numbers of tandem repeats (VNTRs) in exon 5 of *MICA* were identified to affect MICA subcellular localization and serum MICA level [14]. The exon 5 of *MICA* encodes the transmembrane domain and the insertion of an extra G nucleotide in the domain would result in a premature stop codon that would generate MICA protein without a transmembrane domain and subsequently affect sMICA level [14]. However, our previous results indicated that MICA VNTR was not significantly associated with the sMICA level or HCC risk [6]. Therefore, in the current study, we have tried to investigate whether the *MICA* variations would affect the *MICA* transcription in the liver cancer cells. Through the functional analysis of genetic variations in the *MICA* promoter region, we here report a causative SNP rs2596538 that increases the binding affinity of the transcription factor Specificity Protein 1 (SP1) and the risk of progression of the disease.

Materials and Methods

Samples and genotyping

DNA samples for direct sequencing (50 HCV-related HCC cases), imputation analysis (721 HCV-related HCC cases and 5,486 HCV-negative controls), and serum samples for sMICA ELISA (246 HCV-related HCC) were obtained from BioBank Japan [15,16]. Genotyping of SNPs from 1,394 HCC patients and measurement of sMICA expression by ELISA were performed in the previous study [6]. Genotyping of SNP rs2596542 in 1,043 CHC was performed previously in RIKEN using Illumina HumanHap610-Quad BeadChip [17]. All CHC subjects had abnormal levels of serum alanine transaminase for more than 6 months and were positive for both HCV antibody and serum HCV RNA. The SNP rs2596542 in liver cirrhosis samples without hepatocellular carcinoma from BioBank Japan (n = 420) and the University of Tokyo (n = 166) were genotyped using Illumina HumanHap610-Quad BeadChip or invader assay [18]. All subjects were either subjected to liver biopsy or diagnosed by non-invasive methods including hepatic imaging, biochemical data, and the presence/absence of clinical manifestations of portal hypertension [18]. The samples used in the current project were listed in Table S1. Case samples with HBV co-infection were excluded from this study. The subjects with cancers, chronic hepatitis B, diabetes or tuberculosis were excluded from non-HCV controls. All subjects were Japanese origin and provided written informed consent. This research project was approved by the ethical committees of the University of Tokyo and RIKEN.

Imputation study

The imputation study was performed by using a hidden Markov model programmed in MACH [19] and haplotype information from 1000 genomes database [20]. The imputation results were confirmed by direct DNA sequencing in 50 randomly selected samples.

Cell culture

Human liver cancer cell lines HLE and HepG2 were purchased from JHSF (Osaka, Japan) and ATCC. These cells were grown in Dulbecco's modified Eagle's medium (Invitrogen) with 10% fetal bovine serum. Cells were cultured at 37°C with 5% CO_2.

EMSA

HLE cells were grown in 15 cm culture plate until they reached 95% confluency. The plate was then sealed with parafilm and immersed in a water bath at 42.5°C for 1.5 hours [21]. Nuclear extracts from these cells were prepared according to the standard protocol [22]. EMSA was carried out using DIG Gel Shift Kit, 2nd Generation (Roche) according to the manufacturer's instructions. The sequences of the 12 probes were listed in the Table S2. In brief, 30 fmol of labeled probes were hybridized with 5 μg nuclear extract for 15 minutes at room temperature. The mixtures were then loaded into a 6% TBE gel, separated by electrophoresis at 4°C and transferred onto a nylon membrane. The membrane was then hybridized with anti-digoxigenin-AP antibody and developed by CSPD solution. For competition study, nuclear extracts were incubated with non-labeled oligonucleotides first before adding labeled probe. For supershift assay, SP1 antibody (SC-59X, Santa Cruz Biotechnology) was added into the nuclear extract and incubated on ice for 30 minutes first before adding labeled probe. The mixtures were then separated by electrophoresis using 4% TBE gel. All EMSAs were repeated twice for reconfirmation of the results.

ChIP

The HLE cells (G allele homozygote) and HepG2 cells (heterozygote) were used in the ChIP assay. The plasmid pCAGGS-SP1 was transfected into both cells by using FuGENE6 Transfection Reagent (Roche). The ChIP assays were carried out using Chromatin Immunoprecipitation Assay Kit (Millipore) according to the manufacturer's protocol. In brief, the cells were treated with formaldehyde to crosslink DNA-protein complexes at 48 hours post-transfection. DNA-protein complexes were then sheared by sonication and immunoprecipitated by rabbit polyclonal anti-SP1 antibody (SC-59X, Santa Cruz Biotechnology). The resulting DNAs were analyzed by PCR (Table S2). In order to determine the binding specificity of SP1 to the SNP rs2596538 allele, the PCR products from HepG2 cells were further subcloned into pCR 2.1 vector and sequenced to assess G to A ratio in both input DNA and immunoprecipitant.

Dual luciferase reporter assay

Three copies of 31 bp DNA fragments equivalent to the EMSA oligonucleotides of SNP rs2596538 were cloned into pGL3-promoter vector (Promega). The plasmids were co-transfected with pCAGGS-SP1 and pRL-TK plasmids (Promega) into HLE cells by FuGENE6 Transfection Reagent (Roche). The pCAGGS-SP1 plasmid provided the expression of transcription factor SP1, and pRL-TK plasmid served as internal control for transfection efficiency [23]. The cells were lysed at 48 hours post-transfection, and relative luciferase activities were measured by Dual Luciferase Assay System (Toyo B-Net).

Western blotting

Cancer cell lysates were prepared by using pre-chilled RIPA buffer, and 25 μg of each lysate was loaded into the gel and separated by SDS-PAGE. Western blotting was performed according to the standard protocol. Rabbit anti-MICA antibody (ab63709, abcam: 1/1000) and rabbit anti-SP1 antibody (17-601, Upstate Biotechnology: 1/500) were used in the experiment.

Statistical analysis

The case-control association was analyzed by Student's *t*-test and Fisher's exact test as appropriate. The association of allele dependent sMICA expression was studied by Kruskal-Wallis test using R statistical environment version 2.8.1. The LD and coefficients (D' and r^2) were calculated by Haploview version 4.2 [24].

Table 1. Association of rs2596542 with the progression from CHC to LC and HCC.

	Case MAF	Control MAF	P*	OR	95% C.I.
LC vs CHC	0.3797	0.3442	0.04842	1.166	1.01–1.35
HCC vs LC	0.4012	0.3797	0.20296	1.094	0.95–1.26

MAF, minor allele frequency; OR, odds ratio for minor allele. C.I., confidence interval. SNP rs2596542 was analyzed in 1,043 chronic hepatitis C (CHC), 586 liver cirrhosis without hepatocellular carcinoma (LC) and 1,394 HCV-induce hepaticellular carcinoma patients (HCC). *calculated by Armitage trend test.
doi:10.1371/journal.pone.0061279.t001

Results

Analyses of SNP rs2596542 in HCV-infected patients at different disease stages

Since the development of HCC consists of multiple steps, we investigated the role of SNP rs2596542 with disease progression. SNP rs2596542 was genotyped in patients at three different disease categories of CHC (chronic hepatitis C) without liver cirrhosis (LC) or HCC, LC without HCC, and HCC. The statistical analysis indicated that SNP rs2596542 was significantly associated with disease progression from CHC to LC with P-value of 0.048 and odds ratio of 1.17 (Table 1). The risk allele frequency among HCC patients (40.1%) was higher than that among LC patients (38.0%), but the association was not statistically significant (P-value of 0.203 and odds ratio of 1.09). These results suggested the involvement of *MICA* with both liver fibrosis and hepatocellular carcinogenesis.

HCV-HCC risk is not associated with MICA copy number variation

A previous report has indicated the deletion of the entire *MICA* locus in 3.2% of Japanese population [25] and this deletion was shown to be associated with the risk of nasopharyngeal carcinoma (NPC), especially in male [26]. To identify the functional SNP that may affect *MICA* mRNA expression, we analyzed the relation between the *MICA* copy number variation (CNV) and the HCC susceptibility. We quantified this CNV by real-time PCR in 375 HCV-related HCC patients and 350 HCV-negative controls. As shown in Table S3, we found no difference in the copy numbers between HCC cases and controls, indicating that this CNV is unlikely to be causative genetic variation for the risk of HCC.

Direct sequencing of 5′ flanking region of *MICA*

We then focused on the variations in the 5' flanking region of the *MICA* gene which may be associated with its promoter activity. We had conducted direct DNA sequencing of the 5-kb promoter region which included the marker SNP rs2596542 using genomic DNAs of 50 HCC subjects and identified 11 SNPs showing strong linkage disequilibrium with the marker SNP rs2596542 (D'>0.953 and r^2>0.832) (Fig. S1, Table 2).

Allele specific binding of nuclear protein to genomic region including SNP rs2596538

To investigate whether these genetic variations would affect the binding affinity of some transcription factors, we had conducted the electrophoretic mobility shift assay (EMSA) using the nuclear extract of HLE human hepatocellular carcinoma cells. Since MICA is a stress-inducible protein [21], we first treated the cells with heat shock treatment at 42°C for 90 minutes and confirmed significant induction of MICA expression as shown in Fig. 1a. Then we performed EMSA using 24 labeled-oligonucleotides corresponding to each allele of the 12 candidates' SNPs. The results of EMSA demonstrated that an oligonucleotide corresponding to a G allele of SNP rs2596538 exhibited stronger binding affinity to a nuclear protein(s) than that to an A allele (Fig. 1b). We then confirmed the specific binding of nuclear proteins to the G allele by competitor assay using non-labeled oligonucleotides (Fig. 1c). The self (G allele) oligonucleotides inhibited the formation of DNA-protein complex in a dose-dependent manner, but the non-self (A allele) oligonucleotides showed no inhibition effect. Taken together, some nuclear protein(s) in hepatocellular carcinoma cells would interact with a DNA fragment including the G allele of SNP rs2596538.

Table 2. Linkage disequilibrium between 11 candidate SNPs and SNP rs2596542.

SNP ID	Relative position[a]	A1	A1 frequency	D'	r^2
rs2596542	−4815	A	0.36		
rs2428475	−4788	G	0.36	1	1
rs28366144	−4586	T	0.36	1	1
rs2428474	−4387	G	0.39	1	0.88
rs2251731	−4045	A	0.39	1	0.88
rs2844526	−3703	C	0.38	1	0.918
rs2596541	−3572	A	0.38	1	0.918
rs2523453	−3285	G	0.38	1	0.918
rs2544525	−3259	C	0.38	1	0.918
rs2523452	−2870	G	0.34	0.953	0.832
rs2596538	−2778	A	0.34	0.953	0.832
rs2844522	−2710	C	0.34	0.953	0.832

Note: Direct DNA sequence of 5-kb promoter region of MICA from 50 HCV-HCC subjects. D' and r^2 were calculated by comparing the genotypes of these SNPs to the marker SNP rs2596542 by Haploview. A1, minor allele; [a]Relative position to exon 1 of the *MICA* gene.
doi:10.1371/journal.pone.0061279.t002

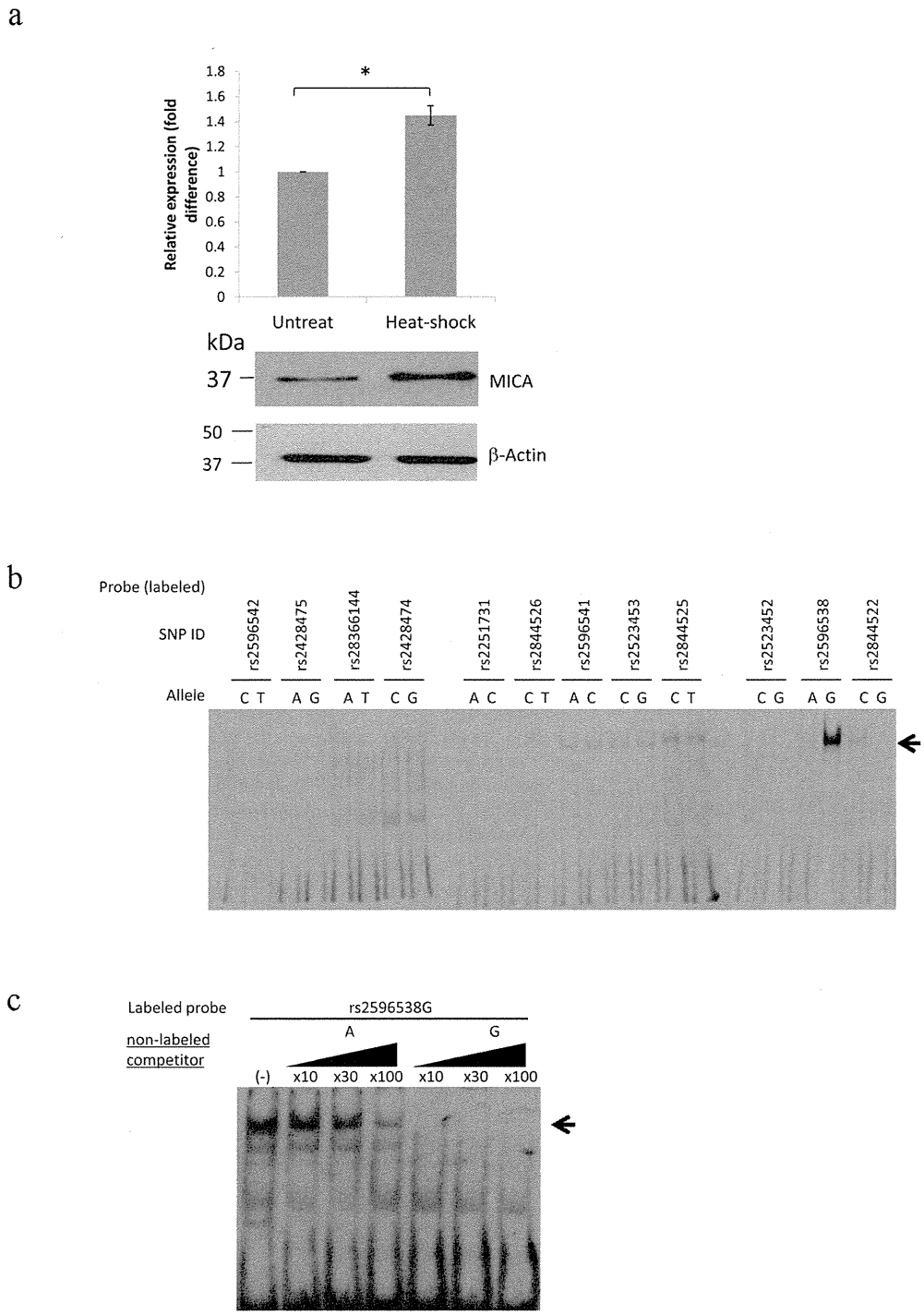

Figure 1. SNP rs2596538 affects the binding affinity of nuclear proteins. (**A**) Real-time quantitative PCR (upper) and Western blotting (lower) of MICA before and after heat shock treatment in HLE cells. *B2M* and β-actin are served as internal and protein loading control. (**B**) EMSA using 31 bp labeled probes flanking each SNP located within the 4.8 kb region upstream of *MICA* transcription start site. A black arrow indicates the shifted band specific to G allele of SNP rs2596538. (**C**) EMSA using the labeled G allele of SNP rs2596538 and nuclear extract from heat treated HLE cells. Non-labeled A or G allele of SNP rs2596538 at different concentrations are used as competitors. Pointed arrow indicates shifted band. *$P<0.05$ by Student's *t*-test.

