Published in final edited form as:

J Neurol Sci. 2015 October 15; 357(0): 35–40. doi:10.1016/j.jns.2015.06.056.

Association of the Charcot-Marie-Tooth disease gene *ARHGEF10* with paclitaxel induced peripheral neuropathy in NCCTG N08CA (Alliance)

Ganesh K. Boora, M.D.^{1,*}, Amit A. Kulkarni, M.D.^{1,*}, Rahul Kanwar, M.S.¹, Peter Beyerlein, Ph.D.², Rui Qin, Ph.D.³, Michaela S. Banck, M.D.¹, Kathryn J. Ruddy, M.D.¹, Josef Pleticha, M.D.¹, Cynthia A. Lynch, M.D.⁴, Robert J. Behrens, M.D.⁵, Stephan Züchner, M.D., Ph.D.⁶, Charles L. Loprinzi, M.D.¹, and Andreas S. Beutler, M.D.^{1,#}

¹Departments of Oncology, Neurology, and Biostatistics, Mayo Clinic, Rochester, MN

²Bioinformatik, Technische Hochschule Wildau, Wildau, Germany

³Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN

⁴Guthrie Corning Cancer Center, Sayre, PA

⁵Iowa Wide Oncology Research Coalition, Des Moines, IA

⁶Departments of Human Genetics and Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA

Abstract

The predisposition of patients to develop polyneuropathy in response to toxic exposure may have a genetic basis. The previous study Alliance N08C1 found an association of the Charcot-Marie-Tooth disease (CMT) gene *ARHGEF10* with paclitaxel chemotherapy induced peripheral neuropathy (CIPN) related to the three non-synonymous, recurrent single nucleotide variants (SNV), whereby rs9657362 had the strongest effect, and rs2294039 and rs17683288 contributed only weakly.

In the present report, Alliance N08CA was chosen to attempt to replicate the above finding. N08CA was chosen because it is the methodologically most similar study (to N08C1) performed in the CIPN field to date. N08CA enrolled patients receiving the neurotoxic chemotherapy agent paclitaxel. Polyneuropathy was assessed by serial repeat administration of the previously validated patient reported outcome instrument CIPN20. A study wide, Rasch type model was used to perform extreme phenotyping in n=138 eligible patients from which “cases” and “controls” were selected for genetic analysis of SNV performed by TaqMan PCR.

[#]Correspondence: Andreas S. Beutler, M.D. Department of Oncology, Mayo Clinic, 200 First St. SW, Rochester, MN 55905, beutler.andreas@mayo.edu, (507) 284-2511.

^{*}These authors have contributed equally.

Conflict of Interest Statement: The authors have not declared any conflict of interest. None of the authors has a financial or personal relationship with a third party whose interests could be positively or negatively influenced by the article's content.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A significant association of *ARHGEF10* with CIPN was found under the pre-specified primary endpoint, with a significance level of $p=0.024$. As in the original study, the strongest association of a single SNV was seen for rs9657362 (odds ratio=3.56, $p=0.018$). To further compare results across the new and the previous study, a statistical “classifier” was tested, which achieved a ROC area under the curve of 0.60 for N08CA and 0.66 for N08C1, demonstrating good agreement.

Retesting of the primary endpoint of N08C1 in the replication study N08CA validated the association of *ARHGEF10* with CIPN.

Introduction

Polyneuropathies rank among the most common neurological disorders. In the majority of cases, polyneuropathy is conceptualized as an “acquired” disorder such as from chronic hyperglycemia in diabetics or from other toxic exposures. However, toxic exposure appears not to be deterministic, because only some exposed patients will develop neuropathy, while others may not. Genetic variability between individuals is a possible explanation raising the possibility that polyneuropathy susceptibility may in some patient populations be a complex genetic trait.

Charcot-Marie-Tooth (CMT) disease refers to a group of hereditary polyneuropathies that are monogenic with high penetrance. Approximately 80 genes have been conclusively identified as CMT genes [1–4].

