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 [1,7,8,9,*]

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

8 Holden Comprehensive Cancer Center, University of Iowa Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA.

9 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.

55 **INTRODUCTION**

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

62

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

71

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

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

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

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

143

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

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

157

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

165 containing the T allele showed significantly higher activity than those with the A allele
166 (p=0.046, **Figure 3B**), suggesting that the variant could impact *DPYD* expression.
167
168 **Impact of rs4294451 genotype on *DPYD* expression and regulation**
169 To directly characterize the impact of rs4294451 alleles on *DPYD* expression at the
170 endogenous locus, we used CRISPR-mediated genome editing to create matched
171 isogenic HCT116 knock-in cell models for each genotype. Consistent with the results
172 from the reporter assay (**Figure 3B**), cells homozygous for the A allele had significantly
173 lower *DPYD* expression compared to cells that were homozygous for the T allele
174 (p=0.0011, **Figure 3C**). Heterozygous rs4294451 A/T cells displayed intermediate
175 expression compared to cells with homozygous A/A and T/T genotypes (p=0.00013 and
176 p=0.028, respectively; **Figure 3C**). At the chromatin level, cells homozygous for the A
177 allele showed reductions in H3K27ac and accumulation of H3K9me at E9 compared to
178 cells that were homozygous or heterozygous for the T allele (**Figure 3D–3E**). This
179 finding is consistent with the A allele being associated with a less active epigenetic state
180 at E9. Cells carrying a heterozygous genotype for rs4294451 displayed an intermediate
181 epigenetic state, which is consistent with one epigenetically active and one inactive
182 copy of E9 (**Figure 3D–3E**). Similar epigenetic differences between rs4294451
183 genotypes were noted at E9 in liver specimens obtained from human donors (**Figure**
184 **S3A–S3B**).
185
186 **Allele-specific expression of *DPYD* transcripts**

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

209

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

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

223

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

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

231 homozygous for the rs4294451 A allele were significantly more sensitive to 5-FU than
232 cells homozygous for the T allele (IC_{50} concentrations were 9.1 and 95.0 μM 5-FU,
233 respectively; $p<0.0001$; **Figure 4E**). Heterozygous T/A cells showed an intermediate
234 IC_{50} value of 29.9 μM 5-FU (**Figure 4E**).
235
236 **Transcription factor binding within E9 is affected by rs4294451 status**
237 Based on ENCODE data, rs4294451 is localized to a region that previously has been
238 shown to be bound by multiple transcription factors (**Figure S2**). To determine if known
239 transcription factor binding sites could be affected by allele status at rs4294451, we
240 computed binding scores for JUND, MYBL2, HNF4A, CEBPB, FOXA1, and FOXA2 for
241 sequences comprising the E9 region containing the T and A alleles at rs4294451.
242 CEBPB, FOXA1, and FOXA2 showed differential predicted binding scores between the
243 rs4294451 A- and T-containing query sequences (data not shown). For CEBPB, the
244 differential binding site had a higher score than other sites in the queried region that did
245 not overlap with rs4294451, suggesting that the SNP could affect CEBPB binding in the
246 region. While predicted binding in the region varied for FOXA1 and FOXA2 depending
247 on rs4294451 genotype, other binding sites that did not overlap with the variant site
248 showed stronger predicted binding, indicating that the SNP was likely not affecting the
249 critical binding site in the region. Binding scores did not differ by genotype for JUND or
250 HNF4A, and no binding site above the threshold score was detected in the region for
251 MYBL2.
252

253 Based on the above, we sought to determine if CEBPB could regulate *DPYD*
254 expression through the E9 region and if regulation was affected by rs4294451
255 genotype. We first used reporter assays in conjunction with CEBPB. CEBPB expression
256 significantly increased reporter activity relative to GFP control for luciferase vectors
257 containing the *DPYD* promoter, indicating that the promoter contains CEBPB
258 recognition sites ($p=9.1x10^{-4}$; **Figure 5A**). In cells overexpressing CEBPB, plasmids
259 containing the promoter and E9 region resulted in higher luciferase activity compared to
260 those containing only the *DPYD* promoter, regardless of the presence of the A or T
261 allele (p=0.0036 and p=0.0017, respectively), suggesting that rs4294451 genotype does
262 not completely disrupt CEBPB regulation through E9 (**Figure 5A**). Comparing results for
263 the A and T alleles at rs4294451 in the presence of overexpressed CEBPB indicates
264 that the T allele is likely more responsive to CEBPB (**Figure 5A**; p=0.036).
265
266 To directly examine CEBPB binding to the E9 region and to determine if binding affinity
267 differs between the rs4294451 A and T alleles, we performed ChIP with CEBPB
268 antibodies in isogenic knock-in HCT116 cells for the rs4294451 A/A, A/T, and T/T
269 genotypes. Cells homozygous for the reference T allele showed significantly higher
270 CEBPB occupancy at both E9 and the *DPYD* promoter than A/T or A/A cells (**Figure**
271 **5B–5C**). These results suggest that rs4294451 genotype determined the binding
272 potential for CEBPB at the E9 enhancer region, which, in turn, affects CEBPB-driven
273 expression of *DPYD* from the promoter.
274

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

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

319 it understood how the regulatory processes can be altered to support the development
320 of 5-FU resistance.
321
322 In the current study, we identified a novel *cis* enhancer region for *DPYD* that is located
323 approximately 9 kb upstream from the gene's transcription start site. We additionally
324 provide evidence that the E9 region directly interactions with the *DPYD* promoter,
325 supporting E9 as a functional enhancer for *DPYD* expression (**Figure 4**). We
326 demonstrated that CEBPB is a critical transcription factor for *DPYD* that binds to this
327 enhancer region, termed E9, promoting enhancer–promoter interactions and increasing
328 *DPYD* expression. We also showed that the allele status of the germline variant
329 rs4294451, located within the E9 region, can affect CEBPB-driven *DPYD* expression
330 and sensitivity/resistance to 5-FU, making it a strong candidate biomarker for 5-FU
331 toxicity risk and potentially tumor resistance to 5-FU–based cancer therapy. These
332 findings are consistent with the recent identification of a haplotype block linked to the
333 rs4294451 T allele that was significantly associated with elevated *DPYD* expression in
334 human liver tissues [9]. In the present manuscript, we show that the rs4294451 T allele
335 is enriched for active chromatin marks in both human liver (**Figure S3**) and in cellular
336 knock-in models (**Figure 3D–E**). Furthermore, using an innovative digital droplet RT-
337 qPCR–based approach, we demonstrate that the rs4294451 T allele is associated with
338 elevated expression of the *cis DPYD* transcript in human liver (**Figure 3F–G**).
339

340 Allele frequency data retrieved from gnomAD suggest that a majority (65-70%,
341 estimated from allele frequencies) of individuals of African ancestry carry the
342 rs4294451-T allele (**Figure S5**), whereas only about 35% of individuals of European
343 ancestry are predicted to be carriers of the T allele. African American patients have
344 worse overall survival in colorectal cancer compared to white patients, owing to
345 biological and non-biological factors. While differential access to healthcare, treatment
346 bias, and socioeconomic factors have been shown to contribute to the poorer prognosis
347 [30], other unrecognized factors also contribute to this difference [31, 32]. Our data
348 support a hypothesis that higher systemic 5-FU catabolism to due elevated liver DPD
349 expression in carriers of the rs4294451-T allele results in lower exposure to active anti-
350 tumor metabolites of 5-FU. Additionally, we demonstrate that colorectal cancer cell lines
351 likely retain the enhancer functions associated with the rs4294451-T allele, suggesting
352 that tumor cells carrying this variant could more readily inactive 5-FU via increased DPD
353 expression. The higher likelihood of carrying the T-allele in individuals of African
354 ancestry would therefore place them at greater risk. Additional studies will be needed to
355 investigate the degree to which rs4294451-T contributes to survival and progression in
356 colorectal cancer and in other solid cancers frequently treated with 5-FU.
357
358 The transcription factor CEBPB is a member of the CCAAT Enhancer Binding Protein
359 family, a group of transcription factors that contain basic leucine zipper (bZIP) domains
360 and is highly expressed in liver [33, 34]. Multiple CEBPB isoforms have been detected,
361 with some acting as transcriptional activators and others as inhibitors. The data

