Utilization of Whole-Exome Next-Generation Sequencing Variant Read Frequency for Detection of Lesion-Specific, Somatic Loss of Heterozygosity in a Neurofibromatosis Type 1 Cohort with Tibial Pseudarthrosis

CrossMark

Rebecca L. Margraf,* Chad VanSant-Webb,[†] David Sant,[‡] John Carey,[§¶] Heather Hanson,[§] Jacques D'Astous,[§] Dave Viskochil,[§¶] David A. Stevenson,[‖] and Rong Mao*,**

From the ARUP Institute for Clinical and Experimental Pathology,* ARUP Laboratories,[†] Salt Lake City, Utah; the Miller School of Medicine,[‡] University of Miami, Miami, Florida; the Shriners Hospital for Children Salt Lake City,[§] Salt Lake City, Utah; the Division of Medical Genetics, Department of Pediatrics[¶] and the Department of Pathology,** School of Medicine, University of Utah, Salt Lake City, Utah; and the Division of Medical Genetics,[‖] Department of Pediatrics, Stanford University, Stanford, California

A subset of neurofibromatosis type 1 patients develop tibial dysplasia, which can lead to pseudarthrosis. The tissue from the tibial pseudarthrosis region commonly has a somatic second hit in *NF1*: single-nucleotide variants, small deletions, or loss of heterozygosity (LOH). We used exome next-generation sequencing (NGS) variant frequency data (allelic imbalance analysis) to detect somatic LOH in pseudarthrosis tissue from three individuals with clinically and diagnostically confirmed neurofibromatosis type 1, and verified the results with microarray. The variant files were parsed and plotted using python scripts, and the NGS variant frequencies between the affected tissue and blood sample were compared. Individuals without somatic single-nucleotide variants or small insertions/deletions were tested for somatic LOH using the NGS variant allele frequencies. One individual's NGS data indicated no LOH in chromosome 17. The other two individuals demonstrated somatic LOH inclusive of *NF1*: one had an LOH region of approximately one million bases and Contra (NGS copy number program) indicated a somatic deletion and the other individual had LOH for most of chromosome 17q and Contra indicated no copy number change (microarray data verified this sample as copy neutral somatic LOH). Both LOH and copy number variation detected by NGS data correlated with microarray data, demonstrating the somatic LOH second hit can be detected directly from the NGS data. *(J Mol Diagn 2017, 19: 468–474; http://dx.doi.org/10.1016/j.jmoldx.2017.01.008)*

Neurofibromatosis type 1 (NF1) is a common autosomal dominant condition that occurs in 1:3000 births. It is associated with a wide range of manifestations, including café au lait macules with intertriginous freckling, neurofibromas, learning disabilities, short stature, malignancies, and distinctive osseous lesions. Tibial dysplasia is a rare manifestation of NF1 (approximately 4%), but can result in high morbidity.[1] Patients display anterolateral bowing of the tibia that can progress to fracture and nonunion (ie, pseudarthrosis). Biopsy specimens of the pseudarthrosis site demonstrate abundant fibroblastic proliferation.[2]

NF1 is caused by mutations in *NF1* located on chromosome 17q11.2. This gene encodes neurofibromin, which is expressed in a wide range of cells and tissues, including maturing chondrocytes, hypertrophic chondrocytes,