Genomic studies of complex traits may gain additional power from leveraging what is already known about mutations in genes that are responsible for Mendelian diseases [5, 6]. Recently, we reported a genetic discovery study on a toxic neuropathy by massively parallel (“nextgen”) sequencing of all CMT genes in 119 patients. Patients were part of the chemotherapy-induced peripheral neuropathy (CIPN) observational clinical trial Alliance N08C1. The lead finding of the study (under the pre-specified primary endpoint) was the association of the CMT gene *ARHGEF10* with susceptibility to polyneuropathy. Three single nucleotide variants (SNV) contributed to the per-gene testing, rs9657362, rs2294039, and rs17683288. Of these, rs9657362 had the strongest effect with an odds ratio of 4.8 and a significance level of $p=4\times 10^{-4}$ [7].

Replication in an independent patient cohort represents the gold standard for validating genetic associations. In the case of CIPN, such validation has not always been successful (e.g. Schneider et al., [8] was refuted by Bergmann et al., [9] and Kulkarni et al. [10]). A dominant reason for poor reproducibility amongst these studies may be related to the poor quality of neurological assessments of polyneuropathy reported in many studies performed in oncology. Oftentimes, polyneuropathy is only sporadically assessed by cancer care providers who are not well-trained in the subtleties of the assessment required. Accordingly, some authors concluded that these studies failed to provide reliable discrimination of who suffered from CIPN because a “phenotype of convenience” was used [11] instead of a prospectively implemented assessment of polyneuropathy with well-validated methods.

Our recent study N08C1 implemented a different approach to polyneuropathy phenotyping. We quantified the rate of progression of polyneuropathy over the study period as the “slope” of symptom progression. The rate (or slope) was derived from up to 12 weekly polyneuropathy assessments conducted while patients were exposed to repeat doses of neurotoxic chemotherapy. Polyneuropathy assessments were performed with a patient reported outcome (PRO) instrument, the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Chemotherapy-Induced Peripheral Neuropathy 20 (CIPN20). This instrument was developed and validated by neurologists for the assessment of peripheral neuropathy in cancer patients [12-15]. The CIPN20 methodology was recently demonstrated to cross-validate well with expert neurological exams [15]. Considering the concerns in the field discussed above, we sought to validate our published findings in Alliance N08C1 in a similarly designed study replicating polyneuropathy phenotyping and patient selection as closely as possible.

Here we report testing of the *ARHGEF10* gene single nucleotide variants (SNV) rs9657362, rs2294039, and rs17683288 in an independent cohort, the CIPN clinical study Alliance N08CA [7].

Methods

Patients

The North Central Cancer Treatment Group (NCCTG, now “Alliance”) study N08CA is a recently reported randomized controlled trial of glutathione for prevention of CIPN in patients with ovarian or lung cancer [16]. Patients in the study were at least 18 years old and had to have a life expectancy of greater than 6 months and an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1. Patients were excluded if they had been diagnosed with peripheral neuropathy (from any cause including diabetes) or fibromyalgia; were to receive concurrent neutrophil colony-stimulating factor therapy; or had previous exposure to paclitaxel or other neurotoxic chemotherapy drugs.

N08CA allowed different chemotherapy regimens stratified as different “arms” of the study. However, the majority of participants received the same regimen, paclitaxel 150-200mg/m² every 3 weeks in combination with carboplatin. To eliminate possible effects related to difference in chemotherapy administration, only patients receiving this treatment were included in the present pharmacogenomic analysis. Patients were included regardless of glutathione treatment assignment, which was found to have no effect on CIPN incidence, as previously reported [16]. Informed consent was obtained from all patients for the serial CIPN phenotype assessment and collection of blood for genetic analysis. The study was reviewed and approved by the Institutional Review Board (IRB) and have been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments (Mayo IRB# 09-002454).

