A machine learning model using SNPs obtained from a genome-wide association study predicts the onset of vincristine-induced peripheral neuropathy

Short title: Machine learning and GWAS for peripheral neuropathy

Hiroki Yamada[1#], Rio Ohmori[1#], Naoto Okada[2], Shingen Nakamura[3], Kumiko Kagawa[4], Shiro Fujii[4], Hirokazu Miki[5], Keisuke Ishizawa[2,6,7], Masahiro Abe[4], Youichi Sato[1*]

# These authors contributed equally to this work.

[1] Department of Pharmaceutical Information Science, Tokushima University Graduate School of Biomedical Sciences, Tokushima, 770-8505, Japan

[2] Department of Pharmacy, Tokushima University Hospital, Tokushima, 770-8503, Japan

[3] Department of Community Medicine and Medical Science, Tokushima University Graduate School of Biomedical Sciences, Tokushima, 770-8503, Japan

[4] Department of Hematology, Endocrinology and Metabolism, Tokushima University Graduate School of Biomedical Sciences, Tokushima, 770-8503, Japan

[5] Division of Transfusion Medicine and Cell Therapy, Tokushima University Hospital, Tokushima, 770-8503, Japan

[6] Department of Clinical Pharmacology and Therapeutics, Tokushima University Graduate School of Biomedical Sciences, Tokushima, 770-8503, Japan

[7] Clinical Research Center for Developmental Therapeutics, Tokushima University Hospital, Tokushima, 770-8503, Japan

*Address correspondence to: Youichi Sato, Ph.D.

Department of Pharmaceutical Information Science, Tokushima University Graduate School of Biomedical Sciences

1-78-1 Sho-machi, Tokushima, Tokushima 770-8505, Japan

Phone: +81-88-633-7253; FAX: +81-88-633-7253

E-mail: youichi.sato@tokushima-u.ac.jp

Abstract

Vincristine treatment may cause peripheral neuropathy. In this study, we identified the genes associated with the development of peripheral neuropathy due to vincristine therapy using a genome-wide association study (GWAS) and constructed a predictive model for the development of peripheral neuropathy using genetic information-based machine learning. The study included 72 patients admitted to the Department of Hematology, Tokushima University Hospital, who received vincristine. Of these, 56 were genotyped using the Illumina Asian Screening Array-24 Kit, and a GWAS for the onset of peripheral neuropathy caused by vincristine was conducted. Using Sanger sequencing for 16 validation samples, the top three single nucleotide polymorphisms (SNPs) associated with the onset of peripheral neuropathy were determined. Machine learning was performed using the statistical software R package "caret." The 56 GWAS and 16 validation samples were used as the training and test sets, respectively. Predictive models were constructed using random forest, support vector machine, naive Bayes, and neural network algorithms. According to the GWAS, rs2110179, rs7126100, and rs2076549 were associated with the development of peripheral neuropathy on vincristine administration. Machine learning was performed using these three SNPs to construct a prediction model. A high accuracy of 93.8% was obtained with the support vector machine and neural network using rs2110179 and rs2076549. Thus, peripheral neuropathy development due to vincristine therapy can be effectively predicted by a machine learning prediction model using SNPs associated with it.

Keywords: genome-wide association study; peripheral neuropathy; vincristine; hematopoietic tumor; machine learning

Introduction

Vincristine is an antineoplastic drug used for the treatment of many cancers, such as leukemia and malignant lymphoma. It exhibits its anticancer effects by binding to tubulin to prevent chromosome segregation, ultimately causing cell death [1]. However, it can cause serious side effects, especially peripheral neuropathy, which has been identified as a dose-limiting toxicity of vincristine [2]. Vincristine can cause peripheral, progressive, and symmetric nerve damage due to the destruction of microtubule structures, inflammatory processes, and axonal dysfunction [3]. Peripheral neuropathy caused by vincristine may decrease the quality of life of patients, cause delay or discontinuation of chemotherapy, and increase medical expenses [4]. Therefore, it is important to identify markers that can predict the occurrence of side effects to avoid them.

Nowadays, genome-wide association studies (GWAS) are being conducted to identify genetic factors that influencing susceptibility to complex diseases. This can statistically analyze the association between gene polymorphisms and specific traits [5]. For instance, Human leukocyte antigen (*HLA*)-A*3101 alleles have been associated with the risk of carbamazepine-induced adverse skin reactions by GWAS in the Japanese population [6], and rs9263726 of psoriasis susceptibility 1 candidate 1 (*PSORS1C1*), an alternative biomarker of *HLA*-B*5801, has been reported as a predictor of allopurinol-related Stevens-Johnson syndrome and toxic epidermal necrolysis [7]. However, such success is rare, and many SNPs discovered by GWAS explain only a small fraction of cases, with few SNPs progressing to clinical application as predictive markers.

In recent years, artificial intelligence (AI) technology has progressed, and its practical applications are gaining attention in the medical industry [8]. Recent examples include using deep learning for computed tomography image analysis in fibrous lung

disease [9] and quantitative structure–activity relationship model for drug discovery and predicting the occurrence of drug-induced liver injury [10]. AI is used in a wide range of medical fields, including drug therapy. Machine learning, which is an AI technology, is a method of letting a machine learn and discover hidden patterns in a given data and consequently predict the results for new data. Algorithms such as random forest (RF), support vector machine (SVM), naive Bayes (NB), and neural networks (NNs) are widely used. Previously, Oyaga-Iriarte et al. developed a machine learning model to predict the occurrence of irinotecan side effects in patients with metastatic colorectal cancer [11]. Additionally, Mo et al. developed a machine learning model to predict the effectiveness of etanercept treatment in juvenile idiopathic arthritis [12]. Thus, machine learning is expected to predict side effect occurrence and drug efficacy in the medical field [13].