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

384 observation that the T allele is associated with cellular resistance to 5-FU (**Figure 4E**).
385 The binding of CEBPB to specific motifs has recently been shown to be cell-type
386 specific and to rely on non-consensus binding motifs with binding strengths that can be
387 modulated by the sequence and structure of surrounding DNA regions [39, 40],
388 providing a possible explanation for the discrepancy between predicted and observed
389 results.
390
391 To our knowledge, this is the first report directly linking CEBPB to 5-FU metabolism and
392 the first mechanistic data demonstrating *cis* effects of a regulatory variant on *DPYD*
393 expression. The role of CEBPB in modulating 5-FU resistance and toxicity is not without
394 precedent. For example, mir-191 is abnormally expressed in several cancers and has
395 been associated with both 5-FU resistance and the regulation of CEBPB expression
396 [41]; however, the CEBPB–*DPYD* regulatory axis has not previously been recognized.
397 Furthermore, CEBPB signaling has been shown to be activated in colorectal cancer
398 cells following treatment with 5-FU [42], suggesting that CEBPB-mediated activation of
399 *DPYD* expression might represent a dynamic response to therapy.
400
401 While the contributions of regulatory variants to 5-FU metabolism have not been widely
402 studied to date, previous studies have explored DPD regulation. Our laboratories
403 previously characterized a *trans*-acting regulatory variant for DPD located within the
404 microRNA mir-27a that was subsequently shown to further increase 5-FU toxicity risk in
405 individuals that carried deleterious nonsynonymous *DPYD* variants [43, 44]. The variant

406 at rs4294451 is in LD with the *DPYD* variants c.85T>C and c.496G>A, which served as
407 coding region proxies for allele-specific expression in our present study (**Figure 3F–G**).
408 The haplotypes defined by these two coding-region variants together with a third variant
409 (c.1129-5923C>G/rs75017182) previously associated with varied levels of systemic
410 DPD activity [45]. A subsequent retrospective analysis indicated that these haplotype
411 differences translate to differential risk of severe 5-FU toxicity [46]; however, additional
412 studies are needed that are powered to evaluate more than the most common
413 haplotypes. Taken together, these results suggest that the differential regulatory effects
414 of rs4294451 alleles could further impact overall DPD activity and, by extension,
415 modulate the risk of severe 5-FU–related toxicity conferred by coding or splice-variants
416 in *DPYD*. Further work is also needed to identify interactions between the linked
417 variants that impact DPD enzyme activity and the regulatory variant to define systemic
418 DPD function and 5-FU toxicity risk.
419
420
421 **METHODS**
422
423 **Cells**
424 HEK293T, HCT116, and HepG2 cell lines were obtained from the American Tissue
425 Collection Center (ATCC, Manassas, VA). All cell lines were cultured in Dulbecco's
426 modified Eagle's medium (DMEM; Corning, Corning, NY) supplemented with 10% fetal
427 bovine serum (FBS; MilliporeSigma, Billerica, MA), 2 mM L-glutamate (Corning), and 1x

428 penicillin/streptomycin solution (MilliporeSigma). Cells were maintained at 37°C with 5%
429 CO_2. To support cell attachment and expansion, HepG2 cells were grown on plates
430 coated with 5% Matrigel (Corning).
431
432 **Liver tissues**
433 Human liver tissues used for ChIP analyses were processed through Dr. Mary Relling's
434 laboratory at St. Jude Children's Research Hospital, part of the Pharmacogenetics of
435 Anticancer Agents Research (PAAR) Group and were provided by the Liver Tissue Cell
436 Distribution System funded by NIH Contract #N01-DK-7-0004/HHSN267200700004C
437 and by the Cooperative Human Tissue Network. The acquisition and use of specimens
438 for this manuscript was conducted with the approval of the University of North Carolina
439 at Chapel Hill IRB (study number 10-2253), which has designated the use of these
440 livers for the current analyses as nonhuman subject research and the need for direct
441 consent for use in this study was waived. Human liver tissues used for allele-specific
442 expression were obtained from the University Clinic of Visceral Surgery and Medicine,
443 Inselspital, Bern, Switzerland. Specimens were from donated material from patients who
444 had undergone liver surgery at Inselspital and signed a written consent form for remnant
445 tissues to be used in research (KEK-BE:2016-02202). Patients with impaired liver
446 function due to cirrhosis or other conditions were excluded from analyses.
447
448 **Quantitative RT-PCR (RT-qPCR)**

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

470 Biosciences) for 1 hour at room temperature. Blots were imaged using the Odyssey
471 infrared imaging system (LI-COR Biosciences).
472
473 **Plasmids**
474 The gRNA expression empty vector lentiGuide-Puro was a gift from Feng Zhang
475 (Addgene_52963). pCMV-FLAG LAP2 was a gift from Joan Massague
476 (Addgene_15738). The oligonucleotides targeting enhancer regions (E9, E16, and E20)
477 were designed using GuideScan [47], hybridized, phosphorylated, and cloned into
478 lentiGuide-Puro via *Bsm*BI sites. The luciferase expression vector pGL4.10 was
479 purchased from Promega (Madison, WI). Lentivirus vectors expressing shRNAs
480 targeting CEBPB were obtained from the University of Minnesota Genomics Center
481 (Sh1: TRCN0000007440 and Sh2: TRCN0000007442).
482
483 **CRISPR inactivation (CRISPRi) and activation (CRISPRa)**
484 HepG2 cell lines that overexpress dCas9-KRAB and dCas9-P300 were generated by
485 lentiviral transduction using lenti-EF1a-dCas9-KRAB-Puro and pLV-dCas9-p300-P2A-
486 PuroR, respectively. To generate lentiviral particles, HEK293T cells were co-
487 transfected with lenti-EF1a-dCas9-KRAB-Puro plasmid (a gift from Kristen Brennand;
488 Addgene_99372) or pLV-dCas9-p300-P2A-PuroR plasmid (a gift from Charles
489 Gersbach; Addgene_83889), psPAX2 (a gift from Didier Trono; Addgene_12260), and
490 pMD2.G (a gift from Didier Trono; Addgene_12259) using TransIT-Lenti Transfection
491 Reagent (Mirus Bio, Madison, WI). A 3:1 ratio of transfection reagent to total plasmid

492 was used for all transfections. For all transfections, medium was changed 14 hours after
493 transfection, and viral supernatants were collected 34 hours later. Supernatants were
494 filtered using 0.45 μm PVDF filters (MilliporeSigma) to remove debris/cells and used
495 directly for transductions. For transductions, HCT116 cells or HepG2 cells were seeded
496 at a density of 4×10^5 cells per well in 6-well plates and incubated with 500 µL virus-
497 containing supernatant, 12.5 µg/mL polybrene (MilliporeSigma), and 1.5 mL fresh
498 DMEM culture medium. Medium was changed after 24 hours. Cells were treated with 1
499 µg/mL puromycin to initiate selection for transduced cells 48 hours after transduction.
500 Expression of dCas9-KRAB and dCas9-P300 was confirmed by western blotting. Guide
501 RNAs (gRNAs) for each target region were identified and designed using GuideScan
502 [47]. Oligonucleotides corresponding to each gRNA were obtained from IDT (Coralville,
503 IA), hybridized, phosphorylated using T4 polynucleotide kinase (New England Biolabs,
504 Ipswich, MA), and ligated into digested BsmBI-digested (New England Biolabs)
505 lentiGuide-Puro vector (a gift from Feng Zhang; Addgene_52963). Cell lines stably
506 expressing dCas9-KRAB or dCas9-P300 were transfected with plasmids encoding
507 gRNAs or lentiGuide empty vector using TransIT-X2 (Mirus Bio). RNA was extracted 2
508 days after transfection, and *DPYD* expression was measured by RT-qPCR.
509

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

511 ChIP assays were performed using the ChIP-IT Express Enzymatic Kit (Active Motif,
512 Carlsbad, CA) following manufacturer's directions. One million cells were harvested,
513 washed, and crossed linked using 1% formaldehyde (Thermo Fisher Scientific) in

