

## ARTICLE

# A recurrent copy number variation of the *NEB* triplicate region: only revealed by the targeted nemaline myopathy CGH array

Kirsi Kiiski<sup>\*1</sup>, Vilma-Lotta Lehtokari<sup>1</sup>, Ari Löytynoja<sup>2</sup>, Liina Ahlström<sup>1</sup>, Jenni Laitila<sup>1</sup>, Carina Wallgren-Pettersson<sup>1</sup> and Katarina Pelin<sup>1,3</sup>

Recently, new large variants have been identified in the nebulin gene (*NEB*) causing nemaline myopathy (NM). NM constitutes a heterogeneous group of disorders among the congenital myopathies, and disease-causing variants in *NEB* are a main cause of the recessively inherited form of NM. *NEB* consists of 183 exons and it includes homologous sequences such as a 32-kb triplicate region (TRI), where eight exons are repeated three times (exons 82–89, 90–97, 98–105). In human, the normal copy number of *NEB* TRI is six (three copies in each allele). Recently, we described a custom NM-CGH microarray designed to detect copy number variations (CNVs) in the known NM genes. The array has now been updated to include all the currently known 10 NM genes. The NM-CGH array is superior in detecting CNVs, especially of the *NEB* TRI, that is not included in the exome capture kits. To date, we have studied 266 samples from 196 NM families using the NM-CGH microarray, and identified a novel recurrent *NEB* TRI variation in 13% (26/196) of the families and in 10% of the controls (6/60). An analysis of the breakpoints revealed adjacent repeat elements, which are known to predispose for rearrangements such as CNVs. The control CNV samples deviate only one copy from the normal six copies, whereas the NM samples include CNVs of up to four additional copies. Based on this study, *NEB* seems to tolerate deviations of one TRI copy, whereas addition of two or more copies might be pathogenic. *European Journal of Human Genetics* (2016) 24, 574–580; doi:10.1038/ejhg.2015.166; published online 22 July 2015

## INTRODUCTION

Nemaline (rod) myopathy (NM; MIM IDs: NEM1 #609284, NEM2 #256030, NEM3 #161800, NEM4 #609285, NEM5 #605355, NEM6 #609273, NEM7 #610687, NEM8 #615348, NEM9 #615731, NEM10 #616165) is one of the most common congenital myopathies. Characteristically, NM presents with congenital proximal muscle weakness, but it is a very heterogeneous disorder including six clinical categories, in which the severity and onset of the disease varies widely.<sup>1,2</sup> Typical histological findings include the presence of nemaline bodies (rods) in the muscle fibres. Nemaline bodies are aggregates of Z-disc and thin-filament proteins of the muscle sarcomere.<sup>3</sup>

Hitherto, 10 different genes have been shown to cause MM (*ACTA1*, *NEB*, *TPM3*, *TPM2*, *TNNT1*, *CFL2*, *KBTBD13*, *KLHL40*, *KLHL41* and *LMOD3*). Variants in the nebulin gene (*NEB*) are the most common cause of recessively inherited NM (NEM2 #256030). *NEB*, is one of the largest genes in the human genome, located in the chromosomal region 2q23.3 (152 341 850–152 591 001; GRCh37/hg19) and consisting of 249 kb of genomic sequence. *NEB* contains 183 exons, but only one region, the donor splice site of intron 32, may be considered a true mutational hotspot. The vast majority, 83% (132/159) of the families included in our recently published mutational update, were heterozygous for two different pathogenic variants.<sup>2</sup> The great variety of different *NEB* variants makes mutation analysis cumbersome, and this is further accentuated by the complexity of *NEB*. The gene includes regions with highly similar exons and

intronic repeat elements such as *Alu* and LINE. It also includes homologous sequences, such as the so-called triplicate region (TRI) in the middle of the gene. TRI consists of eight exons that are repeated three times (exons 82–89, 90–97 and 98–105), all ~10 kb repeats showing high similarity (99%) with each other.<sup>4</sup> This type of homologous DNA segment is difficult to study, and it might even include variants that have been overlooked because of that. For example, the *NEB* TRI is not included in the typical exome sequencing capture kits because of the high similarity of the three copies. Therefore, developing alternative methods is essential for studying the region.

Nebulin is a gigantic protein of the muscle sarcomere, which acts as a molecular ruler regulating thin (actin) filament length, actin–myosin interactions and force generation.<sup>5–7</sup> This highly repetitive protein contains around two hundred 30- to 35-amino-acid-long simple repeats, each containing one actin-binding site. Moreover, the simple repeats are arranged into 22–30 super repeats each containing a putative tropomyosin-binding site. Nebulin has several binding partners and it is an essential part of the properly functioning sarcomere of the striated muscle.<sup>7–10</sup>

Currently, evidence is accumulating that large copy number variations (CNVs) in *NEB* are more common than previously thought. In addition to the 2.5-kb Ashkenazi Jewish deletion of *NEB* exon 55,<sup>11,12</sup> other large deletions have recently been identified, including variations of different sizes, from part of one exon to more than half of the gene.<sup>2,13</sup>

<sup>1</sup>The Folkhälsan Institute of Genetics and The Department of Medical and Clinical Genetics, Medicum, University of Helsinki, Helsinki, Finland; <sup>2</sup>Institute of Biotechnology, University of Helsinki, Helsinki, Finland; <sup>3</sup>Division of Genetics, Department of Biosciences, University of Helsinki, Helsinki, Finland  
\*Correspondence: K Kiiski, The Folkhälsan Institute of Genetics and The Department of Medical and Clinical Genetics, Biomedicum Helsinki, University of Helsinki, PO Box 63 (Haartmaninkatu 8), Helsinki FIN-00014, Finland. E-mail: kirsi.kiiski@helsinki.fi