No effective treatment or prediction of vincristine-induced peripheral neuropathy has been established to date. A marker predicting the onset of peripheral neuropathy will enable the selection of an appropriate drug treatment for each patient and, thus, prevent the onset of peripheral neuropathy. Therefore, we hypothesized that gene polymorphisms associated with side effects could be predictive markers for peripheral neuropathy development. In addition, we considered that machine learning would be useful as a tool for utilizing gene polymorphisms related to side effect expression for side effect prediction. In recent years, gene–gene interaction has attracted attention, and it has been reported that it is possible to make better predictions using multiple single nucleotide polymorphisms (SNPs) than only a single SNP [14]. Prediction by machine learning, which can consider the effects of multiple gene polymorphisms, may yield better results than prediction by individual gene polymorphisms.

In this study, we identified genes related to the onset of peripheral neuropathy by

74 vincristine therapy using GWAS and constructed a model predicting the occurrence of
75 side effects by machine learning based on multiple SNP information that has been shown
76 to be related.
77

Methods

Subjects

This study included 72 patients admitted to the hematology department at Tokushima University Hospital between January 2015 and December 2019 and received vincristine therapy. Patients enrolled in the study were general inpatients receiving vincristine for the treatment of leukemia and malignant lymphoma, among other diseases. Peripheral neuropathy was diagnosed by the treating physician. The severity of neuropathy ranged from grade 1 to 3, but in this study, grade 1 and above was considered to be associated with side effects. These clinical data are from chart views. Other diseases such as diabetes were not investigated. Of the 72 patients, 56 were used for GWAS and 16 were used for validation. This study was approved by the Human Genome, Genetic Analysis Research ethics committees of the Tokushima University (approval reference number: H26-29, date: January 5, 2015), and the Clinical Research Ethics Committee of the Tokushima University Hospital (approval reference number: 2175, date: January 26, 2015). All participants provided written informed consent.

Genotyping and imputation

Genomic DNA was extracted from saliva samples using Oragene OG-500 Saliva Collection Kit (DNA Genotek Inc., Ontario, Canada). GWAS samples were genotyped for 659,184 markers using the Illumina Asian Screening Array V1.0 Kit (Illumina, Tokyo, Japan), following the manufacturer's instructions.

For genotyping imputation analysis, strand correction was performed using the utility program BEAGLE with genotyped data for Asian samples (JPT and CHB). Genotype imputation was performed using BEAGLE V.4.1 [15,16] with the 1000

Genomes Project Phase 3 V.5 as a reference panel. SNPs with a P-value of Hardy–Weinberg equilibrium ≤ 10^{-6}, linkage disequilibrium $R^2 > 0.8$, minor allele frequency < 0.05, and indels were excluded. Finally, 4,340,175 SNPs were used for subsequent association analyses.

For 16 validation samples, rs2110179, rs7126100, and rs2076549 SNPs were genotyped by Sanger sequencing.

Statistical analysis

Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated via logistic regression analysis using the PLINK version 1.07 software package (http://pngu.mgh.harvard.edu/~purcell/plink/) [17]. The Manhattan plot was generated using the qqman package for R software. A regional plot was created by LocusZoom using the 1000 Genomes Project Asian (ASN) data (November 2014) [18]. Significant expression quantitative trait loci (eQTLs) and splicing quantitative trait loci (sQTLs) were searched on the GTEx Portal database (http://www.gtexportal.org/home/] [19], and HaploReg V.4.1 (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php) was used for the functional annotation of nucleotide variants [20].

Machine learning

To predict the occurrence of peripheral neuropathy by vincristine, machine learning was performed using the "caret" package for the R version 3.5.0 software (https://cran.r-project.org/web/packages/caret/) [21]. Fifty-six GWAS samples as a training set and 16 validation samples as a test set were used for machine learning. The accuracy rates for three SNPs (s2110179, rs7126100, and rs2076549), two SNPs-1

126 (rs2110179 and rs7126100), two SNPs-2 (rs2110179 and rs2076549), and two SNPs-3
127 (rs7126100 and rs2076549) were compared using RF, SVM, NB, and NN algorithms.
128 Hyperparameter tuning was performed using 56 GWAS samples as learning data. We
129 performed 10-fold cross-validation (CV) and adopted a combination of hyperparameters
130 with the highest CV accuracy. The accuracy rate, sensitivity, specificity, and positive and
131 negative predictive values were evaluated.
132
133

