Neurofibromatosis Type 1 Gene Mutation Analysis Using Sequence Capture and High-throughput Sequencing

Elina UUSITALO1, Anna HAMMAIS2, Elina PALONEN1, Annika BRANDT1, Ville-Veikko MÄKELÄ1, Roope KALLIONPÄÄ1,2, Eeva-Mari JOUHLAHTI1,2, Minna PÖYHÖNEN2,4, Juhani SOINI3, Juha PELTONEN1 and Sirkku PELTONEN2

1Department of Cell Biology and Anatomy, University of Turku, 2Department of Dermatology, Turku University Hospital and University of Turku, 3Turku University of Applied Sciences, Turku, Finland, 4Department of Biosciences and Nutrition, and Center for Biosciences, Karolinska Institutet, Huddinge, Sweden, 5Department of Clinical Genetics, HUSLAB, Helsinki University Central Hospital, and 6Department of Medical Genetics, University of Helsinki, Helsinki, Finland

Neurofibromatosis type 1 syndrome (NF1) is caused by mutations in the NF1 gene. Availability of new sequencing technology prompted us to search for an alternative method for NF1 mutation analysis. Genomic DNA was isolated from saliva avoiding invasive sampling. The NF1 exons with an additional 50bp of flanking intronic sequences were captured and enriched using the SeqCap EZ Choice Library protocol. The captured DNA was sequenced with the Roche/454 GS Junior system. The mean coverages of the targeted regions were 41× and 74× in 2 separate sets of samples. An NF1 mutation was discovered in 10 out of 16 separate patient samples. Our study provides proof of principle that the sequence capture methodology combined with high-throughput sequencing is applicable to NF1 mutation analysis. Deep intronic mutations may however remain undetectable, and change at the DNA level may not predict the outcome at the mRNA or protein levels. Key words: mutation analysis; neurofibromatosis type 1; next-generation sequencing; pyrosequencing; saliva DNA; target enrichment.

Accepted Jan 23, 2014; Epub ahead of print Mar 25, 2014

Sirkku Peltonen, Department of Dermatology, Turku University Hospital, P.O. Box 52, FI-20521 Turku, Finland.
E-mail: sirkku.peltonen@utu.fi

Neurofibromatosis type 1 (NF1) is an autosomal dominant syndrome with a prevalence of 1:3500. The diagnosis of NF1 is usually based on clinical findings outlined in the NIH criteria (1). Most important of these, café-au-lait macules, skinfold freckles and neurofibromas, are readily visible on skin. However, clinicians often face situations where there are some NF1 symptoms but not sufficient for clinical diagnosis. Since NF1 is a multiorgan disease with frequent complications from various organ systems, the correct early diagnosis is essential. During the 21st century, molecular diagnostics of NF1 has become possible and increasingly important in NF1 diagnosis. Mutation analysis of NF1 has proven valuable especially in young children who may only partially fulfill the clinical criteria. The same holds true for adults with atypical clinical presentation.

The NF1 gene, located on 17q11.2, is challenging to sequence due to its large size and numerous exons. The gene spans ~280 kb of genomic DNA, comprising 57 constitutive and at least 3 alternatively spliced exons. To date, over 1,400 different pathogenic mutations of the NF1 gene have been published (2). The mutations are dispersed throughout the gene and represent various mutation types, including insertions, deletions, substitutions and duplications. Microdeletions refer to large deletions which cover the entire NF1 gene and a number of flanking genes. The type 1 NF1 microdeletion is the most frequent encompassing 1.4 Mb. The type 2 microdeletion spanning 1.2 Mb and type 3 spanning 1.0 Mb are less frequent (3–6). Chromosomal rearrangements affecting one or several exons have also been observed (7). In addition, the human genome contains NF1 pseudogenes in chromosomes 2, 12, 14, 15, 18, 21 and 22 (8–12), which interfere with gDNA-based sequencing methods.

High-throughput methods can yield the sequence of the whole genome in a single analysis, but at costs too high for today’s routine diagnostics. Therefore, targeting the genomic area of interest allows analysing several samples in one run and produces less data for analysis compared to whole genome sequencing. To our knowledge, there is only one report assessing the feasibility of next generation sequencing for the targeted resequencing of the NF1 gene. Chou et al. (13) analysed 2 samples with known NF1 mutations using DNA sequence capture and enrichment by microarray followed by pyrosequencing.

