Effects of *NT5C2* germline variants on 6-mecaptopurine metabolism in children with acute lymphoblastic leukemia

Chuang Jiang^{1,2,3}, Wenjian Yang², Takaya Moriyama², Chengcheng Liu², Colton Smith², Wentao Yang², Maoxiang Qian⁴, Ziping Li³, Morten Tulstrup⁵, Kjeld Schmieglow^{5,6}, Kristine R. Crews², Hui Zhang³, Ching-Hon Pui⁷, William Evans², Mary Relling², Smita Bhatia⁸, Jun J. Yang²

¹Department of Hematology & Oncology, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China; ²Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN, United States; ³Department of Hematology/Oncology, Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong, China; ⁴Institute of Pediatrics and Department of Hematology and Oncology, Children's Hospital of Fudan University, and the Shanghai Key Laboratory of Medical Epigenetics, International Co-laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, Institutes of Biomedical Sciences, Fudan University, Shanghai, China; ⁵Department of Pediatrics and Adolescent Medicine, University Hospital Rigshospitalet, Copenhagen, Denmark; ⁶Institute of Clinical Medicine, Faculty of Medicine, University of Copenhagen, Copenhagen, Denmark; ⁷Department of Oncology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA; ⁸Institute for Cancer Outcomes and Survivorship, University of Alabama at Birmingham, Birmingham, AL, USA.

Corresponding author

Jun J. Yang PhD

Department of Pharmaceutical Sciences, St Jude Children's Research Hospital

MS313, 262 Danny Thomas Place, Memphis, TN 38105-3678

e-mail: jun.yang@stjude.org

Conflict of Interest statement

The authors declared no competing interests for this work.

Funding information

This work was partially supported by National Institutes of Health grants GM118578 GM115279, and CA096670.

Keywords: *NT5C2*; single nucleotide polymorphisms; thiopurine metabolism; acute lymphoblastic leukemia

39 **Abstract**

40 6-mercaptopurine (6-MP) is widely used in the treatment of acute lymphoblastic
 41 leukemia (ALL), and its cytotoxicity is primarily mediated by thioguanine nucleotide
 42 metabolites (TGN). Recent genome-wide association study has identified germline
 43 polymorphisms (e.g., rs72846714) in the *NT5C2* gene associated with 6-MP
 44 metabolism in patients with ALL. However, the full spectrum of genetic variation in
 45 *NT5C2* is unclear and its impact on 6-MP drug activation has not been
 46 comprehensively examined. To this end, we performed targeted sequencing of
 47 *NT5C2* in 588 children with ALL and identified 121 single nucleotide polymorphisms
 48 (SNPs) nominally associated with erythrocyte TGN during 6-MP treatment ($P < 0.05$).
 49 Of these, 61 variants were validated in a replication cohort of 372 children with ALL.
 50 After considering linkage disequilibrium and multivariate analysis, we confirmed two
 51 clusters of variants, represented by rs72846714 and rs58700372, that independently
 52 affected 6-MP metabolism. Functional studies showed that rs58700372 directly
 53 altered the activity of an intronic enhancer, with the variant allele linked to higher
 54 transcription activity and reduced 6-MP metabolism (lower TGN). By contrast,
 55 rs72846714 was not located in a regulatory element and instead its association signal
 56 was explained by linkage disequilibrium with a proximal functional variant rs12256506
 57 that activated *NT5C2* transcription in-*cis*. Our results indicated that *NT5C2* germline
 58 variation significantly contributes to inter-patient variability in thiopurine drug
 59 disposition.

60 **Introduction**

61 6-mercaptopurine (6-MP) is widely used as an anticancer and immunosuppressive
 62 agent [1-8]. In acute lymphoblastic leukemia (ALL), prolonged daily 6-MP exposure
 63 during maintenance therapy is one of the major components of contemporary ALL
 64 treatment regimens [7, 9-13]. 6-MP has a narrow therapeutic index and frequent dose
 65 adjustment is needed to balance between antileukemic efficacy and treatment toxicity
 66 in patients [7, 13, 14]. However, large inter-patient variability in 6-MP disposition
 67 poses a significant challenge to clinically determine the optimal 6-MP dosage for each
 68 individual [15, 16].

69 As a prodrug, 6-MP is converted into thioguanine nucleotides TGN (thioguanosine
 70 monophosphate [TGMP], thioguanosine diphosphate [TGDP], thioguanosine
 71 triphosphate [TGTP]) via a series of enzymatic reactions [8, 17]. TGTP is further
 72 reduced to deoxythioguanosine triphosphate (TdGTP), which is then incorporated into
 73 double-stranded DNA (DNA-TG) to trigger futile mismatch repair and eventually
 74 apoptosis [7, 18-20]. Levels of these intracellular 6-MP metabolites can inform drug
 75 toxicity and response and have the potential to guide treatment individualization. For
 76 example, the NOPHO ALL2008 study demonstrated that DNA-TG concentration is
 77 significantly associated with the risk of leukemia relapse [21]. Variation in erythrocyte
 78 TGN level during maintenance therapy has been associated with relapse-free survival
 79 in some patients with ALL [12]. Therefore, therapeutic drug monitoring of thiopurine
 80 metabolites may be clinically relevant for patients receiving 6-MP.

81 6-MP metabolism is regulated by several key enzymes (e.g., NT5C2, TPMT, and

82 NUDT15), and genetic variants in these genes have been linked to 6-MP toxicity
83 and/or resistance. For example, inherited deficiency of TPMT and/or NUDT15 causes
84 excessive accumulation of 6-MP active metabolites and severe toxicity, and
85 genotype-guided dose reduction can dramatically mitigate side effects of this drug
86 [22-27]. By contrast, somatic *NT5C2* mutations are common in relapsed ALL and
87 encode hyperactive nucleotidase that degrades 6-MP active metabolites thioinosine
88 monophosphate (TIMP) and TGMP, leading to lower TGTP, decreased DNA-TG, and
89 6-MP resistance [28-32]. The presence of sub-clonal ALL *NT5C2* mutations was
90 associated with high risk of treatment failure and inferior outcome after relapse [33]. A
91 recent genome-wide association study (GWAS) reported germline variants at the
92 *NT5C2* locus (e.g., rs72846714) strongly associated with erythrocyte TGN level
93 during ALL therapy [34].

94 Although the genome-wide significant association at the *NT5C2* locus signifies the
95 impact of inherited variation in this gene on 6-MP pharmacokinetics, the full spectrum
96 of genetic polymorphism in *NT5C2* is unclear and its impact on drug activation has not
97 been systematically examined. To this end, we performed targeted sequencing of
98 *NT5C2* to evaluate the contribution of germline *NT5C2* variants to inter-individual
99 variations in 6-MP metabolites in children with ALL.

