dihydropyrimidine dehydrogenase activity:*
898 *Implication for 5-fluorouracil toxicity.* Br J Clin Pharmacol, 2021. **87**(8): p. 3234-
899 3243.

900 46. Medwid, S., T.J. Wigle, and R.B. Kim, *Fluoropyrimidine-associated toxicity and*
901 *DPYD variants c.85T>C, c.496A>G, and c.1236G>A: impact of haplotype.*
902 Cancer Chemother Pharmacol, 2023. **91**(1): p. 97-102.

903 47. Perez, A.R., Y. Pritykin, J.A. Vidigal, S. Chhangawala, L. Zamparo, C.S. Leslie,
904 and A. Ventura, *GuideScan software for improved single and paired CRISPR*
905 *guide RNA design.* Nat Biotechnol, 2017. **35**(4): p. 347-349.

906 48. Machiela, M.J. and S.J. Chanock, *LDlink: a web-based application for exploring*
907 *population-specific haplotype structure and linking correlated alleles of possible*
908 *functional variants.* Bioinformatics, 2015. **31**(21): p. 3555-7.

909 49. Hagege, H., P. Klous, C. Braem, E. Splinter, J. Dekker, G. Cathala, W. de Laat,
910 and T. Forne, *Quantitative analysis of chromosome conformation capture assays*
911 *(3C-qPCR).* Nat Protoc, 2007. **2**(7): p. 1722-33.

912 50. Sarbassov, D.D., D.A. Guertin, S.M. Ali, and D.M. Sabatini, *Phosphorylation and*
913 *regulation of Akt/PKB by the rictor-mTOR complex.* Science, 2005. **307**(5712): p.
914 1098-101.

915 51. Tan, G. and B. Lenhard, *TFBSTools: an R/bioconductor package for transcription*
916 *factor binding site analysis.* Bioinformatics, 2016. **32**(10): p. 1555-6.

917 52. Castro-Mondragon, J.A., R. Riudavets-Puig, I. Rauluseviciute, R.B. Lemma, L.
918 Turchi, R. Blanc-Mathieu, J. Lucas, P. Boddie, A. Khan, N. Manosalva Perez, O.
919 Fornes, T.Y. Leung, A. Aguirre, F. Hammal, D. Schmelter, D. Baranasic, B.
920 Ballester, A. Sandelin, B. Lenhard, K. Vandepoele, W.W. Wasserman, F. Parcy,
921 and A. Mathelier, *JASPAR 2022: the 9th release of the open-access database of*
922 *transcription factor binding profiles.* Nucleic Acids Res, 2022. **50**(D1): p. D165-
923 D173.

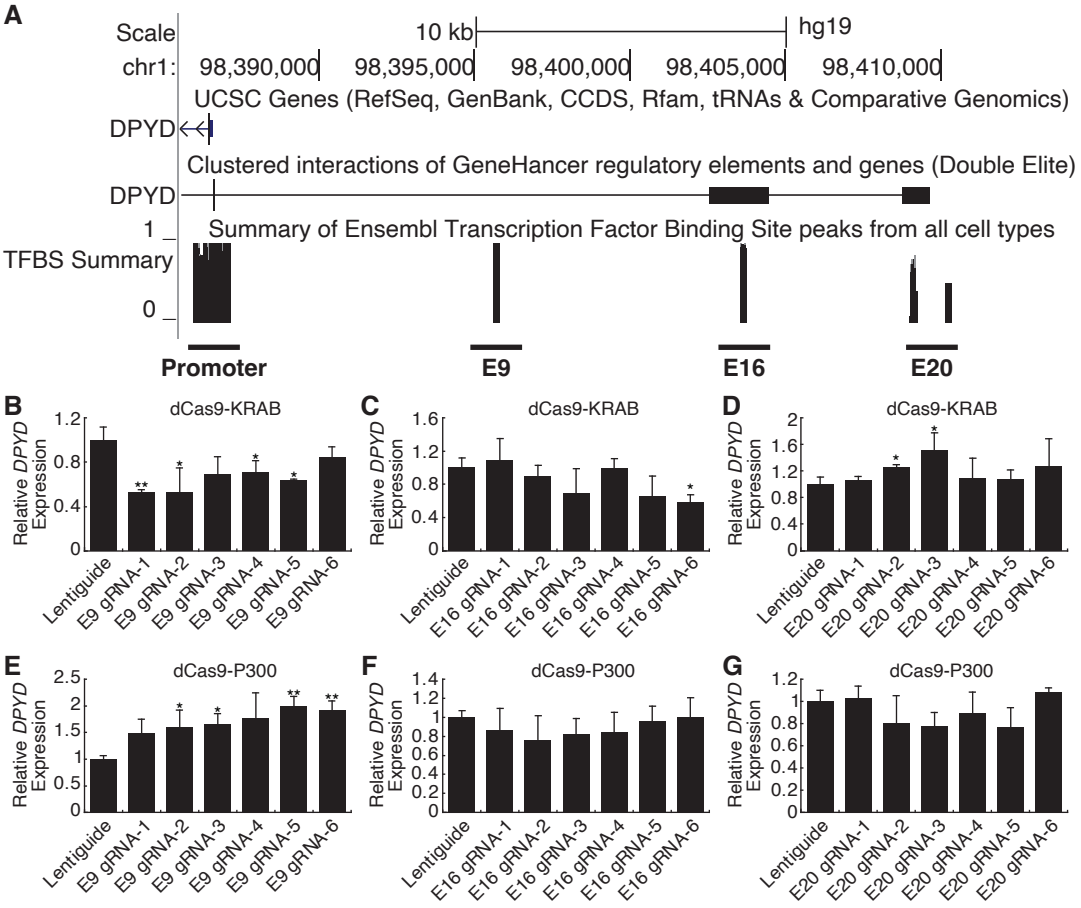
924 53. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human*
925 *genome.* Nature, 2012. **489**(7414): p. 57-74.

