Supplementary Online Content

Diouf B, Crews KR, Lew G, et al. Association of an inherited genetic variant with vincristine-related peripheral neuropathy in children with acute lymphoblastic leukemia. *JAMA*. doi:10.1001/jama.2015.0894

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Supplemental Methods

Genotyping and Imputation

Germline DNA (500 ng) was extracted from normal peripheral blood leukocytes, digested with restriction enzymes, amplified, labeled, and hybridized to the Affymetrix GeneChip Human Mapping 500K array (532552 SNPs) or the SNP 6.0 array (906600 SNPs) (Affymetrix, Santa Clara, CA). SNPs were excluded for genotyping call rates less than 95% and minor allele frequency less than 1% (484623 SNPs from the 500k and 830561 SNPs from the 6.0). SNPs Imputation in the 1000 genome CEU[1] was carried out using Mach/1.0.15[2]. For each imputed SNP, the estimated probability that an average imputed genotype matched an experimental genotype was at least 95%. The total number of imputed SNPs was 21764463. After filtering for call rate and minor allele frequencies, 2179037 SNPs from the SNP 6.0 array and 1091393 SNPs from the 500k were used. Race was genetically determined following a method previously described[3]. Briefly, the genetic ancestry estimates were determined by applying STRUCTURE (version 2.2.3) to 3,000 randomly chosen SNPs from those that passed the quality control described above, a procedure that was repeated 10 times on 10 separate sets of 3,000 SNPs, and the final ancestry was the average of the 10 estimates. HapMap samples from descendants of Northern Europeans (CEU, N=90), West Africans (YRI, N=90), and East Asians (CHB/JPT, N=90) and Native Americans (NA, N=105) were used as ancestral reference populations representing European, African, East Asian, and NA ancestry, respectively. In the input file (genotype file), the reference samples (CEU, YRI, CHB/JPT, and NA references) and patient samples were labeled as 1 and 0 in the PopFlag column, respectively. In the "mainparams" file, the number of SNPs used in each run of genetic ancestry estimation was 3,000 (i.e. NUMLOCI=3000). We assumed there were a maximum of 4 ancestries (European, African, Asian, and NA) in any individual (i.e. MAXPOPS=4). For each ancestry estimation run, we performed 10,000 iterations (i.e.BURNIN=10,000). In the "extraparams" file, both USEPOPINFO and PFROMPOPFLAGONLY were set to 1. Genotype imputation was carried out using Mach/1.0.15 with 50 rounds of Markov sampler[4]. We also turned on --geno option and –greedy option. We first assessed the imputation error rate of MACH1.0.15 with HapMap project data[5] as references in 4 American racial groups (White, Black, Asian and other racial background) determined genomically using STRUCTURE[3,6] in children with leukemia. Germline genotypes were determined using the Affymetrix 100K plus 500K platforms. To impute the missing genotypes for patients with "other" racial background, we made a set of haplotypes using only SNPs common to European (CEU), Asian (JPTCHB) and West African (YRI) populations. For patients whose ancestry was consistent with White, Black or Asian, we used the corresponding haplotypes from HapMap to impute the missing genotypes. To assess the imputation error rate, we masked the SNPs on Affymetrix 100k and imputed these 100k genotypes using the SNP genotypes interrogated on the Affymetrix 500k. For each imputed SNP, the estimated probability that an average imputed genotype will match an experimental genotype has to be at least 95%. Comparing the imputed vs. observed genotypes at these 100k SNPs, the average error rate was 1.6% for Whites, 3.2% for Blacks, 0.9% for Asians, and 1.5% for patients of other racial background. We then used the same procedure to impute genotype except we used all the genotyped data from Affymetrix 500k or Affymetrix SNP 6.0 array as the input and the reference haplotype data (CEU, YRI, CHB/JPT) were downloaded from 1000 genome project 2009 release[7]. The total number of imputed SNPs was 21764463. After filtering for call rate and minor allele frequencies, 2179037 SNPs from the SNP 6.0 array and 1091393 SNPs from the 500k were used.

Statistical analysis

Analyses were conducted using R (http://www.r-project.org/) and SAS (SAS Institute, Cary, NC). Weighted logistic regression model fitting for recurrent events was used to test the associations between SNP genotypes and the phenotype of peripheral neuropathy (i.e., occurrence of grade 2 or higher vincristine-induced neuropathy during continuation therapy), including cumulative vincristine dosage and genetically determined ancestry as covariates. Each episode of grade 2 or higher vincristine-induced neuropathy along with the patient's covariates contributed a data point with a response value of 1 and a full weight of 1, whereas a patient who never developed grade 2 or higher peripheral neuropathy or who stopped treatment early for any reasons (e.g. disease recurrence) during treatment was assigned a response value of 0 and a partial (fractional) weight; the weight was determined from the estimated cumulative occurrence incidence of grade 2 or higher peripheral neuropathy among all patients up to the time the patient's phenotype status was last assessed. The longer the follow up on the patient, the higher the fractional weight because longer follow up is more informative. See Supplement of Kishi et al[8] for additional statistical details.

Cell culture
The human pre-B leukemia (ALL) cell line NALM-6 was obtained from the German Collection of Microorganisms and Cell Cultures. The human T-lineage leukemia (ALL) cell line CEM was obtained from ATCC. SH-SY5Y was generously provided by Dr Jill Lahti at St. Jude Children's Research Hospital. Cells were cultured in RPMI-1640 medium containing 2 mM glutamine and 10% (vol/vol) FBS at 37 °C with 5% CO_2. iCell neurons derived from human induced pluripotent stem cells were purchased from Cellular Dynamics International and maintained per manufacturers's protocol.

Stable short hairpin RNA (shRNA) knockdowns
NALM-6 cells were infected with MISSION lentiviral transduction particles (Sigma-Aldrich) produced from a library of sequence-verified shRNAs targeting human *NKX-6.3* and *CEP72* transcripts. Non-target shRNA control particles (SHC002V) were also purchased from Sigma-Aldrich. Individual cell clones were isolated in medium containing puromycin.

Sequences of shRNA used for the knockdown of *NKX6.3 and CEP72* genes
The sequences of the shRNAs used were:
5'-CCGGGCCTCTTCTCAGAAGTTGGATCTCGAGATCCAACTTCTGAGAAGAGGCTTTTTTG-3' for *CEP72* and 5'-CCGGCGACGAGAAGATCCGCCTGCTCTCGAGAGCAGGCGGATCTTCTCGTCGTTTTT-3' for *NKX6.3*

Western blot analysis
Cell lysates were separated by electrophoresis on a SDS-polyacrylamide gel, and the proteins were then electroblotted onto a Hybond P PVDF membrane. Protein expression was analyzed using different antibodies.

Antibodies
Primary antibodies were purchased for GAPDH, and CEP72 from Santa Cruz Biotechnology. Horseradish Peroxidase-conjugated secondary antibodies were purchased from Dako.

Plasmids and site-directed mutagenesis
A 2612-bp sequence of the of *CEP72* promoter region including the C-allele of rs924607 was amplified using polymerase chain reaction (PCR) with the following primers: Forward 5'-GGTACCCCCAGGAGGAAGAGGCTCAGTGTTTGGGGC-3'; Reverse 5'-CTCGAGCGGCAGAACCGAGAGCCTGGCGGG-3'. The PCR product was cloned into the pGL4.24 luciferase reporter vector using KpnI and XhoI restriction enzymes. The T-allele was generated by site-directed mutagenesis using the quickChange lightning site-directed mutagenesis kit from Stratagene. The following primers were used for this mutagenesis: Forward 5'- CGTGACGATCTGCACTTAGGAATGCTGAGTGTTC-3'; Reverse 5'GAACACTCAGCATTCCTAAGTGCAGATCGTCACG-3'

Luciferase assay
Cells were transfected with pGL4.24 alone or containing the *CEP72* promoter with the C- or T-allele at rs924607 using the amaxa™ Nucleofector system according to the manufacturer's instructions. All transfections were carried out in triplicate. After 24 h of incubation, cells were collected and analyzed for luciferase activity with the Dual-Glo Luciferase Assay System (Promega). Luciferase activities were corrected for transfection efficiency by co-transfection of a plasmid expressing Renilla luciferase.

Genotyping using the TaqMan allelic discrimination assay
Patients were genotyped for the rs924607 (C/T) using TapMan SNP Genotyping Assay (C_8292459_20) from life technologies. The TaqMan SNP Genotyping Assay is used to discriminate between two alleles of a specific SNP. The method involves the use of forward and reverse primers to amplify the polymorphic sequence of interest and two dye-labeled probes for allele specific detection. One probe is labeled with VIC dye, which detects the "Allele C" sequence, while the other one is labeled with FAM dye (6-carboxyfluorescein), which detects the "Allele T" sequence. Assays were performed in accordance with manufactures's protocols[9]. Briefly genomic DNA was amplified in presence of the TaqMan genotyping master mix, the primers and the probes following a PCR protocol of 95°C 10 min, 40 cycles of 15 sec at 92°C and 1 min at 60°C. Fluorescence was quantified using a 7900HT fast

real time PCR system. The data were analysis by the software which makes an automatic call of either AlleleY (homozygous T/T), AlleleX (homozygous C/C) or heterozygous (C/T).