100

101 **Materials and methods**

102 Patients and therapy

103 The discovery cohort consisted of patients prospectively enrolled onto the Children's
 104 Oncology Group (COG) AALL03N1 study. Details of enrollment criteria and study
 105 design have been described previously [16, 26]. Briefly, patients were treated on COG
 106 AALL03N1 frontline protocols for newly diagnosed ALL (or per frontline protocols for
 107 small number of patients), were enrolled onto this study for a total of six months
 108 maintenance therapy that included a daily dose of oral 6-MP. The protocol planned
 109 dose of 6-MP during maintenance phase was 75 mg/m² per day for the COG
 110 AALL03N1 cohort, with provisions for dose adjustment based on the degree of
 111 myelosuppression (white blood cell count) and/or occurrence of infections.

112 The replication cohort consisted of children with ALL treated on the St Jude Children's
 113 Research Hospital Total Therapy XV (St. Jude Total XV) protocol, with the planned
 114 dose of 75 mg/m² in the low-risk arm or 50 mg/m² in the standard- or high-risk arm
 115 during the phase of continuation therapy when TGN was measured [35]. *TPMT* and
 116 *NUDT15* genotyping was performed as described previously [26], and patients with
 117 known risk variants in either one of these two genes were excluded from the main
 118 analyses (N=74 and 47 in the COG AALL03N1 and St. Jude Total XV cohorts,
 119 respectively).

120 This study was approved by respective institutional review boards, and informed
 121 consent was obtained from parents, guardians, or patients, as appropriate.

122 6-MP metabolites measurement

123 Erythrocyte TGN levels were measured monthly for each participant during the six
124 months study period of the COG AALL03N1 study, as described previously [16]. We
125 generated a summarized TGN value for each patient based on their longitudinal TGN
126 data using a linear mixed-effect model adjusting for time point of each measurement
127 and prescribed 6-MP dose recorded monthly prior to the TGN measurement [26].
128 For St. Jude Total XV patients, TGN was measured at the start of Reinduction I during
129 continuation therapy, using methods described previously [36, 37], and total 6-MP
130 dose two weeks prior to sampling was recorded as total mg^2/m^2 . TGN data from
131 patients who received fewer than 12 of the 14 doses of 6-MP prior to drug metabolite
132 measurement (N=18) or those with non-adherence based on self-report (N=4) were
133 excluded.

134 NT5C2 targeted sequencing and SNP genotype imputation

135 Germline DNA was extracted from blood or bone marrow samples collected during
136 clinical remission. Illumina dual-indexed libraries were created from genomic DNA and
137 pooled in sets of 96 before hybridization with customized Roche NimbleGene SeqCap
138 EZ probes (Roche, Roche NimbleGen, Madison, WI, USA) to capture the *NT5C2*
139 genomic region. Quantitative PCR was used to define the appropriate capture product
140 titer necessary to efficiently populate an Illumina HiSeq 2000 flow cell for paired-end 2
141 × 100 bp sequencing. Coverage of greater than 20 x depth was achieved across more
142 than 80% of the targeted regions for nearly all samples. Sequence reads in the
143 FASTQ format were mapped and aligned using the Burrows-Wheeler Aligner (BWA)
144 and genetic variants were called using the GATK pipeline (version 3.1), as previously

described [38-40], and annotated using the ANNOVA program with the annotation databases including RefSeq.

In addition to targeted sequencing, we also performed imputation for variants at the *NT5C2* locus, on the basis of SNP genotype data from Affymetrix SNP6 arrays of COG and St. Jude patients [26]. We employed a large reference panel of human haplotypes from the Haplotype Reference Consortium (HRC r1.1 2016) in Michigan Imputation Server with ShapeIT (v2.r790) as the phasing tool for imputation, following procedures described previously [41]. Imputed variants were filtered based on imputation quality, allele frequency, call rate, and deviation from Hardy-Weinberg equilibrium. Specifically, variants with one of the following features were excluded: imputation quality metric r^2 less than 0.3 (indicating inadequate accuracy of the imputed genotype); minor allele frequency in less than 1% in either one of our two patient cohorts; Hardy-Weinberg equilibrium p value less than 5.00×10^{-4} ; variants located outside *NT5C2* locus (beyond 100kb of the first and last *NT5C2* exons).

Statistical analyses

The association of *NT5C2* genotype with TGN was analyzed in patients with wildtype genotype in *TPMT* and *NUDT15* to avoid confounding effects. A linear regression model was used to test association, with 6-MP dose and genetic ancestry as covariates. R statistical software (version 3.5; <http://www.r-project.org/>) was used for all analyses unless indicated otherwise. Regional association plots were created by Locus zoom (<http://locuszoom.org/>). We also explored association of *NT5C2* variants with TGN in patients with heterozygous genotype at known *TPMT* and/or *NUDT15*

167 variants.

168 NT5C2 recombinant protein production and 5'-nucleotidase activity assay

169 C-terminal Hisx6-tagged full-length cDNA for wildtype *NT5C2* was cloned into the
170 pET30a expression vector using NdeI and HindIII restriction sites. Point mutations, p.
171 Thr3Ala, p.Met15Thr and p.Ile528Val, (hereafter referred to as p.T3A, p.M15T, and
172 p.I528V) were introduced by using the QuickChange II XL Site-Directed Mutagenesis
173 Kit (Agilent Technologies). NT5C2 recombinant protein production and purification
174 were performed using previously described methods [28]. Successful purification of
175 NT5C2 proteins was confirmed by Coomassie brilliant blue staining (**Figure S1**).

176 5'-nucleotidase activity was measured by quantifying the release of phosphate from
177 TIMP (Toronto Research Chemicals), or TGMP (Jena Bioscience), using the
178 SensoLyte Malachite Green Assay kit (AnaSpec) in assay buffer (MgCl₂ 2 mmol/L,
179 NaCl 120 mmol/L, KCl 5 mmol/L, glucose 10 mmol/L, HEPES 20 mmol/L). NT5C2
180 protein (300 ng) was incubated with 40 mmol/L TIMP, or TGMP at 37°C for 30
181 minutes. Experiments were assessed in triplicates.

182 Functional annotation and characterization of non-coding variants in *NT5C2*

183 Chromatin accessibility in human hematopoietic cells were extracted from GSE74912
184 [42]. Transcription factor ChIP-seq clusters and DNase hypersensitive site data were
185 from ENCODE project [43, 44]. All peaks were aligned to GRCh37/hg19 on UCSC
186 genome browser [45-47].