Selection of CIPN “cases” and “controls” by a study-wide statistical model

An ‘extreme phenotyping’ approach was employed to select cases and controls as previously described for N08C1 [7]. Briefly, CIPN patient reported outcomes (PRO) data were

obtained prospectively with a dedicated instrument, the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ) CIPN-20 (CIPN20) a tool that has been well validated [12-15]. The CIPN20 records patient responses to 20 questions (such as “Did you have tingling fingers or hands?”) on a four-tier ordinal scale (1=“Not at All”; 2=“A Little”; 3=“Quite a Bit”; 4=“Very Much”). CIPN20 patient responses were recorded at baseline and before every chemotherapy dose through the course of treatment. Serial repeat assessments of patient neuropathy symptoms were used to estimate the rate of symptom progression using a Rasch type, study-wide statistical model. The result was an overall slope estimate for each patient representing their CIPN susceptibility. Patients' CIPN phenotype was classified CIPN “case”, or “control”, or as “uncategorized” based on the overall slope. CIPN cases and controls thus identified represent the extremes of the CIPN phenotype spectrum.

Genotyping of recurrent SNV in *ARHGEF10*

Recurrent SNV in the *ARHGEF10* gene, rs9657362, rs2294039, and rs17683288 were genotyped as follows. Genomic DNA (gDNA) was isolated from peripheral blood leukocytes using commercially available kits. PCR amplification of gDNA encompassing the loci of interest in the *ARHGEF10* gene was achieved to generate DNA amplicons. TaqMan probes were designed specifically to hybridize to the polymorphic loci. The fluorescence signal from the Taqman PCR reaction was analyzed and allele discrimination was executed with the SDS 2.0 software (Applied Biosystems).

Statistical testing of *ARHGEF10* SNV

Statistical testing of an association of *ARHGEF10* allelic variability with CIPN was performed as in the previous study [7] combining data from all three SNV, which occurred independent of each other, i.e., were not in linkage disequilibrium (LD). The C-alpha test was used for association testing. C-alpha is an over-dispersion test that uses the distribution of variants in the cases versus controls [17]. The C-alpha test was executed within the PLINK-SEQ data analysis suite[18].

“Classifier” algorithm predicting CIPN phenotype from SNV

Different *ARHGEF10* genotypes are possible in a given patient. In principle, each SNV can encode three different genotypes, homozygous major (AA), homozygous minor (aa), and heterozygous (Aa). Accordingly, three SNV (that are not in LD), as found in *ARHGEF10* in the present study, can encode $3^3=27$ genotypes. Each SNV can have different impact on the phenotype and each patient may harbor any combination of risk-and protective alleles. To assess the utility of the entirety of available genetic information (from three SNV loci per patient), we employed a statistical “classifier”, which is an algorithm that predicts the phenotype (CIPN “case” or “control”) from the genotype.

Each patient genotype was encoded as a three-dimensional vector, in which each of the three SNV was encoded as a distance along one of three dimensions using the values (0, 1, 2) for (AA, Aa, aa). The data from N08CA, the present study, was used to define the group of “case” vectors and the group of “control” vectors. For each group, an average vector was computed, which can be conceptualized as the center (or centroid) of the cloud of vectors in

the three-dimensional space. Next the classifier performance was tested by sampling equally-sized case- and control groups, 20 cases and 20 controls, in multiple iterations from the training data set N08CA. For each patient, the Euclidean distance of its genotype vector to the average case- and control vector was computed; the patient was then predicted to belong to the group, to which its vector was closer.

Classifier validation in independent data from N08C1

Classifier development occurs typically in two steps. In the first (described above), the classifier is “trained” and then tested on the same data (the “training data”). In this first step, classifier performance can be overestimated easily due to “over-fitting”. Therefore, it is important to test the classifier without further modification on an independent dataset. This second step was executed in the present study using the previously published data from N08C1 [7].

Receiver operating characteristics (ROC) of the classifier

The above-described approach results in a classifier with optimal accuracy (given equal weighing of the multiple dimensions and use of a given distance measure), i.e., a classifier that yields the smallest sum of false-positive plus false-negative predictions. In a graphical receiver operating characteristic (ROC) curve analysis, this scenario is represented by the point that is farthest from the diagonal. In order to understand the characteristics of a test across the entire spectrum of trade-offs between sensitivity (keeping false-negatives at a minimum) and specificity (keeping false-positives at a minimum) and to compare the strengths between different tests more broadly, it can be helpful to consider a complete ROC curve and to determine the area under the curve (AUC). This was achieved by considering a range of classifier cut-offs K ranging from perfect sensitivity to perfect specificity, which can be formalized as follows:

$$d_{cs} - d_{ct} \geq K \text{ is 'control'}$$

and

$$d_{cs} - d_{ct} < K \text{ is 'case'}$$

whereby d_{cs} is the distance of a vector to the centroid of case-vectors, d_{ct} is the distance to the centroid of control-vectors. $K=0$ is the most accurate classifier (i.e., the shorter distance is chosen), and varying K from -3.47 to 3.47 results in a more sensitive test or more specific test respectively.