Molecular modeling of the binding of NKX-6.3 to DNA duplexes containing rs924607

The binding affinities of the NKX-6.3 HomeoDomain (HD) (KKKHTRPTFTGHQIFALEKTFEQTKYLAGPERARLAYSLGMTESQVWFQNRRTKWR) to target DNA duplexes containing sequences GCACTTA and GCACTCA were evaluated by homology modeling, molecular dynamics, molecular mechanics Generalized Born surface area calculations (MMGBSA) and normal mode calculations. The complex structures of NKX-6.3 HD with target DNA duplexes were constructed from homology modeling according to the crystal structure of the NKX-2.5-DNA complex (PDB ID: 3RKQ). An explicit TIP3P water box with dimensions of 10 angstrom distances from the solute was placed for each complex. Counterions of K+ were added to each system to neutralize the charges. In order to fully equilibrate the two systems, 10 nanoseconds of molecular dynamics simulation with periodic conditions for each system was performed by the SANDER module of the AMBER12 program[10] with the ff03 version of Amber force field[11]. The time step was set to 1 femtosecond and the conformations were collected at each 1 picosecond. MMGBSA calculations were performed for the last 1 nanosecond of each system, with 10 picosecond interval. The normal mode calculations were performed to evaluate the entropic changes.

The relative binding free energy of mutating CG to TA was calculated based on the equation

$$\Delta\Delta G_{TA\text{-}CG}=\Delta G_{TA}-\Delta G_{CG}=\Delta\Delta H_{TA\text{-}CG}-T\Delta\Delta S_{TA\text{-}CG} \qquad \text{(eq. 1)}$$

whereas the binding free energies of NKX-6.3 HD with wild type and mutated DNAs were divided into enthalpic changes and entropic changes. The enthalpic changes were calculated from conformations of Molecular Dynamics (MD) trajectories by MMGBSA. The entropic changes can be further divided as the combination of entropies of individual complexes and DNAs as following

$$T\Delta\Delta S_{TA\text{-}CG}=T(\Delta S_{TA}-\Delta S_{CG})$$

$$=T(S_{TA}^{complex}-S_{TA}^{DNA}-S^{NKX}-S_{CG}^{complex}+S_{CG}^{DNA}+S^{NKX})$$

$$=T(S_{TA}^{complex}-S_{CG}^{complex}-(S_{TA}^{DNA}+S_{CG}^{DNA})) \qquad \text{(eq. 2)}$$

Because entropy contributions estimated by normal mode analysis[12] are of large variations and computationally intensive for macromolecule systems, the individual entropies of complexes and DNAs were calculated from the last 1 nanosecond trajectories of MD with 50 picoseconds interval.

Binding affinity calculations.

The ΔG can be expressed as the logarithm of Kd: $\Delta G=-RT*\ln kd$. Then $\Delta\Delta G$ can be express as the logarithm of kd ratio: $\Delta\Delta G= \Delta G_2-\Delta G_1=-RT*\ln(kd2/kd1)$. The kd ratio can be calculated as: $kd2/kd1=\text{Exp}(-\Delta\Delta G/RT)$.

$\Delta\Delta G$, the calculated relative binding free energy is 7.67 Kcal/mol of TA to CG mutation.

1kcal=4184 joules

R (gas constant)=8.3144621 J/mol.K

T(temperature)=298.13

kd2/kd1= Exp(-7.67*4184/8.314421*298.13)

kd2/kd1 =419301

In vitro sensitivity (MTT) assay
In vitro drug sensitivity of primary ALL cells and of Nalm6 and CEM cell lines (wild-type or each knockdown cell line) for vincristine, was determined using a modification of the MTT (3-[4,5-dimethylthiazol-2-yl]-2, 5-diphenyl-tetrazoliumbromide) assay as we have described previously[13]. Briefly, cells in mid-log phase were exposed to serial dilutions of the different drugs and analyzed after three days of incubation. The concentration of drug required for 50% cell kill (LC_{50}) was used as the measure of relative sensitivity to vincristine.

Vincristine-induced neurotoxicity in neurons derived from human induced pluripotent stem (ips) cells
Neurons derived from human induced pluripotent stem (ips) cells were purchased from Cellular Dynamics International (Madison, WI, USA). iCell Neurons were plated and maintained following manufacturers protocol. Briefly, 1.33×10^5 iCell neurons/mL were plated in 3.3 ug/mL laminin (Sigma-Aldrich) onto 96-well black clear bottom poly-D-lysine Greiner Bio-One plates. 4 h post-plating cells were transfected Accell human *CEP72* SMARTpool siRNA or a non-targeting control pool (ThermoScientific). At 24 h post-transfection, cells were treated with 10 uM of vincristine or PBS control. High content imaging was performed 48 h after vincristine or PBS treatment with an ImageXPress Micro (Molecular Devices LLC) at the Institute for Genomics and Systems Biology Cellular Screening Center at the University of Chicago. Individual cell measurements of total neurite outgrowth (sum of the length of all processes) number of processes and number of branches were calculated in 1000 cells for each replicate experiment (each in triplicate), using the MetaXpress software Neurite Outgrowth Application Module (Molecular Devices LLC). Levels of *CEP72* after knockdown were measured using Life Technologies Cells according to manufacturer's protocol. Percent knockdown of *CEP72* was determined relative to non-targeting control at 48 and 72 h post transfection.

Accession codes. Microarray expression data were obtained from the Gene Expression Omnibus (GEO) under the accession number GSE7761[14].

URLs. LocusZoom, http://csg.sph.umich.edu/locuszoom/; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/; R, http://www.rproject.org/; SAS, http://sas.com; 1000 genome project, ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/2009_04/.

eAppendix 2. Supplemental Results

The presence of the rs924607 T-allele in the promoter of *CEP72* creates a binding sequence (CACT**T**) for the NKX transcription factor[15]. Molecular modeling of NKX-6.3 homeodomain (NKX-6.3 HD) binding to DNA sequences GCACT**T**A (risk T-allele) and GCACT**C**A (C-allele) estimates that NKX-6.3 HD binds to the variant sequence (risk allele) with markedly higher binding affinity. The relative binding free energy for the CG to TA pair (C vs T allele) is -7.67 kcal/mol, which is equivalent to over 400 thousand-fold higher binding affinity when the GC pair is mutated to a TA pair (eFigures 6d and 6e, eTable 10). The C to T transition changes the flexibility of the target DNA duplex, due to fewer hydrogen bonds between base pairs in the T-allele (two hydrogen bonds for the AT pair versus three for the GC pair), thereby enhancing binding affinity with the T-allele. This is in agreement with protein-DNA binding affinity changes being predominantly driven by DNA structural plasticity changes when the C to T transition occurs[16]. Greater binding of NKX-6.3 protein to duplex DNA containing the T-allele was confirmed by electrophoretic mobility shift assay (EMSA) (eFigure 6f).

As depicted in eFigure 6g and 6h, reduction of expression of *NKX-6.3* rescued the lower expression of CEP72 when the promoter contains the T-allele of rs924607. Taken together with the molecular modeling and EMSA results, these data indicate that NKX-6.3 inhibits CEP72 expression by binding with much greater affinity to the CEP72 promoter region containing the T-allele of rs924607.

eTable 1. National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) Version 1.0 Used to Assess Peripheral Neuropathy in the St. Jude Total XIIIB Cohort

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
NCI CTCAE v 1.0 Neuropathy- motor	Subjective weakness, no objective findings	Mild objective weakness, no significant loss of function	Objective weakness with loss of function	Paralysis	Death
NCI CTCAE v 1.0 Neuropathy-sensory	Mild paresthesias, loss of deep tendon reflexes	Mild or moderate objective sensory loss; moderate paresthesias	Severe objective sensory loss or paresthesias that interfere with function	--	Death

eTable 2. Modified ("Balis") Pediatric Scale of Peripheral Neuropathies Used to Assess Peripheral Neuropathy in the COG AALL0433 Cohort

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Balis Neuropathy-motor	Subjective weakness, but no deficits detected on neurological exam, other than abnormal deep tendon reflexes	Weakness that alters fine motor skills (buttoning shirt, coloring, writing or drawing, using eating utensils) or gait without abrogating ability to perform these tasks	Unable to perform fine motor tasks (buttoning shirt, coloring, writing or drawing, using eating utensils) or unable to ambulate without assistance	Paralysis	Death
Balis Neuropathy-sensory	Paresthesias, pain, or numbness that do not require treatment or interfere with extremity function.	Paresthesias, pain, or numbness that are controlled by non-narcotic medications (without causing loss of function), or alteration of fine motor skills (buttoning shirt, writing or drawing, using eating utensils) or gait, without abrogating ability to perform these tasks.	Paresthesias or pain that are controlled by narcotics, or interfere with extremity function (gait, fine motor skills as outlined above), or quality of life (loss of sleep, ability to perform normal activities severely impaired).	Complete loss of sensation, or pain that is not controlled by narcotics.	Death

eTable 3. Patient Characteristics and Relation to Occurrence of Vincristine-Induced Peripheral Neuropathy During Continuation Phase of Treatment in the St Jude Total XIIIB Cohort

Characteristics[a]	Vincristine Neuropathy					p-value[c]
	No. Patients (%)				No. Episodes of Vincristine Neuropathy	
	Total Patients (n=222)	None to Grade 1	1 episode of Grade 2-4	2 or more episodes of Grade 2-4	Episodes of Grade 2-4[b]	
Age at diagnosis, y						
< 1	8 (3.6)	7	1	0	1	
1-10	145 (65.32)	101	27	17	78	0.57
> 10	69 (31.08)	50	12	7	36	
Sex						
Male	128 (57.7)	96	21	11	52	0.02
Female	94 (42.3)	62	19	13	63	
Genetically-determined Race						
White	149 (67.1)	110	25	14	66	
Black	42 (19.8)	32	8	2	15	
Asian	1 (0.45)	1	0	0	0	0.001
Hispanic	20 (9.01)	11	3	6	23	
Other	10 (4.5)	4	4	2	11	
Cumulative vincristine dosage (mg/m2), median [range]	50 [4-120]	50 [4-120]	53 [25-92]	51.5 [28-120]	52 [25-120]	0.04

[a] Data are presented as Number (%) of patients unless otherwise indicated.