At present, molecular diagnostics of NF1 utilise Sanger sequencing with either mRNA and/or genomic DNA (gDNA) as the starting material. The traditional methods can yield excellent results but are laborious and time-consuming. Furthermore, mRNA-based methods usually require fresh blood or tissue sampling. The rapid development of novel sequencing techniques has created visions for a cost-effective and non-invasive method without compromising sensitivity. This is a particularly important pursuit since the availability of information on NF1 has expanded and the demand for molecular diagnostics among patients and physicians is continuously increasing.

The purpose of the present study was to develop an NF1 mutation analysis method, which does not require invasive
sampling and which utilises new sequencing technology. A total of 16 unrelated NF1 patients were investigated.

**PATIENTS AND METHODS** (see Appendix S1)

**RESULTS**

**Sample quality, sequence capture and sequencing**

The gDNA yield of 2.7–28 µg from the saliva samples was sufficient for mutation analysis. The variation in the amount mostly depended on the original volume of saliva. Gel electrophoresis (Fig. S1) showed >10 kb bands consistent with intact gDNA. The sequence capture was successful, as estimated by qPCR using 4 internal control sequences, complying with the manufacturer’s guidelines. Both sequencing runs passed the quality criteria set by the manufacturer.

**Mapping and sequencing coverage**

In the set A of 10 samples, the number of reads per sample was between 8,023 and 16,783. In the set B of 6 samples, 13,984–29,886 reads were obtained per sample. The mean read length across sample sets was 405 bp. For the sets of A and B, the mean proportion of reads that were mapped to the human genome with Bowtie 2 was 96% and 98%, respectively. The number of reads for each sample is listed in Table SII. The distribution of reads into different chromosomes in the Bowtie 2 mapping is shown in Table I. The chromosomes with the most off-target reads are locations of known NF1 pseudogenes. Approximately 32–35% of the reads were mapped to the NF1 gene on chromosome 17.

The overview of the sequencing results is listed in Table SII. The mean coverage of targeted regions was 41× and 74× for the sets A and B, respectively. Exon 1 of the gene on chromosome 17.

The gDNA yield of 2.7–28 µg from the saliva samples was sufficient for mutation analysis. The variation in the amount mostly depended on the original volume of saliva. Gel electrophoresis (Fig. S1) showed >10 kb bands consistent with intact gDNA. The sequence capture was successful, as estimated by qPCR using 4 internal control sequences, complying with the manufacturer’s guidelines. Both sequencing runs passed the quality criteria set by the manufacturer.

**Mapping and sequencing coverage**

In the set A of 10 samples, the number of reads per sample was between 8,023 and 16,783. In the set B of 6 samples, 13,984–29,886 reads were obtained per sample. The mean read length across sample sets was 405 bp. For the sets of A and B, the mean proportion of reads that were mapped to the human genome with Bowtie 2 was 96% and 98%, respectively. The number of reads for each sample is listed in Table SII. The distribution of reads into different chromosomes in the Bowtie 2 mapping is shown in Table I. The chromosomes with the most off-target reads are locations of known NF1 pseudogenes. Approximately 32–35% of the reads were mapped to the NF1 gene on chromosome 17.

The overview of the sequencing results is listed in Table SII. The mean coverage of targeted regions was 41× and 74× for the sets A and B, respectively. Exon 1 was covered poorly in both sets (Fig. S2), with mean coverages of 3× and 6×. Low coverage in the first exon of genes has been previously observed, possibly due to a high GC content (30). This explanation is relevant also in our experiment, as the GC content of the NF1 exon 1 is 71%, while the mean across all NF1 exons is 42%.

**Mutations**

The GATK UnifiedGenotyper program reported 1,420 and 944 preliminary variants in the NF1 gene in the sets of A and B, respectively. The filtering, described in detail in Patients and Methods, resulted in the identification of a total of 63 variants as potential mutations in the sample sets of A and B. Seven variants which were listed in dbSNP database were evaluated individually and their pathogenicity was excluded. In addition, 2 out of the 7 single nucleotide polymorphisms were included in the Finnish database (29). The remaining 39 and 17 variants in sets A and B were assessed individually with respect to homopolymer-related sequencing errors and lack of evidence from reads originating from both the sense and antisense strands. Ten homopolymer-related regions with a potential mutation were selected for Sanger-sequencing (Fig. S3A). These proved to represent false positives.