926 54. Portales-Casamar, E., S. Kirov, J. Lim, S. Lithwick, M.I. Swanson, A. Ticoll, J.
927 Snoddy, and W.W. Wasserman, *PAZAR: a framework for collection and*
928 *dissemination of cis-regulatory sequence annotation.* Genome Biol, 2007. **8**(10):
929 p. R207.

930 55. Hammal, F., P. de Langen, A. Bergon, F. Lopez, and B. Ballester, *ReMap 2022:*
931 *a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions*
932 *from an integrative analysis of DNA-binding sequencing experiments.* Nucleic
933 Acids Res, 2022. **50**(D1): p. D316-D325.

934 56. Jolma, A., J. Yan, T. Whitington, J. Toivonen, K.R. Nitta, P. Rastas, E.
935 Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J.M. Vaquerizas, R.
936 Vincentelli, N.M. Luscombe, T.R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja,
937 and J. Taipale, *DNA-binding specificities of human transcription factors.* Cell,
938 2013. **152**(1-2): p. 327-39.
939

940 **FIGURES AND FIGURE LEGENDS**



941 Zhang et al., Figure 1

942 **Figure 1. Identification of a novel *DPYD* enhancer.** (A) Candidate *DPYD* enhancer
943 regions were selected for further study using data from GeneHancer and Ensembl
944 Regulatory Build. Coordinates are based on GRCh37/hg19. Regions are termed E9,
945 E16, and E20 based on approximate distance upstream of the *DPYD* transcription start
946 site. For CRISPR inhibition (CRISPRi), *DPYD* expression was measured in HepG2 cells
947 expressing dCas9-KRAB following transfection with guide-RNAs specific to the E9 (B),
948 E16 (C), and E20 (D) regions. For CRISPR activation (CRISPRa), *DPYD* expression
949 was measured in HepG2 cells expressing dCas9-P300 following transfection with guide-

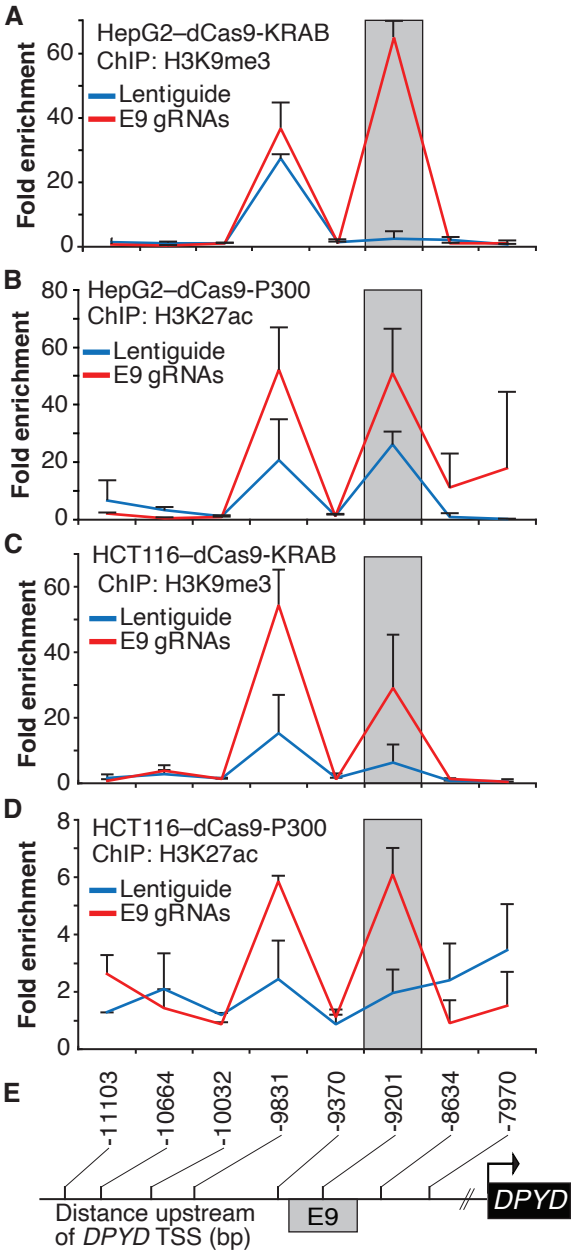
950 RNAs specific to E9 (E), E16 (F), and E20 (G). Data represent the mean of three