doi:10.1371/journal.pone.0061279.g001

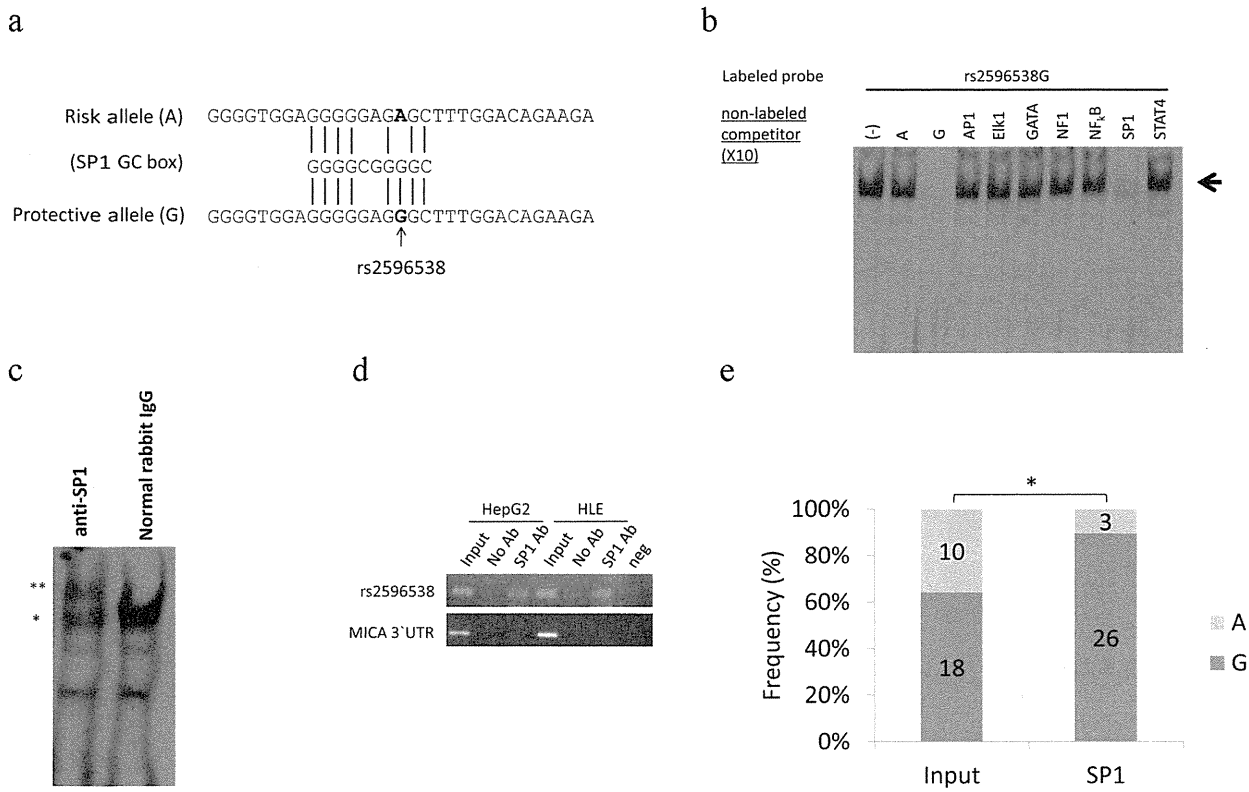

Figure 2. Binding of transcription factor SP1 to G allele of SNP rs2596538. (**A**) Multiple alignment of a GC box and DNA sequence of A or G probe of SNP rs2596538 used in EMSA. (**B**) EMSA using the labeled G allele of SNP rs2596538 and nuclear extract from heat treated HLE cells. Non-labeled consensus oligonucleotides of seven transcription factors are used as competitors. Pointed arrow indicates shifted band. (**C**) EMSA using the labeled G allele of SNP rs2596538 and nuclear extract from heat shock treated HLE cells in the presence of anti-SP1 antibody or normal rabbit IgG. Asterisks on the left side indicate the shifted (*) and super-shifted bands (**). Normal rabbit IgG serves as a negative control. (**D**) ChIP assay using HepG2 and HLE cell lines were ectopically expressed with SP1 protein. DNA-protein complex was immunoprecipitated with anti-SP1 antibody followed by PCR amplification using a primer pair flanking SNP rs2596538. DNAs precipitated without antibody are served as a negative control. PCR primers flanking the 3′ UTR region of *MICA* are served as a negative control. (**E**) Genotype distribution at SNP rs2596538 in PCR fragment amplified from the input genomic DNA and DNA-protein complex immunopurified from HepG2 cells by using anit-SP1 antibody. *$P<0.05$ by Student's *t*-test.
doi:10.1371/journal.pone.0061279.g002

SNP rs2596538 regulates the binding of SP1

Since *in silico* analysis identified a putative GC box in a protective G allele but not in a risk A allele (Fig. 2a), the transcription factor SP1 might preferentially bind to the G allele. Base on this information, we further performed competitor assay using non-labeled oligonucleotides (Table S2) and found that among seven tested oligonucleotides, only SP1-consensus oligonucleotides could effectively inhibit the binding of the nuclear protein(s) to the labeled G allele (Fig. 2b). In addition, we identified that the addition of anti-SP1 antibody caused a supershift of a band corresponding to the DNA-protein complex while control IgG did not cause the band shift (Fig. 2c). This result clearly indicated that the SP1 protein is very likely to be a component of the DNA-protein complex.

Furthermore, we performed chromatin immunoprecipitation (ChIP) assay to confirm the binding of SP1 to this genomic region *in vivo*. We had used two cell lines with different genetic backgrounds at SNP rs2596538 locus: HLE cells carrying the only G allele, while HepG2 cells harboring both A and G alleles. After the introduction of SP1 expression vector (pCAGGS-SP1) into these cell lines, the cell extracts were subjected to ChIP assay using anti-SP1 antibody (Fig. 2d). Subsequent PCR experiments indicated that SP1 bound to a genomic fragment containing the G allele of SNP rs2596538 *in vivo*, while 3′ UTR region of *MICA* (negative control) was not immunoprecipitated with anti-SP1 antibody. To further evaluate the binding ability of SP1 to each allele *in vivo*, we sub-cloned the DNA fragment that amplified from genomic DNA of HepG2 cells before and after immunoprecipitation by anti-SP1 antibody. The subsequent sequencing results showed that 26 out of 29 tested clones contained the G allele, demonstrating the preferential binding of SP1 to the G allele (Fig. 2e).

SP1 over-expression preferentially up-regulates MICA expression at G allele

To further investigate the physiological role of the interaction between SP1 and this genomic region, we performed reporter gene assay. Three copies of 31-bp DNA fragments flanking the candidate functional SNP rs2596538 were subcloned into the multiple cloning sites of the pGL3 promoter vector. The relative luciferase activity of the plasmid including the G allele was significantly higher than that including the A allele (Fig. 3a). Furthermore, over-expression of SP1 in the cells could significantly enhance the luciferase activity of the G-allele vector, while the enhancement of the A-allele vector was relatively modest (Fig. 3a).

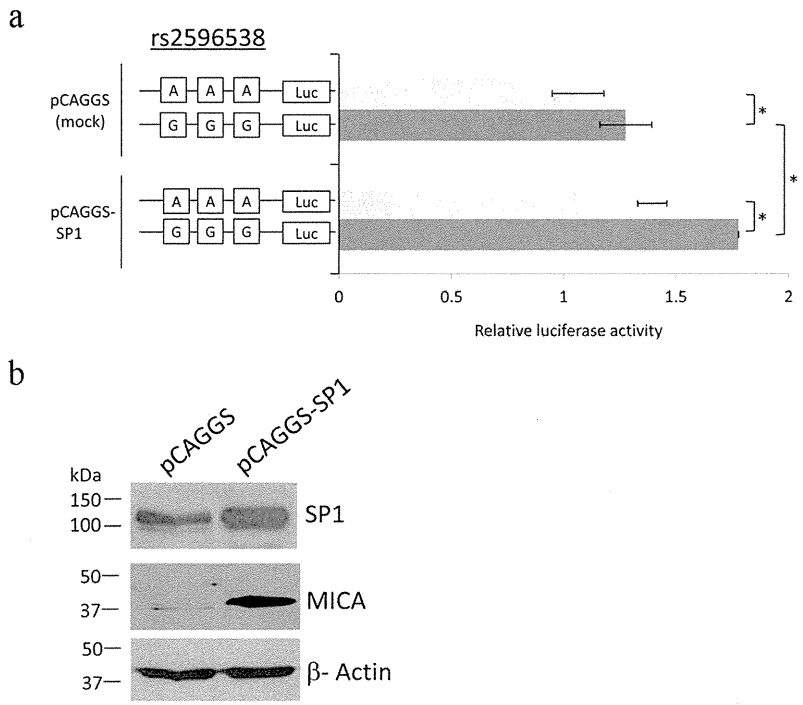

Figure 3. Transcriptional regulation of MICA by SP1 through genomic region including SNP rs2596538. (**A**) Reporter assay using constructs including 3 copies of 31 bp DNA fragment flanking SNP rs2596538. Reporter constructs are transfected into HLE cells with pRL-TK and pCAGGS or pCAGGS-SP1 vector. The value of relative luciferase activity was calculated as the firefly luciferase intensity divided by the renilla luciferase intensity. The data represent the mean ± SD value of 4 independent studies. (*$P<0.05$, Student's *t*-test) (**B**) MICA expression in HLE cells after transfection with pCAGGS or pCAGGS-SP1 vector. β-actin is served as a protein loading control.
doi:10.1371/journal.pone.0061279.g003

We also evaluated the effect of ectopically expressed SP1 on the MICA expression in HLE cells. Western-blot analysis showed that MICA protein expression was significantly increased after the SP1 over-expression (Fig. 3b). These results provided a strong evidence that the G allele has higher transcriptional potential that can be inducible by SP1.

Association of SNP rs2596538 with HCC risk and sMICA level in HCV-induced HCC patients

To further investigate the role of SNP rs2596538 in human carcinogenesis, we investigated the association of SNP rs2596538 with HCV-induced HCC in 721 HCV-HCC cases and 5,486 HCV-negative controls that had been genotyped using Illumina HumanHap610-Quad Genotyping BeadChip in our previous study [6]. We performed imputation analysis by using haplotype data from 1000 genome database [20] and found that an A allele of SNP rs2596538 was considered to be a risk allele for HCV-related HCC (Table 3, odds ratio = 1.343, $P = 1.82\times10^{-5}$). The functional SNP rs2596538 exhibited a stronger association with the HCC risk than the marker SNP rs2596542 (2.46×10^{-5}). We also analyzed the relationship between the SNP rs2596538 and the sMICA level among 246 HCV-induced HCC patients and found a significant association with the P-value of 0.00616 (Fig. 4). These results were concordant with our functional analyses in which the G allele exhibited a higher affinity to SP1 and revealed a higher transcriptional activity.

Table 3. Association of SNP rs2596542 and SNP rs2596538 with HCV-induced HCC.

SNP ID	Relative position[a]	A1	OR	*P* value
rs2596542	−4815	A	1.339	2.46×10^{-5}
rs2596538	−2778	A	1.343	1.82×10^{-5}

Note: Genotype data of 721 HCV-HCC cases and 5,486 HCV-negative controls were imputed using 1000 genomes as reference. A1, risk allele; OR, odds ratio for the risk allele calculated by considering the protective allele as a reference.
[a]Relative position to exon 1 of the *MICA* gene.
doi:10.1371/journal.pone.0061279.t003

Discussion

Approximately 160 million people (2.35% of the worldwide population) are estimated to have HCV infection [27]. Since HCV carriers have an increased risk to develop liver cirrhosis and subsequent HCC [28,29], the prediction of cancer risk is especially important for CHC patients. In our previous study, we have identified that SNP rs2596542 located in the upstream of *MICA* gene was significantly associated with the risk of HCC development among CHC patients as well as the serum level of sMICA [6]. In this study, we found that the genetic variant at SNP rs2596538 strongly affected the binding affinity of SP1. Over-expression of SP1 remarkably induced *MICA* expression in cells carrying the G allele that has a higher affinity to the SP1 binding. These findings are concordant with higher serum sMICA level among HCC patients with the G allele at SNP rs2596538. SP1 is a

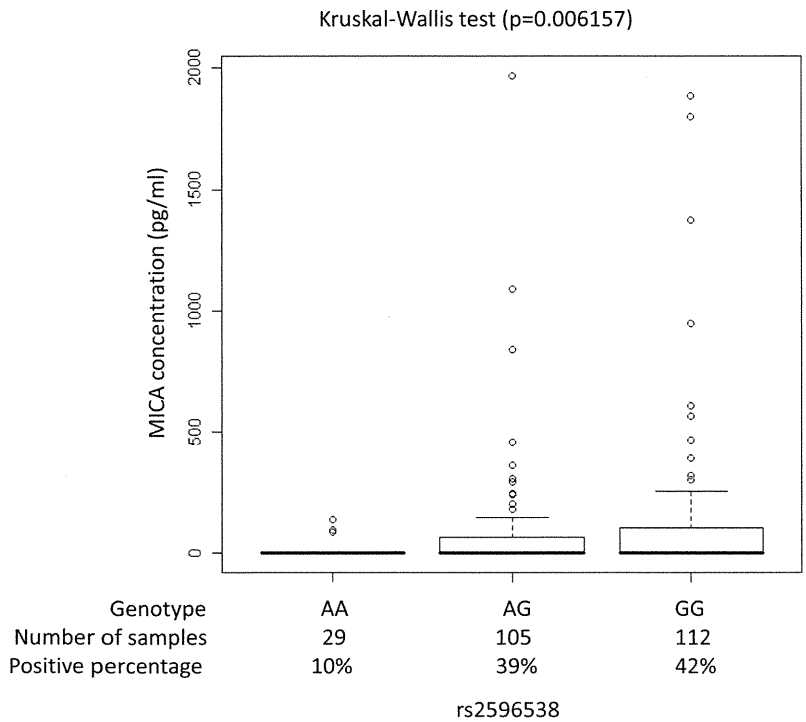

Figure 4. Association between the soluble MICA levels and SNP rs2596538 genotype. The samples were classified into 3 groups according to rs2596538 genotype. The sMICA levels measured by ELISA are indicated in y-axis. The numbers of samples and the proportion of sMICA positive subjects from each group are shown in x-axis. The percentage of the positive sMICA expression in each group are AA = 10%, AG = 39%, and GG = 42%. Statistical significance was determined by Kruskal-Wallis test.
doi:10.1371/journal.pone.0061279.g004

ubiquitously expressed transcription factor which binds to the GC-rich decanucleotide sequence (GC box) and activates the transcription of various viral and cellular genes [30,31]. Phosphorylation of SP1 was shown to be induced by HCV core protein and exhibited higher binding affinity to the promoter region of its downstream targets [32]. From our previous study, we showed a significant difference of sMICA expression between non-HCV individuals and CHC patients. This indicated that sMICA expression was induced after HCV infection [6]. Hence, we here propose the following hypothesis. After HCV infection, the virus core protein enhances the SP1 phosphorylation in hepatocytes, and the phosphorylated SP1 binds to the DNA segment corresponding to the G allele of SNP rs2596538 and then induces *MICA* expression. The membrane-bound MICA (mMICA) serves as a ligand for NKG2D to activate the immune system and results in the elimination of viral-infected cells by NK cells and CD8+ T cells [8,9]. Eventually, HCV-infected individuals with higher MICA level may cause stronger immune response to the infected cells and hence result in a reduced risk for HCC progression. Moreover, the mMICA is then shed by metalloproteinases that are often over-expressed in cancer tissues and convert mMICA to sMICA. This resulted in a significantly increase of sMICA level in the serum of HCV infected patients.

In contrast to HCV-induced HCC, our group had previously identified that higher sMICA level was associated with poor prognosis in HBV-induced HCC patients [33]. Such an opposite effect of *MICA* would be attributable to the difference in downstream pathway between HBV and HCV. HBV virus encodes hepatitis B virus X protein (HBx) that is pathogenic and promotes tumor formation. It had been reported that HBx protein was associated with an elevated expression of MT1-MMP, MMP2, and MMP3 [34,35]. HBx was also shown to transactivate MMP9 through ERKs and PI-3K-AKT/PKB pathway and suppress TIMP1 and TIMP3 activities [36,37]. The activation of metalloproteinases would induce the shedding of mMICA into sMICA, which promotes the tumor formation through the inhibitory effect of sMICA on NK cells. This can explain why high sMICA expression is a marker of poor prognosis for HBV-induced HCC. On the other hand, HCV infection was not associated with metalloproteinases activation, although the expression of sMICA was shown to be proportional to mMICA level. Therefore individuals with high MICA expression are likely to activate natural killer cells and CD8+ T cells to eliminate virus infected cells.