Results

The characteristics of the 72 patients who received vincristine therapy are presented in Table 1. Peripheral neuropathy due to vincristine treatment occurred in 36 of the 56 GWAS subjects and in 14 of 16 validation samples. In the GWAS sample, a significant difference between the age of patients with and without peripheral neuropathy onset was observed. Of the 56 GWAS subjects, 10 had leukemia (acute myeloid leukemia and acute myeloid monoctic leukemia), 39 had malignant lymphoma (diffuse large B-cell lymphoma, Berkitt lymphoma, follicular lymphoma, mucosa-associated lymphoid tissue) lymphoma, peripheral T-cell lymphoma, enteropathy-type T-cell lymphoma, adult T-cell leukemia/lymphoma, intravascular large-cell B-cell lymphoma, and unclassifiable malignant lymphoma), 5 had multiple myeloma, and 1 each had myelodystrophy syndrome and granulocytic sarcoma. Of the 56 patients, 36 received either CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisolone) or R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone) therapy, and the other patients received COP [+R (rituximab), +ETP (etoposide), +THP (pirarubicin)], A-triple V (cytarabine, etoposide, vincristine, and vindesine), VAD (doxorubicin, vincristine, and dexamethasone), or EPOCH (etoposide, prednisolone, vincristine, cyclophosphamide, and doxorubicin) therapy. Of the 16 validation subjects, three had leukemia (acute myeloid leukemia, Philadelphia chromosome-positive acute lymphoblastic leukemia), and 13 had malignant lymphomas (diffuse large B-cell lymphoma, adult T-cell leukemia lymphoma, peripheral T-cell lymphoma, and NK/T-cell lymphoma). Of 16 patients, 10 received either CHOP or R-CHOP therapy, while the others received THP-COP, A-triple V, LVD (L-asparaginase, vincristine, and dexamethasone), R-DA (dose adjusted)-EPOCH, LSG15 (doxorubicin, vincristine, cyclophosphamide, ranimustine, etoposide,

158 vindesine, and carboplatin), or hyper-CVAD (cyclophosphamide, doxorubicin, vincristine,
159 and dexamethasone) therapy. A single dose of vincristine (0.4–2.0 mg) was administered
160 intravenously. The initial onset time observed in patients with peripheral neuropathy was
161 1 day to 3 months after the first administration of vincristine.

162 We conducted a GWAS to identify the loci associated with the onset of peripheral
163 neuropathy due to vincristine therapy. According to study results, the dominant genetic
164 model showed the lowest P-value (Figure 1). The 4p15.2, 11p15.4, and 20q13.12 loci
165 were suggested to be associated with the onset of peripheral neuropathy due to vincristine
166 therapy. The most significant SNPs at each locus were rs2110179 at 4p15.2, rs7126100
167 at 11p15.4, and rs2076549 at 20q13.12. To verify the accuracy of the typing result by
168 imputation, these three SNPs were sequenced for the GWAS samples, and the association
169 was reanalyzed. The strength of the association of each SNPs were as follows: rs2110179
170 (OR = 0.10, 95% CI = 0.029–0.37, $P = 4.3 \times 10^{-4}$), rs7126100 (OR = 14.0, 95% CI = 3.6–
171 53.9, $P = 1.3 \times 10^{-4}$), and rs2076549 (OR = 7.0, 95% CI = 2.1–23.7, $P = 1.7 \times 10^{-3}$) (Table
172 2). rs2110179, rs7126100, and rs2076549 were located downstream of the stromal
173 interaction molecule 2 (*STIM2*) gene, intron of *STIM1* gene, and intron of sulfatase 2
174 (*SULF2*) gene, respectively (Figure 2). The results of searching for eQTLs and sQTLs
175 with these SNPs revealed that rs2110179 was not associated with a significant eQTLs and
176 sQTLs, but the addition of the minor allele of rs7126100 significantly downregulated
177 *STIM1* expression levels in the brain cerebellum (Supplementary Figure S1), whereas that
178 of the minor allele of rs2076549 significantly increased *SULF2* intron excision ratio in
179 the nerve tibia (Supplementary Figure S2). According to the HaploReg database,
180 rs2110179 did not reside in the regulatory motifs, but many SNPs in high linkage
181 disequilibrium (LD) with rs2110179 resided Promoter Histone Marks, Enhancer Histone

Marks, DNase I hypersensitive, Proteins Bound and/or Motifs Changed regions (Supplementary Table S1). In contrast, rs7126100 and rs2076549 resided in the regulatory motifs (Supplementary Tables S2 and S3, respectively).

Next, to verify whether the three SNPs can predict peripheral neuropathy onset due to vincristine therapy, we determined the genotype for 16 validation samples collected separately (Table 3) and calculated the accuracy rate. The accuracy rate of prediction of peripheral neuropathy onset due to vincristine therapy for rs2110179, rs7126100, and rs2076549 in validation samples were 68.8%, 43.8%, and 62.5%, respectively (Table 4). We then used machine learning to construct a better predictive model. Supplementary Table S4 presents the optimized hyperparameter combinations and the CV accuracy for each model. In the model using the two SNPs-2 (rs2110179 and rs2076549), a high accuracy rate of 93.8% (sensitivity, 100%; specificity, 50.0%; positive predictive value, 93.3%; and negative predictive value, 100%) was obtained for SVM and NN algorithms (Table 5).