187 To functionally characterize potential function of non-coding variants in *NT5C2*,
188 356-600bp regions encompassing the SNP of interest were amplified using

189 CloneAmp HiFi PCR Premix (Clontech), and then cloned into the pGL4.23-mini/P
 190 vector including a minimal SV40 promoter upstream of firefly luciferase. The variants
 191 were introduced by site-directed mutagenesis, by using the QuickChange Lightning
 192 Site-Directed Mutagenesis Kit (Agilent). Primers used in this experiment were listed in
 193 **Table S1**. Human erythroleukemic K562 cells were transiently co-transfected with the
 194 pGL4.23 reporter construct and pGL-TK Renilla luciferase construct using Lonza
 195 4D-Nucleofector system (Lonza). Firefly luciferase activity was measured 24 hours
 196 post-transfection by the Dual Luciferase Assay kit (Promega). Results were
 197 normalized to Renilla luciferase activity. These experiments were performed in
 198 triplicated for three independent times.

199

200 **Results**

201 Two cohorts of children with newly-diagnosed ALL receiving 6-MP therapy were
 202 included in this study: the COG AALL03N1 and St. Jude Total XV clinical trials as the
 203 discovery and validation set, respectively (**Table 1**). After removing patients with
 204 known risk variants in *TPMT* and/or *NUDT15* and those without TGN measurements,
 205 588 and 372 patients in the respective cohorts were included in the subsequent
 206 genetic association analyses.

207 **Targeted resequencing of *NT5C2* identified three missense variants**

208 Targeted sequencing of *NT5C2* in the COG AALL03N1 cohort identified three
 209 missense variants, p.T3A (rs10883841), p.M15T (rs751896016), and I528V
 210 (rs756913240) (**Figure 1**). The p.T3A variant was recurrent with minor allele
 211 frequency of 9.3%, and p.M15T and I528V were observed only once in this cohort
 212 (**Figure 2A**). None of the three variants were predicted to be damaging by *in silico*
 213 predication tools SIFT or PolyPhen-2. The common variant p.T3A was not associated
 214 with TGN in patients ($P > 0.05$, **Figure 2A**), and the singleton variants were not
 215 evaluated in the association analysis due to low frequency. Furthermore, we
 216 characterized enzymatic properties of all three variant *NT5C2* proteins with thiopurine
 217 metabolites as substrates. *NT5C2* p.I528V showed a modest increase in
 218 5'-nucleotidase activity against TIMP but not with TGMP (**Figure 2B**), whereas the
 219 other two variants showed a similar activity compared to wildtype *NT5C2* (**Figure 2B**
 220 **and 2C**). Together, these results indicated that these three germline *NT5C2* missense
 221 variants did not affect erythrocyte TGN in children with ALL.

222 **Association of non-coding variants in *NT5C2* with erythrocyte TGN**

223 Non-coding variants in *NT5C2* were determined by targeted sequencing as well as
 224 SNP imputation. A total of 779 non-coding variants in the genomic region
 225 encompassing the *NT5C2* gene and the 200Kb flanking region with a minor allele
 226 frequency of at least 1%, were identified in the COG AALL03N1 cohort and tested for
 227 association with TGN (**Figure 1**). After adjusting for 6-MP dose and genetic ancestry
 228 as covariates, 121 variants were nominally associated with erythrocyte TGN during
 229 6-MP treatment ($P < 0.05$, **Figure 1, Figure S2 and Table S2**). These variants were
 230 then tested for replication in the St. Jude Total XV cohort, and 61 of 121 variants were
 231 validated for association with TGN at the $P < 0.05$ level (**Figure 1, Figure S2 and**
 232 **Table S2**), including rs72846714 which was reported in the previous GWAS of 6-MP
 233 metabolites [34] (**Figure 3A and 3B, Table S2**).

234 Of 61 validated *NT5C2* SNPs, only 22 remained significant in both discovery and
 235 validation cohorts after adjusting for the GWAS hit rs72846714 [34]. This cluster of 22
 236 variants were in strong linkage disequilibrium with each other ($r^2 > 0.99$), and
 237 represented a novel association at the *NT5C2* locus independent of the GWAS signal.
 238 For example, in the discovery cohort COG AALL03N1, there was a significant
 239 decrease in TGN metabolite levels by genotypes at the independent SNP rs58700372
 240 from wildtype (TT, 154.8 ± 61.1) to homozygotes (CC, 108.4 ± 44.4), with the
 241 heterozygotes (TC, 136.2 ± 61.1) exhibiting an intermediate level of TGN ($P = 6.69 \times$
 242 10^{-4} ; **Figure 3C**). This association was replicated in the St. Jude Total XV cohort:
 243 mean levels of the metabolite were 225.5 ± 132.9 , 193.3 ± 86.0 , and 165.5 ± 49.3 in

244 wildtype, heterozygous and homozygous individuals, respectively ($P = 9.58 \times 10^{-5}$.

245 **Figure 3D).**

246 Interestingly, even within patients heterozygous for known *TPMT* and/or *NUDT15*

247 variants, *NT5C2* genotype at rs72846714 and rs58700372 appeared to further

248 distinguish patients with differential 6-MP metabolism although this only reached

249 statistical significance in the St. Jude Total XV cohort (**Figure S3**). The frequency of

250 variant allele at these two *NT5C2* SNPs also differ by race and ethnicity (**Figure S4**).

251 Taken together, we identified two clusters of non-coding variants that independently

252 influenced 6-MP metabolism, represented by rs72846714 and rs58700372.

