Published in final edited form as:

J Neurogenet. 2015 March ; 29(1): 4–7. doi:10.3109/01677063.2014.923884.

Detection of copy number variation by SNP-allelotyping

Brett Parker^{1,*}, Ryan Alexander^{1,*}, Xingyao Wu², Shawna Feely², Michael Shy², Nathalie Schnetz-Boutaud³, and Jun Li^{1,3}

¹Department of Neurology, Vanderbilt University School of Medicine, Nashville, TN, USA

²Department of Neurology, University of Iowa, Iowa City, IA, USA

³Center for Human Genetics Research and Vanderbilt Brain Institute, Vanderbilt University School of Medicine, Nashville, TN, USA

Abstract

Charcot-Marie-Tooth disease type 1A (CMT1A) is caused by an abnormal copy number variation (CNV) with a trisomy of chromosome 17p12. The increase of the DNA-segment copy number is expected to alter the allele frequency of single nucleotide polymorphism (SNP) within the duplicated region. We tested whether SNP allele frequency determined by a Sequenom MassArray can be used to detect the CMT1A mutation. Our results revealed distinct patterns of SNP allele frequency distribution, which reliably differentiated CMT1A patients from controls. This finding suggests that this technique may serve as an alternative approach to identifying CNV in certain diseases, including CMT1A.

Keywords

Charcot-Marie-Tooth disease type-1A; copy number variation; peripheral myelin protein-22; single nucleotide polymorphism

Background

Copy-number variation (CNV) is an alteration of genomic DNA that results in an abnormal number of copies of a specific DNA segment, including deletions, duplications, and insertions (Sebat et al., 2004). This variation is a widespread, common phenomenon that may account for up to 13% of human genomic DNA (Stankiewicz & Lupski, 2010). Certain CNVs have been associated with susceptibility or resistance to diseases. For instance, a lower copy number of CCL3L1 has been associated with higher susceptibility to HIV infection (Gonzalez et al., 2005). CNV has also been associated with autism and schizophrenia (Sebat et al., 2007; Stefansson et al., 2008; Cooper et al., 2011).

© 2014 Informa Healthcare USA, Inc.

Address correspondence to: Jun Li, MD, PhD, Department of Neurology, Vanderbilt University School of Medicine, 1161 21st Avenue South, Nashville, TN 37232, USA. jun.li.2@vanderbilt.edu.

*These two authors contributed equally

Declaration of interest: The authors report no declarations of interest. The authors alone are responsible for the content and writing of the paper.

Supplementary material **available online**

Charcot-Marie-Tooth type-1A (CMT1A) is the most common form of dominantly inherited peripheral nerve disease with a prevalence approaching 1:5000. It is a CNV disorder caused by a trisomy of chromosome 17p12 (c17.12), containing the *PMP22* gene (Lupski, 1992; Suter et al., 1992; Raeymaekers et al., 1992). A patient with CMT1A typically inherits one copy of c17.12 from an unaffected patient and two other copies of c17.12 from an affected parent. Such duplication should alter allelic frequency for a given SNP in the c17.12 region of a CMT1A patient compared to non-CMT1A individuals. We predict that SNP allelotyping may be used to identify the CMT1A duplication.

CNV may be detected by a variety of techniques such as fluorescent in situ hybridization (FISH), array comparative genomic hybridization (aCGH), multiplex RT-PCR or next-generation sequencing. All of these techniques have limitations, however. FISH requires live cells and has relatively low resolutions (about 5~10 Mbp for FISH). The aCGH and multiplex PCR are technically demanding in sample optimization and the latter is not suitable for rapid screening of a large set of samples (Itsara et al., 2010; Lorentzos et al., 2003). The next-generation sequencing can be used to identify CNV with high resolution, but only at the expense of formidable bioinformatics analysis (Mills et al., 2011).

In the present study, we used Sequenom MassArray to identify a CNV by SNP allelotyping. A SNP is a single nucleotide polymorphism, mostly bi-allelic, that differs between humans. In an individual with a CNV, the allelic frequency of a given SNP will change since additional copies or deletion of a DNA region are introduced. Thus, this variability of allelic frequency may be used to detect CNVs. Such an approach has been used to diagnose the trisomy of chromosome 21 (Zhong & Holzgreve, 2009), however, it is still unknown whether this approach can be used to detect an abnormal CNV of a DNA segment on a chromosome, such as the CMT1A mutation.