Received 19 December 2014; revised 28 May 2015; accepted 19 June 2015; published online 22 July 2015

The normal copy number of the *NEB* TRI region has been estimated to be three copies in each allele.<sup>4</sup> Therefore, the normal copy number of the *NEB* TRI is six instead of the typical two copies. Mouse *Neb* contains only one ~7.5 kb copy of the TRI region segments, implying that the triplication has occurred later during evolution. It has been suggested that the TRI region is the result of two tandem duplications through *Alu*-mediated homologous recombination in an ancestor of human and chimpanzee.<sup>14</sup>

One of the most common mechanisms inducing CNVs is thought to be non-allelic homologous recombination (NAHR) caused by misalignment and cross-over of non-allelic homologous DNA segments, such as low copy repeats. CNV breakpoints have also been shown to frequently reside in regions of repeat elements. These include different SINEs (such as different *Alu* repeats), LINES, DNA repeat elements (such as MERs) and long-terminal repeats.<sup>15,16</sup> Repeat elements are known to cause CNVs in the genome in general, but their role in the formation of TRI CNVs requires further investigations.

In the current study, we describe the TRI CNVs that we have identified in our study cohort of 196 families and 60 controls using our custom NM-CGH microarray, and hypothesize on the pathogenicity and possible mechanisms behind this recurrent variation.

## MATERIALS AND METHODS

### Samples

The NM-CGH study included 266 DNA samples from 196 families with patients diagnosed with or suspected to have NM or a related myopathy, in whom one or both pathogenic variants had remained unidentified. In addition, 60 normal control samples, 22 from Finland and 38 from CEPH (Centre d'Étude du Polymorphisme Humain), were studied. The samples were received either as isolated DNA or as blood, cell lines, or muscle or skin biopsies from which DNA extraction was done using appropriate methods.

### Microarray design, protocol and data analysis

The NM-CGH 8x60k microarray (Oxford Gene Technology IP Limited, Oxford, UK) was designed (Human reference sequence, GRCh37/Hg19) as described in Kiiski *et al.*<sup>13</sup> The seven genes causative for NM known at the time (*NEB*, *ACTA1*, *TPM3*, *TPM2*, *TNNT1*, *CFL2*, *KBTBD13*) were densely covered with a tiling approach (with one probe pair starting every 10 bp interval), avoiding the most repetitive regions of these genes. In addition, this study includes samples that were run with the updated NM-CGH microarray version v2, also including the *KBTBD5*, *KBTBD13* and *LMOD3* genes, and v3, also including one unpublished NM-associated gene. In v2 and v3, the control gene

*TTN* has been removed and the probe interval has been reduced from 10 to 20 bp in the intronic regions for every gene except for *NEB*. No other significant modifications were made in the NM-CGH array updates.

The labelling, hybridization, scanning and data analysis were done according to the manufacturer's protocol (Oxford Gene Technology PI Ltd) as previously described in Kiiski *et al.*<sup>13</sup> The CytoSure Interpret Software v.4.2.5-4.6.85 (Hg19) (Oxford Gene Technology Ltd) was used for graphic analysis of the data. The CBS algorithm was used and specific thresholds determined to allow for aberration calling. A setting for 'multiple mappings' was used in the analysis and in the graphic view, that is, each probe is shown in every genomic location where it can be located. Therefore, regarding the *NEB* TRI, most of the probes are shown three times, once for each homologue. The TRI variations were analyzed manually based on the logarithmic scale of the NM-CGH microarray results (Table 1). It was possible to design ~180 unique probe pairs for this region, mostly based on small sequence differences in the TRI introns.

### Exome sequencing and other variant analysis methods

The variants other than *NEB* TRI CNVs were identified using dHPLC and Sanger sequencing as previously described.<sup>13,17</sup> In addition, exome sequencing was used. Exome capture and sequencing were done by Oxford Gene Technology using the Agilent SureSelectXT All Exon 50 Mb target enrichment kit (protocol v1.2; Agilent Technologies, Santa Clara, CA, USA) on an Illumina HiSeq2000 platform using TruSeq v3 chemistry (Illumina Inc, San Diego, CA, USA) and analysis was completed using the Oxford Gene Technology exome sequencing pipeline. A putative pre-mRNA splicing affecting variant identified in family F2 was tested using a *NEB* minigene construct as described previously.<sup>17</sup>

### Variants in the LOVD database

The results have been submitted to the LOVD database (<http://www.LOVD.nl>/NEB) with the submission ID numbers: NEB\_00253-NEB\_00261. The *NEB* cDNA reference sequence NM\_001271208.1 was used and the exon numbering is according to Donner *et al.*<sup>4</sup>

### Statistical analyses

Fisher's exact test was used for determining the statistical significance of the results.

### Bioinformatics methods

The *NEB* TRI region is highly repetitive and many of the tiling array probes match multiple locations. To obtain an even coverage, the measurements for the ambiguous probes were randomly assigned to one of the possible locations. The data were then analyzed using the GLAD package<sup>18</sup> for the R environment<sup>19</sup> and the breakpoints and copy numbers were inferred using the daglad function. Owing to the high probe density and variable signal level across the target region, the default parameters of the function produced unrealistically high numbers of breakpoints. After experimenting with different parameter values, the function was used with options  $\lambda=40$  and  $d=10$ . The resulting breakpoints, along with the original signal intensities and the genome annotations for the target region, were visualized using the GenomeGraphs<sup>20</sup> and rtracklayer<sup>21</sup> packages.