SP1 was previously identified as a transcriptional regulator of both *MICA* and *MICB* [7,9,38]. A polymorphism in the *MICB* promoter region was found to be associated with *MICB* transcription level [7]. To our knowledge, this is the first report showing that *MICA* transcription is directly influenced by functional variant. Moreover, this functional SNP is significantly associated with HCV-induced HCC. Our findings provide an insight that *MICA* genetic variation is a promising prognostic biomarker for CHC patients.

Supporting Information

Figure S1 Pairwise LD map between marker SNP and 11 candidates SNP. Black color boxes represent regions of high pairwise r^2 value. The LD was determined by direct DNA

sequencing of *MICA* promoter region from 50 randomly selected HCV-HCC patients.
(TIF)

Table S1 Characteristics of samples and methods used in this study.
(DOCX)

Table S2 The sequences of each oligo used in the EMSA and ChIP assay.
(DOCX)

Table S3 Copy number variation between HCV-HCC and control samples.
(DOCX)

Author Contributions

Conceived and designed the experiments: PHYL YN KM. Performed the experiments: PHYL YU VK. Analyzed the data: PHYL YU CT. Contributed reagents/materials/analysis tools: KK NK DM KC MK. Wrote the paper: PHYL KM.

References

1. Umemura T, Ichijo T, Yoshizawa K, Tanaka E, Kiyosawa K (2009) Epidemiology of hepatocellular carcinoma in Japan. J Gastroenterol 44 Suppl 19: 102–107.
2. Fassio E (2010) Hepatitis C and hepatocellular carcinoma. Ann Hepatol 9 Suppl: 119–122.
3. Mbarek H, Ochi H, Urabe Y, Kumar V, Kubo M, et al. (2011) A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. Hum Mol Genet 20: 3884–3892.
4. Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, et al. (2009) A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nat Genet 41: 591–595.
5. Zhang H, Zhai Y, Hu Z, Wu C, Qian J, et al. (2010) Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. Nat Genet 42: 755–758.
6. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, et al. (2011) Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nat Genet 43: 455–458.
7. Rodríguez-Rodero S, González S, Rodrigo L, Fernández-Morera JL, Martínez-Borra J, et al. (2007) Transcriptional regulation of MICA and MICB: a novel polymorphism in MICB promoter alters transcriptional regulation by Sp1. Eur J Immunol 37: 1938–1953.
8. Bauer S, Groh V, Wu J, Steinle A, Phillips JH, et al. (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science 285: 727–729.
9. Zhang C, Wang Y, Zhou Z, Zhang J, Tian Z (2009) Sodium butyrate upregulates expression of NKG2D ligand MICA/B in HeLa and HepG2 cell lines and increases their susceptibility to NK lysis. Cancer Immunol Immunother 58: 1275–1285.
10. Sun D, Wang X, Zhang H, Deng L, Zhang Y (2011) MMP9 mediates MICA shedding in human osteosarcomas. Cell Biol Int 35: 569–574.
11. Waldhauer I, Goehlsdorf D, Gieseke F, Weinschenk T, Wittenbrink M, et al. (2008) Tumor-associated MICA is shed by ADAM proteases. Cancer Res 68: 6368–6376.
12. Jinushi M, Takehara T, Tatsumi T, Kanto T, Groh V, et al. (2003) Expression and role of MICA and MICB in human hepatocellular carcinomas and their regulation by retinoic acid. Int J Cancer 104: 354–361.
13. Kohga K, Takehara T, Tatsumi T, Ohkawa K, Miyagi T, et al. (2008) Serum levels of soluble major histocompatibility complex (MHC) class I-related chain A in patients with chronic liver diseases and changes during transcatheter arterial embolization for hepatocellular carcinoma. Cancer Sci 99: 1643–1649.
14. Ota M, Katsuyama Y, Mizuki N, Ando H, Furihata K, et al. (1997) Trinucleotide repeat polymorphism within exon 5 of the MICA gene (MHC class I chain-related gene A): allele frequency data in the nine population groups Japanese, Northern Han, Hui, Uygur, Kazakhstan, Iranian, Saudi Arabian, Greek and Italian. Tissue Antigens 49: 448–454.
15. Nakamura Y (2007) The BioBank Japan Project. Clin Adv Hematol Oncol 5: 696–697.
16. Tanikawa C, Urabe Y, Matsuo K, Kubo M, Takahashi A, et al. (2012) A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. Nat Genet 44: 430–434, S431–432.
17. Miki D, Ochi H, Hayes CN, Abe H, Yoshima T, et al. (2011) Variation in the DEPDC5 locus is associated with progression to hepatocellular carcinoma in chronic hepatitis C virus carriers. Nat Genet 43: 797–800.
18. Urabe Y, Ochi H, Kato N, Kumar V, Takahashi A, et al. (2013) A genome-wide association study of HCV induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at MHC region. J Hepatol.
19. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, et al. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316: 1341–1345.
20. Consortium GP (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.
21. Venkataraman GM, Suciu D, Groh V, Boss JM, Spies T (2007) Promoter region architecture and transcriptional regulation of the genes for the MHC class I-related chain A and B ligands of NKG2D. J Immunol 178: 961–969.
22. Andrews NC, Faller DV (1991) A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res 19: 2499.
23. Hata J, Matsuda K, Ninomiya T, Yonemoto K, Matsushita T, et al. (2007) Functional SNP in an Sp1-binding site of AGTRL1 gene is associated with susceptibility to brain infarction. Hum Mol Genet 16: 630–639.
24. Barrett JC (2009) Haploview: Visualization and analysis of SNP genotype data. Cold Spring Harb Protoc 2009: pdb.ip71.
25. Komatsu-Wakui M, Tokunaga K, Ishikawa Y, Leelayuwat C, Kashiwase K, et al. (2001) Wide distribution of the MICA-MICB null haplotype in East Asians. Tissue Antigens 57: 1–8.
26. Tse KP, Su WH, Yang ML, Cheng HY, Tsang NM, et al. (2011) A gender-specific association of CNV at 6p21.3 with NPC susceptibility. Hum Mol Genet 20: 2889–2896.
27. Negro F, Alberti A (2011) The global health burden of hepatitis C virus infection. Liver Int 31 Suppl 2: 1–3.
28. Cabibbo G, Craxi A (2010) Epidemiology, risk factors and surveillance of hepatocellular carcinoma. Eur Rev Med Pharmacol Sci 14: 352–355.
29. McGlynn KA, London WT (2011) The global epidemiology of hepatocellular carcinoma: present and future. Clin Liver Dis 15: 223–243.
30. Kadonaga JT, Tjian R (1986) Affinity purification of sequence-specific DNA binding proteins. Proc Natl Acad Sci U S A 83: 5889–5893.
31. Suske G (1999) The Sp-family of transcription factors. Gene 238: 291–300.
32. Lee S, Park U, Lee YI (2001) Hepatitis C virus core protein transactivates insulin-like growth factor II gene transcription through acting concurrently on Egr1 and Sp1 sites. Virology 283: 167–177.
33. Kumar V, Yi Lo PH, Sawai H, Kato N, Takahashi A, et al. (2012) Soluble MICA and a MICA Variation as Possible Prognostic Biomarkers for HBV-Induced Hepatocellular Carcinoma. PLoS One 7: e44743.
34. Ou DP, Tao YM, Tang FQ, Yang LY (2007) The hepatitis B virus X protein promotes hepatocellular carcinoma metastasis by upregulation of matrix metalloproteinases. Int J Cancer 120: 1208–1214.
35. Yu FL, Liu HJ, Lee JW, Liao MH, Shih WL (2005) Hepatitis B virus X protein promotes cell migration by inducing matrix metalloproteinase-3. J Hepatol 42: 520–527.
36. Chung TW, Lee YC, Kim CH (2004) Hepatitis B viral HBx induces matrix metalloproteinase-9 gene expression through activation of ERK and PI-3K/AKT pathways: involvement of invasive potential. FASEB J 18: 1123–1125.
37. Kim JR, Kim CH (2004) Association of a high activity of matrix metalloproteinase-9 to low levels of tissue inhibitors of metalloproteinase-1 and -3 in human hepatitis B-viral hepatoma cells. Int J Biochem Cell Biol 36: 2293–2306.
38. Andresen L, Jensen H, Pedersen MT, Hansen KA, Skov S (2007) Molecular regulation of MHC class I chain-related protein A expression after HDAC-inhibitor treatment of Jurkat T cells. J Immunol 179: 8235–8242.

Journal of Human Genetics (2013) 58, 353–361

www.nature.com/jhg

ORIGINAL ARTICLE

Impact of polymorphisms in drug pathway genes on disease-free survival in adults with acute myeloid leukemia

Sook Wah Yee[1], Joel A Mefford[2], Natasha Singh[3], Mary-Elizabeth Percival[3], Adrian Stecula[1], Kuo Yang[1], John S Witte[2], Atsushi Takahashi[4], Michiaki Kubo[4], Koichi Matsuda[5], Kathleen M Giacomini[1,6,7] and Charalambos Andreadis[3,6,7]

Acute myeloid leukemia (AML) is a clinically heterogeneous disease, with a 5-year disease-free survival (DFS) ranging from under 10% to over 70% for distinct groups of patients. At our institution, cytarabine, etoposide and busulfan are used in first or second remission patients treated with a two-step approach to autologous stem cell transplantation (ASCT). In this study, we tested the hypothesis that polymorphisms in the pharmacokinetic and pharmacodynamic pathway genes of these drugs are associated with DFS in AML patients. A total of 1659 variants in 42 genes were analyzed for their association with DFS using a Cox-proportional hazards model. One hundred and fifty-four genetically European patients were used for the primary analysis. An intronic single nucleotide polymorphism (SNP) in *ABCC3* (rs4148405) was associated with a significantly shorter DFS (hazard ratios (HR) = 3.2, $P = 5.6 \times 10^{-6}$) in our primary cohort. In addition, a SNP in the *GSTM1-GSTM5* locus, rs3754446, was significantly associated with a shorter DFS in all patients (HR = 1.8, $P = 0.001$ for 154 European ancestry; HR = 1.7, $P = 0.028$ for 125 non-European patients). Thus, for the first time, genetic variants in drug pathway genes are shown to be associated with DFS in AML patients treated with chemotherapy-based autologous ASCT.
Journal of Human Genetics (2013) **58,** 353–361; doi:10.1038/jhg.2013.38; published online 16 May 2013

Keywords: acute myeloid leukemia; autologous stem cell transplant; busulfan; cytarabine; etoposide; pharmacogenomics

INTRODUCTION

Adult acute myeloid leukemia (AML) is a hematologic malignancy with widely heterogeneous clinical outcomes. New treatments for AML are increasingly being tested in clinical trials of patients with specific tumor cell mutations.[1,2] Although there have been substantial improvements in the number of patients who achieve complete remission, the choice of induction and post-remission therapy for adult AML is still based on the 'one size fits all' principle. Most regimens incorporate antimetabolites (for example, cytarabine, fludarabine), topoisomerase II inhibitors (for example, etoposide, daunorubicin, idarubicin, mitoxantrone) and alkylating agents (for example, busulfan, cyclophosphamide) for the treatment of AML. Prognostic factors for treatment response include age, prior exposure to chemotherapy, cytogenetic markers and expression profiles, and appearance of specific genetic mutations in tumor tissue, such as mutation and translocation of particular genes (for example, *FLT3*, *NPM1*).[3,4] However, these prognostic factors do not adequately capture the wide diversity of clinical outcomes in this disease. The percent of adults with AML who can survive $\geqslant 3$ years and may be cured is ~ 5–70%.[5,6]

One possible explanation for the difference in response to AML treatment is germline genetic variation. Although the pharmacogenomics of AML drug response is an active area of research, there remain large challenges. These include the: (i) poor availability of uniform, well-collected and well-defined drug response phenotype information; (ii) lack of widely available germline DNA not contaminated with tumor cells (myeloblasts); (iii) limited availability of panels of genotype data in large patient cohorts; and (iv) inability to validate findings in replication studies. In this study, we overcome many of these challenges. In particular, to our knowledge, this is the first study that involved the analysis of large numbers of genetic polymorphisms in a cohort of AML patients treated with high-dose chemotherapy followed by autologous stem cell transplantation (ASCT). We present results from testing the association between

[1]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA; [2]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA; [3]Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA; [4]Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan; [5]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan and [6]Institute for Human Genetics, University of California, San Francisco, CA, USA
[7]Co-senior authors.
Correspondence: Dr C Andreadis, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA.
E-mail: candreadis@medicine.ucsf.edu
Received 18 March 2013; accepted 11 April 2013; published online 16 May 2013

germline variants in drug pathway and other genes with disease-free survival (DFS) in adult AML patients. We identified new associations between AML DFS and polymorphisms in several drug pathway genes for cytarabine, etoposide and busulfan, and also replicated single nucleotide polymorphisms (SNPs) previously reported to be associated with the AML response phenotypes.

MATERIALS AND METHODS

Clinical protocol, study criteria and patient cohorts

The population for the current study consists of AML patients who were enrolled in University of California, San Francisco (UCSF) study protocols 9203 or 9303 between 1988 and 2010. The study protocols and patient selection criteria have been previously described.[7–9] This treatment protocol was used in patients with low- and standard-risk *de novo* AML, including acute promyelocytic leukemia in first or second complete remission. It was also used in a small number of patients with high risk AML (that is, with secondary AML) if allogeneic stem cell transplantation was not an option for the patient (for example, unavailable donor).[7] In step 1, patients were treated with consolidation chemotherapy including cytarabine 2000 mg m^{-2} (intravenously) twice daily for 4 days concurrently, with etoposide 40 mg kg^{-1} by intravenous infusion over the 4 days. During the recovery period from chemotherapy, peripheral blood stem cells were collected under granulocyte colony-stimulating factor stimulation. In step 2, patients underwent ASCT, which involved the preparative regimen of busulfan (total dose 16 mg kg^{-1} orally or 12.8 mg kg^{-1} intravenously, over 16 doses in 4 days), followed by etoposide 60 mg kg^{-1} (intravenous bolus) and reinfusion of blood or marrow stem cells. Patients had to be in complete remission for at least 30 days before step 2 (Figure 1). Complete remission was defined as normal bone marrow morphology with <5% blasts, resolution of previously abnormal cytogenetics and no evidence of extramedullary leukemia. In addition, patients must meet criteria for neutrophil and platelet counts, liver and kidney function.[7–9] Detailed procedures of patient enrollment, diagnosis, data collection and follow-up have been previously described.[7–9] Briefly, patients were actively followed-up in the beginning within 6 months of diagnosis, with subsequent annual follow-up by clinic visits. UCSF electronic medical records, the UCSF Blood and Bone Marrow Transplant Clinic database and patients' medical charts were abstracted to determine patients' remission status. The UCSF Committee on Human Research approved the research protocol (institutional review board number 10-00649).

DNA isolation and genotyping

DNA was isolated from peripheral blood stem cells, which were collected during the recovery from step 1 consolidation chemotherapy. As noted in the above section, patients were in complete remission before consolidation chemotherapy, and hence the samples utilized in this step contained <5% leukemic cells. DNA was isolated at the UCSF DNA Banking and Extraction Services Lab. The lab followed standard DNA extraction protocol described in the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). The DNA was then quantified using Picogreen (Life Technologies, Grand Island, NY, USA) and normalized to 50 ng μl^{-1}. For each sample, we genotyped 250 ng of DNA. The Illumina HumanOmniExpress v1.0 Beadchip was used, following the manufacturer's protocols, at the Center for Genomic Medicine, RIKEN, Yokohama, Japan. For quality control of the genotyping, we included one HapMap trio and three duplicates of the DNA samples from the AML patients. A total of 328 distinct DNA samples from patients were genotyped, along with three duplicates and three HapMap samples (trio).