Discussion

The details of the mechanism by which vincristine administration induces peripheral neuropathy are unknown, and no system has been established to predict its occurrence. In this study, we identified genes associated with peripheral neuropathy onset due to vincristine therapy using GWAS, and constructed a prediction model for it by machine learning using the associated SNPs. As a result of GWAS, three SNPs (rs2110179, rs7126100, and rs2076549) associated with peripheral neuropathy onset were identified by the dominant genetic model. rs2110179, rs7126100, and rs2076549 were located downstream of *STIM2*, in the intron of *STIM1*, and in the intron of *SULF2*, respectively. Both STIM1 and STIM2 contribute to intracellular Ca^{2+} influx by store-dependent Ca^{2+} channels by activating the plasma membrane channel calcium release-activated calcium modulator 1 (Orai1) [22, 23]. This store-operated Ca^{2+} influx has also been confirmed in nerve cells [24]. SULF2 regulates the cellular signaling microenvironment by editing the sulfate pattern of heparan sulfate proteoglycans [25]. Elevated *SULF1/2* expression in oligodendrocyte progenitor cells limits myelin sheath remodeling by directly impairing progenitor cell recruitment and subsequent oligodendrocyte production [25]. Vincristine destabilizes microtubules, which play a fundamental role in nerve fiber myelination, and alters oligodendrocytes structure and function, causing abnormal myelination and loss of peripheral sensory fibers [2]. Since these genes are involved in neurotransmission [26-29], it is suggested that they may have some effect on peripheral neuropathy onset caused by vincristine therapy. SNPs are located in the intron or downstream of each gene; however, it remains unknown how they affect these genes. Nevertheless, it has been reported that some SNPs located on introns affect gene expression [30]. eQTLs analysis revealed that the minor allele of rs7126100

221 significantly downregulated *STIM1* expression levels in the brain cerebellum. It has been
222 reported that hereditary cerebellar ataxia causes peripheral neuropathy by affecting the
223 peripheral nerves due to neurodegenerative disorders [31]. It has been suggested that
224 neurotransmission suppression due to decreased *STIM1* expression causes a decrease in
225 cerebellar function and contributes to peripheral neuropathy development [26]. In
226 addition, sQTL indicated that the minor allele of rs2076549 significantly increased the
227 *SULF2* intron excision ratio in the nerve tibia. Thus, the minor allele of rs2076549 results
228 in the expression of a different isoform of SULF2, which affects myelination, suggesting
229 that it causes peripheral neuropathy. rs2110179 was not associated with a significant
230 eQTLs and sQTLs, but many SNPs in high LD with rs2110179 resided regulatory motifs.
231 Therefore, it was suggested that SNPs, which are LD with rs2110179, affect the
232 expression level of *STIM2* and cause peripheral neuropathy due to abnormal
233 neurotransmission.

234 Next, we calculated the predictive rate of peripheral neuropathy onset due to
235 vincristine therapy using three SNPs that were associated with peripheral neuropathy
236 onset in GWAS using 16 validation samples. However, of the three SNPs, the accuracy
237 rate for rs2110179 was the highest (68.8%). Next, a peripheral neuropathy onset
238 prediction model was constructed by machine learning using each algorithm, and its
239 prediction rate was evaluated using 16 validation samples. Results showed that SVM and
240 NN using rs2110179 and rs2076549 had the highest accuracy rate (93.8%). This accuracy
241 rate was superior to the accuracy rate (84.2%) of vincristine-induced peripheral
242 neuropathy by machine learning using metabolite data [32]. The combination of SNPs
243 with the highest accuracy rate did not include rs7126100, which indicated the strongest
244 association with the onset of peripheral neuropathy in GWAS. However, the average

accuracy of hyperparameter tuning by the machine learning model trained with GWAS sample data of three SNPs, including rs7126100, was the highest in both algorithms. This indicates that the machine learning model using three SNPs is more suitable for the GWAS sample data than the model using two SNPs (rs2110179 and rs2076549). In contrast, the model using rs2110179 and rs2076549 predicted the occurrence of peripheral neuropathy in the validation sample with higher accuracy than the average accuracy.

This study had several limitations. The 72 patients used in this study may be too few. For example, for rs2110179, which has the lowest P-value, the power was 0.7 when calculated with OR 14 using the Japanese minor allele frequency 0.281, according to the human genetic variation database. In general, power 0.8 or higher is considered good, but our results are slightly lower. This may be the reason why the predictive rate of a single SNP was low in the validation analysis. A small number of validation samples can also be a factor, but if you have more than 1000 samples, it may find an SNP that is clinically applicable as a predictive marker. To verify the validity of this result, it is necessary to balance case/control for both study data and test data and try to learn and verify further models using more samples and additional sample sets.

In conclusion, a machine-learning predictive model using SNPs that have been associated with vincristine-induced peripheral neuropathy in GWAS will be a useful tool in determining the applicability of vincristine-based chemotherapy by predicting the likelihood of peripheral neuropathy onset in individual patients.

Data availability

The datasets generated and/or analyzed during the current study are available from the

corresponding author on reasonable request.

Acknowledgements

We are thankful to the Prof. Aiko Yamauchi of the Tokushima University for her useful suggestions and discussions. We would like to thank Editage for English language editing.

Contributions

YS conceived and designed the study; HY, RO, and YS performed the experiments and data acquisition; HY and YS analyzed the data; NO, SN, KK, SF, HM, KI, and MA collected information on patient diagnoses and side effects; YS collected the samples; HY, RO, and YS wrote the paper. All authors read and approved the manuscript.

Conflict of Interest

The authors declare no competing interests.

References

[1] Nikanjam M, Sun A, Albers M, Mangalindin K, Song E, Vempaty H, et al. Vincristine-associated Neuropathy With Antifungal Usage: A Kaiser Northern California Experience. J Pediatr Hematol Oncol. 2018;40:e273–e277. doi: 10.1097/MPH.0000000000001220.

[2] Triarico S, Romano A, Attinà G, Capozza MA, Maurizi P, Mastrangelo S, et al. Vincristine-Induced Peripheral Neuropathy (VIPN) in Pediatric Tumors: Mechanisms, Risk Factors, Strategies of Prevention and Treatment. Int J Mol Sci. 2021;22:4112. doi: 10.3390/ijms22084112.