514 serum-free medium for 10 minutes followed by quenching with 125 mM glycine for 5
515 minutes at room temperature. The chromatin was digested using enzymatic shearing
516 cocktail provided by the kit to an average size of 200–1000 bp. Two percent of the
517 sheared chromatin was retained as input control. Approximately 25 µg sheared
518 chromatin was incubated with 2 µg H3K27ac antibody (ab4729; Abcam), 2 µg
519 H3K9me3 antibody (ab8898; Abcam), 2 µg CEBPB antibody (PA5-27244; Thermo
520 Fisher Scientific), or 2 µg control normal Rabbit IgG antibody (antibody 2729; Cell
521 Signaling Technology, Danvers, MA), in the presence of protein G magnetic beads,
522 ChIP buffer, and protease inhibitor cocktail (all Active Motif) for 4 hours at 4°C. Magnetic
523 beads were washed, chromatin was eluted, cross-linking was reversed, and proteinase
524 K treatment performed using reagents provided in the kit following manufacturer's
525 directions. DNA was purified using the QIAquick PCR Purification Kit (Qiagen,
526 Germantown, MD). Purified DNA was used for subsequent qPCR reactions using SYBR
527 Green I Master Mix on a LightCycler 480 System (Roche). Enrichment was calculated
528 using the following formula: (a) % ChIP = $2^{(\text{Input Ct - ChIP Ct})}$ * (dilution factor) (100); (b) %
529 IgG = $2^{(\text{Input Ct - IgG Ct})}$ * (dilution factor) (100); (c) Fold Enrichment = % ChIP ÷ % IgG.
530 Primers used for ChIP-qPCR are listed in **Supplementary Table S1**.
531

532 **Luciferase reporter assays**

533 The *DPYD* promoter region, consisting of the 1154 bp of genomic DNA directly
534 upstream of the *DPYD* TSS, was amplified by PCR, digested with EcoRV and HindIII
535 (New England Biolabs), and cloned into compatible sites on the pGL4.10 vector

536 (Promega). The 1392-bp region comprising the E9 region was PCR amplified from
537 genomic DNA and cloned upstream of the *DPYD* promoter using KpnI and SacI sites
538 (New England Biolabs). A vector containing the A allele of rs4294451 within E9 was
539 confirmed by Sanger sequencing. The vector containing the rs4294451 T allele was
540 generated by site-directed mutagenesis and confirmed by sequencing. All primers used
541 in vector construction are listed in **Supplementary Table S1**. For reporter assays, 10^5
542 HEK293T cells were seeded into 24-well plates and co-transfected with pGL4.10-based
543 plasmids and pRL-SV40 Renilla luciferase plasmid (Promega). After 48 hours,
544 luciferase activity was measured using the Dual-Glo Luciferase Assay (Promega)
545 following manufacturer's recommendations on a Synergy HTX Multimode Plate Reader
546 (Agilent Technologies, Santa Clara, CA).
547
548 **Knock-in cell lines for rs4294451 genotypes**
549 Knock-in cell lines were generated using CRISPR/Cas9 gene editing. Homology-
550 directed repair (HDR) donor templates and target-specific Alt-R crRNA were designed
551 using the Alt-R HDR Design Tool (IDT). Equimolar amounts of crRNA (IDT) and
552 common Alt-R tracrRNA (IDT) were annealed to form the gRNA duplex. RNP
553 complexes were formed by combining gRNA with Alt-R S.p. Cas9 Nuclease V3 (IDT) to
554 a final Cas9:gRNA ratio of 4:4.8. RNA complex and Alt-R HDR Donor Oligos were
555 transfected into HCT116 cells by electroporation (Lonza Nucleofector 96-well Shuttle
556 System; Lonza, Bend, OR) using parameters provided by the manufacturer. Seventy-
557 two hours after transfection, serial dilutions were performed to obtain single-cell clones.

558 Clones were expanded, genomic DNA was isolated, and rs42944551 genotype was
559 determined by rhAmp Genotyping (assay ID: Hs.GT.rs4294451.A.1; IDT) using rhAmp
560 Genotyping Master Mix and universal probe Reporter Mix (both IDT).
561
562 **Allele-specific gene expression**
563 Allele-specific expression of *DPYD* was measured using reverse transcriptase droplet
564 digital PCR (RT-ddPCR) by targeting variants in the coding region of *DPYD* (c.85T>C
565 and c.496A>G). Linkage disequilibrium (LD) between variants was calculated using
566 LDpair implemented within LDlink [48]. DNA was extracted from donor liver tissues
567 using the QIAamp DNA Mini Kit (Qiagen). For RNA extraction, tissues were lysed in
568 QIAzol (Qiagen), and RNA was extracted using the miRneasy Kit (Qiagen) with on-
569 column DNA digestion using RNase-free DNase (Qiagen). RNA quality was assessed
570 using an Agilent 2100 Bioanalyzer running 2100 Expert Software v.B.02.10 using
571 Agilent RNA 6000 Nano kits.
572
573 DNA samples were genotyped for c.85, c.496, and rs4294451 loci using TaqMan SNP
574 Genotyping assays (Applied Biosystems, Waltham, MA). Tissues that were
575 heterozygous for at least one of the coding region SNPs (i.e., c.85T>C and/or
576 c.496A>G) were suitable for allele-specific expression analysis because expression
577 from both alleles could be discriminated using the coding-region markers in mRNA.
578 Allele-specific expression was measured using RT-ddPCR with the One-Step RT-
579 ddPCR Advanced Kit for Probes (Qiagen) on a QX200 ddPCR Droplet Reader (BioRad,

580 Hercules, CA). Fractional abundance was calculated using QuantaSoft software
581 (BioRad). Poisson distributions were determined using Quantasoft and were used to
582 define 95% confidence intervals. To address possible biases associated with differing
583 probe efficacies caused by differential probe binding affinities or specificities, fractional
584 abundances are also reported relative to those measured in DNA.
585

586 **Quantitative analysis of chromatin conformation capture (3C-qPCR)**

587 Quantitative analysis of chromatin conformation capture (3C-qPCR) was performed as
588 described by Hagère *et al*. [49] with minor modifications. Cells were trypsinized using
589 0.25% w/v trypsin-EDTA (Thermo Fisher Scientific) and resuspended in 1% FBS
590 (MilliporeSigma) in DMEM (Corning) for counting by flow cytometry using a NovoCyte
591 3000 RYB system (Agilent Technologies). Cells (10^7) were pelleted by centrifugation at
592 300 × *g* at 22°C for 5 minutes. Supernatant was discarded, and cell pellets were
593 resuspended in 500 μL of 10% FBS in DMEM. Single-cell suspensions were obtained
594 by filtration through a 40-μm cell strainer (Corning). Crosslinking was performed by
595 adding 9.5 mL of 1% formaldehyde in 10% FBS in PBS per 10^7 cells. Reactions were
596 incubated at 22°C while tumbling for 10 minutes. Reactions were transferred to ice and
597 crosslinking was quenched by the addition of ice-cold glycine (MilliporeSigma) to
598 achieve a final concentration of 0.125M. Samples were centrifuged at 300 × *g* at 4°C for
599 5 minutes. Supernatant was removed and discarded. Crosslinked cell pellets were lysed
600 in 5 mL cold lysis buffer (10 mM Tris-HCl, pH7.5; 10 mM NaCl; 5 mM MgCl; 0.1 mM
601 EGTA; 1x Roche Complete protease inhibitor cocktail). Lysis was allowed to proceed for