951 independent biological replicates ± SD. *, $p<0.05$; **, $p<0.005$. P-values were calculated

952 using two-tailed Student's t-test comparing results to those from lentiguide controls.

953

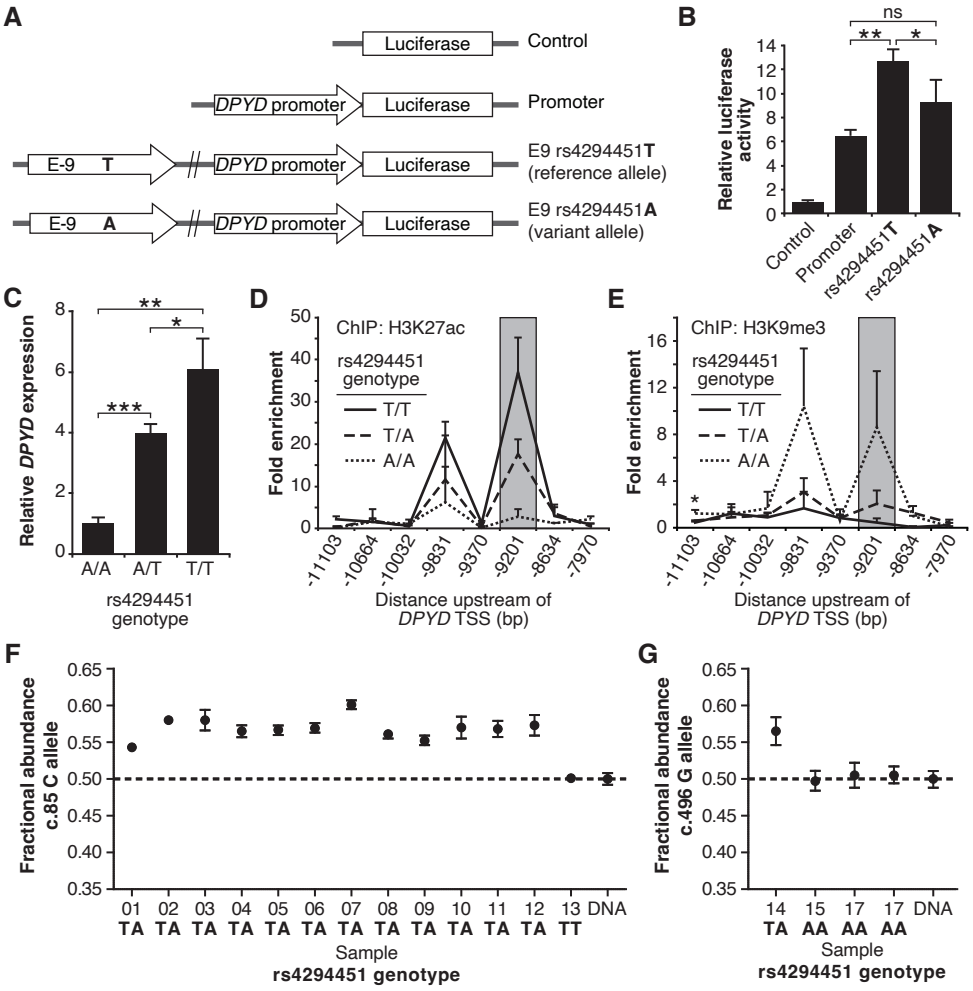
954



Zhang et al., Figure 2

Figure 2. Epigenetic changes at the E9 region induced by CRISPRi/CRISPRa. Lentiguide vectors encoding guide-RNAs targeting the E9 region (E9 gRNAs) or empty vector control (Lentiguide) were transduced into HepG2 cells expressing dCas9-KRAB (A) or dcas9-P300 (B) and HCT116 cells expressing dCas9-KRAB (C) or dcas9-P300 (D). Chromatin immunoprecipitation (ChIP) was performed using antibodies specific to