[3] Carlson K, Ocean AJ. Peripheral neuropathy with microtubule-targeting agents: occurrence and management approach. Clin Breast Cancer. 2011;11:73–81. doi: 10.1016/j.clbc.2011.03.006..

[4] Verma P, Devaraj J, Skiles JL, Sajdyk T, Ho RH, Hutchinson R, et al. A Metabolomics Approach for Early Prediction of Vincristine-Induced Peripheral Neuropathy. Sci Rep. 2020;10:9659. doi: 10.1038/s41598-020-66815-y.

[5] Sim JJ, Chan FM, Chen S, Meng Tan BH, Mi Aung KM. Achieving GWAS with homomorphic encryption. BMC Med Genomics. 2020;13(Suppl 7):90. doi: 10.1186/s12920-020-0717-y.

[6] Ozeki T, Mushiroda T, Yowang A, Takahashi A, Kubo M, Shirakata Y, et al. Genome-

wide association study identifies HLA-A*3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. Hum Mol Genet. 2011;20:1034–41. doi: 10.1093/hmg/ddq537.

[7] Tohkin M, Kaniwa N, Saito Y, Sugiyama E, Kurose K, Nishikawa J, et al; Japan Pharmacogenomics Data Science Consortium. A whole-genome association study of major determinants for allopurinol-related Stevens-Johnson syndrome and toxic epidermal necrolysis in Japanese patients. Pharmacogenomics J. 2013;13:60–9. doi: 10.1038/tpj.2011.41.

[8] Deo RC. Machine Learning in Medicine. Circulation. 2015;132:1920–30. doi: 10.1161/CIRCULATIONAHA.

[9] Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. Lancet Respir Med. 2018;6:837–45. doi: 10.1016/S2213-2600(18)30286-8.

[10] He S, Ye T, Wang R, Zhang C, Zhang X, Sun G, et al. An In Silico Model for Predicting Drug-Induced Hepatotoxicity. Int J Mol Sci. 2019;20:1897. doi: 10.3390/ijms20081897.

[11] Oyaga-Iriarte E, Insausti A, Sayar O, Aldaz A. Prediction of irinotecan toxicity in metastatic colorectal cancer patients based on machine learning models with pharmacokinetic parameters. J Pharmacol Sci. 2019;140:20–5. doi:

10.1016/j.jphs.2019.03.004.

[12] Mo X, Chen X, Ieong C, Zhang S, Li H, Li J, et al. Early Prediction of Clinical Response to Etanercept Treatment in Juvenile Idiopathic Arthritis Using Machine Learning. Front Pharmacol. 2020;11:1164. doi: 10.3389/fphar.2020.01164.

[13] Nguyen DA, Nguyen CH, Mamitsuka H. A survey on adverse drug reaction studies: data, tasks and machine learning methods. Brief Bioinform. 2021;22:164–77. doi: 10.1093/bib/bbz140.

[14] Huang CY, Liao KW, Chou CH, Shrestha S, Yang CD, Chiew MY, et al. Pilot Study to Establish a Novel Five-Gene Biomarker Panel for Predicting Lymph Node Metastasis in Patients With Early Stage Endometrial Cancer. Front Oncol. 2020;9:1508. doi:10.3389/fonc.2019.01508.

[15] Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007;81:1084–97. doi:10.1086/521987.

[16] Browning BL, Browning SR. Genotype Imputation with Millions of Reference Samples. Am J Hum Genet. 2016;98:116–26. doi:10.1016/j.ajhg.2015.11.020.

[17] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am

J Hum Genet. 2007;81:559–75. doi: 10.1086/519795.

[18] Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26:2336–7. doi:10.1093/bioinformatics/btq419.

[19] GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5. doi: 10.1038/ng.2653.

[20] Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):D930–4. doi:10.1093/nar/gkr917.

[21] Kuhn M. Building predictive models in R using the caret package. J. Stat. Software. 2008;28:1–26. Building Predictive Models in R Using the caret Package | Journal of Statistical Software (jstatsoft.org).

[22] Son GY, Subedi KP, Ong HL, Noyer L, Saadi H, Zheng C, et al. STIM2 targets Orai1/STIM1 to the AKAP79 signaling complex and confers coupling of Ca^{2+} entry with NFAT1 activation. Proc Natl Acad Sci U S A. 2020;117:16638–48. doi: 10.1073/pnas.1915386117

[23] Lu T, Zhang Y, Su Y, Zhou D, Xu Q. Role of store-operated Ca2+ entry in

cardiovascular disease. Cell Commun Signal. 2022;20:33. doi: 10.1186/s12964-022-00829-z.

[24] Gemes G, Bangaru ML, Wu HE, Tang Q, Weihrauch D, Koopmeiners AS, et al. Store-operated Ca^{2+} entry in sensory neurons: functional role and the effect of painful nerve injury. J Neurosci. 2011;31:3536–49. doi: 10.1523/JNEUROSCI.5053-10.2011.

[25] Saraswat D, Shayya HJ, Polanco JJ, Tripathi A, Welliver RR, Pol SU, et al. Overcoming the inhibitory microenvironment surrounding oligodendrocyte progenitor cells following experimental demyelination. Nat Commun. 2021;12:1923. doi: 10.1038/s41467-021-22263-4.

[26] Hartmann J, Karl RM, Alexander RP, Adelsberger H, Brill MS, Rühlmann C, et al. STIM1 controls neuronal Ca^{2+} signaling, mGluR1-dependent synaptic transmission, and cerebellar motor behavior. Neuron. 2014;82:635–44. doi: 10.1016/j.neuron.2014.03.027.