602 10 minutes on ice with intermittent gentle pipetting to obtain homogeneous suspensions
603 of nuclei. Nuclei were pelleted at 500 × *g* at 4°C for 5 minutes and resuspended in 0.5
604 mL of 1.2x NEBuffer r2.1 (New England Biolabs) containing 3% sodium dodecyl sulfate
605 (SDS, MilliporeSigma). Samples were incubated at 37°C for 1 hour with shaking.
606 Following incubation, Triton X-100 (MilliporeSigma) was added to a final concentration
607 of 2%. Reactions were incubated at 37°C for 1 hour with shaking. To digest DNA, 600 U
608 of HindIII (New England Biolabs) was added, and reactions were incubated for 16 hours
609 at 37°C with shaking. To inactivate digestion, SDS was added to a final concentration of
610 1.6%, and reactions were incubated at 65°C for 25 minutes with shaking. Excess SDS
611 was sequestered by the addition of Triton X-100 to a final concentration of 1%. Samples
612 were divided into two aliquots, one for ligation and the other for non-ligation control.
613 Ligation was performed on a 7-fold dilution of HindIII-digested chromatin using 100 units
614 of Quick T4 DNA ligase (New England Biolabs) at 16°C for 16 hours, followed by 1 hour
615 at 22°C. Proteinase K (300 *μ*g; New England Biolabs) was added to ligation mixtures
616 and non-ligated controls, and samples were incubated at 65°C for 16 hours to reverse
617 crosslinking. DNA was subsequently purified by adding 300 *μ*g RNase A (Thermo-
618 Fisher) followed by a 30-minute incubation at 37°C and subsequent phenol-chloroform
619 extraction as described [49]. DNA pellets were washed with 70% ethanol and dissolved
620 in 10 mM Tris pH 7.5.
621
622 For qPCR, directional primers were designed within each fragment as depicted in
623 **Figure 4A**. Two anchors, one localized to the fragment containing the *DPYD* promoter

624 and the other to the E9 region, were selected and paired with primers designed to
625 "walk" across the length of the surrounding region. Reactions were carried out in 10 μL
626 reaction volumes, consisting of 5 pmol each primer and 1 μL of a 1:50 dilution of each
627 3C sample. Amplification was performed using LightCycler 480 SYBR Green I Master
628 Mix (Roche) on a LightCycler 480 thermocycler (Roche). PCR conditions were 95°C for
629 10 minutes, and 40 cycles of 95°C for 15 seconds, 65°C for 1 minute, and 72°C for 15
630 seconds. Enrichment was calculated as $2^{-(\text{cp (ligated DNA) - cp (non-ligated DNA)})}$. Enrichment with
631 the fragment containing the nearest primer was used as the control for interaction
632 frequency for further normalization between replicate experiments. Specifically, when
633 primer 5 was used as the anchor, data were normalized to the average of values
634 obtained from using primers 5 and 4. When primer 7 was used as the anchor, data were
635 normalized to the average of values from primers 7 and 8. All experiments were
636 performed in triplicate. Primer sequences and positions relative to the *DPYD*
637 transcription start site (TSS) for qPCR are listed in **Supplementary Table S2**.
638

639 **Cellular sensitivity to 5-FU**

640 Cell viability was monitored using the xCELLigence MP Real-Time Cell Analysis (RTCA)
641 system (Agilent) as previously described [14]. Background impedance values for RTCA
642 E-View plates (Agilent) were obtained using complete DMEM prior to plating cells. Cells
643 were seeded at a density of 5,000 cells per well and incubated for 20 hours, at which
644 time medium was removed and replaced with medium containing serial dilutions of 5-FU
645 ranging from 1.25 µM to 100 µM. Impedance values (expressed as cell index, CI, units)

646 were recorded every 15 minutes over the course of the experiment to monitor
647 proliferation. Results represent the average of three independent cultures. To account
648 for minute differences in plating and potential cell loss during drug addition, relative CI
649 was calculated as the CI measured 48 hours after 5-FU divided by the CI recorded
650 immediately after treatment. Four parameter logistic non-linear regression analysis was
651 used to determine IC_{50} concentrations (GraphPad Prism version 9, GraphPad Software,
652 San Diego, CA).
653
654 **CEBPB knockdown cells**
655 Lentiviral particles for CEBPB knockdown were generated by transfecting shRNA
656 plasmids (TRCN0000007440 and TRCN0000007442) or scramble shRNA control (a gift
657 from David Sabatini; Addgene_1864 [50]), psPAX2 (a gift from Didier Trono;
658 Addgene_12260), and pMD2.G (a gift from Didier Trono; Addgene_12259) into
659 HEK293T cells using TransIT-Lenti (Mirus Bio, Madison, WI). A 3:1 ratio of transfection
660 reagent to total plasmid was used for all transfections. For all transfections, medium
661 was changed 14 hours after transfection, and viral supernatants were collected 34
662 hours later. Supernatants were filtered using 0.45 μm PVDF filters (Millipore) to remove
663 debris/cells and used directly for transductions. For transductions, target cells were
664 seeded at a density of 4×10^5 cells per well in 6-well plates and incubated with 500 µL
665 virus-containing supernatant, 12.5 µg/mL polybrene (MilliporeSigma), and 1.5 mL fresh
666 DMEM culture medium.
667

668 **Analysis of transcription factor binding motifs**

669 Potential transcription factor binding sites within the E9 region were determined using
670 TFBSTools [51]. DNA sequence corresponding to the 101-nucleotide region centered
671 on rs4294451 was retrieved from genome build GRCh38.p13 (NC_000001.11). A
672 second DNA string was created to mimic the sequence corresponding to the rs4294451
673 A allele. Position frequency matrices for each transcription factor were retrieved from
674 JASPAR CORE 2022 [52] and converted to log-scale position weight matrices using the
675 toPWM method implemented in TFBSTools. JASPAR includes binding site information
676 from multiple sources. ENCODE data [53] were available for CEBPB (matrix ID
677 MA04661) and JUND (MA0491.1). In absence of ENCODE data, alternatives including
678 PAZAR [54] (HNF4A, MA0114.2; FOXA1, MA0148.3), REMAP [55] (FOXA2,
679 MA0047.3), and data from an individual publication [56] (MYBL2, MA0777.1) were used.
680 EP300 is a cofactor that does not recognize a specific DNA motif on its own; instead, it
681 interacts with various DNA-binding factors to modify chromatin and facilitate activation
682 of target genes. As such, P300 does not have a DNA-binding motif and was not
683 included in analysis. Nucleotide sequences were scanned using the patterns presented
684 in the position weight matrices to identify putative transcription factor binding sites.
685 Forward and reverse strands were searched, and the 70th percentile between the
686 minimum and maximum possible value for a matrix was used as the minimum threshold
687 score. Empirical p-values for each score were calculated by an exact method using the
688 TFMPvalue R package. R version 4.2.2 was used for analyses
689

690 **Statistical analyses**

691 Significance was defined as $p<0.05$ unless otherwise noted in the text. Pairwise
692 comparisons were performed using unpaired two-tailed Student's t tests calculated
693 using GraphPad Prism version 9. One-way ANOVA statistics were calculated using
694 GraphPad Prism. Summary statistics pertaining to allele-specific expression were
695 calculated using Quantasoft software as described in the Allele-specific gene
696 expression section. Transcription factor binding predictions and associated analyses
697 were performed in R version 4.2.2 as described above.

698

699 **Availability of data and materials**

700 All data generated or analyzed during this study are included in this published article
701 and its supplementary information files.

702

703 **Acknowledgements**

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

710 **REFERENCES**

711

712 1. Ramos, A., S. Sadeghi, and H. Tabatabaeian, *Battling Chemoresistance in*
713 *Cancer: Root Causes and Strategies to Uproot Them.* Int J Mol Sci, 2021.
714 **22**(17).

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

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

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

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

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

737 7. Boix, C.A., B.T. James, Y.P. Park, W. Meuleman, and M. Kellis, *Regulatory*
738 *genomic circuitry of human disease loci by integrative epigenomics.* Nature,
739 2021. **590**(7845): p. 300-307.

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

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

746 *Drug Response, Metabolic, and Sex-Biased Phenotypes.* Clin Pharmacol Ther,
747 2019.

748 10. Sommadossi, J.P., D.A. Gewirtz, R.B. Diasio, C. Aubert, J.P. Cano, and I.D.
749 Goldman, *Rapid catabolism of 5-fluorouracil in freshly isolated rat hepatocytes as*
750 *analyzed by high performance liquid chromatography.* J Biol Chem, 1982.
751 **257**(14): p. 8171-6.