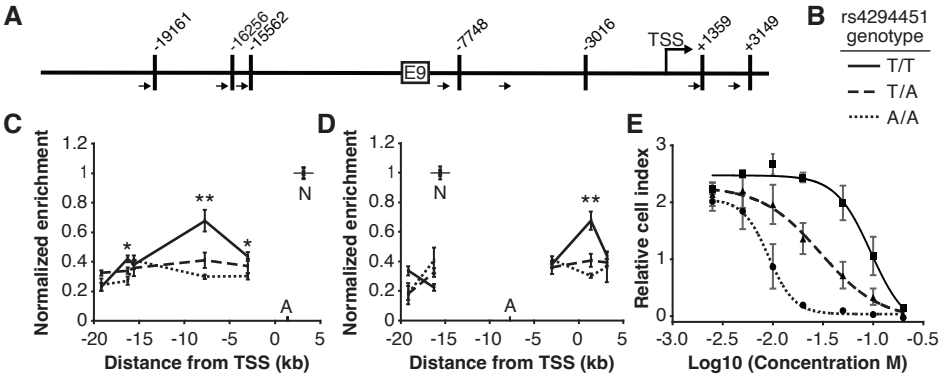
961 H3K9me3 (A, C) or H3K27ac (B, D). Quantitative PCR using primers centered at the
962 indicated regions (E) was used to measure the relative abundance of H3K9me3 and
963 H3K27ac. Data are presented relative to input DNA control and are further normalized
964 to IgG control. Error bars represent the SD of three independent biological replicate
965 experiments.
966



Zhang et al., Figure 3

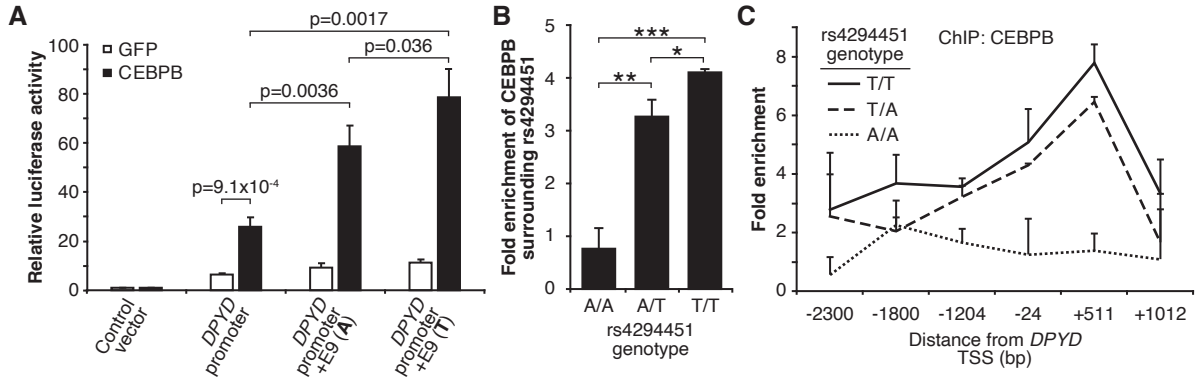
Figure 3. The rs4294451 T allele is associated with elevated *DPYD* expression. (A) Luciferase reporter constructs were generated by cloning the E9 region containing the reference rs4294451 T allele or the variant rs4294451 A allele into reporter vectors containing the *DPYD* promoter. (B) Luciferase reporter activity was measured for vectors shown in panel A. Error bars represent the standard deviation of three independent biological replicates. (C) Relative expression of endogenous *DPYD* measured via RT-qPCR in HCT116 cells engineered using CRISPR-mediated genome editing to contain the depicted genotypes for rs4294451. Chromatin enrichment of

976 H3K27ac (D) and H3K9me3 (E) was measured using ChIP-qPCR in HCT116 cells
977 engineered to contain the indicated genotypes at rs4294451. (F) *DPYD* allele-specific
978 expression was measured in human liver tissues using the C and T alleles at position
979 c.85. (G) Allele-specific expression was measured using the c.496-A and G alleles. All
980 panels: *, $p<0.05$; **, $p<0.005$; ***, $p<0.0005$. P-values were calculated as pairwise
981 comparisons between the indicated groups using two-tailed Student's t-tests.
982