## RESULTS

Using the NM-CGH array, we identified frequent CNVs of the *NEB* TRI, some of the CNVs we interpret to be benign and other pathogenic. *NEB* TRI CNVs were identified in 13% of the studied families (26/196) with NM or NM-related myopathy (referred to as NM families from now on). In more detail, 5% (9/196) showed a loss and 8% (16/196) a gain of the *NEB* TRI (Table 2). In addition, one family (F286) was found to include members with a gain, a loss or the normal copy number of the *NEB* TRI (Figure 1 and Table 4). *NEB* TRI variation was also identified in 10% of the control samples (6/60; Table 2). Based on this study, we suggest that one-copy losses or gains are benign and gains of 2–4 copies may be pathogenic.

**Table 1** The calculated *NEB* TRI variations on the log<sub>2</sub> scale of the NM-CGH array and the corresponding *NEB* TRI copy number

Array log <sub>2</sub> scale	TRI copies	Copy number change
+1	12/6	6 Copy gain
+0.87	11/6	5 Copy gain
+0.73	10/6	4 Copy gain
+0.58	9/6	3 Copy gain
+0.41	8/6	2 Copy gain
+0.22	7/6	1 Copy gain
0	6/6	Normal
-0.23	5/6	1 Copy loss
-0.57	4/6	2 Copy loss
-1	3/6	3 Copy loss
-1.59	2/6	4 Copy loss
-2.55	1/6	5 Copy loss
-Saturated	0/6	6 Copy loss

The zero line corresponds to the normal copy number 6 of the *NEB* TRI. The additional copies are on the plus side and losses on the minus side of the logarithmic scale.

**Table 2** Summary of *NEB* TRI variations identified with the NM-CGH microarray in the NM family and control cohorts

	NM families	Controls
All	100% (196)	100% (60)
<i>NEB</i> TRI variations	13.3% (26)	10.0% (6)
Losses	4.6% (9)	8.3% (5)
1 Copy loss (benign)	4.6% (9)	8.3% (5)
Gains	8.2% (16)	1.7% (1)
1 Copy gain (benign)	4.6% (9)	1.7% (1)
2–4 Copy gains (possibly pathogenic)	3.6% (7)	0.0% (0)
Losses and gains (benign and possibly pathogenic)	0.5% (1)	0.0% (0)

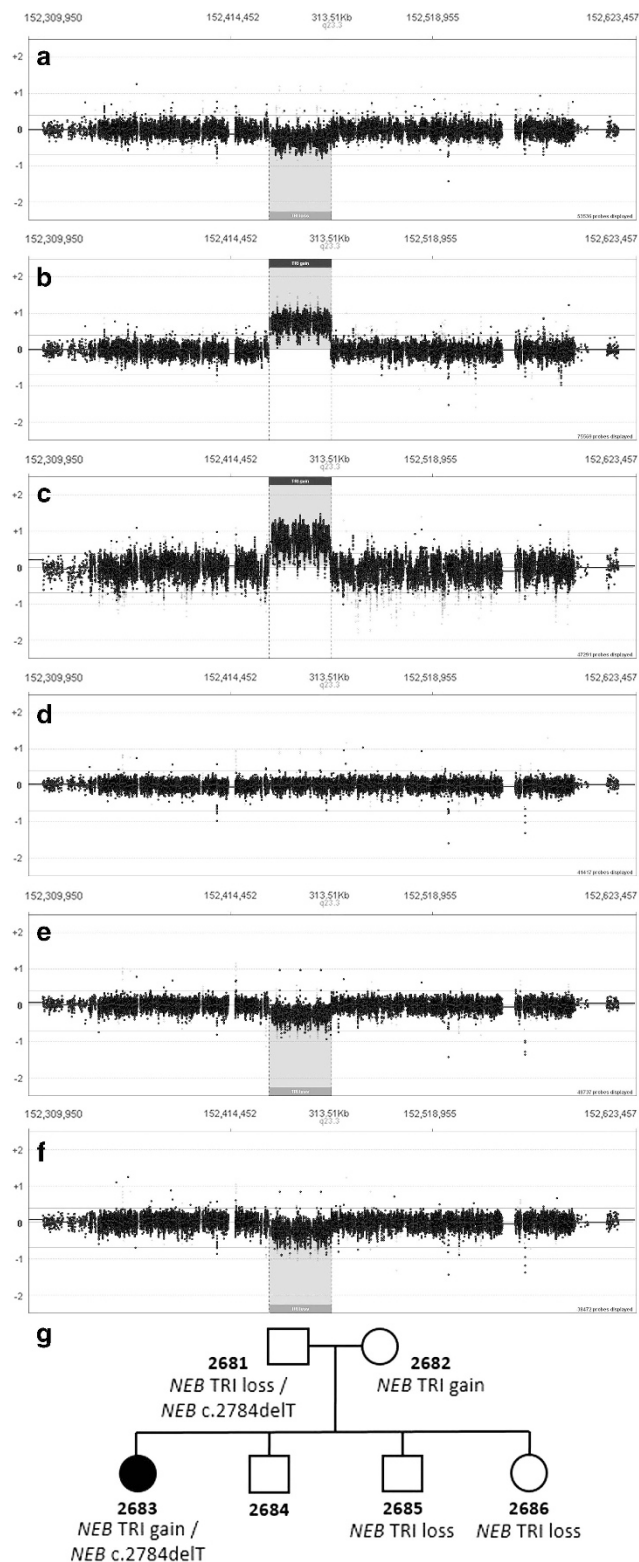
All the *NEB* TRI deletions identified in the NM families and controls were losses of one TRI copy, that is, 5/6 copies in total. One-copy losses were identified in five control samples (5/60) and in nine NM families (9/196). In four of the NM families, the disease-causing variants, either in *NEB* or in another NM-causing gene, had been published previously. Novel disease-causing variants are published here for families F2 and F411. In family F2, the patients have inherited a splice site variant, c.508-7T>A, which was shown on the RNA level to activate a cryptic acceptor splice site in intron 7, causing an insertion of five nucleotides leading to a frameshift p.Val170fs. In family F411, the patient is compound heterozygous for two different frameshift variants, c.24372\_24375del, in exon 172, and c.13134del in exon 86. In three families one or both disease-causing variants remain unknown (Table 3).