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

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

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

764 14. Shrestha, S., C. Zhang, C.R. Jerde, Q. Nie, H. Li, S.M. Offer, and R.B. Diasio,
765 *Gene-Specific Variant Classifier (DPYD-Varifier) to Identify Deleterious Alleles of*
766 *Dihydropyrimidine Dehydrogenase.* Clin Pharmacol Ther, 2018. **104**(4): p. 709-
767 718.

768 15. Offer, S.M., C.C. Fossum, N.J. Wegner, A.J. Stuflesser, G.L. Butterfield, and
769 R.B. Diasio, *Comparative functional analysis of DPYD variants of potential*
770 *clinical relevance to dihydropyrimidine dehydrogenase activity.* Cancer Res,
771 2014. **74**(9): p. 2545-54.

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

775 17. Offer, S.M., N.J. Wegner, C. Fossum, K. Wang, and R.B. Diasio, *Phenotypic*
776 *profiling of DPYD variations relevant to 5-fluorouracil sensitivity using real-time*
777 *cellular analysis and in vitro measurement of enzyme activity.* Cancer Res, 2013.
778 **73**(6): p. 1958-68.

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

782 *mRNA Splicing and Enzyme Activity.* Clin Pharmacol Ther, 2017. **102**(4): p. 662-
783 670.

784 19. Meulendijks, D., L.M. Henricks, G.S. Sonke, M.J. Deenen, T.K. Froehlich, U.
785 Amstutz, C.R. Largiader, B.A. Jennings, A.M. Marinaki, J.D. Sanderson, Z. Kleibl,
786 P. Kleiblova, M. Schwab, U.M. Zanger, C. Palles, I. Tomlinson, E. Gross, A.B.
787 van Kuilenburg, C.J. Punt, M. Koopman, J.H. Beijnen, A. Cats, and J.H.
788 Schellens, *Clinical relevance of DPYD variants c.1679T>G, c.1236G>A/HapB3,*
789 *and c.1601G>A as predictors of severe fluoropyrimidine-associated toxicity: a*
790 *systematic review and meta-analysis of individual patient data.* Lancet Oncol,
791 2015. **16**(16): p. 1639-50.

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

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

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

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

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

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

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

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

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

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

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

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

848 31. Govindarajan, R., R.V. Shah, L.G. Erkman, and L.F. Hutchins, *Racial differences*
849 *in the outcome of patients with colorectal carcinoma.* Cancer, 2003. **97**(2): p.
850 493-8.

851 32. Alexander, D.D., J. Waterbor, T. Hughes, E. Funkhouser, W. Grizzle, and U.
852 Manne, *African-American and Caucasian disparities in colorectal cancer mortality*
853 *and survival by data source: an epidemiologic review.* Cancer Biomark, 2007.
854 **3**(6): p. 301-13.

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

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

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

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

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

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

877 39. Cohen, D.M., H.W. Lim, K.J. Won, and D.J. Steger, *Shared nucleotide flanks*
878 *confer transcriptional competency to bZip core motifs.* Nucleic Acids Res, 2018.
879 **46**(16): p. 8371-8384.

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

883 41. Zhang, X.F., K.K. Li, L. Gao, S.Z. Li, K. Chen, J.B. Zhang, D. Wang, R.F. Tu, J.X.
884 Zhang, K.X. Tao, G. Wang, and X.D. Zhang, *miR-191 promotes tumorigenesis of*
885 *human colorectal cancer through targeting C/EBPbeta.* Oncotarget, 2015. **6**(6): p.
886 4144-58.

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

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

895 44. Amstutz, U., S.M. Offer, J. Sistonen, M. Joerger, R.B. Diasio, and C.R. Largiader,
896 *Polymorphisms in MIR27A Associated with Early-Onset Toxicity in*
897 *Fluoropyrimidine-Based Chemotherapy.* Clin Cancer Res, 2015. **21**(9): p. 2038-
898 44.

899 45. Hamzic, S., D. Scharer, S.M. Offer, D. Meulendijks, C. Nakas, R.B. Diasio, S.
900 Fontana, M. Wehrli, S. Schurch, U. Amstutz, and C.R. Largiader, *Haplotype*
901 *structure defines effects of common DPYD variants c.85T > C (rs1801265) and*
902 *c.496A > G (rs2297595) on dihydropyrimidine dehydrogenase activity:*
903 *Implication for 5-fluorouracil toxicity.* Br J Clin Pharmacol, 2021. **87**(8): p. 3234-
904 3243.

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

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

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

914 49. Hagege, H., P. Klous, C. Braem, E. Splinter, J. Dekker, G. Cathala, W. de Laat,
915 and T. Forne, *Quantitative analysis of chromosome conformation capture assays*
916 *(3C-qPCR).* Nat Protoc, 2007. **2**(7): p. 1722-33.

917 50. Sarbassov, D.D., D.A. Guertin, S.M. Ali, and D.M. Sabatini, *Phosphorylation and*
918 *regulation of Akt/PKB by the rictor-mTOR complex.* Science, 2005. **307**(5712): p.
919 1098-101.

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

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

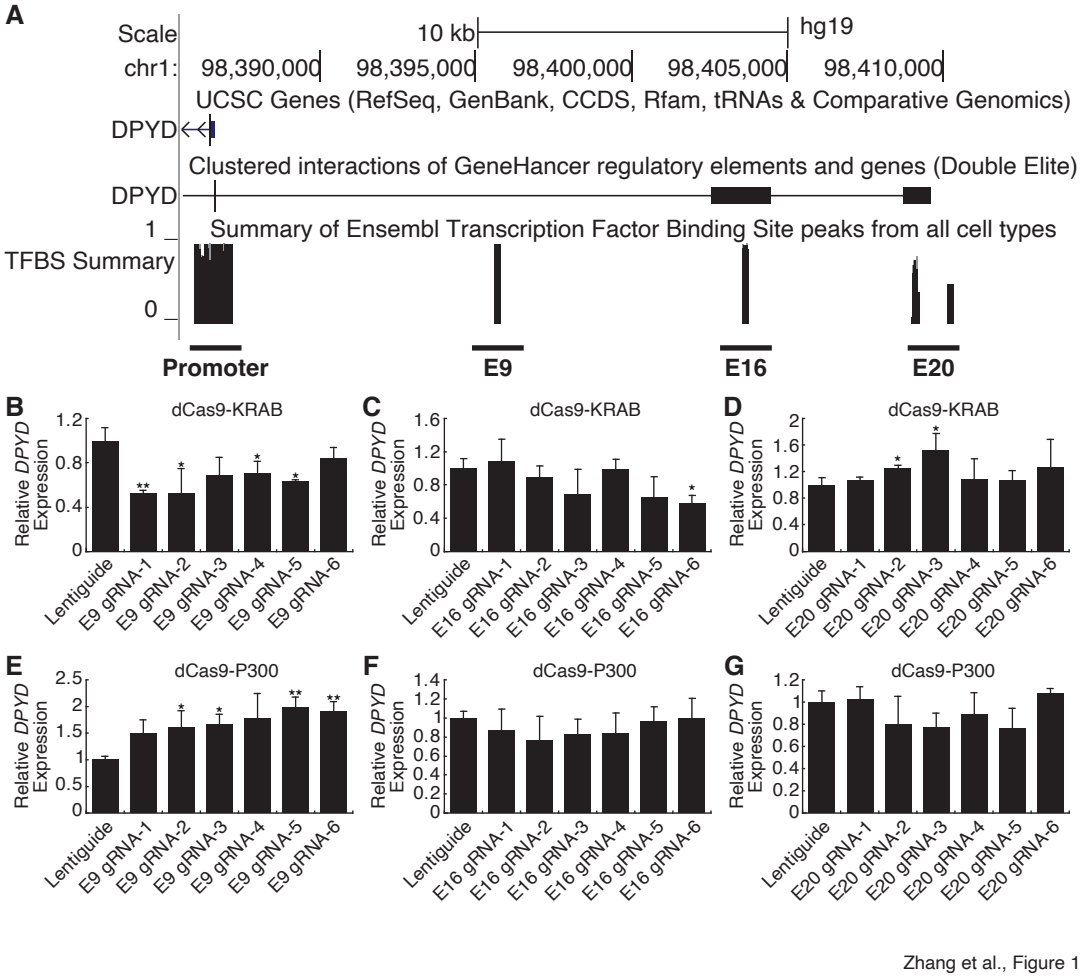
929 53. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human*
930 *genome.* Nature, 2012. **489**(7414): p. 57-74.