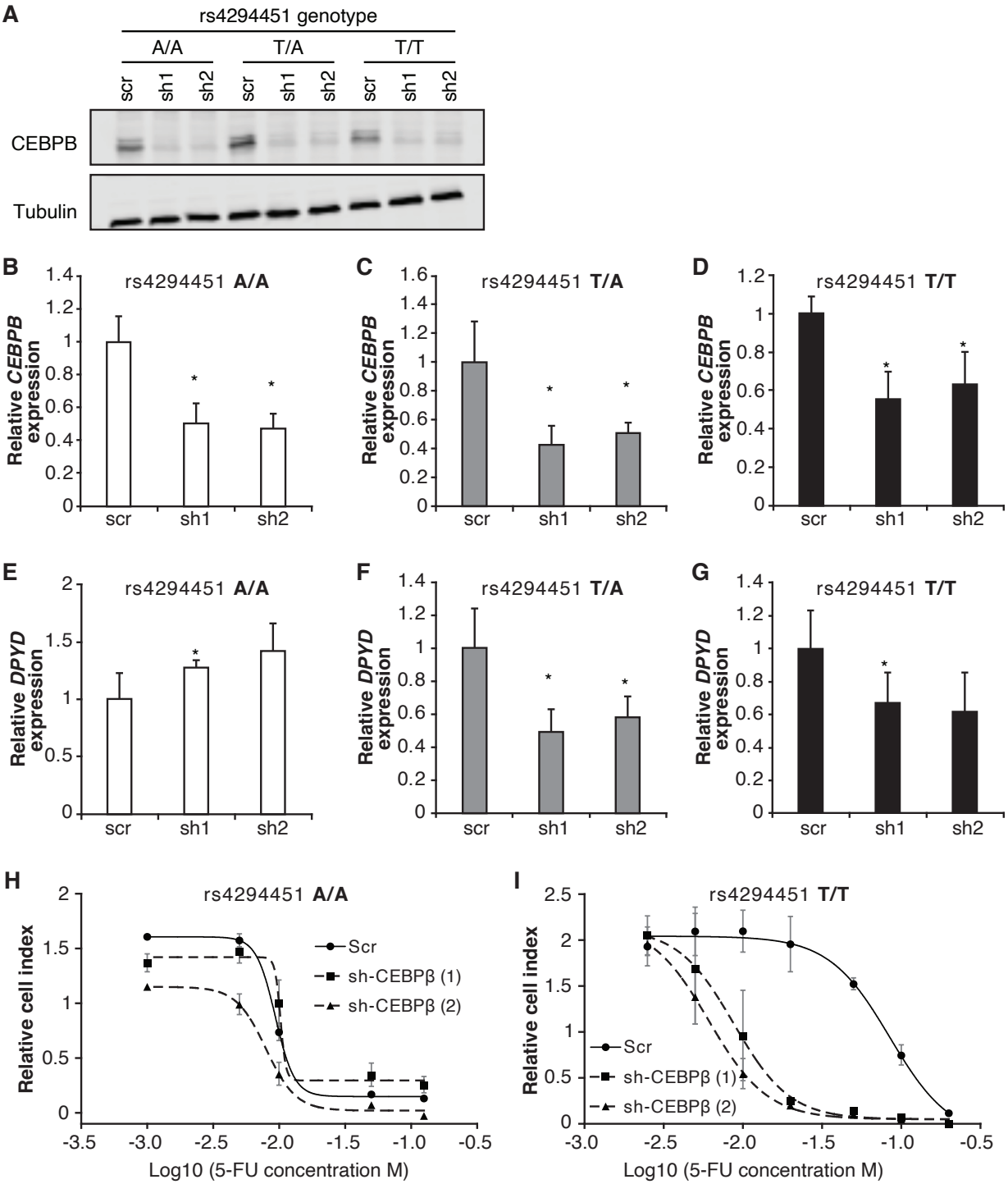
Zhang et al., Figure 4

Figure 4. The rs4294451 T allele is associated with increased interaction between E9 and the *DPYD* promoter. (A) Schematic of *Hin*dIII restriction enzyme sites (vertical bars) and primers (arrows) used for chromatin conformation capture (3C) relative to the *DPYD* transcription start site (TSS) and E9 region. (B) Legend for panels C–E. (C) 3C of chromatin interactions in rs4294451 knock-in HCT116 cells using anchor primer positioned within the digestion fragment containing the *DPYD* promoter. A, location of anchor primer; N, location of primer used for data normalization. (D) 3C of knock-in cells using anchor primer positioned within the fragment containing the E9 region. For panels C–D: one-way ANOVA p: *, $p<0.01$; **, $p<0.001$; all other data points, $p>0.01$. (E) Knock-in cells were treated with dilutions of 5-fluorouracil (5-FU) and viability assessed using real-time cell analysis (RTCA). Cell index is a measure of impedance between electrodes that are arrayed at the bottom of the RTCA plate and is representative of the number of live cells attached to the culture plate. Data are from 48 hours of 5-FU treatment. For all plotted data, the mean ± SD of three independent replicates is presented.