[b] Number of events.

[c] p values are from univariate weighted logistic regression using each episode as an observation.

eTable 4. Patient Characteristics and Relation to Occurrence of Vincristine-Induced Peripheral Neuropathy During Continuation Phase of Treatment in the COG AALL0433 Cohort

a. Combined COG AALL0433 cohort

Characteristics[a]	Vincristine Neuropathy					p-value[c]
	No. Patients (%)				No. Episodes of Vincristine Neuropathy	
	Total Patients (n=99)	None to Grade 1	1 episode of Grade 2-4	2 or more episodes of Grade 2-4	Episodes of Grade 2-4[b]	
Age at diagnosis, y						
< 1	0	0	0	0	0	
1-10	47 (47.5)	36	7	4	25	0.17
> 10	52 (52.5)	41	4	7	48	
Sex						
Male	59 (59.6)	46	8	5	36	0.36
Female	40 (40.4)	31	3	6	37	
Genetically-determined Race						
White	60 (60.6)	46	4	10	57	
Black	1 (1)	1	0	0	0	
Asian	1 (1)	0	1	0	1	0.04
Hispanic	24 (24.2)	19	4	1	13	
Other	13 (13.1)	11	2	0	2	
Cumulative vincristine dosage (mg/m2), median [range]	12 [2-97]	10 [2-97]	12 [8-63]	33 [10-84]	34 [8-84]	0.049

[a] Data are presented as Number (%) of patients unless otherwise indicated.

[b] Number of events.

[c] p values are from univariate weighted logistic regression using each episode as an observation.

eTable 4. Patient Characteristics and Relation to Occurrence of Vincristine-Induced Peripheral Neuropathy During Continuation Phase of Treatment in the COG AALL0433 Cohort (continued)

b. COG AALL0433 Group A (1.5 mg/m^2 vincristine)

Characteristics[a]	Vincristine Neuropathy				
	No. Patients (%)				No. Episodes of Vincristine Neuropathy
	Total Patients (n=51)	None to Grade 1	1 episode of Grade 2-4	2 or more episodes of Grade 2-4	Episodes of Grade 2-4[b]
Age at diagnosis, y					
< 1	0	0	0	0	0
1-10	25 (49.0)	23	2	0	2
> 10	26 (51.0)	25	0	1	4
Sex					
Male	27 (52.9)	25	1	1	5
Female	24 (47.1)	23	1	0	1
Genetically-determined Race					
White	28 (54.9)	26	1	1	5
Black	0 (0)	0	0	0	0
Asian	0 (0)	0	0	0	0
Hispanic	15 (29.4)	15	0	0	0
Other	8 (15.7)	7	1	0	1
Cumulative vincristine dosage (mg/m2), median [range]	14 [4-78]	11 [4-78]	37 [34-40]	22 [22-22]	22 [22-40]

[a] Data are presented as Number (%) of patients unless otherwise indicated.

[b] Number of events.

eTable 4. Patient Characteristics and Relation to Occurrence of Vincristine-Induced Peripheral Neuropathy During Continuation Phase of Treatment in the COG AALL0433 Cohort (continued)

c. COG AALL0433 Group B (2.0 mg/m^2 vincristine)

Characteristics[a]	Vincristine Neuropathy				
	No. Patients (%)				No. Episodes of Vincristine Neuropathy
	Total Patients (n=48)	None to Grade 1	1 episode of Grade 2-4	2 or more episodes of Grade 2-4	Episodes of Grade 2-4[b]
Age at diagnosis, y					
< 1	0	0	0	0	0
1-10	22 45.8)	13	5	4	23
> 10	26 (54.2)	16	4	6	44
Sex					
Male	32 (66.7)	21	7	4	31
Female	16 (33.3)	8	2	6	36
Genetically-determined Race					
White	32 (66.7)	20	3	9	52
Black	1 (2.08)	1	0	0	0
Asian	1 (2.08)	0	1	0	1
Hispanic	9 (18.8)	4	4	1	13
Other	5 (10.4)	4	1	0	1
Cumulative vincristine dosage (mg/m2), median [range]	29 [2-97]	10 [2-97]	10 [8-63]	34.5 [10-84]	36 [8-84]

[a] Data are presented as Number (%) of patients unless otherwise indicated.

[b] Number of events.

eTable 5. Top SNPs Associated With Vincristine-Induced Peripheral Neuropathy

SNP	Chr	Position	Nearest Gene	Allele Major/minor	St Jude Total XIIIB			COGAALL0433			Meta-analysis (univariate)	Meta-analysis (multivariate) [b]
					MAF (%)	OR	P-value	MAF (%)	OR	P-value		
Rs924607[a]	5	663093	CEP72	C/T	36.7	2.43	$1.25x10^{-6}$	36.4	4.1	0.0004	$6.33x10^{-9}$	$4.68x10^{-8}$
Rs17032980	2	67156247	ETAA1	A/G	26.6	3.17	$3.67x10^{-6}$	19.2	10.4	0.0002	$9.77x10^{-9}$	$9.01x10^{-7}$
Rs12786200	11	91618666	MTNR1B	C/T	22.7	0.23	$1.58x10^{-7}$	20.7	0.24	0.016	$2.56x10^{-8}$	$6.30x10^{-7}$
Rs4463516[a]	9	32857481	TMEM215	C/G	33.6	2.89	$6.83x10^{-7}$	24.2	4.94	0.004	$3.02x10^{-8}$	$1.19x10^{-7}$
Rs4878505[a]	9	32859257	TMEM215	G/T	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.19x10^{-7}$
Rs6476384[a]	9	32855610	TMEM215	C/A	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs10758164[a]	9	32856286	TMEM215	A/G	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs7021305[a]	9	32857081	TMEM215	C/A	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs7021527[a]	9	32857203	TMEM215	G/A	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs4626695[a]	9	32857310	TMEM215	G/A	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs1326919[a]	9	32857694	TMEM215	A/G	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs1326920[a]	9	32857718	TMEM215	C/G	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs2031671[a]	9	32858015	TMEM215	T/C	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs2031672[a]	9	32858057	TMEM215	C/T	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs1409693[a]	9	32858571	TMEM215	T/C	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs1409694[a]	9	32858746	TMEM215	C/T	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs2860094[a]	9	32859057	TMEM215	C/T	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs10118373[a]	9	32857694	TMEM215	G/T	33.6	2.89	$6.83x10^{-7}$	23.7	4.94	0.004	$3.14x10^{-8}$	$1.28x10^{-7}$
Rs7818688[a]	8	96093258	NDUFAF6	C/A	12.6	4.26	$3.05x10^{-7}$	14.1	4.59	0.014	$4.46x10^{-8}$	$5.03x10^{-7}$
Rs7832570[a]	8	96092971	NDUFAF6	A/T	12.6	4.26	$3.05x10^{-7}$	14.1	4.59	0.014	$4.46x10^{-8}$	$5.03x10^{-7}$

MAF = minor allele frequency; OR=odds ratio (allelic); Chr= Chromosome; [a] = Imputed snp; [b]= adjusted for genetically defined race and for vincristine dosage

The SNPs were selected based on their univariate meta-analysis p-values and the SNPs with the lowest p-values were represented. These SNPs were among the 5051 typed SNPs and 10195 imputed SNPs that had a p-value <0.05 in each patient cohort when analyzed separately (and with the same effect on neuropathy in both patient cohorts).