In the control sample cohort, a gain of one additional copy, that is, 7/6 copies, was identified in one sample (1/60; Table 2). However, among the NM families, there was a wider spectrum of copy number gains (Tables 2, 3 and 4). Almost half (8/17) of the identified gains in the NM samples included two to four additional copies, that is, 8–10/6 copies in total (Tables 2 and 4). This size category of CNVs was not identified in the control cohort. The patients in nine NM families carried one-copy gains. In one of these families, the disease-causing variants had previously been identified in *LMO3* (Index Patient 3353, Table 3). Using exome sequencing, we found a frameshift variant c.24475\_24479dupCACAA in exon 173 of *NEB*, in another family (F407) segregating on the same allele as the TRI gain (Index Patient 4073, Table 3). The consanguineous healthy parents are both carriers of one TRI copy gain each, as well as a *NEB* frameshift variant, and the patient is homozygous for the frameshift variant and the TRI gain. In another family (F321), we identified, in addition to the *NEB* TRI one-copy gain, a novel complex deletion–insertion rearrangement concerning the *NEB* exons 69–71 and intron 65 (Index Patient 3213, Table 3) using the NM-CGH array. The second variant remained unidentified in this family. One or both disease-causing variants remain unknown altogether in seven families with a gain of one TRI copy.

TRI copy gains of two or more copies were identified in patients of eight NM families. In five of these families, another heterozygous pathogenic *NEB* variant had been identified previously or is published here (Table 4). We had samples from all family members in three of these families and these are described in more detail below.

### Family 268

The index case 2683 (typical NM) has inherited a *NEB* TRI four-copy gain from the mother and a *NEB* exon 28 frameshift variant (c.2784delT) from the father. The unaffected father has the frameshift variant (c.2784delT) on one allele and a TRI deletion (one-copy loss)



**Figure 1** (a–g) NM-CGH profiles of *NEB* in family 268. The index case (c; typical NM) has inherited a *NEB* TRI gain from the mother (b) and a *NEB* (c.2784delT) frameshift mutation from the father (a; not shown in the NM-CGH array). The father and two siblings (e and f) have a deletion and one brother (d) the normal copy number of *NEB* TRI. The index patient has the typical form of NM and all other family members are unaffected. The family pedigree is shown in g.

**Table 3** NM families carrying benign variation of the *NEB* TRI

Family ID	DNA ID	Family member status, myopathy and clinical severity	TRI copy number	Identified mutation(s) in <i>NEB</i> /other genes and additional details
F2	521	Father	5	(1) <i>NEB</i> , int7; c.508-7T>A; p.Val170fs
	522	Mother	6	(2) <i>NEB</i> , ex122; c.19097G>T; p.(Ser6366Ile) <sup>17</sup>
	<b>523</b>	<b>Index, typical NM</b>	5	1 and 2
	524	Sibling	6	
	525	Sibling	6	(2) <i>NEB</i> , ex122; c.19097G>T; p.(Ser6366Ile) <sup>17</sup>
	<b>526</b>	<b>Affected sibling, typical NM</b>	5	1 and 2
F132	<b>1321</b>	<b>Index, typical NM</b>	5	<i>NEB</i> , int156; c.22800+2T>C; p.? <sup>17</sup>
F168	<b>1683</b>	<b>Index, severe NM</b>	5	<i>ACTA1</i> , c.220G>A; p.(Glu74Lys) ( <i>de novo</i> dominant) <sup>23</sup>
F193	<b>1933</b>	<b>Index, severe NM</b>	5	<i>LMOD3</i> , c.1069G>T; p.(Glu357*) (hoz) <sup>24</sup>
F298	2981	Father	5	(1) <i>NEB</i> , ex50; c.6678del; p.(Lys2226fs) <sup>2</sup>
	2982	Mother	6	(2) <i>NEB</i> , ex122; c.19097G>T; p.(Ser6366Ile) <sup>17</sup>
	<b>2983</b>	<b>Index, typical NM</b>	5	1 and 2
F333	<b>3333</b>	<b>Index, NM other forms (mild)</b>	5	<i>NEB</i> , ex130; c.19992_19999dup; p.(Asp6667fs) <sup>2</sup> <i>NEB</i> , del ex43-45; c.5238+335_5764-407del; p.(Arg1747_Asp1921del) <sup>2</sup>
F398	3981	Father	NA	
	3982	Mother	5	
	<b>3983</b>	<b>Index, no data</b>	5	
F403	<b>4033</b>	<b>Index, no data</b>	5	
F411	4111	Father	6	(1) <i>NEB</i> , ex172; c.24372_24375del; p.(Arg8125fs)
	4112	Mother	5	(2) <i>NEB</i> , ex86; c.13134delA; p.(Arg4378fs)
	<b>4113</b>	<b>Index, typical NM</b>	5	1 and 2
F128	1281	Father	7	
	1282	Mother	6	
	<b>1283</b>	<b>Index, severe NM</b>	NA	No sample from the patient
F234	<b>2343</b>	<b>Index, no data</b>	7	
F273	2731	Father	7	
	2732	Mother	6	
	<b>2733</b>	<b>Index, NM other forms (severe)</b>	7	
F321	<b>3213</b>	<b>Index, no data</b>	7	<i>NEB</i> , ex69-int71, c.[10080_10452+2141del;insAATT; ins9414+1085_9414+1189]; p.(Asn3382_Glu3493del)
F335	<b>3353</b>	<b>Index, severe NM</b>	7	<i>LMOD3</i> , c.723_733del; p.(Asp242fs); c.976G>C; p.(Gly326Arg) <sup>24</sup>
F368	<b>3683</b>	<b>Index, severe NM</b>	7	
F377	<b>3773</b>	<b>Index, typical NM</b>	7	No nemaline bodies on biopsy
F404	4041	Father	7	
	<b>4043</b>	<b>Index, no data</b>	7	
	4071	Father	7	<i>NEB</i> , ex173; c.24475_24479dup; p.(Asn8160fs) (hez) <sup>2</sup>
F407	4072	Mother	7	<i>NEB</i> , ex173; c.24475_24479dup; p.(Asn8160fs) (hez) <sup>2</sup>
	<b>4073</b>	<b>Index, severe NM</b>	8 (= 4+4)	<i>NEB</i> , ex173; c.24475_24479dup; p.(Asn8160fs) (hoz) <sup>2</sup>