Zhang et al., Figure 5

Figure 5. Rs4294451 T allele is associated with higher occupancy of CEBPB at E9 and the *DPYD* promoter. (A) Expression plasmids for CEBPB or GFP (control) were co-transfected into HEK293T cells with the luciferase reporter plasmids depicted in Figure 3A. (B) ChIP-qPCR was performed to examine the relative CEBPB enrichment at the DNA fragments within the E9 region in knock-in HCT116 cells for the rs4294451 genotypes indicated. Primers used for E9 are centered on the position at 9201 nucleotides upstream of the *DPYD* TSS. (C) ChIP-qPCR was used to measure CEBPB occupancy surrounding the *DPYD* promoter region. *, $p<0.05$; **, $p<0.005$; ***, $p<0.0005$. P-values were calculated as pairwise comparisons between the indicated groups using two-tailed Student's t-tests. For all panels, error bars represent the SD of three independent replicates.



1013

Zhang et al., Figure 6

1014 **Figure 6. CEBPB-mediated upregulation of *DPYD* is dependent on rs4294451 T**
1015 **allele.** (A) Immunoblot showing knockdown of CEBPB expression in HCT116 cells
1016 carrying different rs4294451 genotypes transduced with lentiviral particles encoding two
1017 independent shRNAs against CEBPB (sh1 and sh2) or a scrambled control shRNA
1018 (scr). *CEBPB* expression was measured by RT-qPCR in HCT116 A/A (B), T/A (C) and
1019 T/T cells (D) transduced with the indicated shRNA lentiviral particles. *DPYD* expression
1020 was measured in shCEBPB and scramble control HCT116 A/A (E), T/A (F) and T/T (G)
1021 cells. The effect of CEBPB knockdown on cell viability in HCT116 A/A cells (H) and
1022 HCT116 T/T cells (I) was measured by RTCA. Data shown are from 48 hours of 5-FU
1023 treatment at the indicated concentrations. *, $p<0.05$, calculated as a pairwise two-sided
1024 Student's t-test comparing the indicated data to that of the associated scr control. Error
1025 bars represent the SD from three independent replicates.
1026

1027 **SUPPLEMENTARY TABLES**

1028

Supplementary Table S1: Primers used for ChIP, cloning, qPCR, and site-directed mutagenesis.

Target	Forward Primer (5'-3')	Reverse Primer (5'-3')
E9 ChIP-1	CCTACCTACCACCCCCAAGA	TGCCCAGGACATTACACATGA
E9 ChIP-2	CATCAGGTTGCTTTTTGCAGC	TTAGCGTGGACTACCAGGGA
E9 ChIP-3	TGGGACTCAAAAAGCGGTTCA	GGTGTTGCGGGTGCTGATTA
E9 ChIP-4	GGTTCAACAGAGGATGCAACAC	AGGGAAAGATTTTCCTGGCCC
E9 ChIP-5	TAAAGATGCACCGAGGTGGG	AGTGTCTGGATTTAAGTAGATGTGC
E9 ChIP-6	AGCAGGAATGAGAAGGAGAGAAG	TCAGTCTCTCACTCCAAACCC
E9 ChIP-7	GCTCTTTTCATTGAAGCCTAAAACA	CCATCATACTTTTTCCAATTGTTGC
E9 ChIP-8	AATGATGAGAGGAAGATGACAAAGT	TGCCTACGCGATGAGTTGTA
DPYD promoter	AGTCGATATCCACAGTGTCTGTGTCTGGC	AGTCAAGCTTGCTCGATGTCCGCCGAG
E9 region	AGTCGGTACCGAATAAAACCAAAATAAAATCCATTTGGACGTTT	AGTCGAGCTCTTTGTGCAAAGGACCTTGGTATTTCC
Rs4294451 A>T	AAAGAAAAATAAATAAAAAAAGGAAAAATCTATAAGC	TTCTGGGGGTTGGTGTTG
Cas9	AACAGCCGCGAGAGAATGAA	CACGGGGTGTTCTTTCAGGA
DPYD	GTAAGGACTCGGCGGACATC	GCCGAAGTGGAACACAGAGT
L32	CCTTGTGAAGCCCAAGATCG	TGCCGGATGAACTTCTTGGT
CEBPB	CGCCGCCTGCCTTTAAATC	AAGCAGTCCGCCTCGTAGTA

1029

Supplementary Table S2. Sequences and positions of the primers used for 3C analysis.