Ten mutations were identified as putative disease-causing mutations (Table II). These included 6 substitutions, an insertion and 3 small deletions. Five previously unknown mutations of patients S47, E66, E71, E396, and S97 were confirmed with Sanger sequencing (Fig. S3A). One previously known mutation in a control sample (patient E39) was excluded in the filtering due to low coverage. However, visual inspection of this area revealed the mutation in 2 out of 9 reads. The known microdeletion of a control sample could not be detected. Mutations of 4 patients thus remained unsolved. To learn why these were not revealed, the 4 DNA samples were sent to an internationally recognised diagnostic laboratory, which sequenced all NF1 exons plus 30 bp intronic sequence and carried out MLPA (Multiplex Ligation-dependent Probe Amplification) analysis. These analyses revealed one additional mutation in patient S49 (c.844C>T, p.Gln282X) in NF1 exon 6. In our experiment, this area of the sample S49 had low coverage of only 11 reads and the mutation was visible in one read and thus could not raise suspicion of a pathogenic mutation. Three mutations remained undiscovered by our protocol, and by an established international diagnostic laboratory.

**DISCUSSION**

Our study of DNA samples from 16 unrelated NF1 patients provides proof of principle that the sequence capture methodology combined with high-throughput sequencing is applicable to NF1 mutation analysis. DNA sampling using a saliva collection kit yielded high-quality DNA without invasive sampling. The samples could be collected by the patients at home, and because of the stability of the samples, they could easily be shipped to the laboratory without need for cold storage. The quality of DNA was evaluated by running the samples on agarose gels, which showed single bands larger than 10 kb. Saliva samples have more commonly been used
Table II. Summary of samples and mutations

<table>
<thead>
<tr>
<th>Sample set</th>
<th>Sample code</th>
<th>Mutation code</th>
<th>Position on NF1 (cDNA, Chromosome 17)</th>
<th>Total depth</th>
<th>Variant depth</th>
<th>Variant frequency</th>
<th>Region</th>
<th>Previously described</th>
<th>Protein or mRNA level change</th>
<th>Control sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>E46</td>
<td>c.7368dupC</td>
<td>Chr17: 2967310</td>
<td>79</td>
<td>0.54</td>
<td>frameshift</td>
<td>Exon 41</td>
<td>yes</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>E51</td>
<td>c.1541_1542delAG</td>
<td>Chr17: 29546036</td>
<td>25</td>
<td>0.36</td>
<td>frameshift</td>
<td>Exon 10c</td>
<td>yes</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>S65</td>
<td>c.4922G&gt;A</td>
<td>Chr17: 29652987</td>
<td>54</td>
<td>0.52</td>
<td>splicing</td>
<td>Exon 28</td>
<td>yes</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>E66</td>
<td>c.2851-1G&gt;A</td>
<td>Chr17: 29556852</td>
<td>34</td>
<td>0.51</td>
<td>frameshift</td>
<td>Intron 16</td>
<td>no</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>E71</td>
<td>c.499_502delTGTT</td>
<td>Chr17: 29496928</td>
<td>37</td>
<td>0.51</td>
<td>frameshift</td>
<td>Exon 4b</td>
<td>no</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>E396</td>
<td>c.3911T&gt;G</td>
<td>Chr17: 29562976</td>
<td>34</td>
<td>0.68</td>
<td>p.R1513X</td>
<td>Exon 23.1</td>
<td>No</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>S594</td>
<td>No mutation found</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>S96</td>
<td>No mutation found</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>S594</td>
<td>No mutation found</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>S97</td>
<td>No mutation found</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>S97</td>
<td>c.1797G&gt;A</td>
<td>Chr17: 295990190</td>
<td>25</td>
<td>0.44</td>
<td>p.Q282X</td>
<td>(Exon 6)</td>
<td>–</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>S97</td>
<td>c.5710G&gt;T</td>
<td>Chr17: 29657477</td>
<td>112</td>
<td>0.09</td>
<td>p.E1904X</td>
<td>(Exon 28)</td>
<td>yes</td>
<td>transcribed ribonucleic acid (cDNA)</td>
<td>–</td>
</tr>
</tbody>
</table>

*The mutations in samples E38 and S49 were not found in this study but were discovered by diagnostic services.

In forensic medicine as a source for DNA (31), and the use of saliva in high throughput sequencing has been elucidated in a recent publication (32). Although the NF1 mutation analysis method described here is not yet validated for clinical application, it paves the way for new approaches in NF1 mutation analysis.