Abbreviations: hez, heterozygous; hoz, homozygous; Index, index patient; NM, nemaline myopathy; NA, not available. Affected family members are indicated with bold text. Total *NEB* TRI copy numbers are shown: 6 is the normal copy number, 5 copies indicates a loss of one copy and 7 copies a gain of one copy. In family 407, the Index case 4073 has inherited a gain of one copy from both parents giving the total of 8 copies. Nebulin mutations are in refseq: NM\_001271208.1. and the exon numbering according to Donner et al.<sup>4</sup> *ACTA1* mutations are in refseq NM\_001100.3 and *LMOD3* mutations in refseq NM\_198271.3.

on the other allele, which suggests that one-copy deletions of TRI would not be pathogenic. We suggest that the TRI four-copy gain carried by the healthy mother, and inherited by the index patient, might be disease causing (Figure 1). The pedigree of the family (Figure 1g) shows that the *NEB* TRI gain segregates with the disorder, and only the combination of the *NEB* TRI gain and the frameshift variant result in the NM phenotype.

#### Family 7

The index case 573 (mild NM) has inherited a *NEB* TRI three-copy gain from the father and a frameshift variant in *NEB* exon 134 (c.20446delC) from the mother. We suggest that the three additional TRI copies might be pathogenic. The parents and sibling are unaffected mutation carriers.

#### Family 272

The index case 2723 (other form of NM) has inherited a *NEB* TRI four-copy gain from the father and a frameshift variant in *NEB* exon 81 (c.12048\_12049delGA) from the mother. We suggest that the four additional TRI copies might be pathogenic. The parents are unaffected mutation carriers.

Exome sequencing of samples from the index patients in all three families (F268, F7 and F272) did not reveal any variants with serious consequences for gene function in any of the other currently known NM-causing genes.

In summary, the copy numbers deviated only by one copy in the control samples, whereas deviations larger than that were only found in NM families. The difference is not, however, statistically significant (Fisher's exact test; Table 2). Based on this study, the *NEB* TRI CNVs are the most common large recurrent *NEB* variation characterized to date.



**Table 4** NM families carrying possibly pathogenic variation of the *NEB* TRI

Family ID	DNA ID	Family member status,		TRI copy number	Identified mutation(s) in <i>NEB</i> /other genes and additional details
		myopathy and clinical severity			
F7	571	Father		9	
	572	Mother		6	<i>NEB</i> , ex134; c.20446del; p.(Arg6816fs)
	<b>573</b>	<b>Index, mild NM</b>		9	<i>NEB</i> , ex134; c.20446del; p.(Arg6816fs)
	574	Sibling		6	<i>NEB</i> , ex134; c.20446del; p.(Arg6816fs)
F125	1251	Father		6	<i>NEB</i> , ex78; c.11770_11787del; p.(Val3924_Asn3929del) <sup>17</sup>
	1252	Mother		9	
	<b>1253</b>	<b>Index, NM other forms (typical)</b>		NA	
F243	<b>2433</b>	<b>Index, severe NM</b>		10	<i>NEB</i> , int61; c.8685+1G>A; p.?
F272	2721	Father		10	
	2722	Mother		NA	<i>NEB</i> , ex81; c.12048_12049del; p.(Lys4017fs)
	<b>2723</b>	<b>Index, NM other forms (typical)</b>		10	<i>NEB</i> , ex81; c.12048_12049del; p.(Lys4017fs)
F315	3151	Father		9	
	3152	Mother		6	
	<b>3153</b>	<b>Index, NM other forms (typical)</b>		9	
F323	<b>3233</b>	<b>Index, ?typical NM</b>		10~11	
F401	<b>4013</b>	<b>Index, no data</b>		8	
F268	2681	Father		5	<i>NEB</i> , ex28; c.2784del; p.(Pro928fs)
	2682	Mother		10	
	<b>2683</b>	<b>Index, severe NM</b>		10	<i>NEB</i> , ex28; c.2784del; p.(Pro928fs)
	2684	Sibling		6	<i>NEB</i> , ex28; c.2784del; p.(Pro928fs)
	2685	Sibling		5	
	2686	Sibling		5	

Abbreviations: Index, index patient; NM, nemaline myopathy; NA, not available.