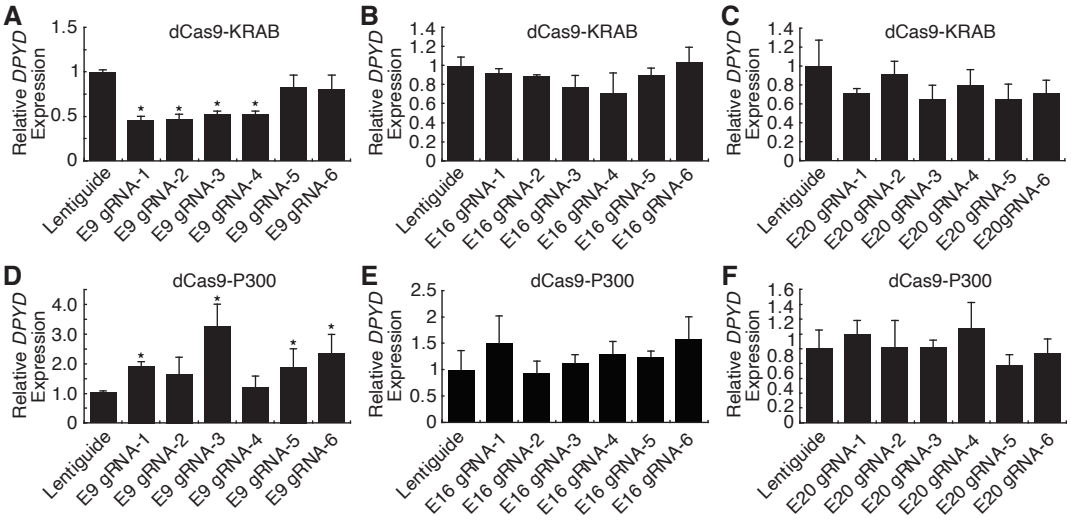
Primer position (TSS: +1)	Sequences (5'-3')
-22822	GGGAAGTTGAGAGAGCTAGGC
-19322	TGCTCTGTCAGCTGAGAAGACCTAGA
-16420	GTCACTACTGGGACTCTGAGAAA
-15616	AAAAGAAATTGCAACCTCTGGCA
-11097	GTTGCTTTTTGCAGCTGGGAT
-8087 (also used as anchor for E9 region)	AGTGCTTGAAGCTGATGAAGGG
-5884	CTGCAGAACAAGAACAGCACAT
1280 (also used as anchor for promoter)	TTAGGGTAGTCTATTCCTTTTTGGT
2826	TGCTTTGTGAGTGTACTGTTTGG
3076	CCTCCACCGGCAAGGATAAT

1030

1031

1032 **SUPPLEMENTARY FIGURES AND FIGURE LEGENDS**

1033



1034 Zhang et al., Figure S1

1035 **Figure S1. CRISPRi and CRISPRa screen to identify *DPYD cis* regulatory elements**

1036 **in HCT116 cells.** For CRISPRi, *DPYD* expression was measured in HCT116 cells

1037 expressing dCas9-KRAB following transfection with guide-RNAs specific to the E9 (A),

1038 E16 (B), and E20 (C) regions. For CRISPRa, *DPYD* expression was measured in

1039 HCT116 cells expressing dCas9-P300 following transfection with guide-RNAs specific

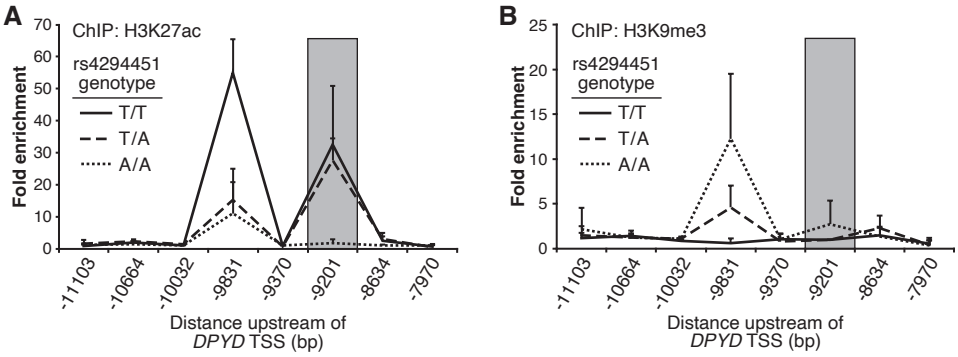
1040 to E9 (D), E16 (E), and E20 (F). Data represent the mean of three independent

1041 biological replicates ± SD. *, $p<0.05$. P-values were calculated using two-tailed