Patients' ancestral origin

The genetic ancestral origin of patients was determined using principal component analysis implemented in Eigenstrat.[10] Genotype information on our 328 AML patients was analyzed in conjunction with SNP data from the HapMap project, which consists of Europeans (CEU and TSI), Asians (JPT and CHB), Africans (TSI and ASW) and Mexican (MXL). From these analyses we were able to distinguish 154 patients of European ancestry on the basis of their close clustering with the European HapMap samples.

Pathway, gene and SNP selection

A total of 42 genes were selected for analysis based on the following criteria: (i) genes in the pharmacokinetic and/or pharmacodynamics pathway of the drugs administered (cytarabine, etoposide and busulfan);[11–15] (ii) genes described in literature as having significant associations with drug cytotoxicity in lymphoblastoid cell lines (LCL);[16,17] (iii) genes involved in DNA mismatch repair;[18] and (iv) genes found previously to be associated with the AML response phenotype[19] (see Supplementary Table 1, Figure 1). After filtering the SNPs with low call rates (<90%) and SNPs with low minor allele frequencies (MAF; MAF<1%) in the 154 European ancestry patients, we selected the SNPs in the candidate genes and within 25 000 bp upstream and downstream flanking regions.

Statistical analysis of the associations

The primary analysis was to estimate the association between SNPs in the selected candidate genes and DFS in 154 patients of European ancestry. A Cox-proportional hazard model was used to estimate the hazard ratios (HR) and 95% confidence limits for the effect of genotype on DFS. An additive coding of genotypes was used in all analyses. The genetic effect estimates were adjusted for levels of a clinical risk score (see Table 1). The SNP associations with $P<0.01$ were also tested in the 125 non-European patients using a Cox-proportional hazard model. In light of the heterogeneous ethnicity of the samples, the genetic effect estimates were adjusted for the first 10 principal components calculated from the genome-wide association study data and the clinical risk score. The Cox-proportional hazards function from the R-project (version 2.15.1) was used. We used 3×10^{-5} (0.05/1659) as the significance level after Bonferroni correction for multiple testing.

Fine mapping of associations *via* imputation

In order to further clarify the association signals, we performed imputation on genes with $P<0.005$ in the primary analysis. For eight genes (*ABCC3*, *DCK*, *GSTM1*, *GSTT1*, *MSH3*, *RRM1*, *SLC22A12*, and *SLC28A3*), the genotypes at polymorphic sites known from the 1000 Genomes Project but not observed on the Illumina HumanOmniExpress v1.0 Beadchip were imputed using IMPUTE version 2 (version 2.3.0 for Mac OS X, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html). The reference panel for the imputation was the 1000 Genomes Phase I-integrated variant set referenced to NCBI build b37 (March 2012 release, retrieved January 20 2013 from http://mathgen.stats.ox.ac.uk/impute/impute_v2.html#reference). Imputed variants with imputation quality scores<0.3 or MAF<0.01 were excluded. The remaining imputed SNPs were each used in Cox-proportional hazards models to predict DFS, as were the genotyped SNPs.

Functional studies

The potential functional effects of the SNPs with $P<0.01$ associated with DFS in European population were examined using the following steps:

(1) All tag-SNPs in linkage disequilibrium to the SNPs with $P<0.01$ in our primary analysis were identified using the Proxy Search in the Broad Institute SNAP (SNP Annotation and Proxy Search; version 2.2), http://www.broadinstitute.org/mpg/snap/ldsearch.php. The search options used in this step were: SNP data set = 1000 Genomes Pilot 1 in CEU population panel, r^2 threshold = 0.8 and distance limit = 500 kbp.

(2) Potential regulatory functions were identified by searching the following databases: Regulome database[20] and eQTL browser (http://eqtl.uchicago.edu, http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/).

(3) Literature searches were used to identify previous reports of associations with drug response phenotypes of any of the SNPs associated with AML response in this study.

RESULTS

This study investigates the potential associations between variants in 42 candidate genes and DFS in adult AML patients treated with a two-step approach to ASCT. High-dose cytarabine, etoposide and busulfan were used in this treatment approach. Table 1 describes

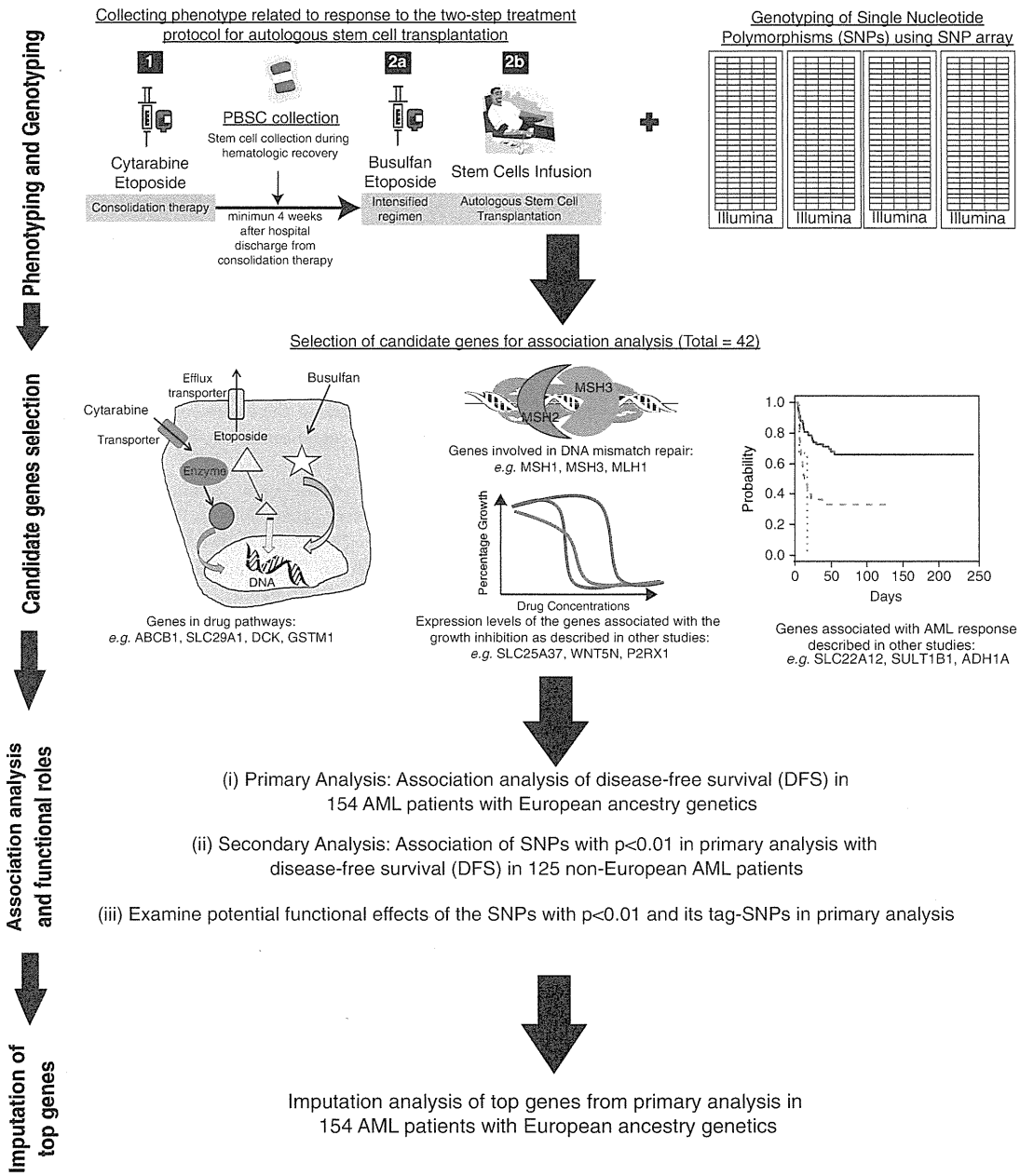

Figure 1 Schematic of workflow applied to determine the association of genetic variations in 42 candidate genes with disease-free survival (DFS) in acute myeloid leukemia (AML) patients treated with a two-step treatment protocol before autologous stem cell transplantation (ASCT). The workflow begins with phenotype data collection and genotyping of DNA samples using a genome-wide SNP array. Candidate genes were selected on the basis of their roles in the: drugs' pharmacokinetic/pharmacodynamics pathway; DNA mismatch repair mechanism; association with the drug cytotoxicity in lymphoblastoic cell lines previously identified in literatures; and pharmacogenomics studies of AML drug response. After association of each SNPs with DFS in the 154 AML patients of European ancestry, the SNPs with $P<0.01$ and their tag-SNPs were examined for their potential functional roles using databases to identify eQTL SNPs and predicted regulatory elements, such as binding sites of transcription factors and for their replications of previous studies. The SNPs with $P<0.01$ were also examined for their associations with DFS in 125 non-European ancestry. Finally, imputation was performed to identify other SNPs with stronger associations with DFS in European ancestry.

the demographic characteristics of this cohort, which consisted of 154 patients of European ancestry and 125 patients of non-European ancestry (African, Asian and Mexican). In the DFS analysis, 55 patients of European ancestry (35.7%) and 47 of non-European ancestry (37.6%) relapsed during the observation period, which extended from 1988 to 2010.

We applied quality control criteria to the SNP data, with a genotype call rate of 0.99 and minor allele frequency $\geqslant 0.01$. After filtering, a total of 42 genes covered by 1659 SNPs were included in this association analysis. Furthermore, none of the DNA samples showed chromosomal abnormality by GenomeStudio (Illumina). Results in Figure 2 and Table 2 showed that among the 40 SNPs

with $P<0.01$, the SNPs in the pharmacokinetic/pharmacodynamic pathway genes have stronger associations compared with SNPs not in the drug pathway. After Bonferroni correction for multiple testing (1659 tests), a SNP in the first intron of *ABCC3* (rs4148405) was significantly associated with DFS, with the minor allele (G) associated with shorter time to relapse (P-unadjusted $= 9.5 \times 10^{-6}$, Figure 3a). Although other SNPs did not reach significance after Bonferroni correction, overall there were 23 SNPs associated with DFS at $P<0.005$ (Table 2). These 23 SNPs are in or within 25 000 bp of eight genes: *SLC28A3*, *DCK*, *RRM1*, *GSTM1*, *ABCC3*, *MSH3*, *GSTT1*, or *SLC22A12*. The majority of the minor alleles were associated with poor outcome (shorter DFS). Kaplan–Meier estimate plots of DFS are shown in Figures 3a–d for four of the top SNPs with MAF $\geqslant 3\%$. Other SNPs in the genes of the cytarabine pathway including *NT5C2* and *RRM2B* were also associated with DFS but with weaker P-values ($P<0.01$, Table 2). Interestingly, SNPs in three out of eight selected genes (*SLC25A37*, *WNT5N* and *P2RX1*), for which expression levels have previously been correlated with either etoposide or cytarabine IC_{50} values in LCL, showed significant but weaker association ($P<0.01$) compared with genes in the drug pathways. Next, we examined the 40 SNPs in patients of non-European ancestry. Only one SNP, in *GSTM1-GSTM5* locus (rs3754446), was significantly associated with DFS in patients of non-European ancestry (HR = 1.7, $P=0.028$). Overall, in the entire cohort the minor allele of the SNP (rs3754446) was significantly associated with shorter DFS (HR = 1.7, $P=0.00027$).

Imputations of the eight candidate genes were performed to determine whether other SNPs in the regions have stronger association with DFS. The results showed that there are 234 imputed SNPs with $P<0.01$ (MAF $\geqslant 1\%$), and among these there are 93 SNPs with improved P-values compared with the genotyped SNPs (Figure 4). Several SNPs in deoxycytidine kinase (*DCK*) have significant P-values $< 1.0 \times 10^{-4}$ (Figure 4), with MAF 5–10%. Although none of these 93 SNPs are in exonic regions, on examination in the eQTL browser, Regulome database and GTEx eQTL browser, we determined that SNPs in GSTM1 (rs929166, rs11101989) and MSH3 (6151896) are associated with their respective gene expression levels in liver[21] or lymphoblastoid cells[22] (data not shown). In the *ABCC3* and *SLC28A3* regions, imputation analysis did not identify other more significant SNPs in addition to the most significant genotyped SNPs, rs4148405 and rs11140500, respectively (Figure 4).

Using *in silico* analysis, we determined whether the 40 SNPs and their tag-SNPs were in known or predicted functional regions of the genome. Several of the SNPs were in DNA regions predicted to have binding sites for transcription factors. Some of these regions appear in

Table 1 Demographic characteristics of patients undergoing autologous stem cells transplantation for AML from 1988 to 2010

Variables	*Patients of European ancestry (N = 154)*	*Patients of non-European ancestry (N = 125)*
Sex, N (%)		
Male	76 (49.3%)	67 (53.6%)
Female	78 (50.6%)	58 (46.4%)
Age (years)		
Median (s.d., range)		
Age at diagnosis	47.0[a] (13.0, 18–72)	40.5[b] (12.8, 17–68)
Age at autologous bone marrow transplantation	47.0 (13.0, 19–72)	41.0 (12.8, 19–69)
Year of transplantation		
1988–1995	36 (23.4%)	38 (30.4%)
1996–2000	41 (26.6%)	31 (24.8%)
2001–2010	77 (50.0%)	56 (44.8%)
De novo/secondary AML, N (%)		
De novo	147 (95.5%)	110 (88.0%)
Secondary	7 (4.5%)	15 (12.0%)
Risk, N (%)[c]		
Acute promyeloid leukemia	15 (9.7%)	12 (9.6%)
Low	18 (11.7%)	18 (14.4%)
Standard	114 (74.0%)	80 (64.0%)
High	7 (4.5%)	15 (12.0%)
DFS (months)		
Median (s.d.)	21.4 (43.9)	16.7 (44.8)

Abbreviations: AML, acute myeloid leukemia; DFS, disease-free survival.
[a]$N=152$.
[b]$N=124$.
[c]Defined by the Cancer and Leukemia Group B (CALGB) criteria.

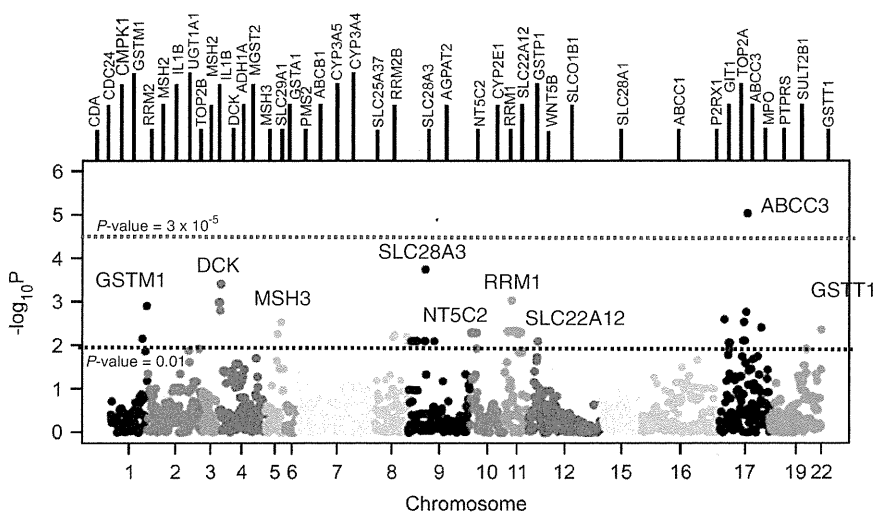

Figure 2 Plot showing the significance ($-\log10$ of the P-value) of associations of the 1659 SNPs with DFS in 154 acute myeloid leukemia (AML) patients of European ancestries. Only SNPs with minor allele frequencies (MAF) of $\geqslant 1\%$ in the selected 42 candidate genes are shown. Each dot represents a SNP. SNPs above the black dotted line are SNPs with $P<0.01$, and the SNP above the red dotted line reached a P-value, which was significant after correction for multiple testing ($P<3 \times 10^{-5}$).