253 **Functional exploration of the germline variants**

254 To gain insight into the biological basis of the association signals at these non-coding

255 *NT5C2* variants, we firstly examined chromatin accessibility inferred from the

256 ATAC-seq of human hematopoietic cells and histone mark ChIP-seq signals from

257 ENCODE project [42-44].

258 For the previous reported GWAS hit rs72846714, we found that it was in the intronic

259 region with minimal overlap with regulatory elements (**Figure 4A**), indicating the lack

260 of transcriptional regulation function of this variant. However, amongst variants in

261 strong LD with rs72846714, we identified five SNPs that showed significant

262 association with TGN in both discovery and validation cohorts and were also in

263 proximity of open chromatin segments in multiple hematopoiesis cells (rs12261294,

264 rs12240508, rs7091462, rs113642369, and rs12256506) (**Figure 4A**). We postulated

265 that one of these variants may directly regulate *NT5C2* transcription and thus causal

266 to the GWAS signal at this locus. We performed luciferase assay to test the effects of
 267 rs72846714 and each of the five LD variants on transcription activity and found that
 268 only rs12256506 substantially influenced enhancer activity (**Figure 4B**). The
 269 substitution of the reference allele with the variant allele at this variant resulted in a
 270 1.48-fold increase in transcription activity ($P = 0.0049$, **Figure 4B**), consistent with
 271 higher expression of *NT5C2* and thus lower TGN in erythrocytes (**Figure 3A and 3B**).
 272 This is in line with the notion that NT5C2 degrades 6-MP active metabolites TIMP and
 273 TGMP, and thus its activity is inversely correlated with TGN.
 274 By contrast, the novel variant rs58700372 was located within an open chromatin
 275 region based on ATAC-seq signal in erythroblast cells (**Figure 5A**). DNase
 276 accessibility and transcription factors ChIP-seq also suggested active transcription in
 277 this region (**Figure 5A**). Using luciferase assay to test its effect on transcription activity,
 278 we found that substitution of the reference allele with the variant allele resulted in a
 279 3.78-fold increase in transcription activity (**Figure 5B**). These findings indicate that
 280 rs58700372 directly potentiates the activity of an intronic enhancer, activates *NT5C2*
 281 transcription, and thereby reduces 6-MP metabolism.

282

283 Discussion

284 Prior pharmacogenetic studies have identified germline variants in *NUDT15* and
 285 *TPMT* associated with 6-MP sensitivity, and treatment modification based on *TPMT*
 286 and *NUDT15* genotype is a clear example of genetically guided precision medicine
 287 [22-27]. However, the majority of inter-individual variability in 6-MP pharmacokinetics
 288 remains unexplained. Recently, a genome-wide association study described a
 289 germline variant located in *NT5C2* intronic region associated with erythrocyte TGN
 290 during 6-MP therapy, pointing to potential effects of *NT5C2* germline variants on this
 291 phenotype [34].

292 *NT5C2* is an ubiquitously expressed cytosolic nucleotidase involved in the
 293 maintenance of intracellular nucleotide pool homeostasis by promoting the clearance
 294 of excess purine nucleotides from cells [48, 49]. *NT5C2* preferentially
 295 dephosphorylates the 6-hydroxypurine monophosphates inosine monophosphate
 296 (IMP), guanosine monophosphate (GMP), and xanthosine monophosphate (XMP), as
 297 well as the deoxyribose forms of IMP and GMP (dIMP and dGMP), facilitating the
 298 export of purine nucleosides. In the context of ALL therapy, *NT5C2* can
 299 dephosphorylate and inactivate 6-MP metabolites TIMP and TGMP, and thus reduce
 300 the amount of TGN available for DNA incorporation and compromise cytotoxicity of
 301 this drug [28]. Acquisition of *NT5C2* mutations in relapsed ALL impedes the formation
 302 of DNA-TGN and is a frequent mechanism of 6-MP drug resistance [29-32].

303 In this study, we comprehensively evaluated *NT5C2* germline variants for their
 304 association with the 6-MP metabolite TGN during ALL therapy. We identified 61

305 *NT5C2* variants significantly associated with TGN after adjusting for dose and genetic
306 ancestry in both discovery and validation cohorts. In line with previous GWAS [34], the
307 variant A allele of rs72846714 was associated with lower TGN levels, indicating that
308 this pharmacogenetic association is robust across different chemotherapy regimens
309 and different patient populations. Importantly, we also identified a novel variant,
310 namely rs58700372, that exhibited independent effect on 6-MP metabolite beyond
311 that of the GWAS hit, suggesting multiple functional variants at this locus. However,
312 neither of these two variants were associated with 6-MP dose intensity in our analyses
313 (**Figure S5**), plausibly because these germline *NT5C2* SNPs only modestly alter drug
314 metabolism to a degree that is insufficient to cause clinical toxicity or drug resistance.
315 It is also worth noting that there was a trend for *NT5C2* variants to be associated with
316 TGN even within patients hemizygously deficient of *TPMT* and/or *NUDT15* (**Figure**
317 **S3**). This is relevant because these individuals exhibit wide variability in 6-MP
318 tolerance and additional biomarkers are needed to precisely predict thiopurine toxicity.
319 It remains possible that many other genetic variants influence 6-MP metabolism with
320 relatively low penetrance, and their combined effects may warrant clinical
321 consideration. There was also a trend that the variant alleles at rs72846714 and
322 rs58700372 were linked to lower MeTIMP in erythrocytes in our ALL cohorts but it did
323 not reach statistical significance (**Figure S6**).
324 Unlike *TPMT* or *NUDT15*, there are no common germline coding variants in *NT5C2*
325 that affect its activity. The majority of pharmacogenetic variants identified in this gene
326 are non-coding. Located in intronic and intergenic regions, these *NT5C2* variants

327 primarily affect gene function by transcription regulation. For example, the association
 328 signal at the previously reported GWAS hit rs72846714 was likely explained by a
 329 proximal variant rs12256506 that activates *NT5C2* transcription in-cis. Our analysis
 330 also identified rs58700372 for its independent association with TGN beyond that of
 331 the GWAS SNP. Our functional study indicated that this variant directly alters the
 332 activity of an intronic enhancer, with the variant allele linked to higher transcription
 333 activity and reduced 6-MP activation (lower TGN). Of note, rs58700372 is located
 334 ~50kb downstream of the coding region of *NT5C2*, and the potential mechanism by
 335 which this variant regulates *NT5C2* transcription may involve chromatin remodeling
 336 and long-distance interaction, which need to be examined in future studies. It remains
 337 unclear whether functional effects of these two *NT5C2* SNPs are restricted to
 338 erythrocytes and future studies are warranted to explore cellular models of a broader
 339 range of tissue origins.

340 In conclusion, we comprehensively analyzed the role of *NT5C2* germline variants in
 341 6-MP metabolism during ALL therapy and identified functional SNPs that
 342 independently and directly regulate *NT5C2* activity and 6-MP metabolism. Our results
 343 indicate that *NT5C2* germline variation contributes to inter-patient variability in
 344 thiopurine drug disposition.

345

346

347 **Study Highlights**

348 What is the current knowledge on the topic?

349 Levels of intracellular 6-MP metabolites can inform drug toxicity and response and
350 potentially inform treatment individualization. Recent GWAS study reported initial
351 evidence linking *NT5C2* germline variants to 6-MP metabolism. However, the full
352 spectrum of genetic polymorphism in *NT5C2* and their impact on inter-patient
353 variability in thiopurine drug disposition are largely unknown.