The sequence capture method was sensitive in enriching the NF1 exons, with the exception of the exon number 1. It should be noted that sequencing of the first exon of the NF1 gene is challenging also in RNA-based protocols (33). In cases where a mutation is not found in the other exons, exon 1 needs to be Sanger sequenced. However, exon 1 is not frequently mutated, since only 6 mutations of the NF1 gene have been described to date. The sequence capture is an independent module of the mutation analysis, allowing sequencing with different platforms. In the current study, the Roche GS Junior sequencing device was used. It utilises the same 454 pyrosequencing technology as the 454 GS FLX device, which is a widely used high-throughput sequencing platform. For the current application, the 454 GS Junior was selected because it has a smaller total capacity, which makes it more applicable to the sequencing of smaller targets such as a single gene instead of the whole genome.

In general, the quality of the sequencing reads was high in our protocol, as shown by the correct reading of the control sequences supplied by the manufacturer. Sequencing errors in homopolymer regions were observed in our data, which is a well-recognised problem of pyrosequencing. To deal with this problem, we have compared the sequences of the homopolymeric regions between different samples. The reads from homopolymeric regions tend to resemble each other in normal samples while real mutations may look different. Thus, putative mutations in homopolymers need to be individually examined, compared to the results for the corresponding position in other samples, and if mutation is still suspected, it needs to be verified by Sanger sequencing. (Fig. S3A’).

Pseudogenes are considered as a challenge in genetic testing and were expected to cause problems also in this method. However, in our approach the correct mapping of the reads either to pseudogenes or the NF1 gene appeared successful. This may be due to the relatively long reads, approximately 400 bp, produced by the sequencing method used. No doubt, the sequence capture methodology suffers from the existence of pseudogenes, in that their sequences are also captured along with the NF1 gene. This reduces the mean coverage of the NF1 exons. However, none of the variants that passed the filters were due to pseudogene sequences falsely mapping to the NF1 gene. Thus, the variant calling was not adversely affected by the existence of pseudogenes, and based on what we have seen, there is no reason to believe that the mapping program used in the analysis would fail in mapping reads correctly to either the NF1 gene, or to its pseudogenes. Therefore, we did not experience problems with the pseudogenes in the data analysis, even though they were originally captured in the sequence capture step.
In the NF1 mutation analysis presented here, substitutions and short insertions/deletions were readily observed in the sequencing data in areas where the coverage was at least 20×. If this coverage was used with variant frequency between 30–70%, about 97% of heterozygous variants could be found, as calculated according to De Leneer et al. (25). However, since the coverage of 20× was not reached in all nucleotides, the sensitivity could be increased by lowering the threshold of frequency from 30% to 20%. This in turn may increase the number of false positives. The best way to increase both sensitivity and specificity would be to increase coverage by sequencing a smaller number of samples per run (25). To avoid missing of known pathogenic mutations because of low coverage, comparison of the variants with previously published mutations will be utilised in the future. Using the mutation information in databases is becoming an increasingly powerful tool since the number of known pathogenic mutations is increasing.

A putative mutation was discovered in 10 samples out of the total 16. Out of the 7 previously analyzed mutations, 5 were readily evident in the data. One control mutation was observed in visual analysis, but was excluded in the filtering due to a total coverage of less than 20×. The known microdeletion could not be detected and in cases when no mutations are found, we recommend combining an MLPA analysis and Sanger sequencing of targets with low coverage. In 4 cases, an NF1 mutation could not be found. Mutation analysis was then carried out in an internationally recognised diagnostic laboratory and this approach, including MLPA, revealed one more mutation which was present in our data in a single read out of 11. However, the NF1 mutation could not be found in 3 cases. It should be noted that these 3 patients clearly fulfilled the NF1 diagnostic criteria for NF1. One of them may represent a case of somatic mosaicism for NF1 because of the clinical features of the patient. In somatic mosaicism, the NF1 mutation is not likely to be found in blood or saliva samples. The 3 mutations remaining undetected may also be deep intronic, or reside outside of the NF1 gene.

ACKNOWLEDGEMENT (SEE APPENDIX S2)

This study was funded by grants from Turku University Hospital (EVO13906), Academy of Finland, The Finnish Cancer Organisations, Centre for Economic Development (ELY Centre, Southwest Finland), Informational and Structural Biology Graduate School (ISB-Graduate School, Åbo Akademi University), and Stiftelsen Liv-och-hälsa.

REFERENCES


Supplementary material to article by E. Uusitalo et al. "Neurofibromatosis Type 1 Gene Mutation Analysis Using Sequence Capture and High-throughput Sequencing"