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

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

939 56. Jolma, A., J. Yan, T. Whitington, J. Toivonen, K.R. Nitta, P. Rastas, E.
940 Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J.M. Vaquerizas, R.
941 Vincentelli, N.M. Luscombe, T.R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja,
942 and J. Taipale, *DNA-binding specificities of human transcription factors.* Cell,
943 2013. **152**(1-2): p. 327-39.
944

945 **FIGURES AND FIGURE LEGENDS**



946

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

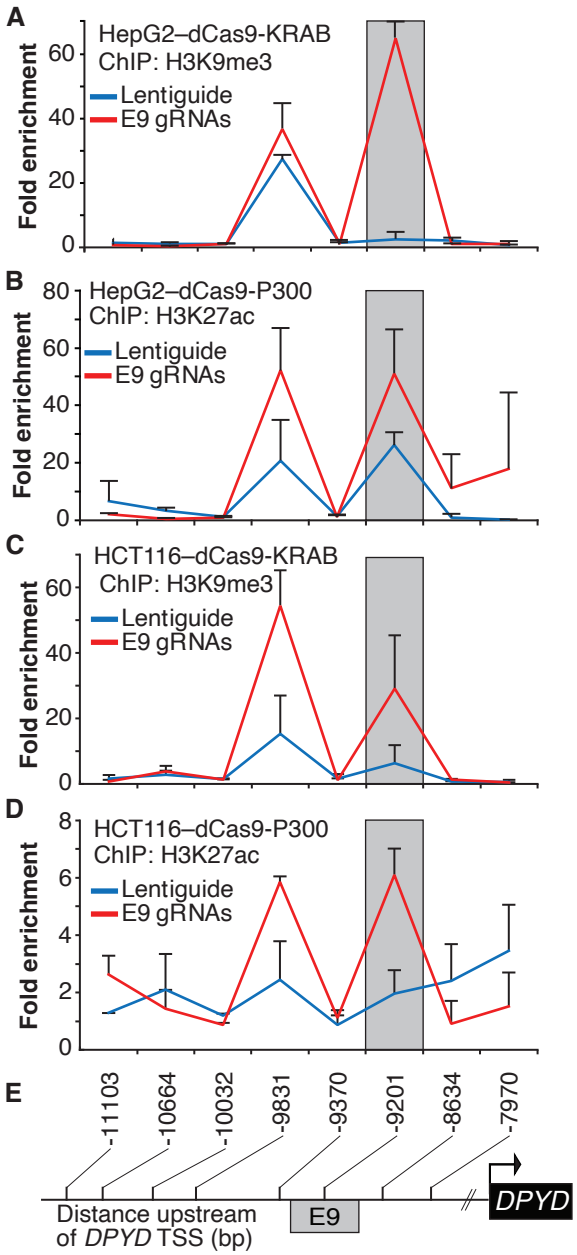
955 RNAs specific to E9 (E), E16 (F), and E20 (G). Data represent the mean of three

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

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

958

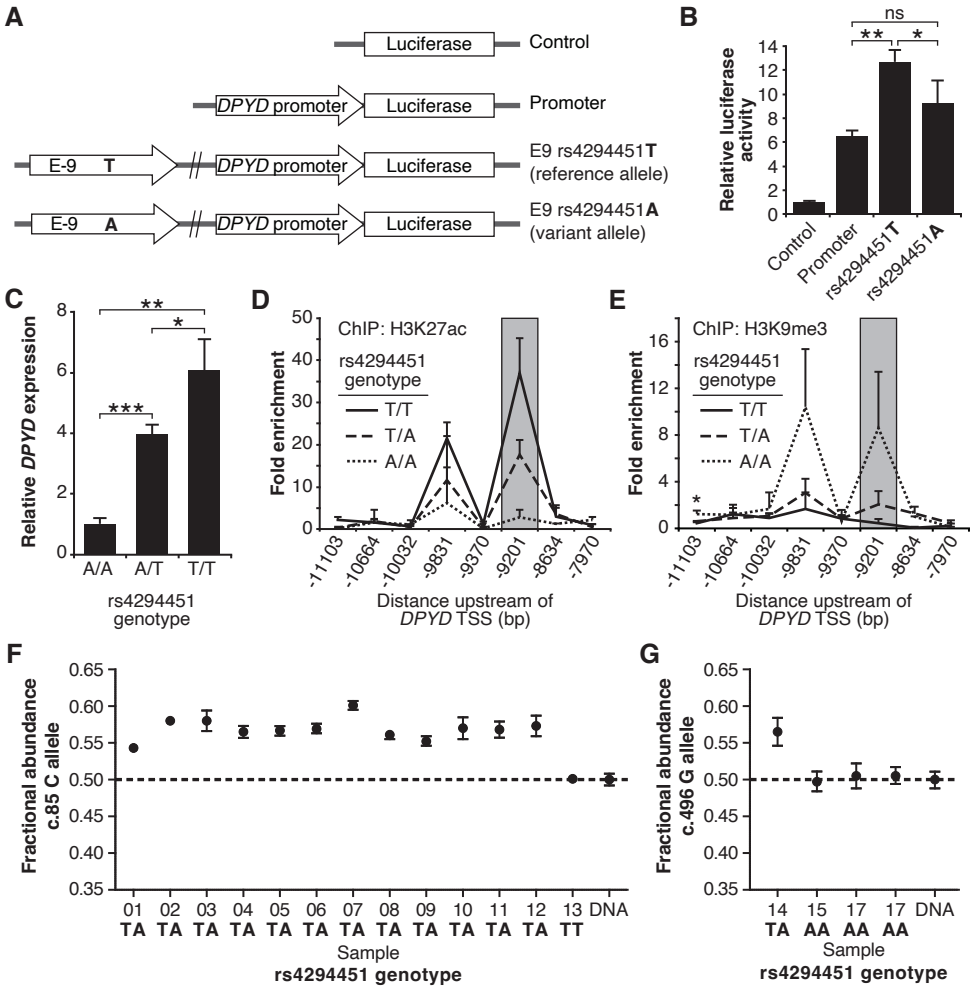
959



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

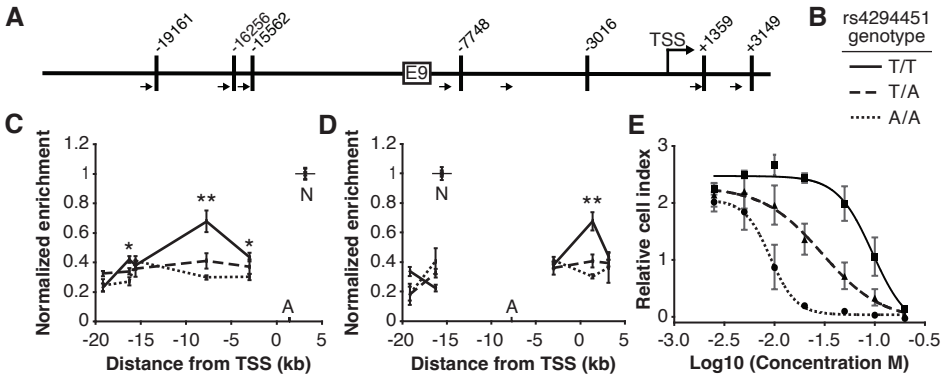
966 H3K9me3 (A, C) or H3K27ac (B, D). Quantitative PCR using primers centered at the
967 indicated regions (E) was used to measure the relative abundance of H3K9me3 and
968 H3K27ac. Data are presented relative to input DNA control and are further normalized
969 to IgG control. Error bars represent the SD of three independent biological replicate
970 experiments.
971