354 What question did this study address?

355 This pharmacogenomics study comprehensively evaluated the contribution of genetic
356 variations in *NT5C2* to 6-MP metabolite TGN in two ALL cohorts and identified two
357 representative SNPs that independently affected 6-MP metabolism. Importantly, we
358 also sought to understand the molecular mechanism by which these non-coding
359 variants modulate *NT5C2* function.

360 What does this study add to our knowledge?

361 This is one of the first studies to describe the full spectrum of genetic variants in
362 *NT5C2* and examined their effects on thiopurine metabolism in large cohorts of
363 patients. We also provided functional evidence of how coding and non-coding variants
364 modulate *NT5C2* activity. Our results firmly established the importance of *NT5C2*
365 germline variation in explaining inter-patient variability in thiopurine drug disposition.

366 How might this change clinical pharmacology or translational science?

367 This study identified novel genetic variants significantly associated with thiopurine
368 drug activation, therefore expanding the list of pharmacogenetic factors that may aid
369 the optimization of pharmacotherapy using this class of drugs.

370 **Acknowledgment**

371 We thank the patients and parents who participated in the clinical studies included in
372 this report. This work was partially supported by National Institutes of Health grants
373 GM118578, GM115279, and CA096670.

374

375 **Author contributions**

376 C.J., H.Z. and J.J.Y. wrote the manuscript; J.J.Y. designed the research; C.J., Z.P.L.
377 and T.M. performed the research; W.J.Y., C.S., W.T.Y., M.X.Q and M.T. analyzed the
378 data; C.C.L., C.H.P., W.E., S.B., K. S., and K.R.C. and J.J.Y. contributed new reagents
379 and/or analytical tools.

380

381

382

383

384

385

386

387

388

389

390

391

392 **Reference**

- 393 1. Goldberg, R. and P.M. Irving, *Toxicity and response to thiopurines in patients*
394 *with inflammatory bowel disease*. Expert review of gastroenterology &
395 hepatology, 2015. **9**(7): p. 891-900.
- 396 2. Karran, P. and N. Attard, *Thiopurines in current medical practice: molecular*
397 *mechanisms and contributions to therapy-related cancer*. Nature Reviews
398 Cancer, 2008. **8**(1): p. 24-36.
- 399 3. Maltzman, J.S. and G.A. Koretzky, *Azathioprine: old drug, new actions*. The
400 Journal of clinical investigation, 2003. **111**(8): p. 1122-1124.
- 401 4. Pui, C.-H., et al., *Biology, risk stratification, and therapy of pediatric acute*
402 *leukemias: an update*. Journal of clinical oncology, 2011. **29**(5): p. 551.
- 403 5. Reinisch, W., et al., *Azathioprine versus mesalazine for prevention of*
404 *postoperative clinical recurrence in patients with Crohn's disease with*
405 *endoscopic recurrence: efficacy and safety results of a randomised,*
406 *double-blind, double-dummy, multicentre trial*. Gut, 2010. **59**(6): p. 752-759.
- 407 6. Vora, A., et al., *Treatment reduction for children and young adults with low-risk*
408 *acute lymphoblastic leukaemia defined by minimal residual disease (UKALL*
409 *2003): a randomised controlled trial*. The lancet oncology, 2013. **14**(3): p.
410 199-209.
- 411 7. Koren, G., et al., *Systemic exposure to mercaptopurine as a prognostic factor in*
412 *acute lymphocytic leukemia in children*. New England Journal of Medicine, 1990.
413 **323**(1): p. 17-21.

- 414 8. Elion, G.B., *The purine path to chemotherapy*. Science, 1989. **244**(4900): p.
415 41-47.
- 416 9. Group, C.A.C., *Duration and intensity of maintenance chemotherapy in acute*
417 *lymphoblastic leukaemia: overview of 42 trials involving 12 000 randomised*
418 *children*. The Lancet, 1996. **347**(9018): p. 1783-1788.
- 419 10. Lilleyman, J.S. and L. Lennard, *Mercaptopurine metabolism and risk of relapse*
420 *in childhood lymphoblastic leukaemia*. The Lancet, 1994. **343**(8907): p.
421 1188-1190.
- 422 11. Schmiegelow, K., et al., *Mercaptopurine/methotrexate maintenance therapy of*
423 *childhood acute lymphoblastic leukemia: clinical facts and fiction*. Journal of
424 pediatric hematology/oncology, 2014. **36**(7): p. 503.
- 425 12. Lennard, L. and J.S. Lilleyman, *Variable mercaptopurine metabolism and*
426 *treatment outcome in childhood lymphoblastic leukemia*. Journal of Clinical
427 Oncology, 1989. **7**(12): p. 1816-1823.
- 428 13. Relling, M.V., et al., *Prognostic importance of 6-mercaptopurine dose intensity in*
429 *acute lymphoblastic leukemia*. Blood, The Journal of the American Society of
430 Hematology, 1999. **93**(9): p. 2817-2823.
- 431 14. Schmiegelow, K., et al., *Maintenance therapy of childhood acute lymphoblastic*
432 *leukemia revisited—Should drug doses be adjusted by white blood cell,*
433 *neutrophil, or lymphocyte counts?* Pediatric Blood & Cancer, 2016. **63**(12): p.
434 2104-2111.
- 435 15. Bhatia, S., et al., *Nonadherence to oral mercaptopurine and risk of relapse in*

- 436 *Hispanic and non-Hispanic white children with acute lymphoblastic leukemia: a*
 437 *report from the children's oncology group.* Journal of Clinical Oncology, 2012.
 438 **30**(17): p. 2094.
- 439 16. Bhatia, S., et al., *Systemic exposure to thiopurines and risk of relapse in*
 440 *children with acute lymphoblastic leukemia: a Children's Oncology Group Study.*
 441 JAMA oncology, 2015. **1**(3): p. 287-295.
- 442 17. Elion, G., *The George Hitchings and Gertrude Elion lecture: the pharmacology*
 443 *of azathioprine.* Ann NY Acad Sci, 1990. **685**: p. 1239-1256.
- 444 18. Li, G.-M., *The role of mismatch repair in DNA damage-induced apoptosis.*
 445 Oncology research, 1999. **11**(9): p. 393-400.
- 446 19. Ling, Y., et al., *2'-Deoxy-6-thioguanosine 5'-triphosphate as a substrate for*
 447 *purified human DNA polymerases and calf thymus terminal*
 448 *deoxynucleotidyltransferase in vitro.* Molecular pharmacology, 1991. **40**(4): p.
 449 508-514.
- 450 20. Swann, P.F., et al., *Role of postreplicative DNA mismatch repair in the cytotoxic*
 451 *action of thioguanine.* Science, 1996. **273**(5278): p. 1109-1111.
- 452 21. Nielsen, S.N., et al., *DNA-thioguanine nucleotide concentration and relapse-free*
 453 *survival during maintenance therapy of childhood acute lymphoblastic*
 454 *leukaemia (NOPHO ALL2008): a prospective substudy of a phase 3 trial.* The
 455 Lancet Oncology, 2017. **18**(4): p. 515-524.
- 456 22. Moriyama, T., et al., *NUDT15 polymorphisms alter thiopurine metabolism and*
 457 *hematopoietic toxicity.* Nature genetics, 2016. **48**(4): p. 367-373.