Table 2 Significant SNPs (P<0.01) associated with DFS in 154 AML patients of European ancestry

SNP	Chromosome	Gene	Feature	5′ flanking gene/ 3′ Flanking gene	Risk allele	MAF	P (unadjusted)	HR	(95% CI) of the HR	Genotype counts	Minor allele	Major allele	Reason for gene selection
rs4148405	17	ABCC3	Intron	CACNA1G/ANKRD40	G	0.12	9.45E-06	3.12	(1.88–5.15)	2/34/117	G	T	Drug pathway
rs11140500	9	SLC28A3	Intron	RMI1/NTRK2	T	0.01	0.00018	9.94	(2.99–33.09)	0/3/151	T	C	Drug pathway
rs10805074	4	DCK	Intron	MOBKL1A/ LOC100128311	A	0.03	0.00039	4.59	(1.98–10.67)	0/9/145	A	G	Drug pathway
rs7684954	4	DCK	Intron	LOC100128311/ LOC727995	A	0.03	0.00039	4.59	(1.98–10.67)	0/9/145	A	G	Drug pathway
rs4593998	11	RRM1	Intron	RRM1/LOC643244	A	0.14	0.00093	2.28	(1.40–3.71)	4/34/116	A	G	Drug pathway
rs6842838	4	DCK	3′UTR of MOB1B	MOB1B/DCK	G	0.05	0.0010	3.21	(1.60–6.45)	0/15/139	G	T	Drug pathway
rs1385985	4	DCK	3′UTR of MOB1B	GRSF1/DCK	C	0.05	0.0010	3.21	(1.60–6.45)	0/15/139	C	T	Drug pathway
rs3754446	1	GSTM1	Near-gene-5 [GSTM5]	GSTM1/GSTM5	G	0.38	0.0012	1.81	(1.26–2.59)	26/66/62	G	T	Drug pathway
rs7689093	4	DCK	Near-gene-5 [DCK]	MOBKL1A/DCK	G	0.01	0.0016	5.34	1.89–15.13)	0/4/150	G	A	Drug pathway
rs1989983	17	ABCC3	Near-gene-5 [ABCC3]	CACNA1G/ABCC3	A	0.11	0.0017	2.33	(1.37–3.96)	2/30/122	A	G	Drug pathway
rs2301835	17	ABCC3	Synonymous variant in coding of CACNA1G	CACNA1G/ABCC3	T	0.06	0.0029	2.60	(1.39–4.89)	1/16/137	T	C	Drug pathway
rs12515548	5	MSH3	Intron	LOC100128458/ RASGRF2	T	0.13	0.0029	2.07	(1.28–3.35)	6/27/121	T	C	DNA mismatch repair genes
rs2277624	17	ABCC3	Synonymous variant	CACNA1G/ANKRD40	A	0.25	0.0040	1.75	(1.20–2.55)	12/54/87	A	G	Drug pathway
rs11090305	22	GSTT1	Near-gene-5 [CABIN1]	GSTTP2/CABIN1	C	0.18	0.0044	1.98	(1.24–3.17)	3/50/101	C	T	Drug pathway
rs7130539	11	RRM1	Intron	STIM1/OR55B1P	C	0.06	0.0047	2.56	(1.33–4.91)	1/16/137	C	T	Drug pathway
rs11031136	11	RRM1	Intergenic	OR55B1P/LOC643244	G	0.06	0.0047	2.56	(1.33–4.91)	1/16/137	G	T	Drug pathway
rs528211	11	SLC22A12	Intergenic	SLC22A11/SLC22A12	G	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	A	G	Other AML study
rs2360872	11	SLC22A12	Intergenic	SLC22A11/SLC22A12	C	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	T	C	Other AML study
rs505802	11	SLC22A12	Near-gene-5 [SLC22A12]	SLC22A11/SLC22A12	A	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	G	A	Other AML study
rs524023	11	SLC22A12	Near-gene-5 [SLC22A12]	SLC22A11/SLC22A12	G	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	A	G	Other AML study
rs9734313	11	SLC22A12	5′UTR	SLC22A11/SLC22A12	T	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	C	T	Other AML study
rs11231825	11	SLC22A12	Synonymous variant	SLC22A11/SLC22A12	C	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	T	C	Other AML study
rs2268166	11	RRM1	Intron	STIM1/OR55B1P	G	0.06	0.0049	2.55	(0.20–0.75)	1/16/136	G	T	Drug pathway
rs11606370	11	SLC22A12	Intron of NRXN2	SLC22A12/RASGRP2	A	0.30	0.0050	0.51	(0.32–0.82)	14/64/76	C	A	Other AML study
rs11191547	10	NT5C2	Intergenic	CNNM2/NT5C2	T	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	T	C	Drug pathway
rs11191549	10	NT5C2	Near-gene-3 [NT5C2]	CNNM2/NT5C2	T	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	T	C	Drug pathway
rs11191553	10	NT5C2	Intron	CNNM2/ LOC100128863	T	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	T	G	Drug pathway
rs10883836	10	NT5C2	Intron	LOC100128863/ LOC729081	C	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	C	T	Drug pathway
rs7095304	10	NT5C2	Intergenic	NT5C2/LOC401648	A	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	A	G	Drug pathway
rs6151816	5	MSH3	Intron	LOC100128458/ RASGRF2	T	0.12	0.0055	2.05	(1.24–3.41)	5/27/122	T	C	DNA mismatch repair genes
rs893006	11	SLC22A12	Intron	SLC22A11/NRXN2	T	0.28	0.0055	0.51	(0.32–0.82)	13/61/80	G	T	Other AML study
rs7818607	8	SLC25A37	Intergenic	SLC25A37/LOC646721	A	0.30	0.0057	1.75	(1.18–2.59)	16/59/79	A	C	Associated with drug cytotoxicy in LBL
rs2853229	8	RRM2B	Intron	NCALD/UBR5	A	0.49	0.0062	1.68	(1.16–2.44)	38/74/42	A	C	Drug pathway
rs8534	8	SLC25A37	Intergenic	SLC25A37/LOC646721	T	0.38	0.0067	1.68	(1.15–2.43)	24/69/61	T	C	Associated with drug cytotoxicy in LBL
rs8079740	17	ABCC3	Intron of CACNA1G	SPATA20ANCC3	G	0.31	0.0078	0.56	(0.37–0.86)	20/56/78	A	G	Drug pathway
rs757420	17	ABCC3	Intergenic	CACNA1G/ABCC3	T	0.30	0.0079	0.55	(0.36–0.86)	18/56/80	T	C	Drug pathway
rs2010851	12	WNT5B	Near-gene-3 [WNT5B]	WNT5B/ LOC100132548	A	0.30	0.0082	0.52	(0.32–0.84)	13/67/74	C	A	Associated with drug cytotoxicy in LBL
rs4995289	17	P2RX1	Intergenic	P2RX1/ATP2A3	T	0.28	0.0088	0.51	(0.31–0.85)	14/59/81	C	T	Associated with drug cytotoxicy in LBL
rs1516801	17	P2RX1	Intergenic	P2RX1/ATP2A3	G	0.28	0.0088	0.51	(0.31–0.85)	14/59/81	T	G	Associated with drug cytotoxicy in LBL
rs2607662	8	RRM2B	Intron of UBR5	NCALD/UBR5	T	0.46	0.0095	1.64	(1.13–2.37)	31/79/44	T	C	Drug pathway

Abbreviations: AML, acute myeloid leukemia; CI, confidence interval; DFS, disease-free survival; HR, hazards ratio; LBL, lymphoblastoid cell lines; MAF, minor allele frequencies; SNPs, single nucleotide polymorphisms.
The classifications near-gene-5 and near-gene-3 label SNPs that are outside transcribed regions, but within 2000 bp of a transcription region. Near-gene-5 includes upstream promoter region and untranslated 5′ mRNA.

the ENCODE Chip-Seq and DNase I peaks (Supplementary Table 2), suggesting that they could have regulatory functions. Interestingly, the GTEx (Genotype-Tissue Expression) eQTL browser (http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi) and eQTL browser (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/) showed that the SNP in *GSTM1*, rs3754446, is associated with *GSTM1* and GSTM5 expression levels in the liver,[21] brain and LCL[22–24] (Supplementary Table 2).

DISCUSSION

Previous pharmacogenomic studies of AML response to chemotherapy have been limited to a small number of candidate genes.[2,25–27] Although studies related to genes in drug pathways have been performed, many have not been replicated. In addition, some AML pharmacogenomics studies have been conducted using DNA from blast cells, which may have included somatic cell mutations in addition to germline polymorphisms.[28–30]

To determine whether germline genetic variations are associated with AML response, we designed our own pharmacogenomic study in AML patients treated with a chemotherapeutic regimen consisting of cytarabine, etoposide and busulfan, followed by ASCT. Our study was focused primarily on 154 AML patients of European ancestry. The analysis was centered on 42 genes related to the pharmacokinetic and pharmacodynamic pathways of the chemotherapy. A few other genes that had previously been associated with drug response in AML were also included (Supplementary Table 1, Figure 1 and Figure 2). The goals of this study were to identify new associations with DFS in AML patients and to determine whether SNPs previously reported to be associated with AML response could be replicated.

Overall, the most significant SNP in our analysis was the intronic variant in *ABCC3*, rs4148405 (HR = 3.1, $P = 9.5 \times 10(-6)$). ABCC3 is a multidrug resistance-associated protein, which is known to transport the etoposide metabolite, etoposide glucuronide.[31] Several lines of evidence support a role for *ABCC3* in DFS in AML patients. First, in a previous study a promoter variant in *ABCC3*, rs4793665 was associated with a shorter survival time in Israeli AML patients.[29] Although this variant was not associated with AML response in the current study, the data support a role of the transporter in response to chemotherapy in AML. Second, following etoposide administration to *Abcc2* −/−, *Abcc3* −/− mice, higher etoposide glucuronide levels were observed in the liver,[32] consistent with a potentially important role of *ABCC3* in etoposide pharmacokinetics. Third, cell lines transfected with *ABCC3* show greater resistance to etoposide.[33] Finally, higher *ABCC3* expression levels in leukemia cells are associated with poor outcome in children with leukemia.[34,35]

Data from Regulome database suggest that the *ABCC3* variant, rs4148405, is in a functional location in the genome, as this region demonstrates direct evidence of binding through ChIP-seq studies. Transcription factors that have a role in hepatic gene regulation (for example, *CEBPB*, *USF1*, *FOXA1*) have DNA response elements within this gene region (see http://regulome.stanford.edu/snp/chr17/48713567).[20,36] On the basis of the results of our primary analysis with DFS, we speculate that the minor allele of rs4148405 is associated with higher expression levels of *ABCC3* in the liver and/or leukemia cells, and thus reduced levels of etoposide in the tumor cells. Although the SNP rs4148405 is found at a MAF of >10% in non-European populations, this SNP was not significantly associated with DFS in the AML patients of non-European ancestry. It is possible that different linkage disequilibrium patterns between rs4148405 and potential causative SNPs may have confounded the analysis.

In addition to a SNP in *ABCC3*, we also identified 23 SNPs in seven other genes (*SLC28A3*, *DCK*, *RRM1*, *GSTM1*, *GSTT1*, *MSH3* and *SLC22A12*) that were associated with DFS in the European AML patients (with $P < 0.005$, Table 2). Interestingly, expression levels of these genes or other SNPs in these genes have been previously associated with response to chemotherapy in AML or other cancers.[19,29,37–39] Expression levels or SNPs in these genes have also been associated with IC_{50} values of various chemotherapy agents in cell lines.[40,41] Genetic polymorphisms in glutathione-S-transferases,

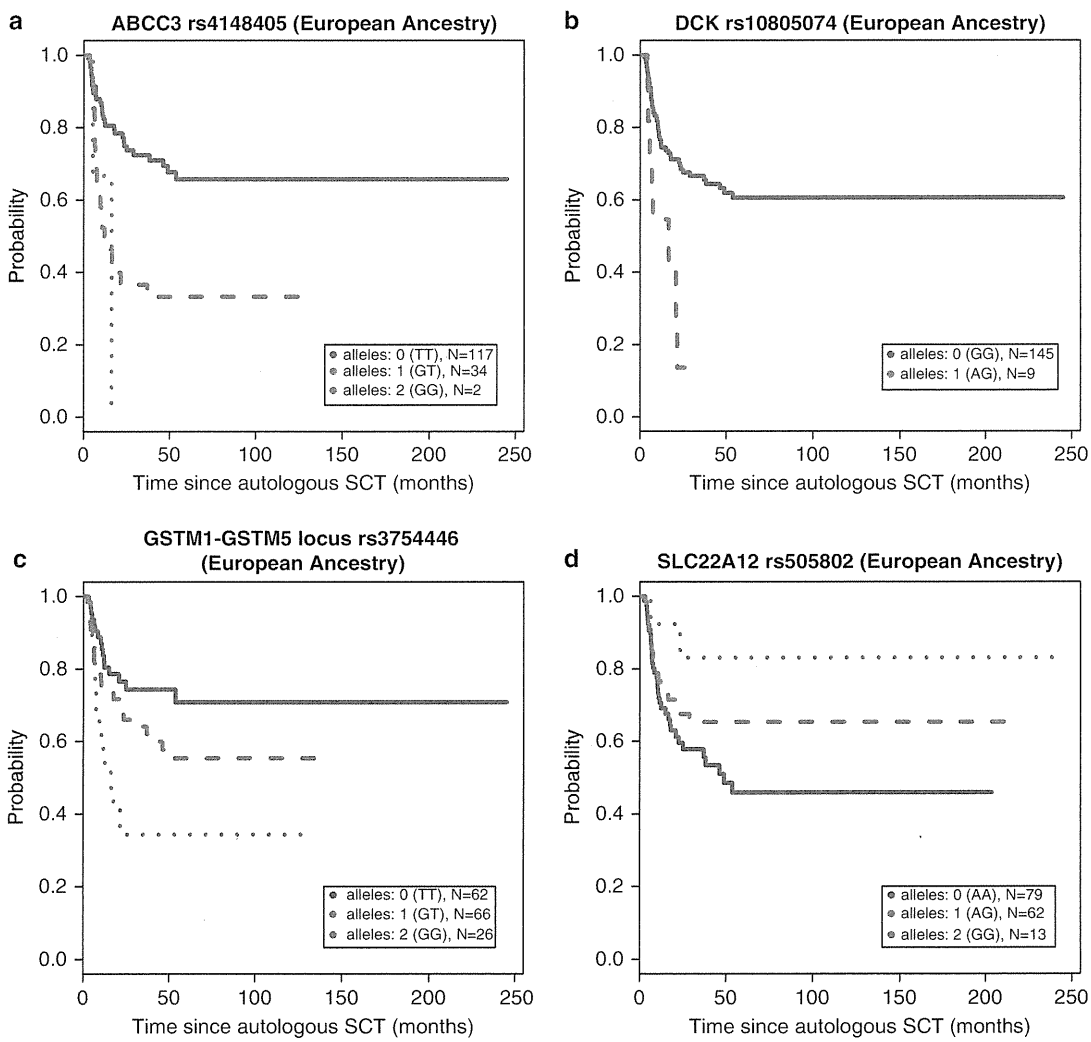

Figure 3 (**a**–**d**) Kaplan–Meier estimate of disease-free survival (DFS) stratified by the top SNPs (**a**) rs4148405 *ABCC3* (**b**) rs10805074 *DCK* (**c**) rs3754446 *GSTM1-GSTM5* locus and (**d**) rs505802 *SLC22A12* genotypes in patients of European ancestry.