1042 Student's t-test comparing results to those from lentiguide controls.

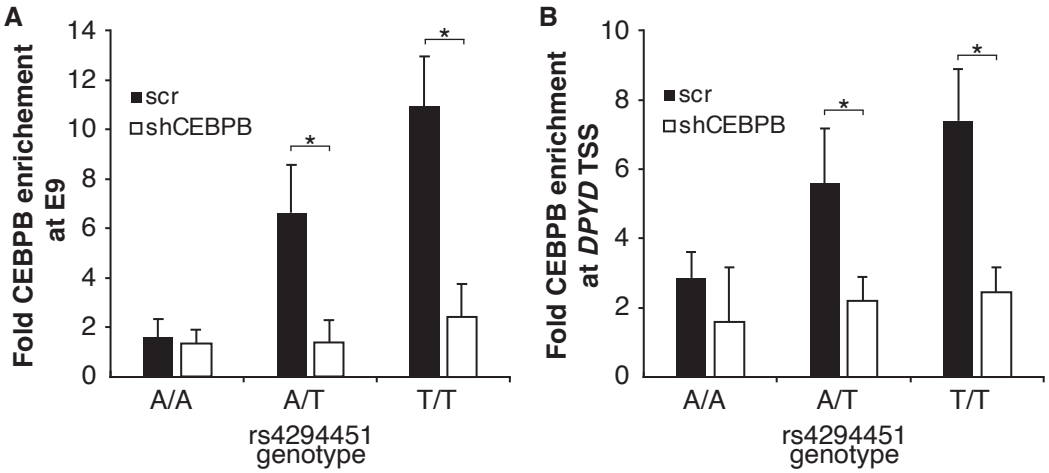
Zhang et al., Figure S2

Figure S2. Genomic context of rs4294451. Rs4294451 is located within a putative enhancer region showing evidence for regulatory activity in Ensembl Regulatory Build data and within transcription factor binding sites in ENCODE Factorbook data.



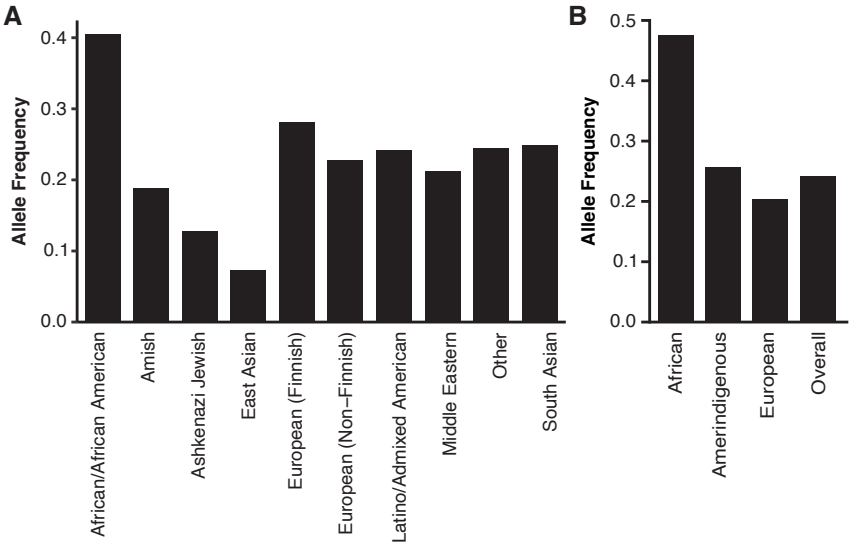
Zhang et al., Figure S3

Figure S3. Rs4294451 A allele is associated with epigenetic repression at the E9 region in human liver specimens. Chromatin enrichment of H3K27ac (A) and H3K9me3 (B) was measured using ChIP-qPCR of liver specimens obtained from human donors carrying different rs4294451 genotypes. Data represent three independent measurements from a single liver specimen with each indicated genotype ± SD.



Zhang et al., Figure S4

Figure S4. Disruption of CEBPB reduces enrichment at the E9 region and *DPYD* promoter in rs4294451 T/T and A/T cells, but not in T/T cells. CEBPB enrichment at the E9 region (A) and the *DPYD* promoter (B) was measured by ChIP-qPCR in CEBPB knockdown and scramble (scr) control knock-in HCT116 cells containing the indicated rs4294451 genotype. Data represent the mean of three independent replicates ± SD. *, $p<0.05$. P-values were calculated using two-tailed Student's t-test.



1062

1063 **Figure S5. Allele frequency for rs4294451-T allele in global populations.** (A) Allele

1064 frequencies in various populations for the rs4294451-T allele was retrieved from the

1065 gnomAD browser v3.1.2. (B) Local ancestry-informed frequency data was retrieved for

1066 the rs4294451-T allele within Latino-Admixed American samples of gnomAD v3.1