Affected family members are indicated with bold text. Total *NEB* TRI copy numbers are shown; 6 is the normal copy number, 8–11 copies indicate a gain of 2–5 copies. In family 268, there are also family members with benign variation of 5 copies (a loss of one copy). Nebulin mutations are in refseq: NM\_001271208.1. and the exon numbering according to Donner et al.<sup>4</sup>

### Breakpoint analysis

Breakpoint analyses were done for all samples that showed TRI CNVs. The results achieved using the NM-CGH array indicate that the locations of the breakpoint regions are the same for both gains and losses (Figure 2). All variants start in intron 105, but the exact breakpoint cannot be determined due to a probe gap within this intron because of repeat elements, across which no unique probes could be designed. The stop region seems to differ slightly between different samples but it is always inside intron 81. Repeat elements, that is, LINEs (CR1 and L2) reside in introns 89, 97 and 105 and long-terminal repeats (ERVL) reside in introns 82, 90 and 98. In addition, DNA transposon fossils (MER113A) are present in introns 88, 96 and 104.

### DISCUSSION

In the current study, 13% (26/196) of the studied NM families showed a *NEB* TRI CNV. The identified losses were deletions of one copy, which did not seem to segregate with the disorder and were found in the control population (8%) more frequently than in the NM families (5%). We estimate that the loss of one copy would cause a 1458 bp shortening of various transcripts and a 486-amino-acid shortening of the translated protein. One TRI copy encodes 486 amino acids corresponding to two nebulin super repeats.<sup>4</sup> The two remaining TRI copies would be enough for the allele to produce a functional protein.

*NEB* transcripts of many different lengths are known to be produced in normal muscle, and therefore, this alteration might not cause a drastic effect on the thin actin filament structure of the sarcomere.<sup>22</sup> In family 268, the father showed a one-copy loss of TRI and a frameshift variant on the other allele, and he is healthy. In six families, the TRI deletion has also been found to be accompanied by

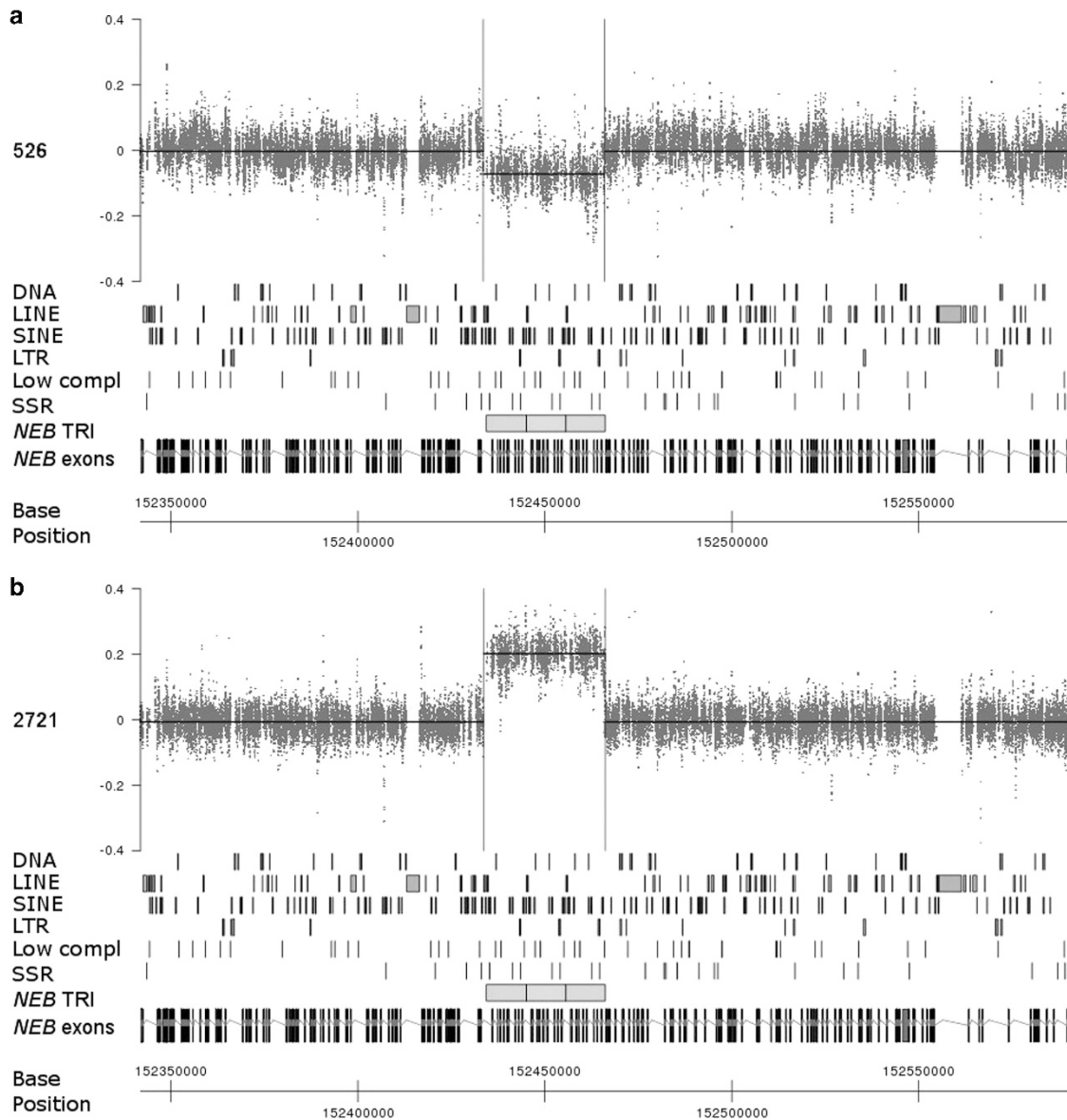
two identified pathogenic variants (Table 3). In addition, it is also frequent in controls (8%). Thus, one-copy deletions of *NEB* TRI are unlikely to have a serious impact on protein function.