eTable 6. Allele Frequencies of rs924607 in the Different Race Groups

a. Combined cohort

Patients	Total XIIIB cohort					COG cohort				
	White	Black	Asian	Hispanic	Others	White	Black	Asian	Hispanic	Others
C allele frequency(%)	179(60.05)	69(82.1)	2(100)	20 (50)	11(55)	69(57.5)	2 (100)	2(100)	37(77)	16 (61.5)
T allele frequency(%)	119(39.95)	15(17.9)	0(0)	20 (50)	9(45)	51(42.5)	0 (0)	0(0)	11(23)	10(38.5)
Total number of patients	149	42	1	20	10	60	1	1	24	13
Total alleles	298	84	2	40	20	120	2	2	48	26

b. HapMap populations

Individuals	CEU	GIH	CHB	CHD	JPT	MEX	YRI	ASW	LWK	MKK
C allele frequency(%)	207(63)	131(64.85)	156(56.9)	158(72.5)	161(71.24)	125(72.7)	372(91.6)	158(90.8)	210(95.45)	334(90.76)
T allele frequency(%)	123(37)	71(35.15)	118(43.1)	60(27.5)	65(28.76)	47(27.3)	34(8.4)	16(9.2)	10(4.55)	34(9.24)
Total number of Individuals	165	101	137	109	113	86	203	87	110	184
Total alleles	330	102	274	218	226	172	406	174	220	368

CEU=Utah Residents with Northern and Western European Ancestry; GIH= Gujarati Indians in Houston, Texas; CHB= Han Chinese in Beijing, China; CHD=Chinese in Metropolitan,Denver, Colorado; JPT= Japanese in Tokyo, Japan; MEX= Mexican ancestry in Los Angeles; YRI =Yoruba in Ibadan, Nigeria; ASW = African ancestry in Southwest USA; LWK= Luhva in Webuve, Kenya; MKK= Maasai in Kinyama, Kenya

The allele frequencies (in %) of the C-allele and the risk (T) allele of rs924607 are presented for the combined St. Jude Total XIIIB and for COG AALL0433 cohort (a) and for the HapMap populations (b).

eTable 7. Hazard Ratio Estimates by Cox Regression Model (With Time to Failure Defined as Time to the First Episode of Grade 2 or Higher Vincristine-Induced Neuropathy) in Combined Cohort of Total XIIIB and COG AALL0433 Group A (1.5 mg/m^2 Vincristine), Along With Number of Patients With Neuropathy and Number of Neuropathy Episodes in Each Factor Level

Factor	Hazard Ratio	95% CI of Hazard Ratio	*P* value	Number of patients with neuropathy (at least 1 episode)	Total neuropathy episodes
Rs924607 TT (n=38)	3.58	2.10-6.10	<.0001	21	52
Rs924607 CT or CC (n=235)	Reference (1)	Reference (1)		46	69
Cumulative dose of vincristine	1.00	0.99-1.02	0.57	67	121
White (n=177)	Reference (1)	Reference (1)	Reference (1)	41	71
Asian (n=19)	0.20	0.00-1.94	0.64	10	15
Black (n=42)	0.95	0.44-2.06	0.91	7	12
Hispanic (n=35)	1.45	0.11-1.87	0.78	9	23

Rs924607 CT or CC is used as reference compared to Rs924607 TT; white is used as reference compared to the other race groups. Reference has a value of 1.0. n, the number of patients. The cumulative dose of vincristine is a continuous variable expressed in mg/m^2 with reference increment of 1mg/m^2.

eTable 8. Accelerated Failure Time (AFT) Regression Modeling of Time to the First Episode of Grade 2 or Higher of Vincristine Related Neuropathy by the rs924607 Genotypes (TT vs. CT or CC) Adjusting for Cumulative Dose and Genetically Determined Ancestry

Factor	Effect Estimate[a]	Z statistic	*P* value
Rs924607, TT	0.78	-4.52	$6.19.\ 10^{-6}$
Rs924607, CT or CC	Reference (1)	Reference (1)	
Cumulative dose	1.00	-1.10	0.27
Black	1.11	1.10	0.27
Hispanic	1.23	0.25	0.80
White	1.12	1.42	0.15
Asian/other	Reference (1)	Reference (1)	Reference (1)

[a]Exponential of the regression coefficient estimates from log-transformed data, representing the factor's multiplicative effect on median onset time. Rs924607 CT or CC is used as reference compared to rs924607 TT; Asian/other is used as reference compared to the other race groups. Reference has a value of 1.0.

eTable 9. Association of *CEP72* SNP With Severity (Grade) of Vincristine-Induced Neuropathy

Parameter	Effect Estimate[a]	Wald 95% Confidence Intervals		Wald Chi-Square	*P*-value
		Est_Lower	Est_Upper		
rs924607 TT	2.42	1.59	3.69	16.99	<0.0001
rs924607 CT or CC	Reference (1)	Reference	Reference	Reference	
Cumulative dose	1.02	1.01	1.02	10.15	0.001
Asian	0.00	0.00	4640.6	0.96	0.33
Black	1.02	0.55	1.91	0.01	0.94
Hispanic	1.08	0.14	8.45	0.01	0.94
White	Reference (1)	Reference	Reference	Reference	Reference

[a] Exponential of the regression coefficient estimates from Poisson regression model (log of mean), representing the factor's multiplicative effect on mean

CEP72 SNP genotype, cumulative vincristine dose, genetic ancestry information were used as covariates. Rs924607 CT or CC is used as reference compared to Rs924607 TT; white is used as reference compared to the other race groups. Reference has a value of 1.0.

eTable 10. Analysis of Free Energy Components of the NKX-6.3 Homeodomain Bound With Wild Type (C) and Mutated (T) DNA Molecules

	GCACTTA(mutated)	GCACTCA(wild-type)
[a] ΔH	-225.20(12.59)	-230.02(18.1)
[b] S_{DNA}	3223.91(11.36)	3226.85(12.59)
[b] $S_{Complex}$	7947.44(15.01)	7908.46(16.63)
[b] $\Delta\Delta S_{TA\text{-}CG}$	-41.90	
[a] $\Delta\Delta H_{TA\text{-}CG}$	4.82	
[a] $T\Delta\Delta S_{TA\text{-}CG}$	12.49	
[a] $\Delta\Delta G_{TA\text{-}CG}$	-7.67	

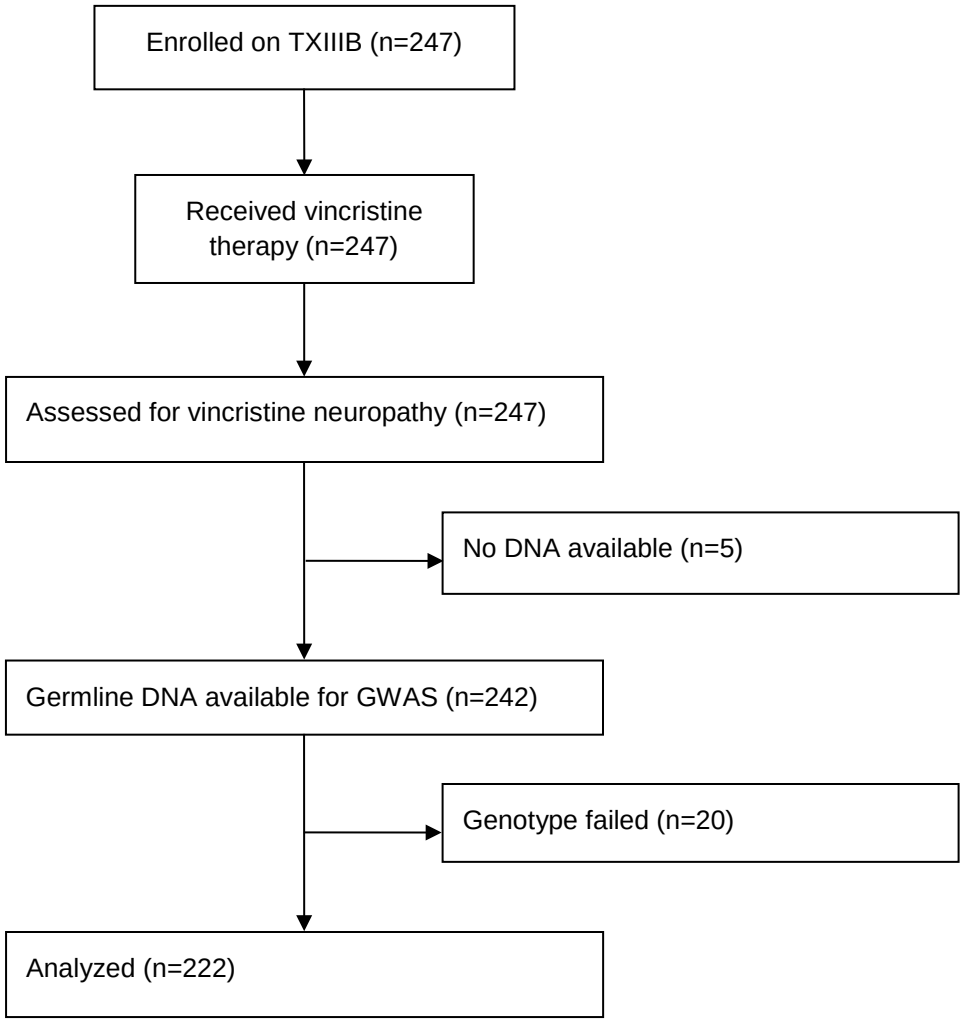
[a] All units are kcal/mol

[b] All units are cal/(mol.K)