[27] Ryu C, Jang DC, Jung D, Kim YG, Shim HG, Ryu HH, et al. STIM1 Regulates Somatic C2+Signals and Intrinsic Firing Properties of Cerebellar Purkinje Neurons. J Neurosci. 2017;37:8876–94. doi: 10.1523/JNEUROSCI.3973-16.2017.

[28] Chanaday NL, Nosyreva E, Shin OH, Zhang H, Aklan I, Atasoy D, et al. Presynaptic store-operated Ca^{2+} entry drives excitatory spontaneous neurotransmission and

augments endoplasmic reticulum stress. Neuron. 2021;109:1314–32.e5. doi: 10.1016/j.neuron.2021.02.023.

[29] Kalus I, Salmen B, Viebahn C, von Figura K, Schmitz D, D'Hooge R, et al. Differential involvement of the extracellular 6-O-endosulfatases Sulf1 and Sulf2 in brain development and neuronal and behavioural plasticity. J Cell Mol Med. 2009;13:4505–21. doi: 10.1111/j.1582-4934.2008.00558.x.

[30] Hiramoto M, Udagawa H, Watanabe A, Miyazawa K, Ishibashi N, Kawaguchi M, et al. Comparative analysis of type 2 diabetes-associated SNP alleles identifies allele-specific DNA-binding proteins for the KCNQ1 locus. Int J Mol Med. 2015;36:222–30. doi: 10.3892/ijmm.2015.2203.

[31] Anheim M, Tranchant C. Neuropathies périphériques associées aux ataxies cérébelleuses héréditaires [Peripheral neuropathies associated with hereditary cerebellar ataxias]. Rev Neurol (Paris). 2011;167:72–6. French. doi:10.1016/j.neurol.2010.07.041.

[32] Verma P, Devaraj J, Skiles JL, Sajdyk T, Ho RH, Hutchinson R, et al. A Metabolomics Approach for Early Prediction of Vincristine-Induced Peripheral Neuropathy. Sci Rep. 2020;10:9659. doi: 10.1038/s41598-020-66815-y.

Figure and Table Legends

Figure 1. Manhattan plot of associations from the GWAS of peripheral neuropathy

The negative $\log_{10}$-transformed *P* values (Y axis) of genotyped and imputed SNPs are shown according to their positions on chromosome. The horizontal line represents suggestive thresholds.

Figure 2. Regional association plot for a peripheral neuropathy-associated locus

The negative $\log_{10}$-transformed P-values (Y axis) of genotyped and imputed SNPs located in the 400 kb upstream and downstream regions of the GWAS-lead SNP rs2110179 (A), rs7126100 (B), and rs2076549 (C) are shown according to their chromosomal positions. Purple diamond and circles represent the lead SNP and other SNPs within the region, respectively, with the color of each circle indicating the range of pairwise r^2 value with lead SNP. The right Y axis shows the recombination rates estimated from the 1000 Genomes project Asian (ASN) data (Nov 2014). The RefSeq gene within the region is shown in the panel below.

Table 1. Characteristics of subjects

Case refers to patients who developed peripheral neuropathy due to vincristine treatment, and control refers to patients who did not.

Data are presented as mean ± standard deviation. *P* values of age and BMI were obtained using unpaired Student's t-test, and *P* values of the male sex were obtained using Fisher's exact test.

Table 2. Top SNP in each loci identified the association ($P < 10^{-4}$) in GWAS for vincristine-induced peripheral neuropathy

Chr, chromosome; SNP, single-nucleotide polymorphism; AF, allele frequency

Table 3. Allele frequency in the validation samples

SNP, single-nucleotide polymorphism; AF, allele frequency

Table 4. Accuracy rate of prediction of the onset of peripheral neuropathy due to vincristine therapy of rs2110179, rs7126100 and rs2076549 in validation samples

Table 5. Accuracy rate of prediction of the onset of peripheral neuropathy due to vincristine therapy using machine learning in validation samples