The identified gains have been categorized into two groups: gain of one TRI copy, which we interpret to be non-pathogenic, and gains of two to four TRI copies, which we interpret to be pathogenic. In one family (Index Patient 3353, Table 3) with a one-copy gain, two other causative variants in another NM-causing gene, *LMOD3*, had previously been identified, and in another family (Index Patient 4073, Table 3), we found a frameshift variant in *NEB* exon 173 segregating with the one-copy gain. Also, one control sample was identified with one TRI copy gain (1/60). Thus, it seems that deviations of one TRI copy in an allele, be it loss or gain, are tolerated.

However, 4% of the NM families had patients with gains of more than one TRI copy (8–10/6 copies present). Both disease-causing variants in *NEB* had not previously been identified in these families. In those families with parental samples available for analysis, the TRI CNV segregated with the disorder. Furthermore, this type of aberration was not encountered among the control samples. We speculate that gains of more than one copy may disrupt the stability or secondary structure of the mRNA, and also the translation process. If so, no nebulin protein would be produced from this allele.

If, on the other hand, RNA processing and translation were to take place, the TRI gain would cause an increase in the length of the transcript and further of the nebulin protein. Each TRI gain would add 1458 bases to the mRNA, and 486 amino acids to the protein. Therefore, a gain of four copies would add 5.8 kb to the mRNA, and 1944 amino acids or eight super repeats to the protein.

The breakpoints of the TRI variations are difficult to characterize further. Normally, the entire TRI region is ~32 kb in size and each



**Figure 2** (a and b) Breakpoint analysis of *NEB* TRI copy number loss and gain. Sample 526 from Family 2 shows a one-copy loss (a) and sample 2721 from Family 272 shows a four-copy gain (b) of the *NEB* TRI. Breakpoint analysis shows repeat elements in the breakpoint regions. In the upper part of the Figure, the inferred breakpoints are shown with vertical lines and the inferred copy number of the corresponding fragment is indicated with horizontal lines. The tracks in the lower part of the Figure show the different repeat elements, the *NEB* triplicate region and the gene structure. DNA, DNA repeat elements; LTR, long-terminal repeat; Low Compl, low complexity DNA sequences; SSR, simple sequence repeats = microsatellite DNA.

one-copy gain would add ~10 kb to the gene. The identified losses of one TRI copy would set the region at around 20 kb in size. The homology of the TRI region has proven to be very challenging especially for PCR studies. The last introns of each TRI repeat (introns 89 and 97 and 105) contain repeat elements, such as *Alu* and *LINE* repeats. These transposable elements are known to be capable of being involved in NAHR, thought to be the most common underlying mechanism for recurrent CNVs.<sup>15,16</sup> The lack of further knowledge of these TRI variations makes it difficult to estimate the exact causative mechanism but NAHR seems plausible. It is thought that the human *NEB* TRI has emerged from two duplication events. This is based on the fact that the mouse *Neb* gene only contains one copy of this region (exons 82–89) and is lacking the *LINE*-L2 elements as well as the exons corresponding to human *NEB* exons 90–106.<sup>4</sup> These repeat

elements might explain the susceptibility of the *NEB* TRI region to the recurrent copy number changes discovered in this study.

The NM-CGH microarray is an effective and easy method for revealing *NEB* TRI CNVs. However, if there is a deletion in one allele and a duplication in the other, the copy number would look normal and these cases would thus remain unidentified. Because the normal copy number of the *NEB* TRI is six instead of two, it can be difficult to differentiate between subtle differences in size, for example, between gains of three copies and gains of four copies, especially if the DNA quality is poor. Also, the NM-CGH array cannot determine which allele harbours the CNV. Therefore, studying parental samples is necessary to verify the origin of the identified variant. Regarding duplications, the identification of the location or the orientation of the duplicated region requires alternative methods.

## CONCLUSIONS AND FUTURE PROSPECTS

The NM-CGH microarray has revealed a novel, recurrent CNV of the *NEB* triplicate region, apparently the most common large recurrent variant in *NEB*. The NM-CGH array is a very easy and effective method for detecting these CNVs. We will continue to use the NM-CGH array as a first tier diagnostic method for new NM families, as well as analyzing samples from families and patients whose causative variants have remained unidentified.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank Professor Dimitrios Zafeiriou and Dr Sascha Vermeer for sending us samples from one nemaline myopathy family each, Professor Johan T den Dunnen for submitting variants into the LOVD database and help with the HVGs nomenclature, and Marilotta Turunen for excellent technical assistance. This study was supported by grants from the Sigrid Jusélius Foundation, the Association Française contre les Myopathies, the Finska Läkaresällskapet and the Medicinska understödsföreningen Liv och Hälsa.