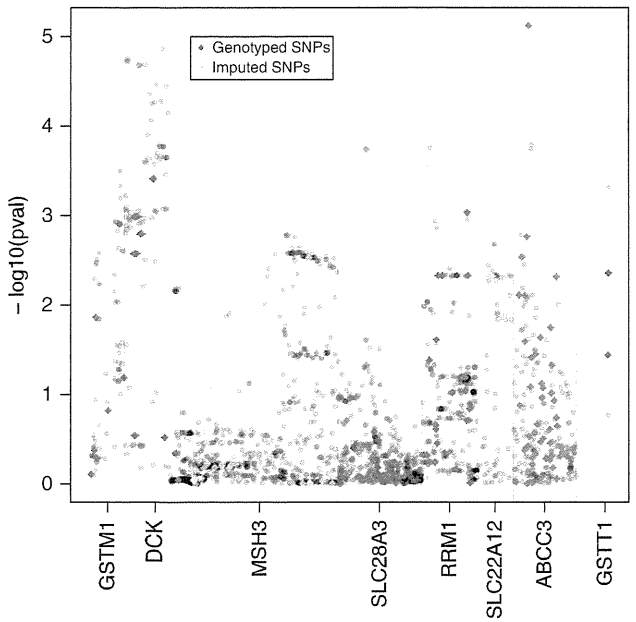

Figure 4 Plot showing the association of the genotyped and imputed SNPs with MAF $\geqslant 1\%$ in the selected top eight candidate genes (*ABCC3*, *DCK*, *SLC28A3*, *SLC22AA12*, *MSH3*, *RRM1*, *GSTM1*, *GSTT1*). Each colored diamond (not gray or black) represents a genotyped SNP and the gray/black dots represent imputed SNPs.

such as *GSTT1* and *GSTM1*, have been widely studied for their associations with drug toxicity, drug response and disease risk in leukemia patients.[29,39,42] Notably, *GSTM1* and *GSTT1* deletions have been implicated in various phenotypes associated with leukemia, including drug response,[43] busulfan pharmacokinetics[28] and disease risk.[44] Although we did not examine the effect of the *GSTM1* deletion in this study, we identified several SNPs (genotyped and imputed) in the *GSTM1-GSTM5* locus (rs3754446, rs929166 and rs11101989) associated with DFS in individuals of European ancestry that have not been previously reported (Table 2, Figure 2b). One of the SNPs in this locus, rs3754446, was also associated with AML response in the individuals of non-European ancestry (HR = 1.7, $P = 0.028$). Overall, in the entire cohort this minor allele SNP, rs3754446, was significantly associated with shorter DFS (HR = 1.7, $P = 0.00027$). Perusal of eQTL databases (Regulome database, GTEx browser and eQTL browser) suggests that the SNPs in the *GSTM1-GSTM5* locus are associated with expression levels of *GSTM1* and/or *GSTM5* in liver,[21] brain and LCL.[22–24] Thus, the SNPs in these genes, which are involved in drug metabolism, could affect AML response by affecting the pharmacokinetics of the drugs used in the treatment of AML. We examined 21 AML patients from our overall cohort, where we have their first-dose busulfan area under the curve. Interestingly, we observed a significant association between rs3754446 and reduced busulfan area under the curve in these 21 AML patients ($P = 0.03$, Supplementary Figure 1). The minor allele, G, in rs3754446, was associated with lower busulfan plasma levels (area under the curve), which was consistent with our observation that patients with the G allele had shorter DFS (Figure 3c). As higher busulfan plasma levels have been associated with busulfan liver toxicity,[14,45] future studies are needed to determine whether the SNPs in *GSTM1-GSTM5* are associated with liver toxicity. Collectively, these data suggest that *GSTM1* could have an important role in determining busulfan drug levels, drug response and/or drug toxicity.

Recently, using a drug-metabolizing enzyme/transporter gene SNP array, a synonymous variant in *SLC22A12* (rs11231825) was found to be associated with response in 94 AML patients treated with a combination drug regimen of gemtuzumab-ozogamicin with fludarabine-cytarabine-idarubicin.[19] Among the SNPs in the drug-metabolizing enzyme/transporter genes that were found to be significantly associated with AML response, we were able to replicate the SNP (rs11231825) in the uric acid transporter, *SLC22A12*. Notably, this synonymous variant, which is in linkage disequilibrium with a SNP upstream of *SLC22A12*, rs505802, has been found in various genome-wide association studies to be associated with uric acid levels.[46,47] In these genome-wide association studies,[46,47] the minor allele T has been associated with higher uric acid levels. Although speculative, our study and the previous study,[19] which demonstrated that patients with the T allele have a better response to chemotherapy, suggest that higher uric acid levels may be associated with longer DFS time. In our study, the T allele, which is associated with higher uric acid levels,[48,49] was associated with longer DFS time (see Figure 3d). Uric acid is a potent antioxidant, and it is possible that higher levels are beneficial for survival in AML patients.

Previous studies have shown that the nucleoside transporter, *SLC28A3* (CNT3) has a role in cytarabine cytotoxicity and resistance.[40,50,51] In this study, we observed several low-allele frequency variants (MAF 1%) in *SLC28A3* are associated with DFS in AML patients on cytarabine and other chemotherapy. Therefore, we hypothesized that CNT3 may transport cytarabine. Supplementary Figure 2a shows that radiolabeled cytarabine was taken up into CNT3 stably expressing cells, and the uptake (over empty vector cells) was significantly enhanced in cell lines exposed to the equilibrative nucleoside transporter inhibitor *S*-(4-nitrobenzyl)-6-thioinosine, which reduced background uptake of cytarabine in the cells. Notably, cytarabine uptake decreased significantly in CNT3 stable cells treated with the *SLC28A3* inhibitor (phloridzin) or with both inhibitors (Supplementary Figure 2a). Fludarabine, a known substrate of CNT3 was used as a positive control in this study (Supplementary Figure 2b).[52] As low-allele frequency variant in CNT3 was found to be significantly associated with DFS (Table 2), we interpret the results with caution. Although our finding that *SLC28A3* transports cytarabine supports the association, functional studies of variants in this region and/or a larger sample size are required to determine whether these uncommon variants are associated with cytarabine response.

DCK has an important role in activating cytarabine to its active metabolite, cytarabine triphosphate. Two tag-SNPs, rs4308342 and rs3775289, in *DCK* that were associated with DFS in our AML patients of European ancestry have been previously associated with the IC_{50} of another nucleoside analog, gemcitabine, in LCL.[40] Furthermore, a previous study demonstrated that the level of cytarabine triphosphate in AML blast cells correlates with the ratio in expression levels of the cytarabine-activating enzyme, *DCK* and the -inactivating enzyme, 5′-nucleotidase, cytosolic II (*NT5C2*).[48] In our study, in addition to the two tag-SNPs in *DCK* associated with DFS in AML patients, five SNPs in the *NT5C2* region were associated with DFS (Table 2). Although several of these SNPs are eQTLs (see Supplementary Table 2), further studies are required to determine whether these SNPs have important roles in determining the levels of cytarabine triphosphate in AML blast cells. Other genes, which have important roles in the cytarabine pharmacodynamics pathway are ribonucleotide reductase, *RRM1* and *RRM2*, which are considered

targets of nucleoside drugs such as cytarabine. The role of this enzyme is to regulate intracellular pools of ribonucleotides, such as deoxycytidine triphosphate, which is important in building blocks for DNA replication. Studies have shown that AML blasts cells with high levels of deoxycytidine triphosphate are resistant to cytarabine, and that there is a significant correlation between *RRM1* and *RRM2* gene expression levels and deoxycytidine triphosphate levels after cytarabine treatment in AML blast cells.[15] Consistent with previous studies showing that SNPs in *RRM1* are associated with response or toxicity to gemcitabine-based chemotherapy in lung and breast cancer patients,[49,53] our findings suggest that the SNPs in *RRM1* are associated with AML response to chemotherapy that include cytarabine.

Although our current association analysis supports the important roles of drug pathway genes, mainly transporters and enzymes, in AML response, we also selected eight genes that have been associated with cytarabine or etoposide IC_{50} values in LCL.[16,17] A few SNPs in the eight genes were significantly associated with response in AML patients, suggesting that genes identified in *in vitro* assays in LCLs may also be important predictors of *in vivo* drug response in AML patients.[16,17]

In summary, in this genetic association study of DFS in AML patients, we identified polymorphisms that have not been previously associated with AML response, including SNPs in *ABCC3*, *DCK*, *GSTM1*, *MSH3*, *NT5C2*, *RRM1* and *SLC28A3*. A SNP in *ABCC3*, rs4148405, which remained significant after multiple testing, was associated with DFS, suggesting an important role of *ABCC3* in determining etoposide levels in the liver and other tissues, and hence AML response. Many of the significant SNPs or their tag-SNPs were eQTLs or located in functional regions in the genome. Finally, we determined for the first time that *SLC28A3* (CNT3) transported cytarabine into cells, suggesting an important role of this transporter in cytarabine cytotoxicity.

ACKNOWLEDGEMENTS

This research was supported in part by a grant from NIH NIGMS (GM61390). This study was supported by the NIH Pharmacogenomics Research Network (PGRN)-RIKEN Center for Genomic Medicine (CGM) Strategic Alliance. We are grateful to the UCSF AML Tissue Bank staff (Ms Joy Cruz, Ms Christine Cheng). SWY thanks Chris Gignoux and Pär Mattson for their help in creating plots in the R-project.

1 Daver, N. & Cortes, J. Molecular targeted therapy in acute myeloid leukemia. *Hematology* **17** (Suppl 1), S59–S62 (2012).
2 Patel, J. P., Gonen, M., Figueroa, M. E., Fernandez, H., Sun, Z., Racevskis, J. *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N. Engl. J. Med.* **366,** 1079–1089 (2012).
3 Buccisano, F., Maurillo, L., Del Principe, M. I., Del Poeta, G., Sconocchia, G., Lo-Coco, F. *et al.* Prognostic and therapeutic implications of minimal residual disease detection in acute myeloid leukemia. *Blood* **119,** 332–341 (2012).
4 Martelli, M. P., Sportoletti, P., Tiacci, E., Martelli, M. F. & Falini, B. Mutational landscape of AML with normal cytogenetics: Biological and clinical implications. *Blood Rev.* **27,** 13–22 (2013).
5 Kroger, N., Brand, R., van Biezen, A., Cahn, J. Y., Slavin, S., Blaise, D. *et al.* Autologous stem cell transplantation for therapy-related acute myeloid leukemia and myelodysplastic syndrome. *Bone Marrow Transplant.* **37,** 183–189 (2006).
6 Novitzky, N., Thomas, V., du Toit, C. & McDonald, A. Is there a role for autologous stem cell transplantation for patients with acute myelogenous leukemia? A retrospective analysis. *Biol. Blood Marrow Transplant.* **17,** 875–884 (2011).
7 Linker, C. A., Damon, L. E., Ries, C. A., Navarro, W. A., Case, D. & Wolf, J. L. Autologous stem cell transplantation for advanced acute myeloid leukemia. *Bone Marrow Transplant.* **29,** 297–301 (2002).
8 Linker, C. A., Owzar, K., Powell, B., Hurd, D., Damon, L. E., Archer, L. E. *et al.* Auto-SCT for AML in second remission: CALGB study 9620. *Bone Marrow Transplant.* **44,** 353–359 (2009).
9 Linker, C. A., Ries, C. A., Damon, L. E., Sayre, P., Navarro, W., Rugo, H. S. *et al.* Autologous stem cell transplantation for acute myeloid leukemia in first remission. *Biol. Blood Marrow Transplant.* **6,** 50–57 (2000).
10 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
11 Yang, J., Bogni, A., Schuetz, E. G., Ratain, M., Dolan, M. E., McLeod, H. *et al.* Etoposide pathway. *Pharmacogenet. Genomics* **19,** 552–553 (2009).
12 Li, L., Schaid, D. J., Fridley, B. L., Kalari, K. R., Jenkins, G. D., Abo, R. P. *et al.* Gemcitabine metabolic pathway genetic polymorphisms and response in patients with non-small cell lung cancer. *Pharmacogenet. Genomics* **22,** 105–116 (2012).
13 Hassan, M. & Andersson, B. S. Role of pharmacogenetics in busulfan/cyclophosphamide conditioning therapy prior to hematopoietic stem cell transplantation. *Pharmacogenomics* **14,** 75–87 (2013).
14 Abbasi, N., Vadnais, B., Knutson, J. A., Blough, D. K., Kelly, E. J., O'Donnell, P. V. *et al.* Pharmacogenetics of intravenous and oral busulfan in hematopoietic cell transplant recipients. *J. Clin. Pharmacol.* **51,** 1429–1438 (2011).
15 Lamba, J. K. Genetic factors influencing cytarabine therapy. *Pharmacogenomics* **10,** 1657–1674 (2009).
16 Huang, R. S., Duan, S., Bleibel, W. K., Kistner, E. O., Zhang, W., Clark, T. A. *et al.* A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc. Natl Acad. Sci. USA* **104,** 9758–9763 (2007).
17 Hartford, C. M., Duan, S., Delaney, S. M., Mi, S., Kistner, E. O., Lamba, J. K. *et al.* Population-specific genetic variants important in susceptibility to cytarabine arabinoside cytotoxicity. *Blood* **113,** 2145–2153 (2009).
18 Hewish, M., Martin, S. A., Elliott, R., Cunningham, D., Lord, C. J. & Ashworth, A. Cytosine-based nucleoside analogs are selectively lethal to DNA mismatch repair-deficient tumour cells by enhancing levels of intracellular oxidative stress. *Br. J. Cancer* **108,** 983–992 (2013).
19 Iacobucci, I., Lonetti, A., Candoni, A., Sazzini, M., Papayannidis, C., Formica, S. *et al.* Profiling of drug-metabolizing enzymes/transporters in CD33+ acute myeloid leukemia patients treated with Gemtuzumab-Ozogamicin and Fludarabine, Cytarabine and Idarubicin. *Pharmacogenomics J.* (e-pub ahead of print 15 May 2012; doi:10.1038/tpj.2012.13).
20 Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22,** 1790–1797 (2012).
21 Schadt, E. E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P. Y. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6,** e107 (2008).
22 Stranger, B. E., Nica, A. C., Forrest, M. S., Dimas, A., Bird, C. P., Beazley, C. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39,** 1217–1224 (2007).
23 Veyrieras, J. B., Kudaravalli, S., Kim, S. Y., Dermitzakis, E. T., Gilad, Y., Stephens, M. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4,** e1000214 (2008).
24 Gibbs, J. R., van der Brug, M. P., Hernandez, D. G., Traynor, B. J., Nalls, M. A., Lai, S. L. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6,** e1000952 (2010).
25 Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286,** 531–537 (1999).
26 Gaidzik, V. I., Paschka, P., Spath, D., Habdank, M., Kohne, C. H., Germing, U. *et al.* TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. *J. Clin. Oncol.* **30,** 1350–1357 (2012).
27 Paschka, P., Marcucci, G., Ruppert, A. S., Whitman, S. P., Mrozek, K., Maharry, K. *et al.* Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *J. Clin. Oncol.* **26,** 4595–4602 (2008).
28 Kim, S. D., Lee, J. H., Hur, E. H., Lee, J. H., Kim, D. Y., Lim, S. N. *et al.* Influence of GST gene polymorphisms on the clearance of intravenous busulfan in adult patients undergoing hematopoietic cell transplantation. *Biol. Blood Marrow Transplant.* **17,** 1222–1230 (2011).
29 Muller, P., Asher, N., Heled, M., Cohen, S. B., Risch, A. & Rund, D. Polymorphisms in transporter and phase II metabolism genes as potential modifiers of the predisposition to and treatment outcome of de novo acute myeloid leukemia in Israeli ethnic groups. *Leuk. Res.* **32,** 919–929 (2008).
30 Emadi, A. & Karp, J. E. The clinically relevant pharmacogenomic changes in acute myelogenous leukemia. *Pharmacogenomics* **13,** 1257–1269 (2012).
31 Zelcer, N., Saeki, T., Reid, G., Beijnen, J. H. & Borst, P. Characterization of drug transport by the human multidrug resistance protein 3 (ABCC3). *J. Biol. Chem.* **276,** 46400–46407 (2001).
32 Lagas, J. S., Fan, L., Wagenaar, E., Vlaming, M. L., van Tellingen, O., Beijnen, J. H. *et al.* P-glycoprotein (P-gp/Abcb1), Abcc2, and Abcc3 determine the pharmacokinetics of etoposide. *Clin. Cancer Res.* **16,** 130–140 (2010).
33 Belinsky, M. G., Dawson, P. A., Shchaveleva, I., Bain, L. J., Wang, R., Ling, V. *et al.* Analysis of the in vivo functions of Mrp3. *Mol. Pharmacol.* **68,** 160–168 (2005).
34 Steinbach, D., Wittig, S., Cario, G., Viehmann, S., Mueller, A., Gruhn, B. *et al.* The multidrug resistance-associated protein 3 (MRP3) is associated with a poor outcome in childhood ALL and may account for the worse prognosis in male patients and T-cell immunophenotype. *Blood* **102,** 4493–4498 (2003).
35 Steinbach, D., Lengemann, J., Voigt, A., Hermann, J., Zintl, F. & Sauerbrey, A. Response to chemotherapy and expression of the genes encoding the multidrug