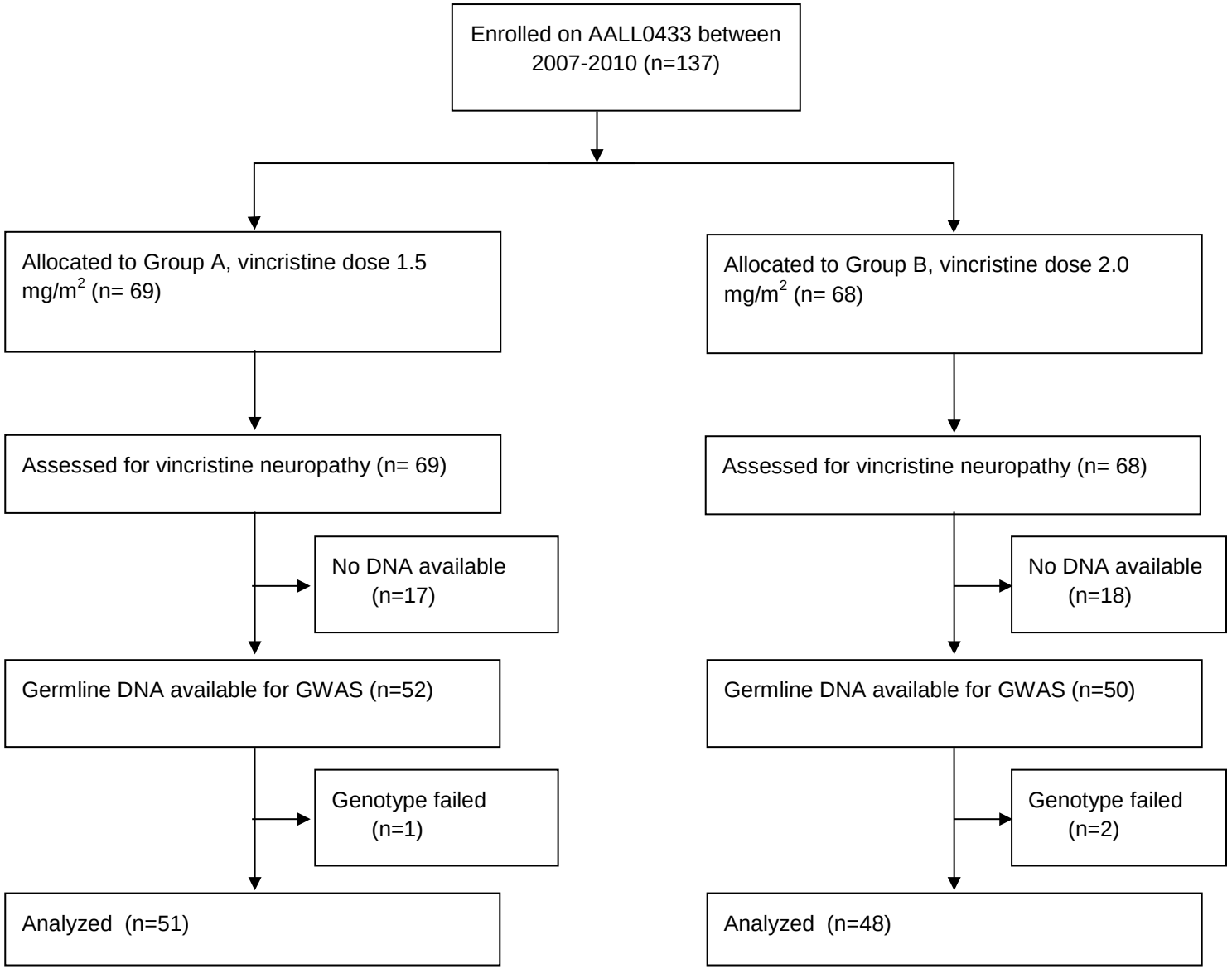
ΔH =change of enthalpy; S_{DNA} =entropy of DNA; $S_{Complex}$ =entropy of the complex; $\Delta\Delta S_{TA\text{-}CG}$ =difference of the binding entropy change; $\Delta\Delta H_{TA\text{-}CG}$ =difference of the binding enthalpy change; $T\Delta\Delta S_{TA\text{-}CG}$ =multiplication of temperature with difference of the binding entropy change; $\Delta\Delta G_{TA\text{-}CG}$ = relative binding free energy

Standard deviations calculated from ensembles are shown in parentheses.

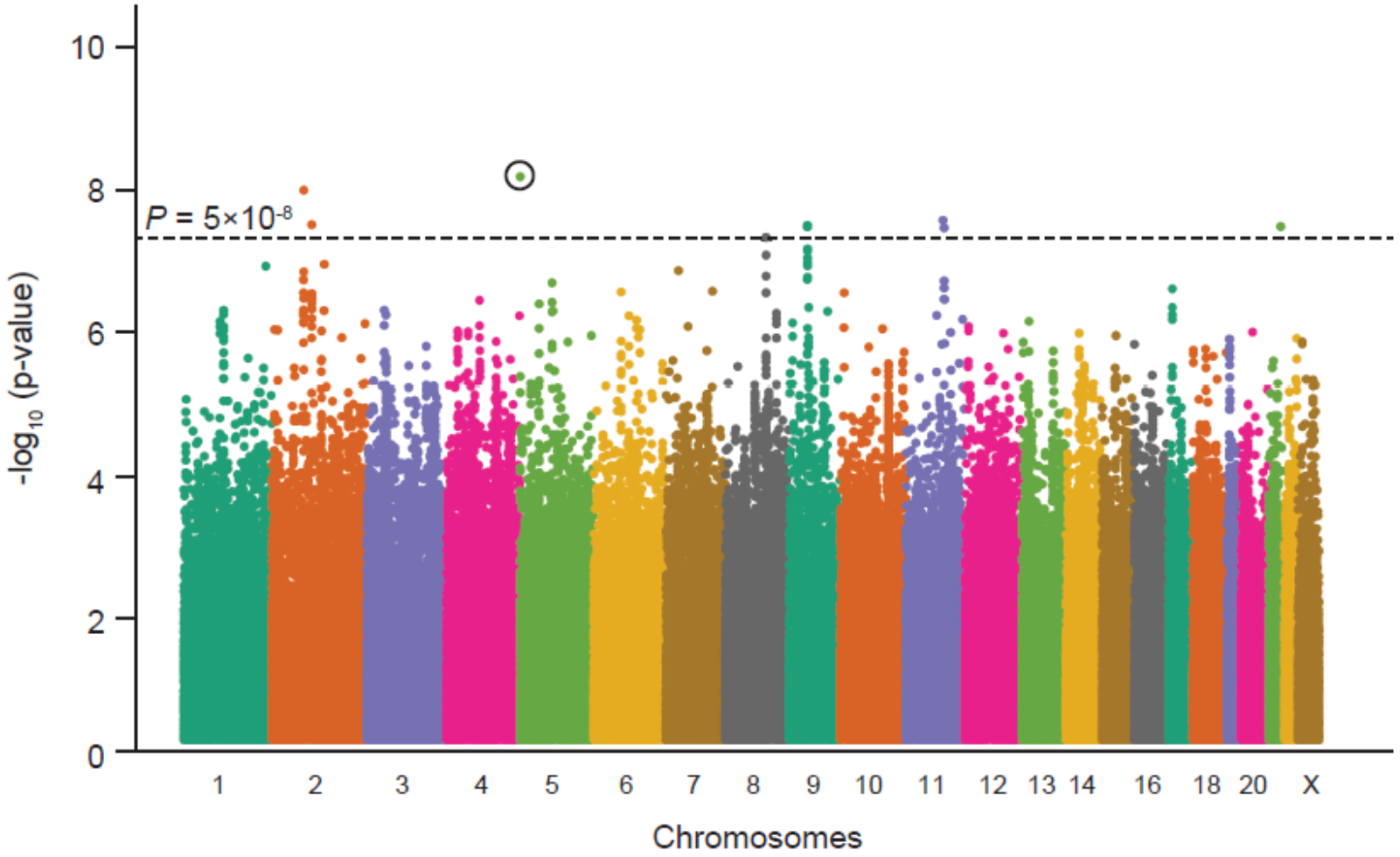
eFigure 1. CONSORT Flow Diagram of St Jude Total XIIIB Cohort