## AUTHOR CONTRIBUTIONS

KK designed the original as well as the updates of the NM-CGH microarray, set up the NM-CGH method in the laboratory, ran half of the array hybridizations, carried out analysis and interpretation of the data, performed exome sequencing analyses and prepared the manuscript. V-LL selected samples for the NM-CGH analyses, analyzed the dHPLC and Sanger sequencing results and wrote parts of the manuscript. AL did the bioinformatic analyses and wrote parts of the manuscript. LA ran half of the NM-CGH hybridizations. JL performed the minigene experiment for family F2. CW-P was responsible for the project as PI. Her contribution included collection of samples and clinical correlates, as well as writing parts of the manuscript. KP was responsible for the project as Co-PI. She participated in the design of the microarray, interpretation of results and writing parts of the manuscript.

- 1 Wallgren-Pettersson C, Laing NG: Report of the 70th ENMC International Workshop: nemaline myopathy, 11-13 June 1999, Naarden, The Netherlands. *Neuromuscul Disord* 2000; **10**: 299-306.
- 2 Lehtokari VL, Kiiski K, Sandaradura SA *et al*: Mutation update: the spectra of nebulin variants and associated myopathies. *Hum Mutat* 2014; **35**: 1418-1426.

- 3 Wallgren-Pettersson C, Jasani B, Newman GR *et al*: Alpha-actinin in nemaline bodies in congenital nemaline myopathy: immunological confirmation by light and electron microscopy. *Neuromuscul Disord* 1995; **5**: 93-104.
- 4 Donner K, Sandbacka M, Lehtokari VL, Wallgren-Pettersson C, Pelin K: Complete genomic structure of the human nebulin gene and identification of alternatively spliced transcripts. *Eur J Hum Genet* 2004; **12**: 744-751.
- 5 Bang ML, Li X, Littlefield R *et al*: Nebulin-deficient mice exhibit shorter thin filament lengths and reduced contractile function in skeletal muscle. *J Cell Biol* 2006; **173**: 905-916.
- 6 Witt CC, Burkart C, Labeit D *et al*: Nebulin regulates thin filament length, contractility, and Z-disk structure in vivo. *EMBO J* 2006; **25**: 3843-3855.
- 7 Chandra M, Mamidi R, Ford S *et al*: Nebulin alters cross-bridge cycling kinetics and increases thin filament activation: a novel mechanism for increasing tension and reducing tension cost. *J Biol Chem* 2009; **284**: 30889-30896.
- 8 Pfuhl M, Winder SJ, Castiglione Morelli MA, Labeit S, Pastore A: Correlation between conformational and binding properties of nebulin repeats. *J Mol Biol* 1996; **257**: 367-384.
- 9 Labeit S, Gibson T, Lakey A *et al*: Evidence that nebulin is a protein-ruler in muscle thin filaments. *FEBS Lett* 1991; **282**: 313-316.
- 10 Wang K: Titin/connectin and nebulin: giant protein rulers of muscle structure and function. *Adv Biophys* 1996; **33**: 123-134.
- 11 Anderson SL, Ekstein J, Donnelly MC *et al*: Nemaline myopathy in the Ashkenazi Jewish population is caused by a deletion in the nebulin gene. *Hum Genet* 2004; **115**: 185-190.
- 12 Lehtokari VL, Greenleaf RS, DeChene ET *et al*: The exon 55 deletion in the nebulin gene—one single founder mutation with world-wide occurrence. *Neuromuscul Disord* 2009; **19**: 179-181.
- 13 Kiiski K, Laari L, Lehtokari VL *et al*: Targeted array comparative genomic hybridization—a new diagnostic tool for the detection of large copy number variations in nemaline myopathy-causing genes. *Neuromuscul Disord* 2013; **23**: 56-65.
- 14 Bjorklund AK, Light S, Sagit R, Elofsson A: Nebulin: a study of protein repeat evolution. *J Mol Biol* 2010; **402**: 38-51.
- 15 Vissers LE, Bhatt SS, Janssen IM *et al*: Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum Mol Genet* 2009; **18**: 3579-3593.
- 16 Gu W, Zhang F, Lupski JR: Mechanisms for human genomic rearrangements. *Pathogenetics* 2008; **1**: 4.
- 17 Lehtokari VL, Pelin K, Sandbacka M *et al*: Identification of 45 novel mutations in the nebulin gene associated with autosomal recessive nemaline myopathy. *Hum Mutat* 2006; **27**: 946-956.
- 18 Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E: Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* 2004; **20**: 3413-3422.
- 19 R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2013.
- 20 Durinck S, Bullard J, Spellman PT, Dudoit S: GenomeGraphs: integrated genomic data visualization with R. *BMC Bioinformatics* 2009; **10**: 2.
- 21 Lawrence M, Gentleman R, Carey V: rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 2009; **25**: 1841-1842.
- 22 Laitila J, Hanif M, Paetau A *et al*: Expression of multiple nebulin isoforms in human skeletal muscle and brain. *Muscle Nerve* 2012; **46**: 730-737.
- 23 Laing NG, Dye DE, Wallgren-Pettersson C *et al*: Mutations and polymorphisms of the skeletal muscle alpha-actin gene (ACTA1). *Hum Mutat* 2009; **30**: 1267-1277.
- 24 Yuen M, Sandaradura SA, Dowling JJ *et al*: Leiomodion-3 dysfunction results in thin filament disorganization and nemaline myopathy. *J Clin Invest* 2014; **124**: 4693-4708.