While K can be varied at any number of intervals, the number of distinct classifier results available to construct an ROC curve is finite because it is limited by the number of vector dimensions and distinct values that each dimension can assume and is in practice further constrained by the size of the experimental dataset, because rare genotypes may not be represented (such as for example homozygous minor alleles for all three SNV in the same patient). Accordingly, seven points could be computed from the available data to construct ROC curves for N08CA and N08C1.

Results

CIPN cases and controls

Patient selection for the present study is shown in a Consort diagram (Fig. 1). 138 patients were eligible based on the having received the same paclitaxel chemotherapy regimen and having a minimum of three serial CIPN symptom assessments available. Of these patients, 63 were “uncategorized” because they had slopes of symptom progression that were not significantly different from the group median. The slopes (and standard errors) for these patients is shown in Fig. 1 (left) depicting that error bars for all “uncategorized” patients overlapped the group study median (indicated by the dotted line). The remaining 75 patients experienced either significantly slower symptom progression than the group median, 28 controls, or significantly faster symptom progression, 47 cases, i.e., were deemed relatively resistant or susceptible to CIPN. Thereby these 75 patients represented the extremes of the CIPN phenotype in N08CA. Fig. 1 (right side) shows individual CIPN susceptibility slopes (with standard errors) for each patient in the control and patient group. These patients were subjected to the genetic analysis. Clinical characteristics of these 75 patients are provided in Table 1.

Genetic association testing of the *ARHGEF10* gene with CIPN

Taqman PCR testing achieved a genotyping rate of 99.5% in the 75 patients at the three *ARHGEF10* SNV studied. One patient sample at rs9657362 failed the PCR reaction. All SNV passed quality control tests. The observed minor allele frequency (MAF) of *ARHGEF10* SNV in N08CA was similar to the prior study N08C1 and to the standard reference datasets by the 1000 Genome Project (1KG) and the NHLBI (Fig. 2a). The distribution of genotypes was consistent with Hardy Weinberg equilibrium. All tested SNV were non-synonymous, i.e., predicted to alter the protein sequence (Fig. 2b). Gene level testing with C-alpha found a significant association of the *ARHGEF10* gene with the CIPN phenotype with a significance level of $p=0.024$.

Effect of the three SNV in *ARHGEF10*

Breakdown of the individual SNV contribution suggested a phenomenon similar to N08C1. In N08C1, the association signal was dominated by rs9657362 with an odds-ratio (OR) of 4.8. In the present replication study, N08CA, rs9657362 also had the highest OR at 3.56 with a significance level of $p=0.018$. As is the original study, the direction of the effect of rs9657362 was protective against CIPN. Consistent with weaker effects of rs17683288 and rs2294039 observed in the previous study, there was only a weak contribution of rs17683288 and rs2294039. In the present study, rs17683288 had an OR of 1.79 compared with an OR of 2.29 in N08C1. The effect of rs2294039 was indeterminate in N08CA. The magnitude and direction of effect of the *ARHGEF10* SNV in N08CA were thereby overall similar to the previous study, while they were also consistent with the regression to the mean effect typically observed in replication studies. OR results for N08CA were calculated for a dominant model to make them directly comparable to the N08C1 results in our previous report [7].

Classifier to estimate probability of CIPN

Eight different genotype combinations were possible for the three alleles in *ARHGEF10*. A classifier was used to determine, how effectively this genetic information could help to make a clinical prediction. For the N08CA dataset, on which the classifier was initially “trained”, it predicted the “case” or “control” status correctly in 61.8% percent of patients. Comparing this result with a rate of 50% expected by chance (“flipping a coin”), the genetic information added a correct prediction for 11.8% of patients. Testing the classifier on the previously reported data from N08C1, it predicted the “case” versus “control” label correctly in 64.9% of patients.