eFigure 2. CONSORT Flow Diagram of COG AALL0433 Cohort

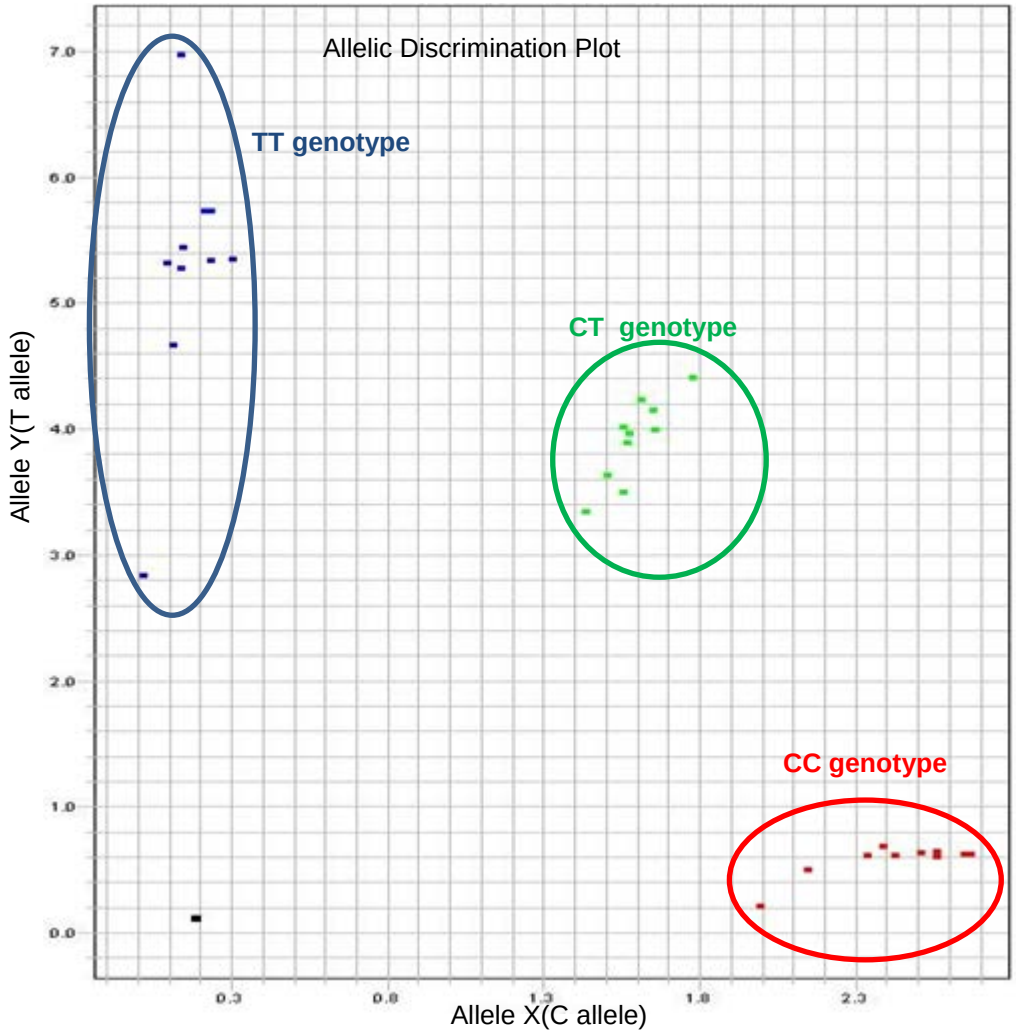


eFigure 3. Significance of SNPs Associated With Vincristine-Induced Peripheral Neuropathy



Single nucleotide polymorphism (SNP) Manhattan plot depicting the $-\log_{10} P$ value for the meta-analysis showing the association of SNPs with the occurrence of vincristine-induced neuropathy during treatment for childhood acute lymphoblastic leukemia, with SNPs displayed according to chromosomal location. The horizontal line marks the threshold for genome-wide statistical significance. The SNP (rs924607) with the lowest meta-analysis *P* value is circled.

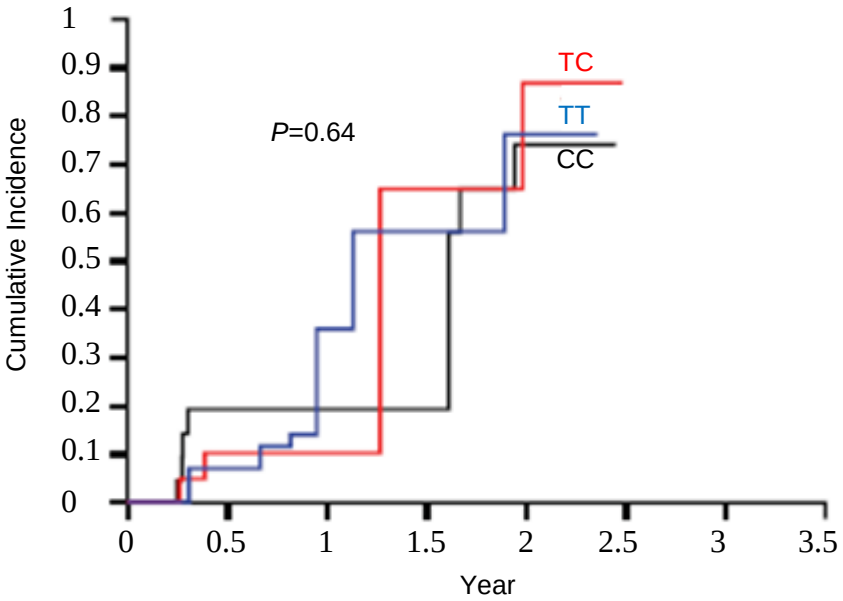
eFigure 4. Genotyping Using the Allelic Discrimination Assay



CC genotype (10 dots corresponding to 10 patients)
CT genotype (10 dots corresponding to 10 patients)
TT genotype (10 dots corresponding to 10 patients)
Nontemplate control (one dot)

Plot of the fluorescence was obtained during the plate read on an ABI PRISM 7900HT Sequence detection system. The Y and X axis scale represent the absolute fluorescence intensity. Each dot represents each well indicating the alleles in each sample. Blue, green and red represent respectively the samples with the TT, CT, or CC genotype corresponding to the 30 patients with the imputed genotypes with 10 for each genotype.

eFigure 5. Cumulative Incidence of Vincristine-Induced Neuropathy in the COG Cohort Group B



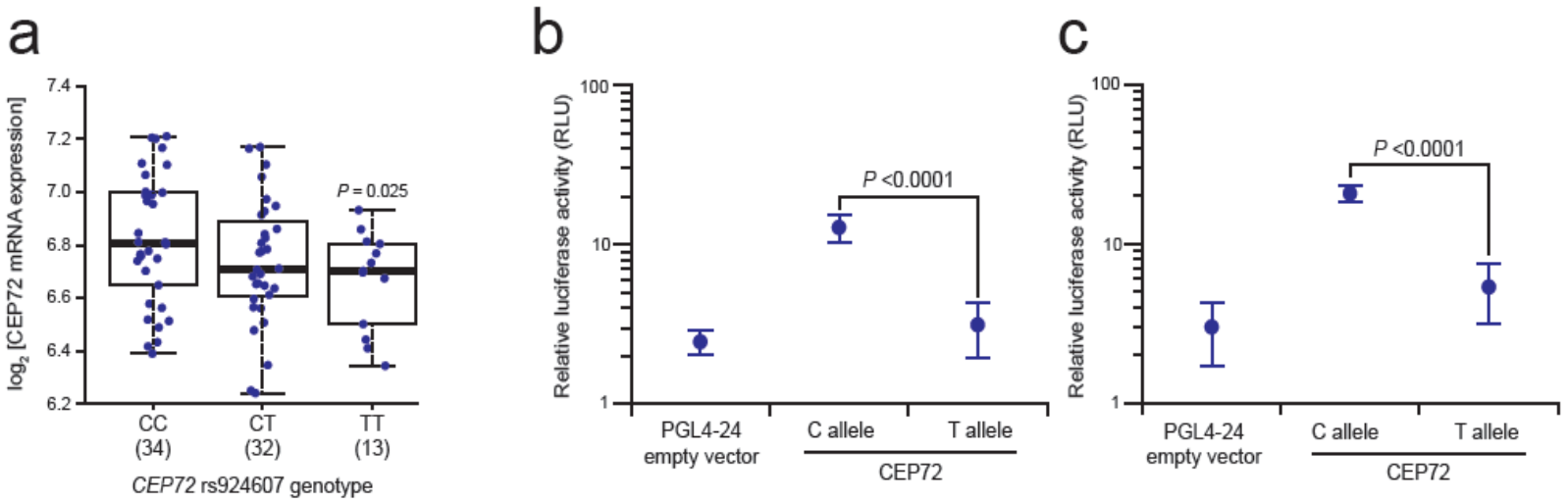
Number at risk:

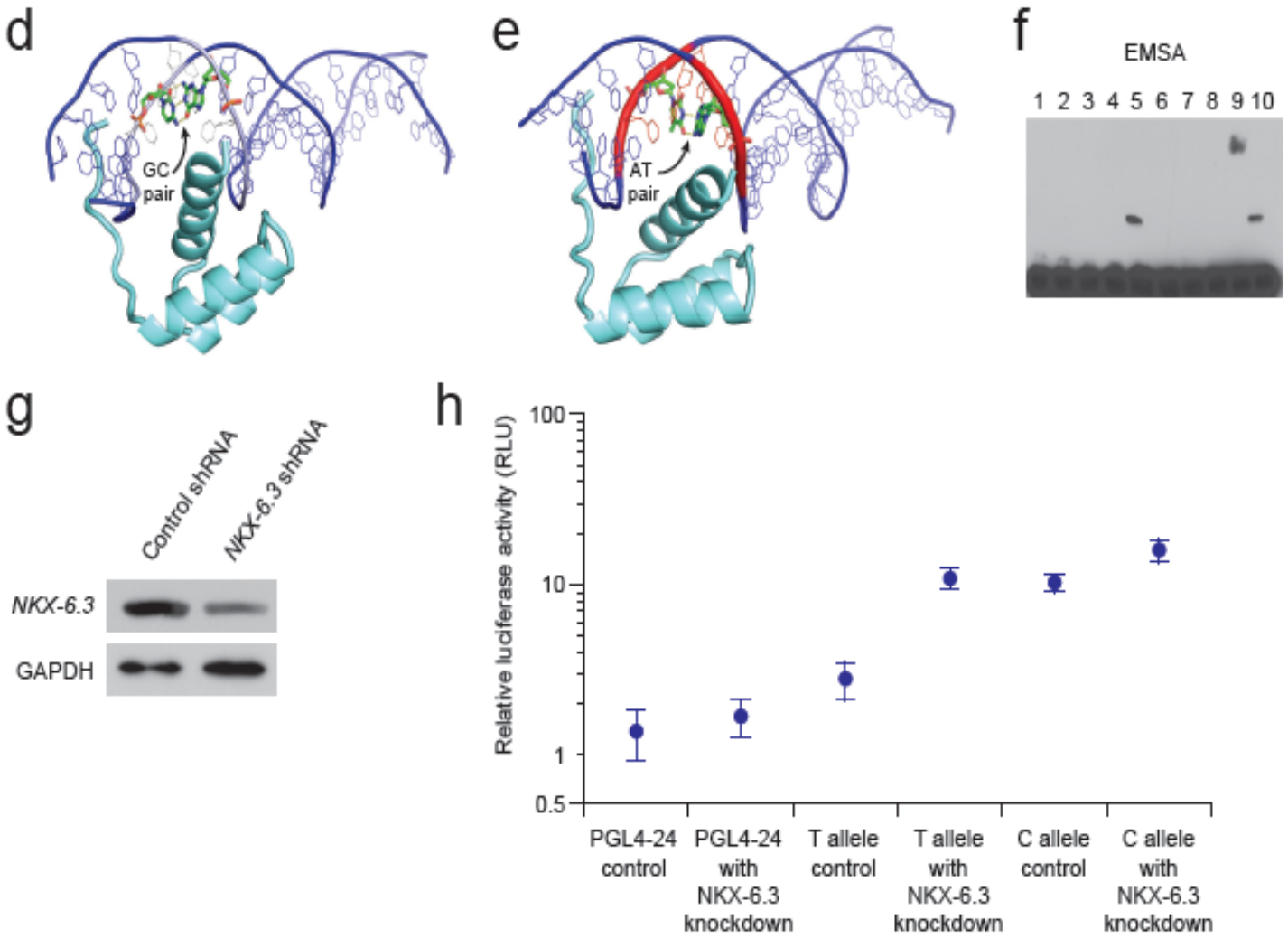
CC 21	7	3
CT 15	5	2
TT 12	3	1

P value was computed based on weighted logistic regression adjusting for ancestry and accounting for multiple episodes of neuropathy. The horizontal axis represents the years since the start of the continuation treatment period, and time 0 is the beginning of this continuation treatment that includes multiple doses of vincristine. The number of patients evaluable and at risk of neuropathy is provided for each *CEP72* genotype (diplotype).

eFigure 6. *CEP72* Promoter SNP Genotype (rs924607) and *CEP72* Expression

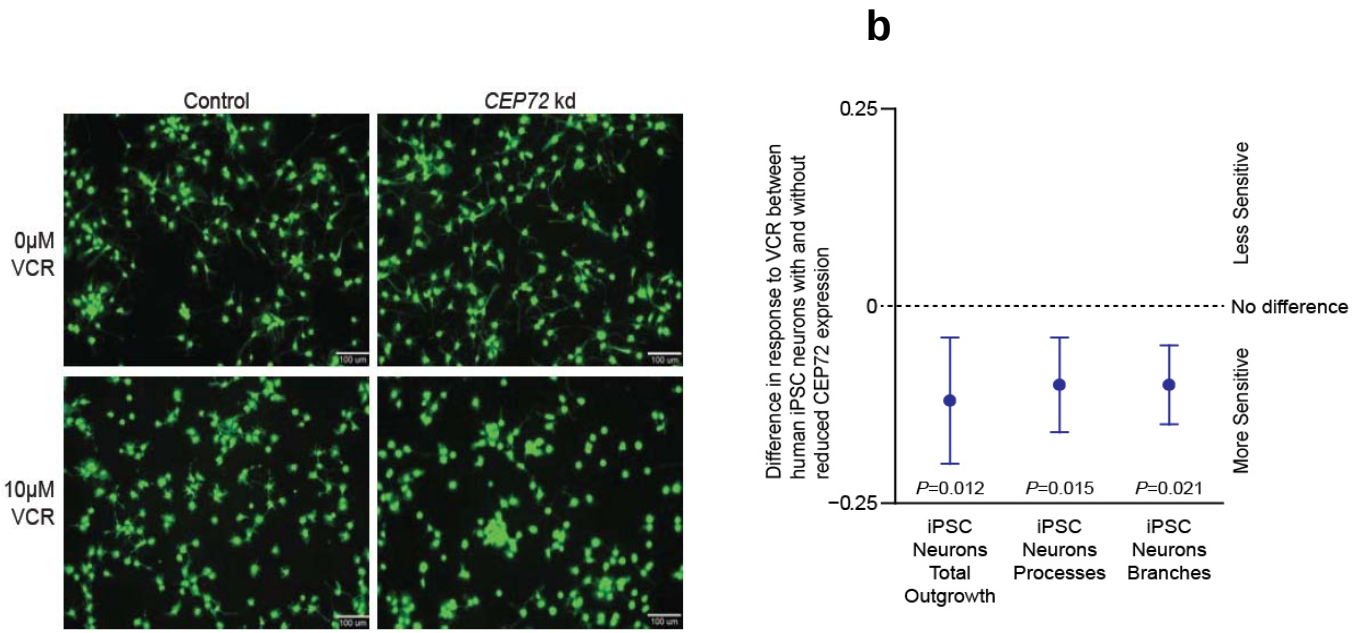
(a) In HapMap cell lines (CEU), eQTL analysis shows significant association between rs924607 genotype and expression of *CEP72* in CEU HapMap cell lines ($p < 0.05$). The horizontal line inside each box depicts the median; the upper and lower limits of the box are the 75th and 25th percentiles, respectively; the vertical bars above and below each box indicate the maximum and minimum values, respectively. The blue circles depict individual expression values. The *CEP72* promoter construct with the C- or T-alleles of rs924607 was generated and sub-cloned into the pGL4.24 vector encoding the luciferase reporter gene *luc2P.* The pGL4.24 firefly luciferase alone or with the *CEP72* promoter containing the C-allele or the T-allele of rs924607 were co-transfected with the renilla luciferase in human neuroblastoma SHSY5Y cells (b) or human leukemia Nalm6 cells (c). The relative luciferase activity was obtained after normalizing the firefly signal by the renilla luciferase. The relative luciferase activity is presented as the mean of three replicate experiments, using the mean of each experiment run in triplicate. The error bars represent 95% CI. Student's t test was used to examine the differences in luciferase activity between the *CEP72* C-allele (reference allele) and T-allele. The structure of the NKX-6.3 homeodomain (green) with target DNA duplex (dark blue) containing either the C-allele (d) or the T-allele (e) of rs924607 was constructed from homology modeling according to the crystal structure of the NKX-2.5-DNA complex. The C to T mutation causes changes in flexibility as depicted by the sausage representation, with the larger radii depicted by the red color with the T-A duplex. (f) Electrophoretic mobility shift assay for rs924607 was performed using biotin-labeled duplex containing either the C-allele (lines 1, 2, 3, 7 and 8) or the T-allele (lines 4, 5, 6, 9 and 10). NKX.6-3 protein was added to the assay (lines, 2, 3, 5, 6, 7, 8, 9, 10). Excess of unlabeled duplexes with C-allele (line 3) or T-allele (line 6) was used to compete with biotin-labeled duplexes. Antibody (lines 7 and 9) or IgG control (lines 8 and 10) were added to the reactions. The super-shifted band (line 9) indicates that NKX-6.3 binds with a higher affinity to the T allele compared to the C allele. (g) *NKX-6.3* expression was reduced by using shRNA (compared to the control using no targeting shRNA) in Nalm6 leukemia cells. GAPDH was used as a loading control. (h) Luciferase activity was determined after knockdown of *NKX-6.3* or in presence of the shRNA control (scrambled sequence). The relative luciferase activity was obtained after normalizing the firefly signal by the renilla luciferase. The relative luciferase activity is presented as the mean of three different experiments and each experiment was performed in triplicate. The error bars represent 95% CIs.

