

Appendix S1

Patients and DNA sampling. Samples of 16 unrelated NF1 patients were collected through the NF1 clinic, Department of Dermatology, Turku University Hospital, Finland, or through the Finnish NF Patient Association. The study was carried out with the approval of the Ethics Committee of the Hospital District of Southwest Finland and the research permission of Turku University Hospital. The samples were collected with the informed consent of the patients. All patients were of Finnish ancestry and fulfilled the NIH diagnostic criteria for NF1. Of all patients, 4 had an inherited mutation, while 12 patients did not have a known family history of NF1. The *NF1* mutations of 6 patients were previously established for clinical indications in diagnostic laboratories abroad. In addition, one patient had a previously detected *NF1* type 2 microdeletion.

Genomic DNA was isolated from saliva using a saliva sampling kit (OG-575, Oragene, DNA Genotek, Inc., Ottawa, Ontario, Canada). Contaminating RNA was degraded using RNase treatment at 37°C for 30 min (RNase Cocktail Enzyme mix, Ambion Inc., Applied Biosystems, Foster City, CA, USA, AM2286). The DNA was further purified using a gDNA isolation kit (NucleoSpin® Tissue, Macherey-Nagel GmbH & Co., Düren, Germany, 740952). The quality of the DNA was evaluated on agarose gels, the NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA, ND-1000), and 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA, G2939AA). The gDNA was quantified using the Quant-iT Picogreen assay (Life Technologies, Eugene, Oregon, USA, P11496) and Plate Chameleon V fluorometer (Hidex, Turku, Finland, 425-106).

Sample library preparation. A total of 16 indexed Rapid libraries were prepared according to the manufacturer's instructions starting from 500 ng of gDNA per sample (Rapid Library Preparation Method Manual, GS Junior Titanium Series, May 2010 and March 2012, Roche). The gDNA was fragmented by nebulisation and the fragments were end-repaired, leaving a single A overhang in the 5' end of each strand. Subsequently, indexed adaptors were ligated to the fragment ends. The library preparation was completed by removing small DNA fragments with the Agencourt Ampure XP reagent (Ordior Inc, Helsinki, Finland, A63880). The final preparation contained less than 10% of fragments < 350 bp.

Sequence capture and sequencing. The data presented in this article is derived from 2 separate sequence captures and sequencing runs, referred to sets A and B. Set A contained 10 and Set B 6 libraries per set, respectively. The sample libraries were amplified and pooled into 2 sets prior to the sequence capture. The total amount of DNA in both sets was 1 µg. Samples with known mutations were used as control samples for the new sequencing approach. Set A contained 3 controls and 7 samples with unknown mutations, and set B had 4 controls and 2 samples with unknown mutations. The 2 different sets were sequenced to search for the ideal number of samples per sequencing run.

The *NF1* exome was enriched using a sequence capture with custom probe design, and a modified NimbleGen SeqCap EZ Choice Library protocol optimised for the enrichment of target areas < 100 kb (Roche NimbleGen SeqCap EZ Rapid Library Small Target Capture LR, December 2009, Roche Nimblegen Inc., Madison, WI, USA). The targeted regions included 58 *NF1* exons and an additional 50 bp of flanking upstream and downstream intronic sequences, resulting in a total target size of 16 kb. The target region did not include alternatively spliced exon 10a2, central nervous system-specific exon 9b, or exon 48a expressed in muscles. To our knowledge, disease-causing mutations have not been discovered in these regions.

The enrichment of the target region in sequence capture was assessed by qPCR by measuring the relative abundance of 4 control targets in amplified sample library and amplified captured DNA. The control assays have been validated on NimbleGen Sequence Capture arrays, but the control sequences are not revealed to the

customer. The captured DNA was quantified using the Picogreen reagent, and the fragment sizes were measured with BioAnalyzer before and after capture. The average lengths of captured fragments were 757 bp and 727 bp, in the sets of A and B, respectively. The emulsion PCR and sequencing were performed separately for each set according to the manufacturer's instructions (emPCR Amplification Method Manual - Lib-L, and Sequencing Method Manual, GS Junior Titanium Series, May 2010 and March 2012, Roche) using 2720 Thermal cycler (Life Technologies, Eugene, Oregon, USA, 435965), and the 454 GS Junior instrument (454 Life Sciences, Branford, CT, USA). Following the sequencing run, the reads were automatically quality filtered using the default settings of GS Run Processor. Only reads that passed all the quality filters were used in the analysis. The number of reads passing the quality-filtering exceeded 100,000 which is the cut-off limit for a successful run (Sequencing Method Manual, GS Junior Titanium Series, May 2010 and March 2012, Roche). Furthermore, over 60% of the control reads had less than 5% errors in the first 400 bp, consistent with the criteria for a successful sequencing run.

Data analysis. Control samples were analysed blindly in parallel with samples of which the mutations were unknown. Sequencing reads in sff (standard flowgram format) file format obtained from the 454 GS Junior sequencing run were converted to fastq format using a freely available sff-extract script (version 0.3.0) (14). The reads were mapped to the whole human genome (UCSC hg19, NCBI build 37) using the Bowtie 2 alignment program (version 2.0.0-beta7) (15). The mapping was performed as an end-to-end alignment using the "very-sensitive" option in order to minimise the risk of mapping pseudogene sequences to the *NF1* gene. Samtools (version 0.1.18) (16), Picard (version 1.78) (17) and Biopython (version 1.60) (18) software packages were used for file handling and making files compatible with the different programs used in the analysis pipeline. The sequencing coverage across target and non-target areas was calculated using BEDTools (version 2.16.2) (19) and R (version 2.15.1) (20, 21). The alignments were visualised with the Integrative Genomics Viewer (IGV, version 2.0.34 1623) (22).

Identification of mutations. Putative mutations in the region of the *NF1* gene were identified from the sequence alignments using GATK UnifiedGenotyper (version 2.1-11) (23) with settings that allow for both single nucleotide polymorphisms (SNPs) and insertions/deletions to be called. The main parameters influencing the power of detecting heterozygous variants are the coverage and selection of the used sample variant frequency. The variants were further filtered by the following criteria: 1) Minimum coverage of 20× or higher, 2) sample variant frequency between 30–70%. If the combination of values 20× coverage and minimum variant frequency 30% are used, 96.8% of the heterozygous variants are detected (24, 25), 3) Variant in a targeted region (Fig. S2¹), and 4) variant present in only one sample within a sample set with a frequency of 30% or higher. The fourth criterion was based on the expectation that 16 unrelated NF1 patients were not likely to have shared mutations. This criterion also excluded possible SNPs present in the Finnish population, as well as a majority of the false positive insertions/deletions in homopolymer sequences, which are a common type of error in 454 sequencing data (26, 27). The variants which passed all criteria were compared to the dbSNP database build 135 (28) and Finnish "The Sequencing Initiative Suomi" database (29). Variants were also checked whether they were supported by evidence from reads mapped to both the sense and antisense strands.

Sanger sequencing. Verification of discovered variants was carried out using Sanger sequencing. Fragments with putative mutations or potential mutations in homopolymer regions were amplified by PCR and sequenced with Applied Biosystems 3130xl Genetic Analyzer in DNA Sequencing Service of Turku Centre for Biotechnology, Turku, Finland.