Receiver operating characteristic (ROC) curve comparison in N08CA and N08C1

Receiver-operating characteristic (ROC) curves were generated by varying the discriminating threshold (detailed in Methods). The area under the curve (AUC) for N08CA was 0.60. The AUC for N08C1 was 0.66 again underscoring the robustness of the validation and comparability of results in the two studies.

Discussion

The present study was designed to allow independent validation of a recently published report on CIPN genetics [7]. The original study, Alliance N08C1, was the first in the field to integrate cutting-edge next generation sequencing and reliable CIPN phenotyping by serial PRO with the CIPN20 instrument. The study concluded that mutations in non-CMT alleles, i.e., SNV, in the CMT gene *ARHGEF10* were associated with CIPN. The present study successfully replicated this observation in an independent cohort, Alliance N08CA. N08CA lent itself as an ideal replication cohort for two reasons. Firstly, like N08C1, N08CA is a CIPN trial in patients receiving a similar regimen of paclitaxel and, secondly, neuropathy was phenotyped using the CIPN20 instrument repeatedly throughout the course of chemotherapy administration. Because of these methodological novelties, validation in an independent trial is an important result.

Alliance N08CA was smaller in size than the original study N08C1. Its statistical power was therefore smaller and the significance level of the positive result was—as expected—more moderate. Given the study design, this did not compromise the interpretation of the present study, because the endpoint for the analysis of N08CA was predefined and therefore was significant, at $p=0.024$, without a need to adjust for additional tests.

The present report also goes beyond the original publication, by exploring the utility of the genetic association results using a “classifier”. We found that the improvement of predicting clinical outcomes on the basis of the single gene studied, which harbors three SNV with only small to moderate effect sizes, could be considered moderate at best. These results are not clinically actionable but might serve in the future as one component of more predictors.

The possibly most interesting outcome of the classifier analysis was how stably it performed when the algorithm that was “trained” on N08CA, was retested with the previously reported dataset N08C1. Classifiers can be affected by “over fitting”, which becomes apparent when results are best on training data and subsequently degrade when testing is performed on a

dataset, for which the classifier was not optimized. In the present case, the correct prediction rate for both studies was similar and the ROC AUC remained stable between training- and testing steps (Fig. 3).

While many authors refer to *ARHGEF10* as a CMT gene (as this report), it is important to note that the only reported familial *ARHGEF10* variant results in a trait of slow nerve conduction velocities without clinical phenotype (REF Verhoeven et al., 2003).

The susceptibility of patients to toxic neuropathies may be conceptualized as a complex genetic trait. In the present study, CIPN from paclitaxel served as a model to explore this notion. CIPN is also a clinical problem in its own right because it affects a large number of patients [19]. Genetic determinants reported for CIPN to date have only moderate effect sizes and are therefore not clinically actionable. Furthermore, only few results have been validated in independent studies. Therefore, the present report may serve as an encouraging addition to the literature. Additional studies in larger cohorts of well-selected patients will be required to discover additional genetic determinants of CIPN and to develop a more sophisticated genetic predictor that may eventually provide guidance for clinical recommendations.

Acknowledgments

Research reported in this publication was supported by the National Institute Of Nursing Research (NINR) under Award Number R01NR015259 (to A.S.B.), by the National Cancer Institute (NCI) under Award Numbers U10CA180821, U10CA180882, and U10CA189823 (to the Alliance for Clinical Trials in Oncology), and by the National Center for Advancing Translational Sciences (NCATS) under Award Number KL2TR000136 (to Mayo Clinic). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH).