Zhang et al., Figure 3

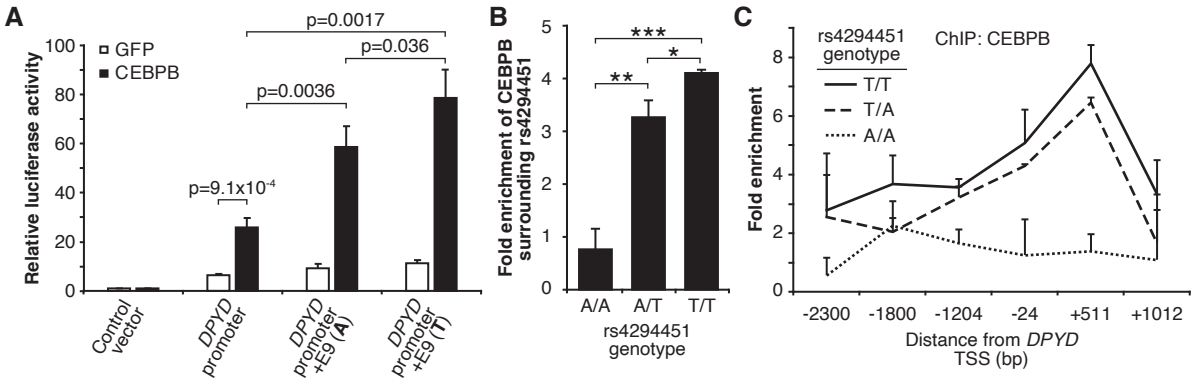
Figure 3. The rs4294451 T allele is associated with elevated *DPYD* expression. (A) Luciferase reporter constructs were generated by cloning the E9 region containing the reference rs4294451 T allele or the variant rs4294451 A allele into reporter vectors containing the *DPYD* promoter. (B) Luciferase 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

981 H3K27ac (D) and H3K9me3 (E) was measured using ChIP-qPCR in HCT116 cells
982 engineered to contain the indicated genotypes at rs4294451. (F) *DPYD* allele-specific
983 expression was measured in human liver tissues using the C and T alleles at position
984 c.85. (G) Allele-specific expression was measured using the c.496-A and G alleles. All
985 panels: *, $p<0.05$; **, $p<0.005$; ***, $p<0.0005$. P-values were calculated as pairwise
986 comparisons between the indicated groups using two-tailed Student's t-tests.
987



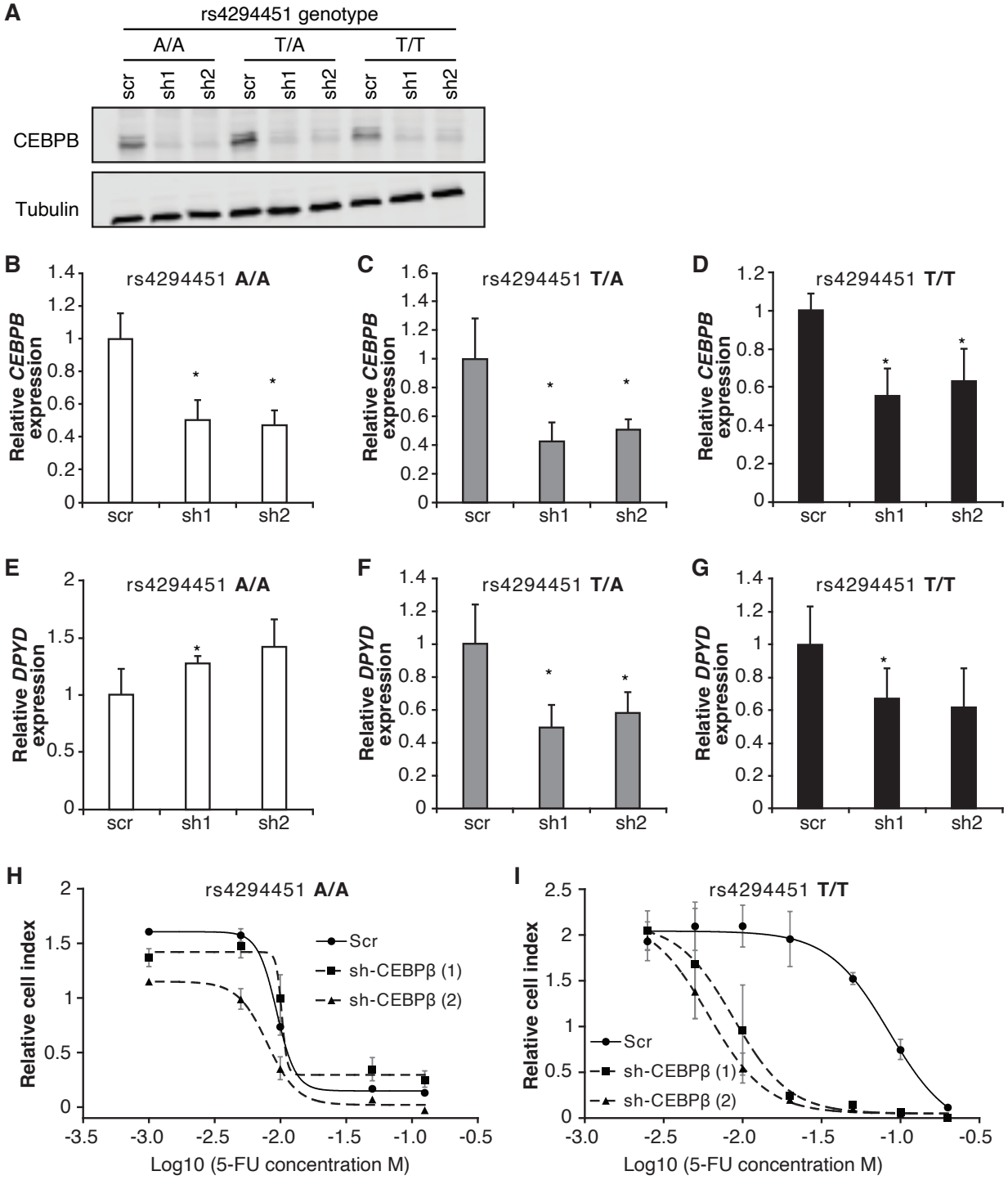
Zhang et al., Figure 4

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.



Zhang et al., Figure 6

1019 **Figure 6. CEBPB-mediated upregulation of *DPYD* is dependent on rs4294451 T**
1020 **allele.** (A) Immunoblot showing knockdown of CEBPB expression in HCT116 cells
1021 carrying different rs4294451 genotypes transduced with lentiviral particles encoding two
1022 independent shRNAs against CEBPB (sh1 and sh2) or a scrambled control shRNA
1023 (scr). *CEBPB* expression was measured by RT-qPCR in HCT116 A/A (B), T/A (C) and
1024 T/T cells (D) transduced with the indicated shRNA lentiviral particles. *DPYD* expression
1025 was measured in shCEBPB and scramble control HCT116 A/A (E), T/A (F) and T/T (G)
1026 cells. The effect of CEBPB knockdown on cell viability in HCT116 A/A cells (H) and
1027 HCT116 T/T cells (I) was measured by RTCA. Data shown are from 48 hours of 5-FU
1028 treatment at the indicated concentrations. *, $p<0.05$, calculated as a pairwise two-sided
1029 Student's t-test comparing the indicated data to that of the associated scr control. Error
1030 bars represent the SD from three independent replicates.
1031

1032 **SUPPLEMENTARY TABLES**

1033

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	AGTCAAGCTTGCTCGATGTCCGCC GAG
E9 region	AGTCGGTACCGAATAAAACCAAAATAAAAT CCATTTGGACGTTT	AGTCGAGCTCTTTGTGCAAAGGAC CTTGGTATTTCC
Rs4294451 A>T	AAAGAAAAATAAATAAAAAAAGGAAAAATC TATAAGC	TTCTGGGGGTTGGTGTTG
Cas9	AACAGCCGCGAGAGAATGAA	CACGGGGTGTTCTTTCAGGA
DPYD	GTAAGGACTCGGCGGACATC	GCCGAAGTGGAACACAGAGT
L32	CCTTGTGAAGCCCAAGATCG	TGCCGGATGAACTTCTTGGT
CEBPB	CGCCGCCTGCCTTTAAATC	AAGCAGTCCGCCTCGTAGTA