resistance-associated proteins MRP2, MRP3, MRP4, MRP5, and SMRP in childhood acute myeloid leukemia. *Clin. Cancer Res.* **9,** 1083–1086 (2003).
36 Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
37 Kim, K. I., Huh, I. S., Kim, I. W., Park, T., Ahn, K. S., Yoon, S. S. *et al.* Combined interaction of multi-locus genetic polymorphisms in cytarabine arabinoside metabolic pathway on clinical outcomes in adult acute myeloid leukaemia (AML) patients. *Eur. J. Cancer* **49,** 403–410 (2013).
38 Shi, J. Y., Shi, Z. Z., Zhang, S. J., Zhu, Y. M., Gu, B. W., Li, G. *et al.* Association between single nucleotide polymorphisms in deoxycytidine kinase and treatment response among acute myeloid leukaemia patients. *Pharmacogenetics* **14,** 759–768 (2004).
39 Xiao, Z., Yang, L., Xu, Z., Zhang, Y., Liu, L., Nie, L. *et al.* Glutathione S-transferases (GSTT1 and GSTM1) genes polymorphisms and the treatment response and prognosis in Chinese patients with de novo acute myeloid leukemia. *Leuk. Res.* **32,** 1288–1291 (2008).
40 Li, L., Fridley, B. L., Kalari, K., Jenkins, G., Batzler, A., Weinshilboum, R. M. *et al.* Gemcitabine and arabinosylcytosin pharmacogenomics: genome-wide association and drug response biomarkers. *PloS one* **4,** e7765 (2009).
41 Mitra, A. K., Crews, K. R., Pounds, S., Cao, X., Feldberg, T., Ghodke, Y. *et al.* Genetic variants in cytosolic 5'-nucleotidase II are associated with its expression and cytarabine sensitivity in HapMap cell lines and in patients with acute myeloid leukemia. *J. Pharmacol. Exp. Ther.* **339,** 9–23 (2011).
42 Mossallam, G. I., Abdel Hamid, T. M. & Samra, M. A. Glutathione S-transferase GSTM1 and GSTT1 polymorphisms in adult acute myeloid leukemia; its impact on toxicity and response to chemotherapy. *J. Egypt. Natl Canc. Inst.* **18,** 264–273 (2006).
43 Voso, M. T., Hohaus, S., Guidi, F., Fabiani, E., D'Alo, F., Groner, S. *et al.* Prognostic role of glutathione S-transferase polymorphisms in acute myeloid leukemia. *Leukemia* **22,** 1685–1691 (2008).
44 Das, P., Shaik, A. P. & Bammidi, V. K. Meta-analysis study of glutathione-S-transferases (GSTM1, GSTP1, and GSTT1) gene polymorphisms and risk of acute myeloid leukemia. *Leuk. Lymphoma* **50,** 1345–1351 (2009).
45 Zhang, H., Graiser, M., Hutcherson, D. A., Dada, M. O., McMillan, S., Ali, Z. *et al.* Pharmacokinetic-directed high-dose busulfan combined with cyclophosphamide and etoposide results in predictable drug levels and durable long-term survival in lymphoma patients undergoing autologous stem cell transplantation. *Biol. Blood. Marrow Transplant.* **18,** 1287–1294 (2012).
46 Takeuchi, F., Yamamoto, K., Isono, M., Katsuya, T., Akiyama, K., Ohnaka, K. *et al.* Genetic Impact on Uric Acid Concentration and Hyperuricemia in the Japanese Population. *J. Atheroscler. Thromb.* **20,** 351–367 (2012).
47 Kolz, M., Johnson, T., Sanna, S., Teumer, A., Vitart, V., Perola, M. *et al.* Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet.* **5,** e1000504 (2009).
48 Yamauchi, T., Negoro, E., Kishi, S., Takagi, K., Yoshida, A., Urasaki, Y. *et al.* Intracellular cytarabine triphosphate production correlates to deoxycytidine kinase/cytosolic 5'-nucleotidase II expression ratio in primary acute myeloid leukemia cells. *Biochem. Pharmacol.* **77,** 1780–1786 (2009).
49 Kim, S. O., Jeong, J. Y., Kim, M. R., Cho, H. J., Ju, J. Y., Kwon, Y. S. *et al.* Efficacy of gemcitabine in patients with non-small cell lung cancer according to promoter polymorphisms of the ribonucleotide reductase M1 gene. *Clin. Cancer Res.* **14,** 3083–3088 (2008).
50 Errasti-Murugarren, E., Pastor-Anglada, M. & Casado, F. J. Role of CNT3 in the transepithelial flux of nucleosides and nucleoside-derived drugs. *J. Physiol.* **582,** 1249–1260 (2007).
51 Sarkar, M., Han, T., Damaraju, V., Carpenter, P., Cass, C. E. & Agarwal, R. P. Cytosine arabinoside affects multiple cellular factors and induces drug resistance in human lymphoid cells. *Biochem. Pharmacol.* **70,** 426–432 (2005).
52 Badagnani, I., Chan, W., Castro, R. A., Brett, C. M., Huang, C. C., Stryke, D. *et al.* Functional analysis of genetic variants in the human concentrative nucleoside transporter 3 (CNT3; SLC28A3). *Pharmacogenomics J.* **5,** 157–165 (2005).
53 Rha, S. Y., Jeung, H. C., Choi, Y. H., Yang, W. I., Yoo, J. H., Kim, B. S. *et al.* An association between RRM1 haplotype and gemcitabine-induced neutropenia in breast cancer patients. *Oncologist* **12,** 622–630 (2007).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

The AMPK-related kinase SNARK regulates hepatitis C virus replication and pathogenesis through enhancement of TGF-β signaling

Kaku Goto[1,2,3], Wenyu Lin[1], Leiliang Zhang[1], Nikolaus Jilg[1], Run-Xuan Shao[1], Esperance A.K. Schaefer[1], Hong Zhao[1], Dahlene N. Fusco[1], Lee F. Peng[1], Naoya Kato[2], Raymond T. Chung[1,*]

[1]*Gastrointestinal Unit, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA;*
[2]*The Advanced Clinical Research Center, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan;*
[3]*Japan Society for the Promotion of Science, Tokyo 102-8472, Japan*

Background & Aims: Hepatitis C virus (HCV) is a major cause of chronic liver disease worldwide. The biological and therapeutic importance of host cellular cofactors for viral replication has been recently appreciated. Here we examined the roles of SNF1/AMP kinase-related kinase (SNARK) in HCV replication and pathogenesis.
Methods: The JFH1 infection system and the full-length HCV replicon OR6 cell line were used. Gene expression was knocked down by siRNAs. SNARK mutants were created by site-directed mutagenesis. Intracellular mRNA levels were measured by qRT-PCR. Endogenous and overexpressed proteins were detected by Western blot analysis and immunofluorescence. Transforming growth factor (TGF)-β signaling was monitored by a luciferase reporter construct. Liver biopsy samples from HCV-infected patients were analyzed for SNARK expression.
Results: Knockdown of SNARK impaired viral replication, which was rescued by wild type SNARK but not by unphosphorylated or kinase-deficient mutants. Knockdown and overexpression studies demonstrated that SNARK promoted TGF-β signaling in a manner dependent on both its phosphorylation and kinase activity. In turn, chronic HCV replication upregulated the expression of SNARK in patients. Further, the SNARK kinase inhibitor metformin suppressed both HCV replication and SNARK-mediated enhancement of TGF-β signaling.
Conclusions: Thus reciprocal regulation between HCV and SNARK promotes TGF-β signaling, a major driver of hepatic fibrogenesis. These findings suggest that SNARK will be an attractive target for the design of novel host-directed antiviral and antifibrotic drugs.

Keywords: SNARK; NUAK2; HCV; Metformin; Fibrosis; TGF-beta; SMAD; Kinase.
Received 5 December 2012; received in revised form 3 June 2013; accepted 19 June 2013; available online 2 July 2013
* Corresponding author. Address: Gastrointestinal Unit, Massachusetts General Hospital, Boston, MA 02114, USA. Tel.: +1 617 724 7562; fax: +1 617 643 0446.
E-mail address: rtchung@partners.org (R.T. Chung).
Abbreviations: SNARK, sucrose-non-fermenting protein kinase 1/AMP-activated protein kinase-related protein kinase; TGF-β, transforming growth factor beta; JFH1, Japanese fulminant hepatitis 1; HTAs, host-targeting antivirals; HCC, hepatocellular carcinoma; SVR, sustained virological response; CsA, cyclosporin A.

Introduction

Chronic infection with hepatitis C virus (HCV) is a major cause of chronic liver disease and hepatocellular carcinoma (HCC) and the leading reason for liver transplantation worldwide. HCV infects approximately 170 million individuals worldwide [1]. Current therapy with pegylated interferon (IFN)-α in combination with ribavirin produces sustained virological response (SVR) in fewer than half of the patients infected with genotype 1 HCV, and does so with high rates of often unacceptable side effects [2]. In recent years, it has become increasingly evident that HCV propagation is highly dependent on host cellular cofactors which in turn represent promising antiviral targets [3]. It is hoped that these strategies will lead to rational host-targeting antivirals (HTAs) [4]. Moreover, HCV interferes with host cellular signaling pathways causing pathogenic effects such as insulin resistance (IR), diabetes, and alterations in host lipids. Indeed, mounting evidence supports hepatitis C as a metabolic disorder [5]. Hence, intervention against key host cellular factors critical for both HCV replication and viral pathogenesis may yield anti-hepatitis C therapies that both halt replication and abrogate other pathogenic effects of HCV, which might be further thought of as host-directed antiviral and antipathogenic therapies. Based on this assumption, we have previously conducted a functional genomic screen for host cellular factors supporting HCV replication using an HCV replicon system [6]. Among positive hits in our original screen was sucrose-non-fermenting protein kinase 1 (SNF1)/AMP-activated protein kinase (AMPK)-related protein kinase (SNARK), the fourth member of 14 mammalian AMPK-related kinases [7], which has been consistently found in other screens [8,9]. Although SNARK function is not well understood, SNARK heterozygote knockout mice displayed elevated serum triglyceride concentrations, hyperinsulinemia, glucose intolerance [10], and impaired contraction-stimulated glucose transport [11], implying that SNARK operates to maintain glucose and lipid homeostasis in a

manner analogous to AMPK, which was recently described to inhibit HCV replication [12].

Transforming growth factor (TGF)-β is a pleiotropic cytokine partaking in cell proliferation, differentiation, apoptosis, migration [13], and the major cytokine responsible for fibrosis in tissues including the liver. In HCV-infected persons, levels of TGF-β are elevated [14], and TGF-β was exhibited to promote the viral replication in a replicon model and was correlated with accelerated liver fibrosis in an *in vivo* model [15,16]. Intriguingly, a prior high-throughput mapping study of protein-protein interaction (PPI) identified an association of SNARK with SMADs [17], implying a direct link of SNARK to TGF-β signaling. Therefore, we sought to examine the significance and potential of SNARK as a therapeutic target in HCV replication and pathogenesis and its contribution to TGF-β signaling. We report that the phosphorylation and phosphotransferase activities of SNARK are required for HCV replication. Furthermore SNARK was demonstrated to enhance TGF-β signaling, and finally chronic HCV infection upregulated the expression of SNARK in patients. SNARK has pleiotropic functions including pro-TGF-β signaling activities in addition to the previously described AMPK-like properties. The finding of a reciprocal regulation between HCV and SNARK suggests that SNARK could be an effective host cellular target not only for an antiviral but also antipathogenic strategy.

Materials and methods

Compounds, antibodies, cells, and viruses

Metformin, TGF-β, and CsA were purchased from EMD chemicals USA (Gibbstown, NJ), Fitzgerald (North Acton, MA), and Sigma-Aldrich (St. Louis, MO), respectively. Antibodies to SNARK, FLAG, and β-actin were obtained from Sigma-Aldrich, and antibodies to HCV NS5A and phosphothreonine were obtained from BioFront Technologies (Tallahassee, FL) and Cell Signaling Technology (Danvers, MA), respectively. HuH7.5.1 and OR6 replicon cells were cultured as described previously [18], and HeLa cells were cultured in DMEM with 10% FBS. JFH1 virus infection was performed as described previously [19].

Further Materials and methods are described in the Supplementary Material section.

Results

Functional SNARK enhances HCV replication

To assess the contribution of SNARK to HCV replication, we first knocked down endogenous SNARK expression (Supplementary Fig. 1) with siRNAs in the Japanese fulminant hepatitis 1 (JFH1) virus infection system. HuH7.5.1 cells were transfected with SNARK-targeted siRNAs, which was followed by JFH1 infection. Reduced levels of *SNARK* mRNA were associated with impaired viral replication (Fig. 1A). We then constructed plasmids encoding the siRNA-resistant *SNARK* open reading frame (ORF) bearing synonymous mutations that are not recognized by *SNARK* siRNAs. The over expression of these siRNA-resistant SNARK proteins successfully rescued *SNARK* RNAi-impaired HCV replication (Fig. 1A, rSN-1 and rSN-7). We also tested the effects of SNARK knockdown and overexpression in the genotype 1 OR6 replicon system, and found that the decreased level of HCV RNA replication was also rescued by overexpression of siRNA-resistant forms of SNARK (Fig. 1B). Thus, SNARK was demonstrated to specifically support HCV replication in both a *bona fide* infection system and replicon model.

Next we sought to identify the function(s) responsible for SNARK's contribution to HCV replication. We introduced single

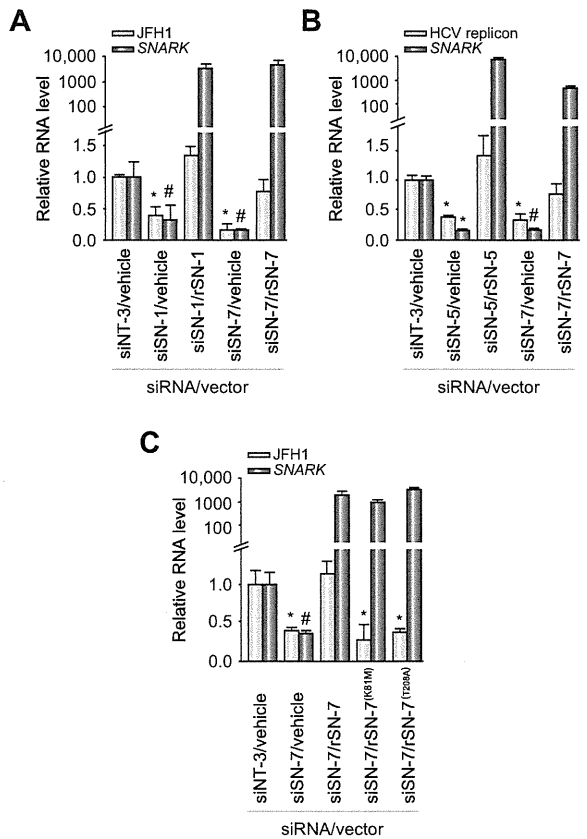

Fig. 1. SNARK supports HCV replication. (A) HuH7.5.1 cells were transfected with either non-targeting (siNT-3) or *SNARK*-targeted siRNAs (siSN-1 and siSN-7), followed by transfection of either empty or siRNA-resistant SNARK expression vectors (rSN-1 and rSN-7) 48 h later. The cells were infected with JFH1 on the next day and total RNA was harvested 48 h later. Relative JFH1 RNA and *SNARK* mRNA levels were quantified by real-time PCR analysis and normalized to *GAPDH*; $^{*}p$ <0.05 or $^{\#}p$ <0.01 *vs.* siNT-3 empty control. (B) OR6 replicon cells were transfected with siRNAs, followed by the transfection of empty or siRNA-resistant SNARK expression vectors 48 h later. Then total RNA was harvested 72 h later. Relative replicon RNA and *SNARK* mRNA levels were quantified by real-time PCR analysis and normalized to *GAPDH*. $^{*}p$ <0.01 or $^{\#}p$ <0.05 *vs.* siNT-3 empty control. (C) The rescue assay was conducted as described in (A). Here expression vectors of siRNA-resistant SNARK with either kinase-deficient mutation (rSN-7 K81M) or phosphorylation-deficient mutation (rSN-7 T208A) were used. Relative JFH1 RNA and *SNARK* mRNA levels were quantified by real-time PCR and normalized to *GAPDH*. $^{*}p$ <0.01 or $^{\#}p$ <0.05 *vs.* siNT-3 empty control.

mutations that abrogate either its phosphotransferase activity in the enzymatic pocket (K81M) or its phosphorylation at the phosphoacceptor site (T208A) in the siRNA-resistant *SNARK* ORF and overexpressed them in the rescue assay system used above with JFH1. In contrast to the rescue effects by wild type SNARK on viral replication, both functionally deficient mutants failed to recover impaired HCV replication by SNARK depletion (Fig. 1C and Supplementary Fig. 2). This result suggested that both the phosphorylation and kinase activities of SNARK are essential for its support of HCV replication.

