

Research Article

A Versatile Method for Gene Dosage Quantification: Genotyping of the Spinal Muscular Atrophy Related Locus Case Study

Slobodanka Radovic^{1*}, Giorgia Dubsky de Wittenau², Federica Cesca², Nina Mandl³, Francesco Curcio^{2,4}, Incoronata Renata Lonigro^{2,4}, Michele Morgante^{5,6}

¹* Department of Immunology, University of Udine, Italy

² Department of Medical, University of Udine, Italy

³ Medical and Pharmaceutical Biotechnology, University of Applied Sciences, Austria

⁴ Department of Laboratory Medicine, SOC Institute of Clinical Pathology, Italy

⁵ Applied Genomics Institute, Luigi Danieli Technology Park, Italy

⁶ Department of Agricultural and Environmental Sciences, University of Udine, Italy

***Corresponding author:** Slobodanka Radovic, Department of Immunology, University of Udine, IGA Technology, Udine, Italy.
Tel: +390432629783; Fax: +390432603887; Email: sradovic@igatechnology.com

Citation: Radovic S, Wittenau GD, Mandl N, Cesca F, Curcio F, et al. (2017) A Versatile Method for Gene Dosage Quantification: Genotyping of the Spinal Muscular Atrophy Related Locus Case Study. Biomark Applic: BMAP- 106. DOI: 10.29011/BMAP-106. 100106

Received Date: 27 June, 2017; **Accepted Date:** 4 July, 2017; **Published Date:** 11 July, 2017

Abstract

Objectives: A comparison of the individual genomes within a species demonstrates that structural variation, including Copy Number Variation (CNV), is a major contributor to phenotypic diversity and evolutionary adaptation. CNVs lead to the deregulation of gene expression and could account for the development of a number of genomic disorders. Thus, the development of efficient, rapid and accurate CNV screening is of fundamental importance. We report a method that enables the simultaneous determination of the copy numbers of different genetic targets as well as the discrimination among highly similar/almost identical DNA sequences that differ by only one single nucleotide variant.

Methodology: The PCR co-amplification and single-base extension technologies are used to identify the copy number of target sequences, the primary spinal muscular atrophy-determining gene, SMN1, and the disease modifier gene, SMN2, in a cohort of 160 subjects previously genotyped with MLPA-based and qPCR-based techniques. The copy numbers of SMN1/ SMN2 were relative to a reference sequence of known genomic copy number (Albumin and Factor VIII genes).

Results: We developed an efficient and accurate quantification platform which can be adopted as an alternative to other technologies for CNV evaluation, like as MLPA-based or qPCR-based techniques. In addition, our method has proved effective in resolving a diagnostic event by questionable result in using the previously mentioned technologies.

Conclusions: The reliability, low-cost and potential for high-throughput make our method suitable for screening large populations as well as for use as a tool in clinical settings for genetic diagnosis/prognosis.

Keywords: Copy Number Variation; Gene Dosage; Hereditary Diseases; Multiplex-PCR; SNPs identification; Single Base Extension

Introduction

Copy Number Variations (CNVs) are a form of genomic diversity involving DNA sequences >50 bp that are present in the genome in a variable number of copies [1]. In plants, CNVs are involved in resistance to biotic [2,3] and biotic stresses [4] and underlie the plant's growth and development [5,6]. In domesticated animals, CNVs have been associated with different morphological traits as well as with a variety of diseases and developmental disorders [7]. In humans, CNVs are a primary source of genetic variation [8], with up to 20% of the genome exposed to CNV [9,10], and approximately 35% of the genes encompassed totally or partially by a CNV [11]. Many human disease conditions have been identified that are either caused by CNVs or that have a relative risk that is increased by CNVs [12-16]. The extent to which CNVs are likely to contribute to the diversity of human phenotypes, including "Single Gene Defects", genomic disorders and complex diseases has been increasingly recognised, and CNVs are now widely recruited for genome-wide association studies, with the aim of assessing their influence on human disease causation/susceptibility [12,17,18]. A number of different tools have been developed to assess CNV. Nevertheless, some of the current techniques that are used for gene dosage determination in molecular diagnostics have several limitations and disadvantages. Cytogenetic analysis and Southern blotting are time-consuming and require a dedicated laboratory and large amounts of DNA. The shortcomings of other approaches, including Fluorescent In Situ Hybridisation (FISH) [19], Comparative Genomic Hybridisation-Based Microarray Approaches (CGH) [20,21] quantitative real-time PCR [22], Fluorescence-Labelled CE [23] and Multiplex Ligation-Dependent Probe Amplification (MLPA) [24] usually manifest significant false-positive rates and require an increased handling time and cost, while the recently introduced Next Generation Sequencing (NGS) approach presents substantial computational and bioinformatics challenges [1].