- 458 23. Nishii, R., et al., *Preclinical evaluation of NUDT15-guided thiopurine therapy*
459 *and its effects on toxicity and antileukemic efficacy*. Blood, 2018. **131**(22): p.
460 2466-2474.
- 461 24. Relling, M.V., et al., *Clinical pharmacogenetics implementation consortium*
462 *guideline for thiopurine dosing based on TPMT and NUDT 15 genotypes: 2018*
463 *update*. Clinical Pharmacology & Therapeutics, 2019. **105**(5): p. 1095-1105.
- 464 25. Suiter, C.C., et al., *Massively parallel variant characterization identifies NUDT15*
465 *alleles associated with thiopurine toxicity*. Proceedings of the National Academy
466 of Sciences, 2020. **117**(10): p. 5394-5401.
- 467 26. Yang, J.J., et al., *Inherited NUDT15 variant is a genetic determinant of*
468 *mercaptopurine intolerance in children with acute lymphoblastic leukemia*.
469 Journal of clinical oncology, 2015. **33**(11): p. 1235.
- 470 27. Yang, S.-K., et al., *A common missense variant in NUDT15 confers*
471 *susceptibility to thiopurine-induced leukopenia*. Nature genetics, 2014. **46**(9): p.
472 1017.
- 473 28. Moriyama, T., et al., *Mechanisms of NT5C2-mediated thiopurine resistance in*
474 *acute lymphoblastic leukemia*. Molecular cancer therapeutics, 2019. **18**(10): p.
475 1887-1895.
- 476 29. Dieck, C.L., et al., *Structure and mechanisms of NT5C2 mutations driving*
477 *thiopurine resistance in relapsed lymphoblastic leukemia*. Cancer Cell, 2018.
478 **34**(1): p. 136-147. e6.
- 479 30. Meyer, J.A., et al., *Relapse-specific mutations in NT5C2 in childhood acute*

- 480 *lymphoblastic leukemia*. *Nature genetics*, 2013. **45**(3): p. 290.
- 481 31. Tzoneva, G., et al., *Clonal evolution mechanisms in NT5C2 mutant-relapsed*
482 *acute lymphoblastic leukaemia*. *Nature*, 2018. **553**(7689): p. 511-514.
- 483 32. Tzoneva, G., et al., *Activating mutations in the NT5C2 nucleotidase gene drive*
484 *chemotherapy resistance in relapsed ALL*. *Nature medicine*, 2013. **19**(3): p.
485 368-371.
- 486 33. Barz, M.J., et al., *Subclonal NT5C2 mutations are associated with poor*
487 *outcomes after relapse of pediatric acute lymphoblastic leukemia*. *Blood, The*
488 *Journal of the American Society of Hematology*, 2020. **135**(12): p. 921-933.
- 489 34. Tulstrup, M., et al., *NT5C2 germline variants alter thiopurine metabolism and*
490 *are associated with acquired NT5C2 relapse mutations in childhood acute*
491 *lymphoblastic leukaemia*. *Leukemia*, 2018. **32**(12): p. 2527-2535.
- 492 35. Pui, C.-H., et al., *Treating childhood acute lymphoblastic leukemia without*
493 *cranial irradiation*. *New England Journal of Medicine*, 2009. **360**(26): p.
494 2730-2741.
- 495 36. Su, Y., et al., *Assay of 6-mercaptopurine and its metabolites in patient plasma*
496 *by high-performance liquid chromatography with diode-array detection*. *Journal*
497 *of Chromatography B: Biomedical Sciences and Applications*, 1999. **732**(2): p.
498 459-468.
- 499 37. Dervieux, T., et al., *HPLC determination of thiopurine nucleosides and*
500 *nucleotides in vivo in lymphoblasts following mercaptopurine therapy*. *Clinical*
501 *chemistry*, 2002. **48**(1): p. 61-68.

- 502 38. Churchman, M.L., et al., *Germline genetic IKZF1 variation and predisposition to*
503 *childhood acute lymphoblastic leukemia*. *Cancer cell*, 2018. **33**(5): p. 937-948.
504 e8.
- 505 39. Moriyama, T., et al., *Germline genetic variation in ETV6 and risk of childhood*
506 *acute lymphoblastic leukaemia: a systematic genetic study*. *The Lancet*
507 *Oncology*, 2015. **16**(16): p. 1659-1666.
- 508 40. Qian, M., et al., *TP53 germline variations influence the predisposition and*
509 *prognosis of B-cell acute lymphoblastic leukemia in children*. *Journal of Clinical*
510 *Oncology*, 2018. **36**(6): p. 591.
- 511 41. Qian, M., et al., *Genome-wide association study of susceptibility loci for T-cell*
512 *acute lymphoblastic leukemia in children*. *JNCI: Journal of the National Cancer*
513 *Institute*, 2019. **111**(12): p. 1350-1357.
- 514 42. Corces, M.R., et al., *Lineage-specific and single-cell chromatin accessibility*
515 *charts human hematopoiesis and leukemia evolution*. *Nature genetics*, 2016.
516 **48**(10): p. 1193-1203.
- 517 43. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human*
518 *genome*. *Nature*, 2012. **489**(7414): p. 57-74.
- 519 44. Davis, C.A., et al., *The Encyclopedia of DNA elements (ENCODE): data portal*
520 *update*. *Nucleic acids research*, 2018. **46**(D1): p. D794-D801.
- 521 45. Karolchik, D., et al., *The UCSC Table Browser data retrieval tool*. *Nucleic acids*
522 *research*, 2004. **32**(suppl_1): p. D493-D496.
- 523 46. Kent, W.J., et al., *The human genome browser at UCSC*. *Genome research*,

524 2002. **12**(6): p. 996-1006.