osteoblasts, osteocytes, and osteoclasts.[3] Neurofibromin is a RasGAP that negatively regulates the activity of an intracellular signaling molecule p21[ras]. Haploinsufficiency of *NF1* results in a dose-dependent elevation in Ras signal transduction activity. The pathogenic mechanism responsible for dysplasia of the tibia is not well understood, but several reports show that tissue from the tibial pseudarthrosis region commonly has a lesion-specific somatic second hit in *NF1* in addition to the germline mutation, resulting in the biallelic loss of the neurofibromin protein.[2,4,5] These somatic second hits were shown to be single-nucleotide variants, small insertions/deletions (indels), or large regions of loss of heterozygosity (LOH) across the long arm of chromosome 17. Although the somatic second hit is not prognostic, it is commonly used for research purposes in attempts to better understand the pathophysiology of pseudarthrosis. In a report by Paria et al,[4] five individuals had a detected somatic single-nucleotide variants (SNV) or small indels in the pseudarthrosis tissue region, whereas six individuals had large regions of somatic LOH as detected by genome-wide microarray analysis. However, one individual (individual 12 in Paria et al[4]) had a germline heterozygous SNV at 57% variant read frequency and the same variant had 95% frequency in the pseudarthrosis tissue sample, hypothesizing a LOH or loss of the *NF1* wild-type allele for the second hit.[4] Also, several other SNVs within and surrounding the *NF1* gene showed similar allelic imbalances in next-generation sequencing (NGS) variant read percentage in the whole-exome data (similar deviation from the expected 50% read frequency for heterozygous germline variants). On the basis of these findings as well as the LOH being a common somatic mutation in the pseudarthrosis samples (found 5 of 16 individuals in our previous report[4]), a protocol using a python script to parse and filter the NGS variant data was developed to detect these regions of somatic LOH/allelic imbalance using the NGS variant allele frequencies. This identifies large somatic chromosome 17 LOH regions covering *NF1* directly from the NGS data without the need for additional microarray testing. This method of LOH detection was used on whole-exome NGS data for individuals who did not have a detected second hit somatic SNV or small indels from DNA extracted from the pseudarthrosis region.

Materials and Methods

Samples and Exome Sequencing

Individuals with a clinical diagnosis of NF1 and tibial pseudarthrosis were enrolled into this study through a University of Utah Institutional Review Board–approved protocol. Individuals 1 and 2 in this report had germline *NF1* mutations published previously as individuals 12 and 13.[4] Although a second hit somatic mutation was not detected in either sample at the time, individual 1 was hypothesized to have a somatic LOH because of the nearly homozygous allelic imbalance of the germline *NF1* mutation in the pseudarthrosis sample (57% germline versus 95% in pseudarthrosis tissue).[4] The other individuals are new study participants. Blood was obtained through an intravenous blood draw. After surgery, discarded tibial pseudarthrosis site tissues were collected. DNA was extracted from peripheral blood samples using the Puregene DNA Extraction automated system (Qiagen, Venlo, the Netherlands). DNA was extracted from frozen tissue from the pseudarthrosis site by previously described methods.[2] Briefly, the frozen samples taken from the pseudarthrosis site were mechanically sheared and DNA was extracted using the Chemagic Tissue 10 kit (PerkinElmer Chemagen Technologie Gmbh, Baesweiler, Germany).

Whole-exome libraries were prepared, sequenced, and analyzed, as published previously.[4] Briefly, whole-exome capture was performed using the SeqCap EZ Human Exome Library version 3.0 (Roche NimbleGen Inc., Madison, WI) and sequenced using the paired-end 100-bp protocol on the Illumina HiSeq 2500 instrument (Illumina, Inc., San Diego, CA), according to the manufacturer's recommendations. Sequence reads were mapped to the human genome (GRCh37), and variants were called as described previously.[2] The VarBin method was used to filter false-positive variants from the NGS data set files.[6] Variants from the blood and pseudarthrosis samples were analyzed using the Golden Helix SVS software version 7.7.3 (Golden Helix, Inc., Bozeman, MT) and filtered as described previously,[2] for detection of germline or somatic *NF1* mutations. *NF1* somatic mutations were identified after comparison of variants from the pseudarthrosis samples to matched blood samples, whereas germline variants were present in both samples. Any detected somatic or germline SNVs and small indels were verified by Sanger sequencing.

NGS Allelic Imbalance Analysis for LOH Detection

NGS variant read frequencies (the number of variant reads divided by the total read per position) were calculated from the vcf data and filtered for variant quality using a python script. Only variants that passed the default GATK filters [QD (QualByDepth) < 2.0, MQ (RMSMappingQuality) < 40.0, FS (FisherStrand) > 60.0, HaplotypeScore > 13.0, MQRankSum < -12.5, and ReadPosRankSum < -8.0] were used in the analysis, to keep only the most confident variant data. The NGS variant read frequencies were plotted per chromosome (*x* axis was chromosome position, *y* axis was % variant reads). The data for the blood sample (germline) were compared to the pseudarthrosis tissue sample, to visually detect allelic imbalance by shifts in the % variant read frequency from the expected 50% allelic frequencies for the heterozygous germline variants present in the blood sample. Focus was on chromosome 17, because *NF1* is located at 17q11.2. Large somatic LOH inclusive of *NF1* (or lack of LOH) detected by this allelic imbalance protocol was verified using clinical