1034

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

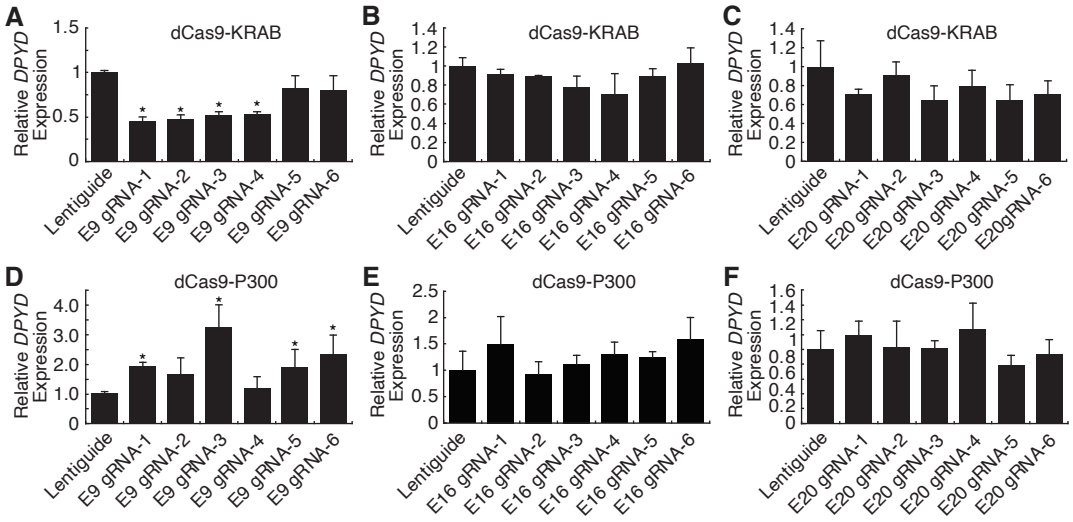
Primer position (TSS: +1)	Sequences (5'-3')
-22822	GGGAAGTTGAGAGAGCTAGGC
-19322	TGCTCTGTCAGCTGAGAAGACCTAGA
-16420	GTCACTACTGGGACTCTGAGAAA
-15616	AAAAGAAATTGCAACCTCTGGCA
-11097	GTTGCTTTTTGCAGCTGGGAT
-8087 (also used as anchor for E9 region)	AGTGCTTGAAGCTGATGAAGGG
-5884	CTGCAGAACAAGAACAGCACAT
1280 (also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

1035

1036

1037 **SUPPLEMENTARY FIGURES AND FIGURE LEGENDS**

1038



1039 Zhang et al., Figure S1

1040 **Figure S1. CRISPRi and CRISPRa screen to identify *DPYD cis* regulatory elements**

1041 **in HCT116 cells.** For CRISPRi, *DPYD* expression was measured in HCT116 cells

1042 expressing dCas9-KRAB following transfection with guide-RNAs specific to the E9 (A),

1043 E16 (B), and E20 (C) regions. For CRISPRa, *DPYD* expression was measured in

1044 HCT116 cells expressing dCas9-P300 following transfection with guide-RNAs specific

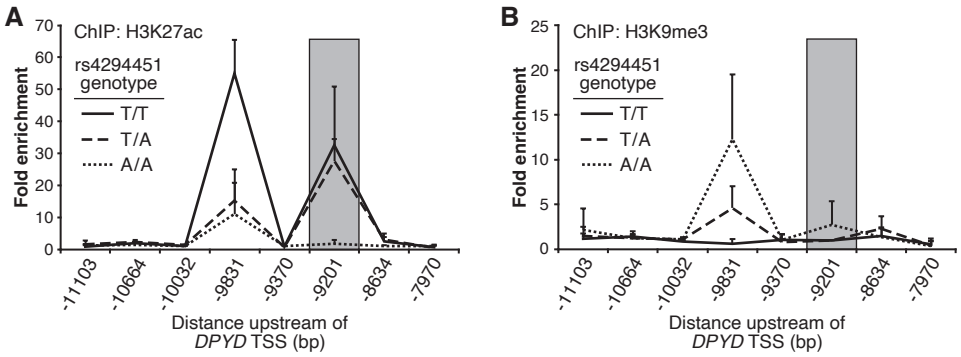
1045 to E9 (D), E16 (E), and E20 (F). Data represent the mean of three independent

1046 biological replicates ± SD. *, $p<0.05$. P-values were calculated using two-tailed

1047 Student's t-test comparing results to those from lentiguide controls.

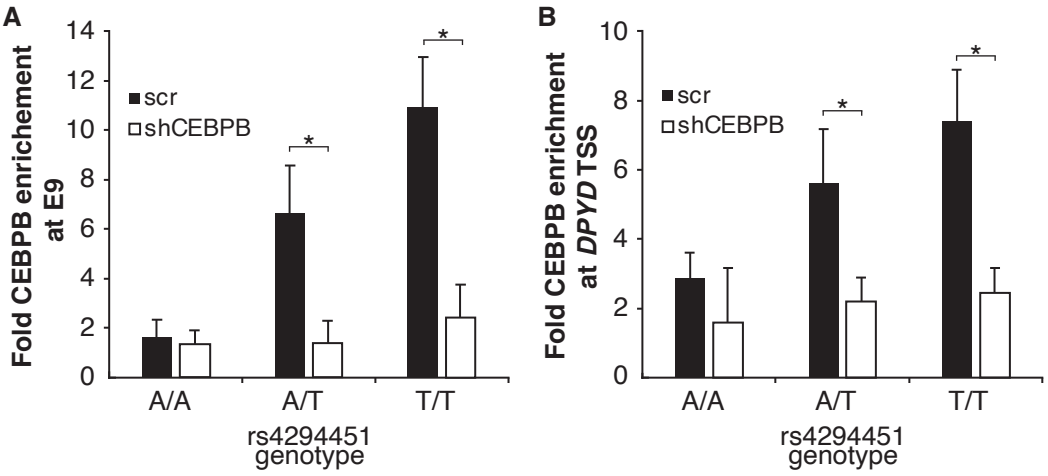
Zhang et al., Figure S2

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.



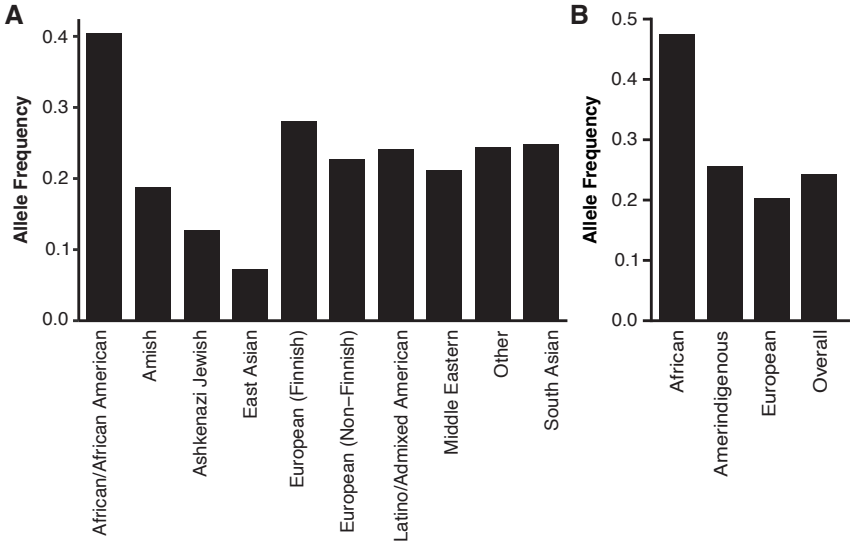
Zhang et al., Figure S3

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.



1067

1068 **Figure S5. Allele frequency for rs4294451-T allele in global populations.** (A) Allele

1069 frequencies in various populations for the rs4294451-T allele was retrieved from the

1070 gnomAD browser v3.1.2. (B) Local ancestry-informed frequency data was retrieved for

1071 the rs4294451-T allele within Latino-Admixed American samples of gnomAD v3.1