525 47. Rosenbloom, K.R., et al., *ENCODE data in the UCSC Genome Browser: year 5*
526 *update*. Nucleic acids research, 2012. **41**(D1): p. D56-D63.

527 48. Gazzola, C., et al., *Cytosolic high Km 5'-nucleotidase and*
528 *5'(3')-deoxyribonucleotidase in substrate cycles involved in nucleotide*
529 *metabolism*. Journal of Biological Chemistry, 2001. **276**(9): p. 6185-6190.

530 49. Sychala, J., V. Madrid-Marina, and I. Fox, *High Km soluble 5'-nucleotidase*
531 *from human placenta. Properties and allosteric regulation by IMP and ATP*.
532 Journal of Biological Chemistry, 1988. **263**(35): p. 18759-18765.

533 **References for figures and tables**

534 34. Tulstrup, M., et al., *NT5C2 germline variants alter thiopurine metabolism and*
535 *are associated with acquired NT5C2 relapse mutations in childhood acute*
536 *lymphoblastic leukaemia*. Leukemia, 2018. **32**(12): p. 2527-2535.

537 40. Qian, M., et al., *TP53 germline variations influence the predisposition and*
538 *prognosis of B-cell acute lymphoblastic leukemia in children*. Journal of Clinical
539 Oncology, 2018. **36**(6): p. 591.

540 41. Qian, M., et al., *Genome-wide association study of susceptibility loci for T-cell*
541 *acute lymphoblastic leukemia in children*. JNCI: Journal of the National Cancer
542 Institute, 2019. **111**(12): p. 1350-1357.

543 42. Corces, M.R., et al., *Lineage-specific and single-cell chromatin accessibility*
544 *charts human hematopoiesis and leukemia evolution*. Nature genetics, 2016.
545 **48**(10): p. 1193-1203.

546

547

548 **Figure legends**

549 **Figure 1. Discovery of *NT5C2* variants associated with 6-MP metabolism in** 550 **children with ALL.**

551 Targeted sequencing coupled with SNP imputation in the discovery COG AALL03N1
552 cohort (588 patients) identified 782 variants at the *NT5C2* locus: three missense
553 variants and 779 non-coding variants. The association of *NT5C2* SNP genotype with
554 erythrocyte TGN was tested using a linear mixed-effects model with dose intensity
555 and genetic ancestry as covariates. This analysis was restricted to patients with
556 wildtype genotype for *TPMT* and *NUDT15*. To identify novel TGN associated variants,
557 we also performed multivariate analysis conditioning on index variant rs72846714
558 identified from previous GWAS [34]. Validation was performed in the independent St.
559 Jude Total XV cohort (372 patients), following the same analysis strategy.
560 #non-coding variants with minor allele frequency<1% were excluded from the
561 association analyses due to statistical consideration. * All 22 variants are in linkage
562 disequilibrium with pair-wise LD $r^2 > 0.99$.

563 **Figure 2. Functional characterization of *NT5C2* missense variants.**

564 (A) Targeted sequencing of *NT5C2* in the COG AALL03N1 cohort identified three
565 missense variants. Chromosome location was aligned to GRCh37/hg19. SIFT and
566 PolyPhen2 prediction tools were used to predict mutation function. MAF, minor allele
567 frequency. (B) and (C) 5'-nucleotidase activity was determined of wildtype and variant

NT5C2, with thiopurine metabolites as substrates. Purified NT5C2 proteins were incubated with TIMP or TGMP, and 5'-nucleotidase activity was measured by quantifying the release of free phosphate using the Malechite Green assay. Bars represent the mean values; the error bars represent the SD from triplicate. Ref allele, reference allele; Var allele, variant allele; #, P value of association of genotype and TGN was analyzed using a linear regression model with 6-MP dosage and genetic ancestry as covariates; *, $P < 0.05$ (Student t test). TGMP thioguanosine monophosphate; TIMP, thioinosine monophosphate.

Figure 3. Association of non-coding NT5C2 variants with erythrocyte TGN in the COG AAL03N1 and St. Jude Total XV cohorts.

Effects of NT5C2 SNP genotype on TGN were shown for two non-coding variants, rs72846714 and rs58700372, selected to represent two independent association signals at this locus. (A) and (B) TGN was significantly lower for patients heterozygous or homozygous for variant allele at rs72846714 in the COG AALL03N1 cohort (A) and the St. Jude Total XV cohort (B), compared with reference allele. Genotype-TGN association was presented in a similar format in (C) and (D) for rs58700372 in the COG AALL03N1 cohort (C) and the St. Jude Total XV cohort (D). P value was estimated using linear regression model with dosage and genetic ancestry as covariates. Each box included data between 25th and 75th percentiles, with horizontal line indicating median. Whiskers indicated maximal and minimal observations within 1.5x length of box.

589 **Figure 4. Functional characterization of *NT5C2* non-coding variant rs72846714**
590 **and its LD SNPs.**

591 A total of six SNPs were presented including rs72846714 and five SNPs that are in
592 strong LD with the index variant (ALL cohorts, $r^2 > 0.6$) and also annotated to putative
593 regulatory DNA. (A) Functional annotation of regulatory DNA at the *NT5C2* locus.
594 Genomic positions, scale and gene structure for GRCh37/hg19 were shown on the
595 top. rs72846714 and its LD SNPs were marked in the middle panel. ATAC-seq
596 signals of hematopoietic cells, DNase and TF ChIP clusters were also included, using
597 publicly available datasets [42-44]. (B) Enhancer activity of cis-elements
598 encompassing these SNPs was tested using luciferase reporter assay in K562 cells.
599 Relative luciferase activity indicated the ratio between the value of variant allele over
600 that from the reference allele. All experiments were performed in triplicate and
601 repeated three times. Statistical significance was evaluated by using two-sided
602 Student's *t* test. Error bars indicate SD. ns, not significant; *, $P < 0.05$; **, $P < 0.01$.

603 **Figure 5. Functional characterization of *NT5C2* non-coding variant rs58700372.**

604 (A) Functional annotation of regulatory DNA within the genomic region encompassing
605 rs58700372. Genomic positions, scale and gene structure for GRCh37/hg19 were
606 shown on the top. rs58700372 was marked in the middle panel. ATAC-seq signals of
607 hematopoietic cells, DNase and TF ChIP clusters were also included, using publicly
608 available datasets [42-44]. (B) Luciferase assay was performed to determine
609 enhancer activity of the DNA segment encompassing rs58700372 in K562 cells.
610 Relative luciferase activity indicated the ratio between the value of the variant allele

611 and that from the reference allele. This experiment was performed in triplicate and
 612 repeated three times. Statistical significance was evaluated by using two-sided
 613 Student's *t* test. Error bars indicate SD; **, $P < 0.01$.