Methods

Human DNA Samples

De-identified DNA samples were collected from 94 patients with CMT1A and 94 controls with Amyotrophic Lateral Sclerosis (ALS) for initial CNV study. All 94 CMT1A patients had nerve conduction studies, and 41 individuals of the 94 patients had a DNA test showing c17p12 duplication. The CMT1A mutation was detected by multiplex PCR technique in 38 patients and by FISH in the remaining 3 patients. The other 53 patients were diagnosed based on family history of CMT1A and abnormal nerve conduction study results. For the prospective study, 34 samples were from Vanderbilt's BioVU DNA bank.

SNP Design

A total of 126 SNPs were identified within c17p12 and regions flanking c17p12 using UCSC Genome Browser and NCBI's dbSNP database. All SNPs were highly variable with a minor allele frequency (MAF) >0.43. The location of all SNPs is illustrated in Figure 1 and listed in Supplementary Table 1 available online at <http://informahealthcare.com/doi/abs/10.3109/01677063.2014.923884>

SNP Allelotyping

SNP Genotyping—SNP allelotyping was performed in the Vanderbilt VANTAGE core using a Sequenom MassArray (MALDI-TOF) platform. This technology uses single base extension reactions paired with mass spectrometry to quantitatively determine the identity of the amplified nucleotide based on the mass differential of the amplified nucleotides. The Sequenom MassARRAY Typer software was used for the MALDI-TOF data analysis. SNPs were genotyped in accordance with Sequenom standard protocol (Gabriel et al., 2009). Primers were designed using the MassArray Assay Design software. Amplified DNA was purified by SAP (Shrimp Alkaline Phosphatase). The primer was extended by one mass-modified nucleotide (ddNTP) specific for each assay designed. The cleaned products by resin were plated onto a 384-well PectroCHIP using a nanodispenser robot. The chip was read by the Sequenom MassARRAY system through matrix assisted laser desorption ionization (MALDI) – time of flight (TOF) mass spectrometry processing.

Allele Frequency Determination—The software (MassARRAY Assay Design Software v3.1) was configured to calculate the individual allele frequency for each SNP based on mass differential of each allele. The results were expressed as a fraction of each allele frequency with both alleles totaling 1.0 (Herbon et al., 2003).

Results

SNP Heterozygosity Demarked the DNA Segment Known to be Duplicated in Patients with CMT1A

We first determined whether SNPs can be used to define the duplicated DNA segment of c17.12. We identified a total of 126 highly variable SNPs that covered the entire c17.12 segment and two regions flanked the c17.12 (Figure 1). For the controls, each SNP of a given SNP allele should have a roughly equal chance for homozygosity (50%) or heterozygosity (50%). Due to an additional copy of c17.12 in patients with CMT1A, the chance of heterozygosity for each SNP allele was expected to be close to 75%, and confirmed in patients with CMT1A. The region with the increased heterozygosity matched with the region known to be duplicated in CMT1A (Figure 2).

Patterns of SNP Allele Frequency Distribution Differentiate Patients with CMT1A from Controls

Data displayed in Figure 3 shows examples of allele frequency distribution pattern for two SNPs among CMT1A patients and controls (Figure 3). A subset of SNPs, such as rs230950 in Figure 3A and B, showed clear differential patterns between patients with CMT1A and controls. These SNPs were clustered at the predicted 1, 0.67, 0.33 and 0 of allele frequency in patients with CMT1A; while the SNPs in the controls were clustered at 1, 0.5, and 0 of allele frequency. We noticed that a subset of SNPs was not clear-cut (rs2610065 in Figure 3C and D), presumably due to skewed primer binding affinity to DNA.

After this analysis, 9 SNPs with clear differential patterns were selected (rs230950, rs230941, rs231015, rs230967, rs230968, rs230895, rs8072625, rs767680, rs28649950) for the following study.