Figure 1

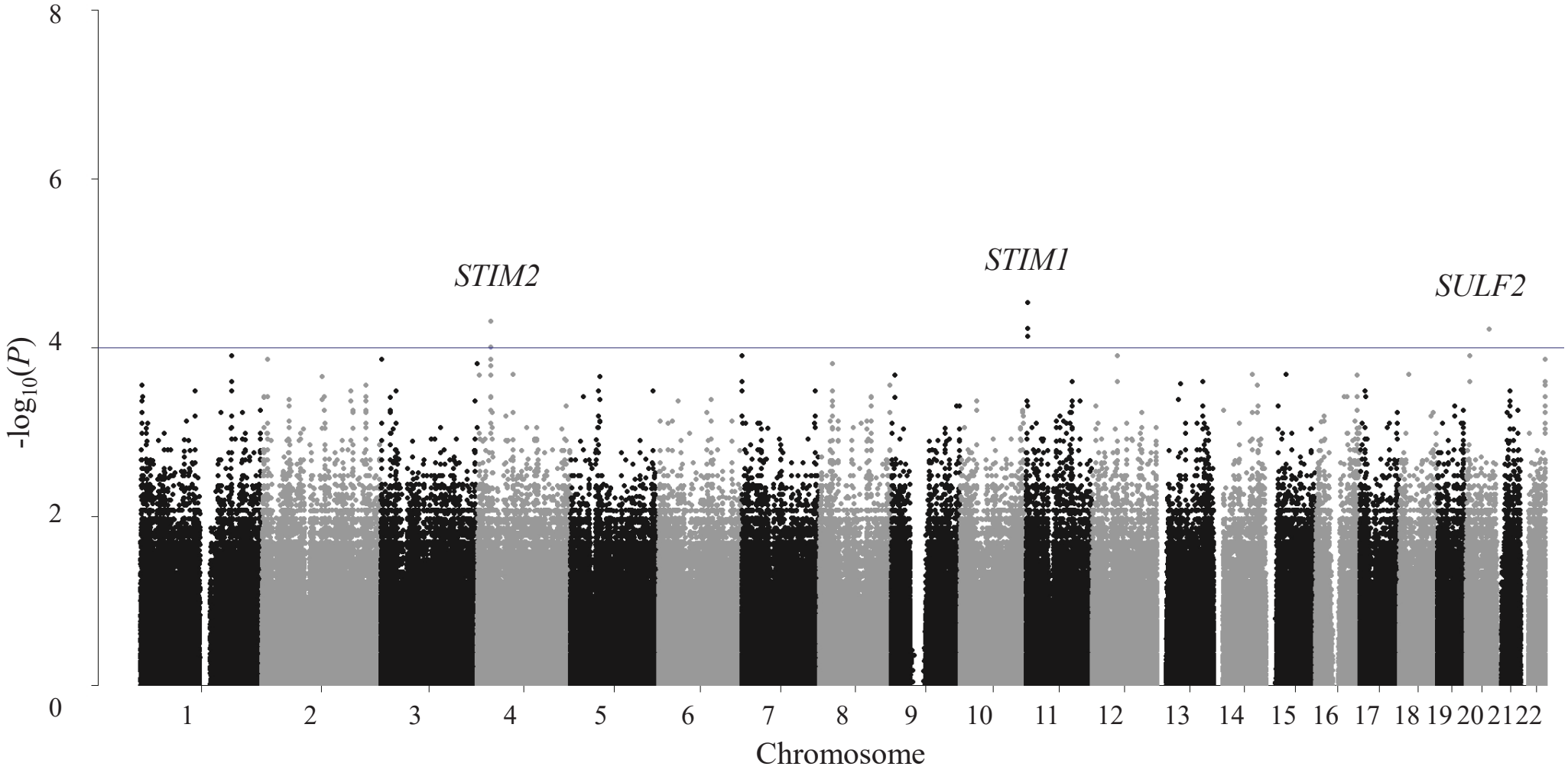
8
6
4
2
0
-log10(P)
STIM2
STIM1
SULF2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

Figure 2

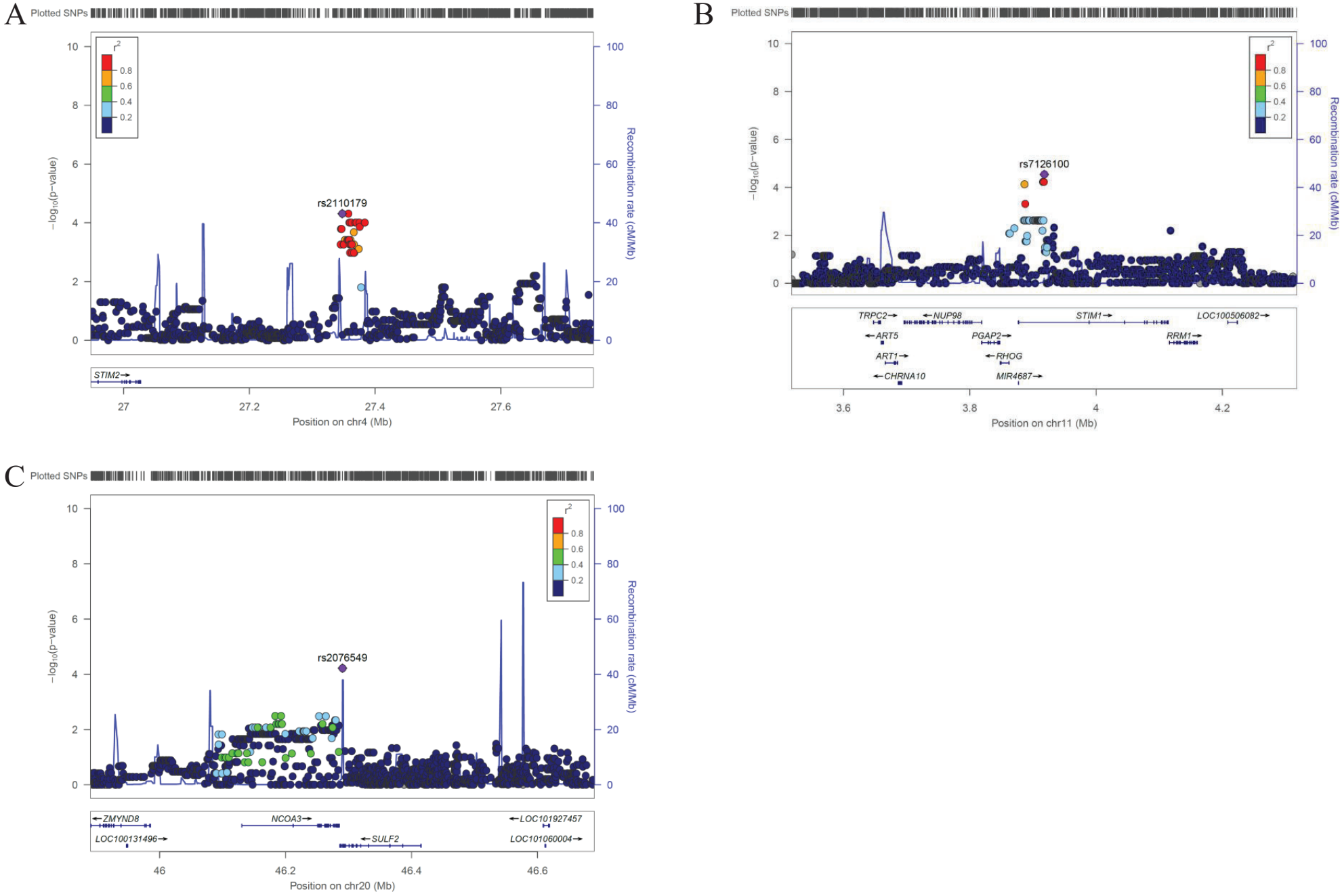
A
Plotted SNPs
rs2110179
STIM2
Position on chr4 (Mb)
B
Plotted SNPs
rs7126100
TRPC2
NUP98
STIM1
LOC100506082
ART5
PGAP2
RRM1
ART1
RHOG
CHRNA10
MIR4687
Position on chr11 (Mb)
C
Plotted SNPs
rs2076549
ZMYND8
NCOA3
LOC101927457
LOC100131496
SULF2
LOC101060004
Position on chr20 (Mb)
Recombination rate (cM/Mb)