SNARK phosphotransferase activity can be targeted

In a human hepatocarcinoma cell line, the kinase activity of SNARK was previously reported to be inhibited by metformin

Research Article

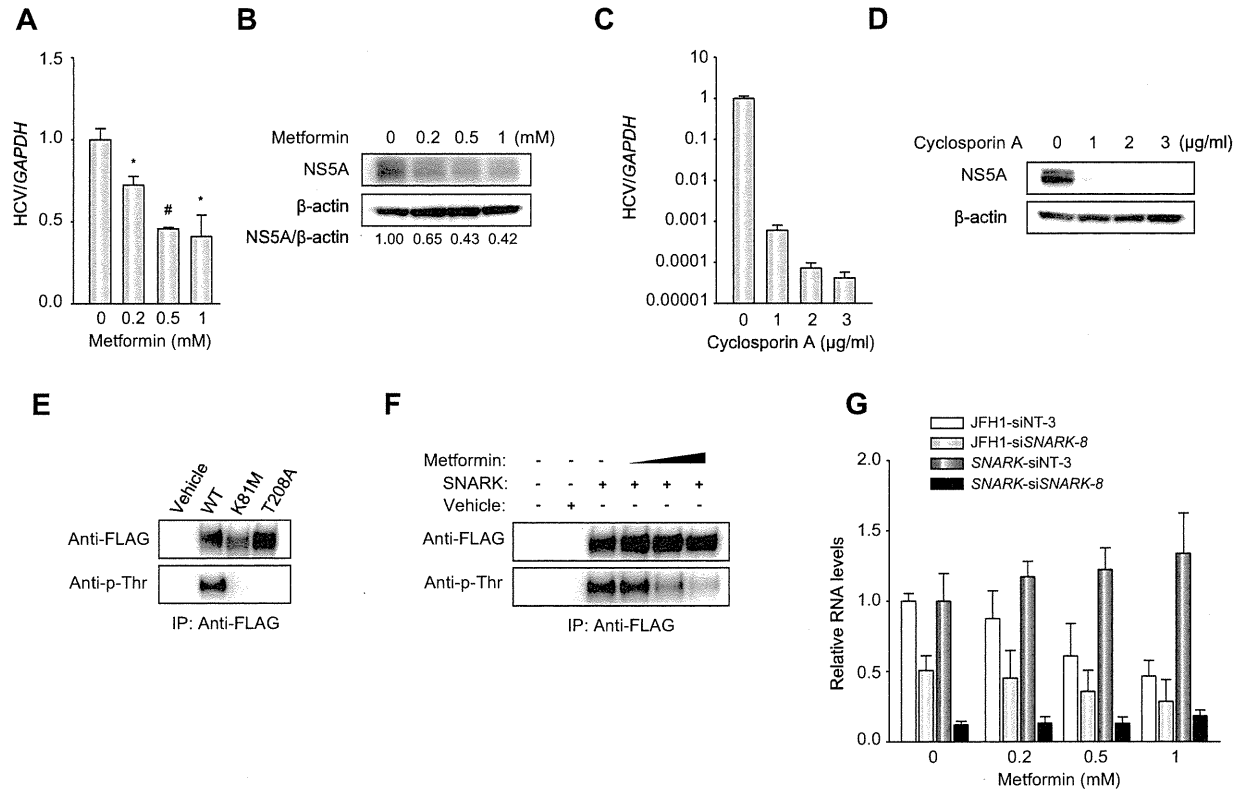

Fig. 2. Metformin suppressed HCV replication. HuH7.5.1 cells were infected with JFH1 and then treated with either metformin (A and B) or CsA (C and D) for 3 days. Densitometric values of NS5A normalized to those of β-actin were given as NS5A/β-actin. No significant cytotoxicity was observed by the compounds at the indicated concentrations. Relative HCV RNA level was quantified by real-time PCR and normalized to *GAPDH*; $^{*}p$ <0.05 or $^{\#}p$ <0.01 *vs.* untreated control. (E) In HuH7.5.1 cells either wild type or mutants (K81M or T208A) of FLAG-tagged SNARK were overexpressed and immunoprecipitated, followed by the detection with anti-FLAG or anti-phosphothreonine antibodies. (F) Wild type FLAG-tagged SNARK was overexpressed in the presence of metformin at 0.2, 0.5, and 1 mM in HuH7.5.1 cells and the levels of threonine phosphorylation were examined as indicated in (E). (G) HuH7.5.1 cells were transfected with either non-targeting (siNT-3) or *SNARK*-targeted (si*SNARK-8*) siRNAs. 48 hours later, cells were infected with JFH1 and treated with the indicated concentrations of metformin for 72 h. Relative mRNA levels for JFH1 or *SNARK* were quantified by real-time PCR and normalized to *GAPDH*.

[20], a well-known type 2 diabetes drug. In that setting, the kinase activity of SNARK was measured by incorporation of phosphate into SAMS peptide substrate, which was demonstrated to be significantly decreased by metformin treatment in the cell line. Here, we treated JFH1-infected HuH7.5.1 cells with metformin and observed moderate antiviral effects (Fig. 2A and B) with no cytotoxicity within the indicated dose range (data not shown). As a positive control, cyclosporin A (CsA) [21] strongly inhibited viral replication (Fig. 2C and D). To assess the phosphorylation level of SNARK subsequently, FLAG-tagged SNARK was overexpressed in HuH7.5.1 cells and immunoprecipitated for Western blot to monitor phosphothreonine levels [22]. While wild type SNARK was detected to be phosphorylated at threonine residue(s), neither K81M nor T208A mutant was (Fig. 2E), implying the autophosphorylation [23] and importance of threonine 208 as a phosphorylated site as reported [7]. Then we performed the assay using the wild type SNARK in the presence of metformin, which resulted in the reduced levels of phosphorylation dose-dependently (Fig. 2F). Here the data indicated that metformin suppressed SNARK phosphorylation, potentially interfering the phosphotransferase activity. The dose response of JFH1 to metformin was next examined when SNARK was knocked down by siRNAs. Metformin exerted dose-dependent antiviral effects in the cells transfected with non-targeting siRNAs, which was blunted by the siRNA-mediated reduction of SNARK expression (Fig. 2G). These data indicate that metformin's antiviral effect is mediated by inhibition of activated SNARK, bringing its full kinase activity, which may be a pharmacologic target for anti-HCV activity, and that metformin by itself could be a plausible component of a combination regimen targeting HCV.

SNARK is involved in TGF-β signaling

To investigate the possible roles of SNARK in viral pathogenesis based upon the induction of mRNA expression over the viral replication in cell culture (Supplementary Fig. 3), we also explored its involvement in downstream cellular signaling pathways. Intriguingly, SNARK appeared as an interactor with SMAD proteins in a high-throughput protein-protein interaction mapping study [17], which raised the distinct possibility that SNARK is involved in TGF-β signaling, the major profibrogenic pathway in the liver. Therefore, we first knocked down SNARK in HuH7.5.1 cells and assessed alterations in TGF-β signaling using an expression construct (PAI/L) encoding a luciferase reporter gene driven by promoter sequences of plasminogen activator inhibitor 1 (PAI-1), a transcriptional target of TGF-β [24]. siRNAs against *SNARK*

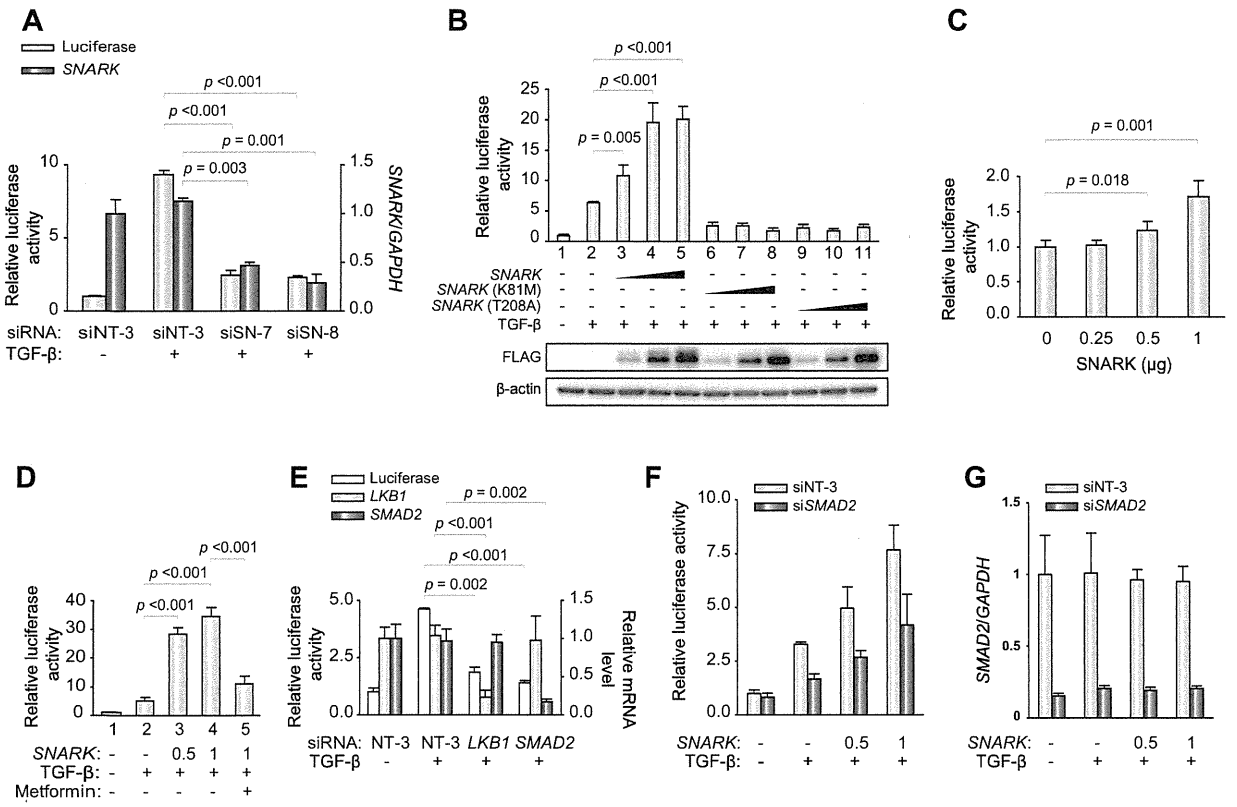

Fig. 3. SNARK accentuated TGF-β signaling. (A) HuH7.5.1 cells were transfected with siRNAs, which was followed by transfection of PAI/L reporter with pRL-TK 72 h later. On the next day, the cells were treated with TGF-β and lysed 24 hours later. The firefly and *Renilla* luciferase activities were measured and knockdown of *SNARK* was confirmed by real-time PCR with normalization to *GAPDH*. *P* Values were calculated as indicated. (B) Naïve HuH7.5.1 cells were transfected with either wild type (lanes 3–5) or mutant (lanes 6–11) SNARK expression plasmid at increasing doses (0.25, 0.5, and 1 μg) together with PAI/L reporter and pRL-TK. At 24 h post transfection, the cells were treated with TGF-β at 7 ng/ml for 20 h and lysed for dual luciferase assay. *P* Values were calculated as indicated. (C) HuH7.5.1 cells were transfected with increasing amount of wild type SNARK expression plasmid for dual luciferase assay as described in (A) without TGF-β. *P* Values were calculated as indicated. (D) HuH7.5.1 cells were transfected with increasing amount of SNARK expression plasmids (0.5 and 1 μg) together with PAI/L reporter and pRL-TK. On the next day, the cells were treated with TGF-β at 7 ng/ml and metformin at 1 mM (lane 5) for 20 h, and then lysed for dual luciferase assay. *P* Values were calculated as indicated. (E) HuH7.5.1 cells transfected with siRNAs against either *LKB1* or *SMAD2* for 72 h were transfected with PAI/L reporter and pRL-TK, which was followed by TGF-β treatment at 7 ng/ml for 20 h. Then the cells were lysed for dual luciferase assay and knockdown of targeted genes expression was confirmed by real-time PCR with normalization to *GAPDH*. *P* Values were calculated as indicated. (F) HuH7.5.1 cells transfected with siRNAs for 72 h were transfected with increasing amounts of wild type SNARK expression plasmid (0.5 and 1 μg) together with PAI/L reporter and pRL-TK. 24 h later, the cells were treated with TGF-β for 20 h, and then lysed for dual luciferase assay. (G) *SMAD2* knockdown levels were quantified by real-time PCR with normalization to *GAPDH*.

markedly reduced PAI/L luciferase activity and SNARK expression (Fig. 3A) in parallel, suggesting that SNARK is an important regulator of TGF-β signaling.

In order to elucidate the function of SNARK responsible for its contribution to TGF-β signaling, we assessed the effects of the overexpressed either wild type or mutant SNARK on TGF-β-stimulated PAI/L activity. In contrast to the dose-dependent increase of PAI/L activity by wild type SNARK, either kinase-dead K81M or unphosphorylated T208A mutant suppressed TGF-β-driven PAI/L activity (Fig. 3B). Moreover, the overexpression of SNARK in the absence of TGF-β moderately induced luciferase activity in HuH7.5.1 cells (Fig. 3C). These data demonstrate that both kinase activity and phosphorylation of SNARK are required for TGF-β signaling. Thereupon, in the same setting we treated HuH7.5.1 cells with metformin and found that SNARK-mediated stimulation of TGF-β signaling was inhibited by metformin (lane 5, Fig. 3D) though that was not the case in the absence of SNARK overexpression (Supplementary Fig. 5A) and the basal level of procollagen mRNA was not affected by metformin alone in HuH7.5.1 cells (Supplementary Fig. 5B), again underscoring that the kinase activity of SNARK is important for TGF-β signaling, and additionally raises the possibility that metformin may have utility as an anti-fibrotic agent in SNARK-facilitated pathogenesis.

We next examined regulators upstream and downstream of SNARK, depleting either liver kinase B1 (LKB1)/serine threonine kinase 11 (STK11), an upstream kinase of SNARK, or SMAD2, and assessed TGF-β-dependent PAI-1 luciferase activity. We found that knockdown of LKB1 abrogated PAI/L stimulation by TGF-β to the same extent as did SMAD2 knockdown (Fig. 3E). In addition, in contrast to luciferase activities in the presence of non-targeting siRNAs, overexpression of SNARK failed to rescue PAI/L activity in cells knocked down for either SMAD2 (Fig. 3F and G) or LKB1 (Supplementary Fig. 4). These data indicate that SNARK-mediated stimulation depends on SMAD2, and also that phosphorylation of SNARK by LKB1 and ensuing SNARK-mediated phosphorylation of downstream substrates, potentially in conjunction with SMAD2, are critical for TGF-β signaling.