References

1. Klein CJ, Duan X, Shy ME. Inherited neuropathies: clinical overview and update. *Muscle Nerve*. 2013; 48:604–22.10.1002/mus.23775 [PubMed: 23801417]
2. Saporta ASD, Sottile SL, Miller LJ, et al. Charcot-Marie-Tooth disease subtypes and genetic testing strategies. *Ann Neurol*. 2011; 69:22–33.10.1002/ana.22166 [PubMed: 21280073]
3. Züchner S, Vance JM. Mechanisms of disease: a molecular genetic update on hereditary axonal neuropathies. *Nat Clin Pract Neurol*. 2006; 2:45–53.10.1038/ncpneuro0071 [PubMed: 16932520]
4. Timmerman V, Strickland AV, Züchner S. Genetics of Charcot-Marie-Tooth (CMT) Disease within the Frame of the Human Genome Project Success. *Genes (Basel)*. 2014; 5:13–32.10.3390/genes5010013 [PubMed: 24705285]
5. Goldstein DB, Allen A, Keebler J, et al. Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet*. 2013; 14:460–70.10.1038/nrg3455 [PubMed: 23752795]
6. Morrison AC, Voorman A, Johnson AD, et al. Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nat Genet*. 2013; 45:899–901.10.1038/ng.2671 [PubMed: 23770607]
7. Beutler AS, Kulkarni AA, Kanwar R, et al. Sequencing of Charcot-Marie-Tooth disease genes in a toxic polyneuropathy. *Ann Neurol*. 2014; 76:727–37.10.1002/ana.24265 [PubMed: 25164601]
8. Schneider BP, Li L, Miller K, et al. Genetic associations with taxane-induced neuropathy by a genome-wide association study (GWAS) in E5103. *ASCO Meet Abstr*. 2011; 29:1000.
9. Bergmann TK, Vach W, Feddersen S, et al. GWAS-based association between RWDD3 and TECTA variants and paclitaxel induced neuropathy could not be confirmed in Scandinavian ovarian cancer patients. *Acta Oncol*. 2013; 52:871–4.10.3109/0284186X.2012.707787 [PubMed: 22877241]

10. Kulkarni AA, Boora G, Kanwar R, et al. RWDD3 and TECTA variants not linked to paclitaxel induced peripheral neuropathy in North American trial Alliance N08C1. *Acta Oncol.* 2014;1–3.10.3109/0284186X.2014.985388
11. Wheeler HE, Maitland ML, Dolan ME, et al. Cancer pharmacogenomics: strategies and challenges. *Nat Rev Genet.* 2013; 14:23–34.10.1038/nrg3352 [PubMed: 23183705]
12. Postma TJ, Aaronson NK, Heimans JJ, et al. The development of an EORTC quality of life questionnaire to assess chemotherapy-induced peripheral neuropathy: the QLQ-CIPN20. *Eur J Cancer.* 2005; 41:1135–9.10.1016/j.ejca.2005.02.012 [PubMed: 15911236]
13. Lavoie Smith EM, Barton DL, Qin R, et al. Assessing patient-reported peripheral neuropathy: the reliability and validity of the European Organization for Research and Treatment of Cancer QLQ-CIPN20 Questionnaire. *Qual Life Res.* 2013; 22:2787–99.10.1007/s11136-013-0379-8 [PubMed: 23543373]
14. Cavaletti G, Cornblath DR, Merkies ISJ, et al. The chemotherapy-induced peripheral neuropathy outcome measures standardization study: from consensus to the first validity and reliability findings. *Ann Oncol.* 2013; 24:454–62.10.1093/annonc/mds329 [PubMed: 22910842]
15. Alberti P, Rossi E, Cornblath DR, et al. Physician-assessed and patient-reported outcome measures in chemotherapy-induced sensory peripheral neurotoxicity: two sides of the same coin. *Ann Oncol.* 2014; 25:257–64.10.1093/annonc/mdt409 [PubMed: 24256846]
16. Leal AD, Qin R, Atherton PJ, et al. North Central Cancer Treatment Group/Alliance trial N08CA-the use of glutathione for prevention of paclitaxel/carboplatin-induced peripheral neuropathy: a phase 3 randomized, double-blind, placebo-controlled study. *Cancer.* 2014; 120:1890–7.10.1002/cncr.28654 [PubMed: 24619793]
17. Neale BM, Rivas MA, Voight BF, et al. Testing for an unusual distribution of rare variants. *PLoS Genet.* 2011; 7:e1001322.10.1371/journal.pgen.1001322 [PubMed: 21408211]
18. [Accessed 22 Feb 2015] PLINK/SEQ genetics library. <https://atgu.mgh.harvard.edu/plinkseq/>
19. Cavaletti G, Zanna C. Current status and future prospects for the treatment of chemotherapy-induced peripheral neurotoxicity. *Eur J Cancer.* 2002; 38:1832–7. [PubMed: 12204664]