Differentiation between Controls and Patients with CMT1A can be Achieved Using Just 9 SNPs

To test whether the 9 SNPs could be used to reliably identify CMT1A patients in a cohort of mixed DNA samples, a simple criterion was set: positive for CMT1A if the allele frequency of two SNPs out of the 9 SNPs falls in (0.28–0.38) or (0.61–0.71). All SNP allele frequencies from 94 controls and 94 patients with CMT1A were entered into an Excel sheet in a random order. Each individual case was identified by a code with no diagnosis available to the experimenter. All individuals who matched the criterion were marked as CMT1A. This procedure identified 92 CMT1A patients. None of 94 control subjects was false positive.

There were two samples that were marked as CMT1A patients who failed to match our CMT1A criterion of SNP allele frequency. DNA from the two was subsequently sent to a commercial lab (Medical Neurogenetics Lab, Atlanta, GA) for analysis (by multiplex PCR) and no CMT1A mutation was identified.

Prospective Study

In order to test the diagnostic criterion in prospective, we acquired 34 additional DNA samples (de-identified) from the Vanderbilt BioVU DNA Bank using two ICD-9 codes (356.0 for CMT and 335.2 for ALS). These samples were examined for the same SNPs using the Sequenom procedure. We identified 12 individuals that matched our SNP criterion of CMT1A. The results were compared with the diagnoses recorded in the DNA sample-linked medical records. There were 7 patients with CMT1A that were confirmed by DNA testing. The remaining 5 patients had no medical record showing a DNA testing done to confirm the CMT1A diagnosis, but were clinically consistent with CMT1A (uniform slowing of conduction velocity, positive family history). The other 23 patients who were not diagnosed with CMT1A by our SNP study had several recorded diagnoses: 4 with ALS and 19 with acquired peripheral neuropathies.

Discussion

Our data show that SNP allelotyping by DNA MassArray can reliably identify c17.12 duplications in patients with CMT1A. This is supported by the following facts. Once our allele frequency analysis was completed, we reviewed the medical records of all 94 patients with CMT1A. Forty-one of the 94 patients had DNA testing that confirmed the c17.12 duplication. All 41 of these patients were also found to have CMT1A by our allele frequency testing. In contrast, all 94 controls with the diagnosis of ALS were negative for the c17.12 duplication by the allele frequency test. Together, these findings would yield a sensitivity of 100% and a specificity of 100%. The 100% sensitivity and specificity were also found in the prospective study since all 7 CMT1A cases were positive and all 23 non-CMT1A cases were negative by our Sequenom test.

There were two samples that were initially labeled as CMT1A, but found to be negative for CMT1A mutation by our Sequenom test. The medical records of these individuals showed that the two cases never had DNA testing, but were diagnosed with CMT1A based on family

history and abnormal nerve conduction studies. When the DNA of these two individuals was subsequently sent for analysis (by multiplex PCR in a commercial lab), no CMT1A mutation was identified. Presumably, they were misdiagnosed due to a false family history of CMT1A. It should be clarified that some cases of CMT1A are not caused by *PMP22* duplication but by *PMP22* triplication in rare cases (Liu et al., 2014). This Sequenom method has not been shown yet to detect the triplication and will not detect missense mutations.

This technique appears to be simple and rapid in comparison with the current standard test, multiplex PCR. Both multiplex PCR and the Sequenom test have a turn-around time of 3 days (DNA extraction in day one, followed by two reactions in day two and three). However, the former typically tests a maximum of 20 patients' samples, whereas the Sequenom technique can run 384 samples as a batch. In communicating with a DNA diagnostic laboratory (Dr. Trey Langley at Medical Neurogenetics Lab, Atlanta, GA), we found that costs between the two tests are comparable (\$279.25 per patient for multiplex PCR and \$274 per patient for Sequenom). Therefore, as long as SNPs are determined, the Sequenom test would be suitable for screening a large set of DNA samples to identify CNVs in human genome.