Table 1. Characteristics of subjects

	GWAS (N=56)			Validation (N=16)		
	Control (N=20)	Case (N=36)	*P*-value	Control (N=2)	Case (N=14)	*P*-value
Age (years)	66.5±8.0	59.5±9.6	0.010	53.5±1.5	61.6±9.5	0.82
Male sex – no. (%)	14 (70.0)	19 (52.8)	0.26	2 (100)	5 (35.7)	0.18
BMI (kg/m^2)	21.9 ± 3.6	22.1 ± 2.8	0.86	19.0 ± 0.3	19.6 ± 3.4	0.28

Case refers to patients who developed peripheral neuropathy due to vincristine treatment, and control refers to patients who did not. Data are presented as mean ± standard deviation. *P* values of age and BMI were obtained using unpaired Student's t-test, and *P* values of the male sex were obtained using Fisher's exact test.

Table 2. Top SNP in each loci identified the association ($P < 10^{-4}$) in GWAS for vincristine-induced peripheral neuropathy

Chr.	SNP	Gene locus	Allele	Control (N=20)		Case (N=36)		OR (95%CI)	P-value
				Genotypes	AF	Genotypes	AF		
4	rs2110179	*STIM2* downstream	G/A	6/11/3	0.43	29/5/2	0.13	0.10 (0.029-0.37)	4.3×10^{-4}
11	rs7126100	*STIM1* intron	A/T	16/4/0	0.10	8/22/6	0.47	14.0 (3.6-53.9)	1.3×10^{-4}
20	rs2076549	*SULF2* intron	C/T	14/3/3	0.23	10/16/10	0.50	7.0 (2.1-23.7)	1.7×10^{-3}

Chr, chromosome; SNP, single-nucleotide polymorphism; AF, allele frequency

Table 3. Allele frequency in the validation samples

SNP	Allele	Control (N=2)		Case (N=14)	
		Genotypes	AF	Genotypes	AF
rs2110179	G/A	0/2/0	0.50	9/3/2	0.25
rs7126100	A/T	1/1/0	0.25	8/5/1	0.25
rs2076549	C/T	1/1/0	0.25	5/7/2	0.39

SNP, single-nucleotide polymorphism; AF, allele frequency

Table 4. Accuracy rate of prediction of the onset of peripheral neuropathy due to vincristine therapy of rs2110179, rs7126100 and rs2076549 in validation samples

SNP	Accuracy rate (%)	Sensitivity (%)	Specificity (%)	Positive predictive values (%)	Negative predictive values (%)
rs2110179	68.8	64.3	100	100	28.6
rs7126100	43.8	42.9	50.0	85.7	11.1
rs2076549	62.5	64.3	50.0	90.0	16.7

Table 5. Accuracy rate of prediction of the onset of peripheral neuropathy due to vincristine therapy using machine learning in validation samples

SNP	Algorithm	Accuracy rate (%)	Sensitivity (%)	Specificity (%)	Positive predictive values (%)	Negative predictive values (%)
<3 SNPs> rs2110179 rs7126100 rs2076549	RF	62.5	57.1	100	100	25
	SVM	62.5	57.1	100	100	25
	NB	81.3	78.6	100	100	40
	NN	62.5	57.1	100	100	25
<2 SNPs-1> rs2110179 rs7126100	RF	37.5	28.6	100	100	16.7
	SVM	37.5	28.6	100	100	16.7
	NB	75	78.6	50	91.7	25
	NN	37.5	28.6	100	100	16.7
<2 SNPs-2> rs2110179 rs2076549	RF	68.8	64.3	100	100	28.6
	SVM	93.8	100	50	93.3	100
	NB	87.5	100	0	87.5	NA
	NN	93.8	100	50	93.3	100
<2 SNPs-3> rs7126100 rs2076549	RF	68.8	78.6	0	84.6	0
	SVM	43.8	42.9	50	85.7	11.1
	NB	43.8	42.9	50	85.7	11.1
	NN	68.8	78.6	0	84.6	0