Highlights

- N08CA studied 138 patients undergoing neurotoxic paclitaxel chemotherapy
- Polyneuropathy phenotyped by serial patient reported outcomes using CIPN20
- Gene *ARHGEF10* found to be associated with paclitaxel polyneuropathy
- Study validates result of previous report in the paclitaxel trial N08C1
- Toxic neuropathies such as from paclitaxel may have a genetic component

Figure 1. Alliance N08CA patient selection and extreme phenotyping

195 patients were enrolled in Alliance N08CA. Patients were excluded if they did not provide a blood sample for genomic analysis; if they received paclitaxel under a schedule other than the most commonly used regimen; or if fewer than 3 serial CIPN20 assessments were collected. For the resulting eligible n=138 patients, CIPN susceptibility was quantified as the slope of symptom progression (rate of symptom worsening over time) using a study-wide, Rasch-type statistical model. Because the slope was derived from multiple (serially obtained) CIPN20 assessments, a standard error could be computed for each slope indicating the reliability of CIPN phenotyping for each patient. Patients with average slopes or with a large standard error were excluded from the genetic analyses because they represent unreliable patient responses or intermediate phenotypes ('Uncategorized' patients shown on the left in grey). Patients with slopes significantly smaller or significantly larger than the study median were included in the study and were classified as "Controls" (blue) or "Cases" (red).

Figure 2. Quality control of *ARHGEF10* genotyping

a. The minor allele frequency (MAF) observed in the N08CA genomic study population (solid-colored bars) is shown for the three non-synonymous, recurrent SNV in *ARHGEF10*. The observed MAF was similar to the MAF previously observed in N08C1 and to the MAF reported for large reference datasets from the 1000 genomes project (1KG) and the NHLBI 4550 exome cohort (NHLBI). **b.** The relative proportion of homozygous major, heterozygous, and homozygous minor genotypes for each SNV were consistent with Hardy Weinberg equilibrium expectations.

Figure 3. Receiver operating characteristic (ROC) curve of the classifier

A classifier was developed in two steps: training using the data from N08CA and validation using the data from N08C1. In the first step, classifier performance may be overestimated due to “over-fitting”. Therefore, it is important to test the classifier without further modification on an independent dataset. This second step was executed in a previously published data set, N08C1 [7]. The classifier was trained to achieve optimal accuracy (given equal weighing of the multiple dimensions and use of a given distance measure), i.e., a classifier that yields the smallest sum of false-positive plus false-negative predictions. In a graphical receiver operating characteristic (ROC) curve shown, this scenario is represented by the point that is farthest from the diagonal. In order to understand the characteristics of a test across the entire spectrum of trade-offs between sensitivity (keeping false-negatives at a minimum) and specificity (keeping false-positives at a minimum) and to compare the strengths between different tests more broadly, it can be helpful to consider a complete ROC curve and to determine the area under the curve (AUC). Seven points could be computed from the available data to construct ROC curves for N08CA and N08C1. The AUC was similar for the training- and the validation data sets demonstrating reproducibility.

Table 1
Patient characteristics

Demographics		N08CA	
		Control (N=28)	Case (N=47)
Age	Mean	59.2	61.6
Sex	Female	92.9%	85.1%
	Male	7.1%	14.9%
Ethnicity	Caucasian	92.9%	87.2%
	African American	7.1%	8.5%
	Others/unknown	0.0%	4.3%
Primary site	Breast	0.0%	2.1%
	Ovarian/Fallopian	42.9%	40.4%
	Lung	32.1%	25.5%
	Head and neck	0.0%	0.0%
	Endometrial	21.4%	19.1%
	Others	3.6%	12.8%
Tumor status	Recurrent	3.6%	6.4%
	Resected with known residual	14.3%	27.7%
	Resected with no residual	39.3%	36.2%
	Unresected	42.9%	29.8%