CMT1A is a well-documented disease with abnormal CNV. It provides an excellent population to validate our approach of allele frequency. We view our study as an initial step that shows the feasibility of the Sequenom technique in detecting CNV. Because this test can be done rapidly in a large set of DNA samples, we envision that this approach might be useful in screening CNVs for other diseases as well. Moreover, SNP allelotyping DNA MassArray may serve as an approach complementary to the existing diagnostic tools. For instance, when one of the techniques reveals no mutation in a patient with suspected CMT1A, a different technique may be used to confirm or reject the negative finding.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Authors would like to thank Dr. Trey Langley for providing valuable information about multiple PCR technique, and also wish to thank the technical assistance from Ms. Audra Hamilton, Ms. Sezgi Arpag, Ms. Chelsea Bacon and Vanderbilt Technologies for Advanced Genomics (VANTAGE) core.

This study is, in part, supported by grants from NINDS (NS066927 and NS081364), the National Center for Advancing Translational Sciences (CTSA award UL1TR000445) to J.L. and NINDS/ORD 5U54NS065712 to M.E.S.

References

- Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. *Nat Genet.* 2011; 43:838–846. [PubMed: 21841781]
- Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr Protoc Hum Genet.* 2009; Chapter 2 Unit.

- Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*. 2005; 307:1434–1440. [PubMed: 15637236]
- Herbon N, Werner M, Braig C, Gohlke H, Dutsch G, Illig T, et al. High-resolution SNP scan of chromosome 6p21 in pooled samples from patients with complex diseases. *Genomics*. 2003; 81:510–518. [PubMed: 12706109]
- Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, Eichler EE. De novo rates and selection of large copy number variation. *Genome Res*. 2010; 20:1469–1481. [PubMed: 20841430]
- Liu P, Gelowani V, Zhang F, Drory VE, Ben-Shachar S, Roney E, et al. Mechanism, prevalence, and more severe neuropathy phenotype of the Charcot-Marie-Tooth type 1A triplication. *Am J Hum Genet*. 2014; 94:462–469. [PubMed: 24530202]
- Lorentzos P, Kaiser T, Kennerson ML, Nicholson GA. A rapid and definitive test for Charcot-Marie-Tooth 1A and hereditary neuropathy with liability to pressure palsies using multiplexed realtime PCR. *Genet Test*. 2003; 7:135–138. [PubMed: 12885335]
- Lupski JR. An inherited DNA rearrangement and gene dosage effect are responsible for the most common autosomal dominant peripheral neuropathy: Charcot-Marie-Tooth disease type 1A. *Clin Res*. 1992; 40:645–652. [PubMed: 1486731]
- Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011; 470:59–65. [PubMed: 21293372]
- Raeymaekers P, Timmerman V, Nelis E, Van Hul W, De Jonghe P, Martin JJ, Van Broeckhoven C. Estimation of the size of the chromosome 17p11.2 duplication in Charcot-Marie-Tooth neuropathy type 1a (CMT1a). HMSN Collaborative Research Group. *J Med Genet*. 1992; 29:5–11. [PubMed: 1552545]
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007; 316:445–449. [PubMed: 17363630]
- Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004; 305:525–528. [PubMed: 15273396]
- Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010; 61:437–455. [PubMed: 20059347]
- Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008; 455:232–236. [PubMed: 18668039]
- Suter U, Moskow JJ, Welcher AA, Snipes GJ, Kosaras B, Sidman RL, et al. A leucine-toproline mutation in the putative first transmembrane domain of the 22-kDa peripheral myelin protein in the trembler-J mouse. *Proc Natl Acad Sci U S A*. 1992; 89:4382–4386. [PubMed: 1374899]
- Zhong XY, Holzgreve W. MALDI-TOF MS in prenatal genomics. *Transfus Med Hemother*. 2009; 36:263–272. [PubMed: 21049077]

Figure 1.

Location of the 126 SNPs within and flanking the classical CMT1A duplication region (grey bar) on chromosome 17p11.2. The 9 SNPs tested in the prospective analysis are in *bold*.

Figure 2.
Percentage of heterozygosity for each SNP allele in patients with CMT1A compared to controls.

Figure 3.

Allele frequency distribution pattern for two individual SNPs among CMT1A patients (A and C) and controls (B and D). SNP rs230950 (A and B), shows clear differential patterns clustered at the predicted 1, 0.67, 0.33 and 0 of allele frequency in patients with CMT1A and a pattern clustered at 1, 0.5, and 0 of allele frequency in controls. In contrast, the pattern in some SNPs is not clear-cut (C and D).