614

615 Supplementary Information Titles

616 Figure S1.

617 Figure S2.

618 Figure S3.

619 Figure S4.

620 Figure S5.

621 Figure S6.

622 Table S1. Primers used in this study.

623 Table S2. TGN significantly associated *NT5C2* SNPs in COG AALL03N1 cohort.

624

Table 1. Patient characteristics of the COG AALL03N1 and the St. Jude Total XV cohorts, used for discovery and validation analyses, respectively.

	AALL03N1 (discovery cohort)	St. Jude Total XV (Validation cohort)
Age at diagnosis, median (range), y	5.0 (1-19)	5.8 (1.0-18.9)
Sex		
Male (N=)	399 (67.9%)	209 (56.2%)
Female (N=)	189 (32.1%)	163 (43.8%)
Genetically determined race/ethnicity		
European (N=)	189 (32.2%)	250 (67.2%)
African (N=)	97 (16.5%)	47 (12.6%)
Hispanic (N=)	186 (31.6%)	56 (15.1%)
Other (N=)	116 (19.7%)	19 (5.1%)
6-MP dosage (mg/m ² /day), median (range)	65.8 (12.7-159.9)	63.2 (17.9-119.7)

Figure 1. Discovery of *NT5C2* variants associated with 6-MP metabolism in children with ALL. Targeted sequencing coupled with SNP imputation in the discovery COG AALL03N1 cohort (588 patients) identified 782 variants at the *NT5C2* locus: three missense variants and 779 non-coding variants. The association of *NT5C2* SNP genotype with erythrocyte TGN was tested using a linear model with genetic ancestry as covariate. This analysis was restricted to patients with wildtype genotype for *TPMT* and *NUDT15*. To identify novel TGN associated variants, we also performed multivariate analysis conditioning on index variant rs72846714 identified from previous GWAS [34]. Validation was performed in the independent St. Jude Total XV cohort (372 patients), following the same analysis strategy. #non-coding variants with minor allele frequency < 1% were excluded from the association analyses due to statistical consideration. * All 22 variants are in linkage disequilibrium with pair-wise LD $r^2 > 0.99$.

A

Location (Chr10)	Ref allele	Var allele	Observed MAF	TGN significance	Amino acid substitution	gnomAD MAF	SIFT	PolyPhen2
104934709	T	C	0.093	$P > 0.05^{\#}$	3 Thr > Ala	0.118	Tolerated	Benign
104934672	A	G	0.00073	NA	15 Met > Thr	0.0000199	Tolerated	Benign
104849533	T	C	0.00073	NA	528 Ile > Val	0.000053	Tolerated	Benign

B

C

Figure 2. Functional characterization of *NT5C2* missense variants. (A) Targeted sequencing of *NT5C2* in the COG AALL03N1 cohort identified three missense variant. Chromosome location was aligned to GRCh37/hg19. SIFT and PolyPhen2 prediction tools were used to predict mutation function. MAF, minor allele frequency. (B) and (C) 5'-nucleotidase activity was determined of wildtype and variant *NT5C2*, with thiopurine metabolites as substrates. Purified *NT5C2* proteins were incubated with TIMP or TGMP, and 5'-nucleotidase activity was measured by quantifying the release of free phosphate using the Malechite Green assay. Bars represent the mean values; the error bars represent the SD from triplicate. Ref allele, reference allele; Var allele, variant allele; #, P value of association of genotype and TGN was analyzed using a linear mixed-effects model with genetic ancestry, time point of TGN measurement and prescribed 6-MP dose recorded monthly prior to TGN measurement as covariates. *, $P < 0.05$ (Student t test). TGMP thioguanosine monophosphate; TIMP, thioinosine monophosphate.

Figure 3. Association of non-coding *NT5C2* variants with erythrocyte TGN in the COG AAL03N1 and St. Jude Total XV cohorts. Effects of *NT5C2* SNP genotype on TGN were shown for two non-coding variants, rs72846714 and rs58700372, selected to represent two independent association signals at this locus. (A) and (B) TGN was significantly lower for patients heterozygous or homozygous for variant allele at rs72846714 in the COG AALL03N1 cohort (A) and the St. Jude Total XV cohort (B), compared with reference allele. Genotype-TGN association was presented in a similar format in (C) and (D) for rs58700372 in the COG AALL03N1 cohort (C) and the St. Jude Total XV cohort (D). *P* value was estimated using linear model as described. Each box included data between 25th and 75th percentiles, with horizontal line indicating median. Whiskers indicated maximal and minimal observations within 1.5x length of box.

Figure 4. Functional characterization of *NT5C2* non-coding variant rs72846714 and its LD SNPs. A total of six SNPs were presented including rs72846714 and five SNPs that were in strong LD with the index variant (ALL cohorts, $r^2 > 0.6$) and also annotated to putative regulatory DNA. (A) Functional annotation of regulatory DNA at the *NT5C2* locus. Genomic positions, scale and gene structure for GRCh37/hg19 were shown on the top. rs72846714 and its LD SNPs were marked in the middle panel. ATAC-seq signals of hematopoietic cells, DNase and TF ChIP clusters were also included, using publicly available datasets [40-42]. (B) Enhancer activity of cis-elements encompassing these SNPs was tested using luciferase reporter assay in K562 cells. Relative luciferase activity indicated the ratio between the value of variant allele over that from the reference allele. All experiments were performed in triplicate, and repeated three times. Statistical significance was evaluated by using two-sided Student's *t* test. Error bars indicate SD. ns, not significant; *, $P < 0.05$; **, $P < 0.01$.

Figure 5. Functional characterization of *NT5C2* non-coding variant rs58700372. (A) Functional annotation of regulatory DNA within the genomic region encompassing rs58700372. Genomic positions, scale and gene structure for GRCh37/hg19 were shown on the top. rs58700372 was marked in the middle panel. ATAC-seq signals of hematopoietic cells, DNase and TF ChIP clusters were also included, using publicly available datasets [40-42]. (B) Luciferase assay was performed to determine enhancer activity of the DNA segment encompassing rs58700372 in K562 cells. Relative luciferase activity indicated the ratio between the value of the variant allele and that from the reference allele. This experiment was performed in triplicate, and repeated three times. Statistical significance was evaluated by using two-sided Student's *t* test. Error bars indicate SD. ; **, $P < 0.01$.