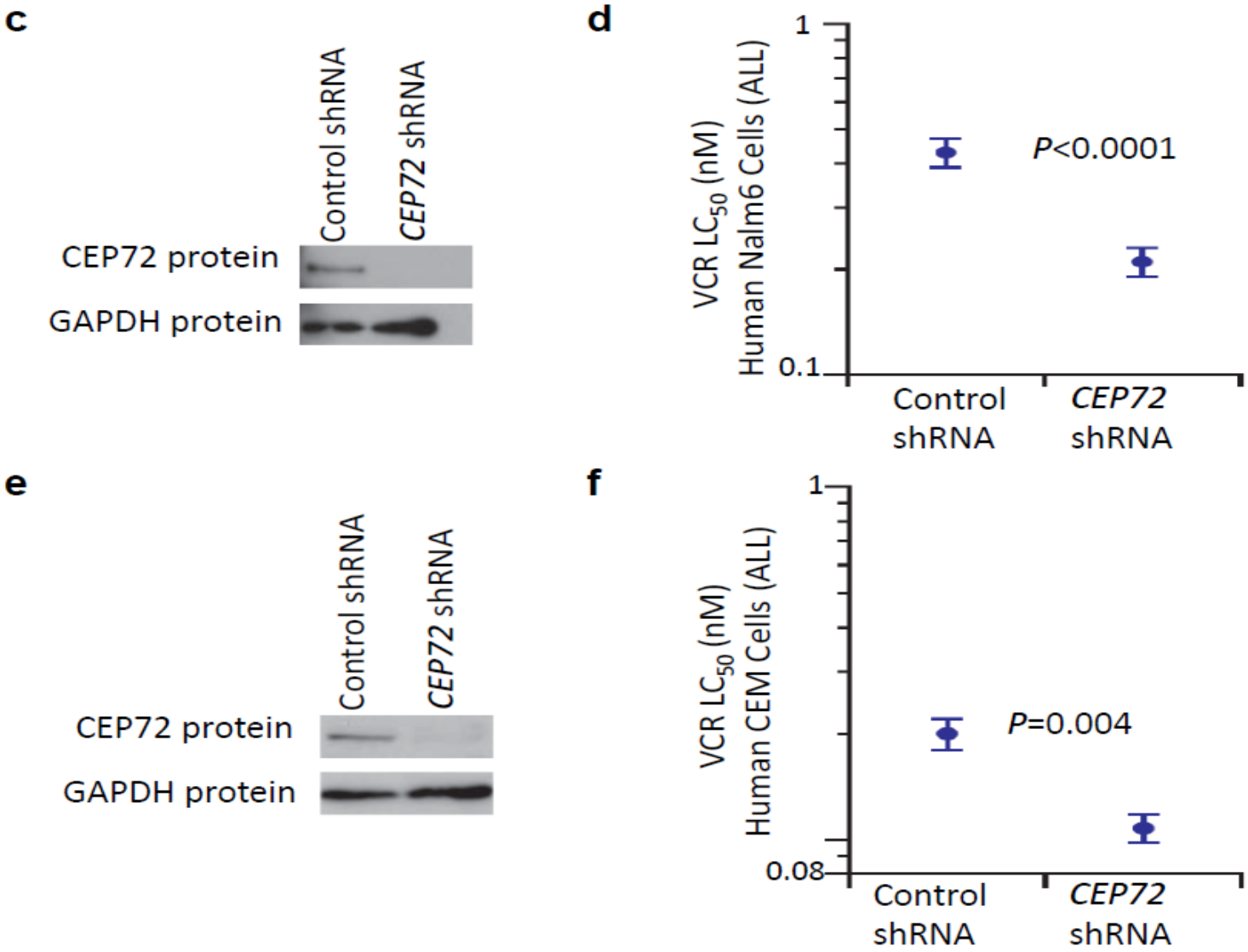




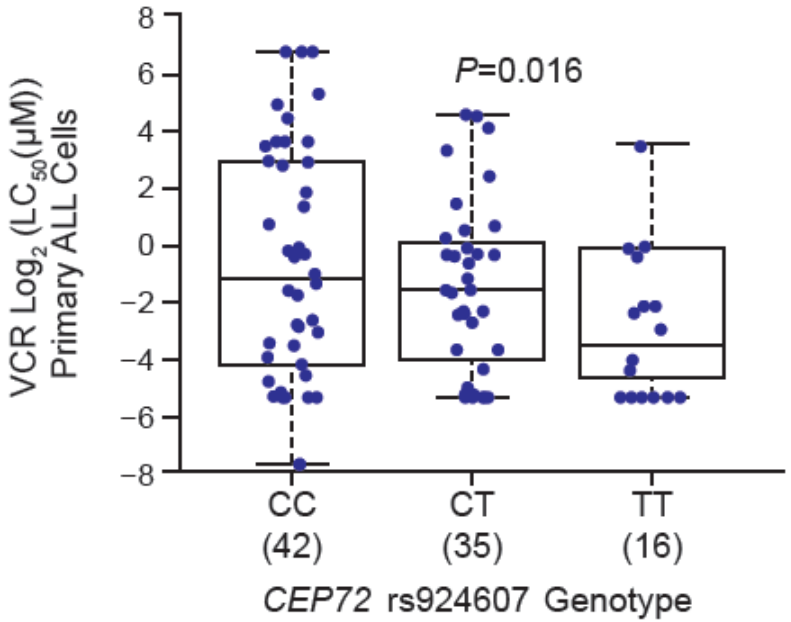
eFigure 7. CEP72 and Vincristine Sensitivity in iPSC Neurons and Human Leukemia Cells

(a) Vincristine-induced changes in neurite outgrowth phenotypes in human iPSC-derived neurons are enhanced by lower *CEP72* expression. Representative images of neurons from the *CEP72* siRNA experiment. Clockwise from upper left: Control (small interfering RNA (siRNA) with no vincristine (VCR) treatment, *CEP72* siRNA (knockdown = kd) with no VCR treatment, *CEP72* siRNA (kd) with 10 μM VCR, and control siRNA with 10 μM VCR. Cells were treated with VCR in PBS (or PBS along for controls) for 48 h, stained with Calcein AM, Hoechst 33343 and imaged at 10X. (b) Point estimates and 95% confidence intervals of differences in each neuron characteristic (i.e., total outgrowth, branches, processes) after vincristine treatment of iPSC neurons in which *CEP72* expression was reduced (by siRNA) compared to control neurons in which *CEP72* expression was not reduced. For each experiment (performed in triplicate), neuron characteristics were measured in 1000 cells. Values below zero (mean and 95% CI) indicate greater vincristine effects in neurons with reduced *CEP72* expression. P-values were determined by a paired t test. (c) Reduce expression (Knockdown) of *CEP72* in human Nalm6 leukemia cells or (e) in human CEM leukemia cells increased sensitivity of leukemia cells to vincristine (d) (f). Nalm6 cells and CEM cells in which *CEP72* was knocked down were treated for 72 hours in the presence of increasing concentrations of vincristine. Cell viability was determined by the MTT assay. Lethal concentration to 50% of cells (LC_{50}) values represent at least three independent experiments and error bars represent 95% CIs. (g) Primary ALL cells isolated from bone marrow of St Jude newly diagnosed patients with different genotypes of *CEP72* rs924607 were tested for sensitivity to vincristine using the MTT assay. The horizontal lines in each boxplot indicate the mean LC_{50} (μM) values for each genotype and the bottom and top of the boxes depict the 25th and 75th percentile. The bars represent the maximum and the minimum values.





g.



eReferences

1. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature.* Oct 28 2010;467(7319):1061-1073.
2. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology.* Dec 2010;34(8):816-834.
3. Yang JJ, Cheng C, Devidas M, et al. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nature genetics.* Mar 2011;43(3):237-241.
4. Li YaA, G. Rapid haplotype reconstruction and missing genotype inference. *American journal of human genetics.* 2006;S79(2290).
5. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* Oct 18 2007;449(7164):851-861.
6. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* Jun 2000;155(2):945-959.
7. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* Nov 1 2012;491(7422):56-65.
8. Kishi S, Cheng C, French D, et al. Ancestry and pharmacogenetics of antileukemic drug toxicity. *Blood.* 2007;109(10):4151-4157.
9. Schleinitz D, Distefano JK, Kovacs P. Targeted SNP genotyping using the TaqMan(R) assay. *Methods in molecular biology (Clifton, N.J.).* 2011;700:77-87.
10. Case DA, Darden TA, Cheatham I, T.E. , et al. AMBER 12, University of California, San Francisco. 2012.
11. Duan Y, Wu C, Chowdhury S, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of computational chemistry.* Dec 2003;24(16):1999-2012.
12. Kuhn B, Kollman PA. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *Journal of medicinal chemistry.* Oct 5 2000;43(20):3786-3791.
13. Holleman A, Cheok MH, den Boer ML, et al. Gene-expression patterns in drug-resistant acute lymphoblastic leukemia cells and response to treatment. *N Engl J Med.* . 2004;351(6):533-542.
14. Huang RS, Duan S, Shukla SJ, et al. Identification of genetic variants contributing to cisplatin-induced cytotoxicity by use of a genomewide approach. *American journal of human genetics.* Sep 2007;81(3):427-437.
15. Jolma A, Yan J, Whitington T, et al. DNA-binding specificities of human transcription factors. *Cell.* Jan 17 2013;152(1-2):327-339.
16. Seldeen KL, McDonald CB, Deegan BJ, Bhat V, Farooq A. DNA plasticity is a key determinant of the energetics of binding of Jun-Fos heterodimeric transcription factor to genetic variants of TGACGTCA motif. *Biochemistry.* Dec 29 2009;48(51):12213-12222.