Here, we describe a method for determining the CNV within a genome. This method combines multiplex-PCR and Single Base Extension (SBE) genotyping, providing an efficient and reliable method to analyse gene deletions and duplications. The specific multiplex PCR protocol involves the amplification of the target locus/loci with an unknown copy number and reference locus/loci with a known copy number in a single reaction followed by SBE genotyping and Capillary Gel Electrophoresis. The CNVs are determined by calculating the ratios of the target signal over reference signal for a test sample and by comparing these ratios to those obtained using a control sample(s) for which all of the copy numbers are known. A comparison of these relative ratios results in a dosage quotient, indicating the copy number of the target region in

the test sample. We introduce this method and verify its robustness and versatility by applying it as a diagnostic protocol for detecting SMN1 deletion/conversion and for further determining the copy numbers of the SMN1 and SMN2 genes that are involved in Spinal Muscular Atrophy (SMA). SMA is a severe neuromuscular disease characterized by the degeneration of alpha motor neurons in the spinal cord, which results in progressive muscle weakness and paralysis [25]. The SMN1 gene that is located in a complex region of chromosome 5q13 [26] in 95% of SMA cases is deleted [27]. A nearly identical gene, SMN2, which differs due to a few nucleotide changes [28], plays no role in SMA. Nevertheless, because SMN2 codes for the same protein product as SMN1, although at much lower rate, and can be present in multiple copies within the genome, SMN2 is considered to be a disease modulator that decreases the severity of the SMA phenotype in a dose-dependent manner [29]. We selected the SMA genetic locus to validate this method because it is a well-characterized disease-causing locus for which SNP identification and CNV quantification have relevant diagnostic and/or prognostic roles. Furthermore, we take advantage of a DNA cohort of SMA-affected subjects and carriers that most of them have already been well-genotyped in our laboratory by other semi-quantitative and reference methods [30,31]. This new method of gene copy numbers evaluation gave undoubted results for all individuals tested, including a male subject for whom the Real-Time and the MLPA methods had given questionable results.

Materials and Methods

Subjects

We genotyped the DNA samples of 160 subjects. All of the DNAs of the subjects were previously characterised for SMN1 and SMN2 copy numbers by multiplex real-time PCR [30] and by MLPA [31].

Nucleic Acid Extraction

The genomic DNA was isolated from the peripheral blood leukocytes by a PurageneTM DNA Purification Kit (Gentra Systems, Milan, Italy), according to the manufacturers protocol. The DNA concentration was determined using a Nano Drop 1000 Spectrophotometer (Thermo Scientific, Milano, Italy).

Primer Design

The PCR primers that flanked the marker polymorphisms in exon 7 and in exon 8 of the target SMN genes as well as the primers that were used to amplify reference genes Albumin (ALB) on chromosome 4 and Coagulation Factor VIII (F8) on chromosomes X, were designed by using Primer3 (<http://www-genome.wi.mit.edu/cgi-bin/primer/primer-3www.cgi>). The primers were designed to have similar melting temperatures and different PCR product lengths, as is optimal for multiplex PCR. The SBE primers were designed to terminate amplification one nucleotide before the nucleotide of interest and with different lengths to avoid peak

overlaps and interferences. Both of the SBE oligonucleotide that were used to detect nucleotide substitutions between the SMN1 and SMN2 genes were reverse primers designed on the (-)-strand. Thus, the oligonucleotide that was used to detect the C and T residues in exon 7 incorporated G for SMN1 and A for SMN2, while the oligonucleotide that was used to detect the G and A residues in exon 8 incorporated C for SMN1 and T for SMN2. All of the PCR and SBE primer sequences are included in (Table 1).

Primer name	5'-3' sequence
SMN_exon7_for	TGTCTTGAAACAAAATGCTT
SMN_exon7_rev	CACTTCATAATGCTGGCAGA
SMN_exon8_for	CCACTAAAGAACGATCAGACAGA
SMN_exon8_rev	TTTCTCAACTGCCTCACCA
ALB_for	CCGTGGTCTGAACCAGTTA
ALB_rev	CTCCTTCTCAGAAAGTGTGCAT
F8_for	TGGCATGGAAGCTTATGTC
F8_rev	AGGGAGCATAGTCCCAGTCC
SMN_exon7_sbe	AATGTGAGCACCTCCCTCTTTGATTTGTCT
SMN_exon8_sbe	GTGCTGGCCTCCCACCCCCACC
ALB_sbe	CTGCACAGAATCCTTGGTAACAGGGGAC
F8_sbe	AGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAAC CCAAC

Table 1: PCR and SBE primer sequences.

Multiplex PCR and SBE

The multiplex PCRs were performed in a final volume of 25 μ l. We combined the template (30 ng of genomic DNA) with the KAPA2G Fast HS Ready Mix PCR Kit Premix (Kappa Bio systems, Wilmington, MA, USA) and locus-specific primers (10 μ M each; Sigma Aldrich), as recommended by the KapaTaq protocol. The thermo-cycling conditions consisted of an initial denaturation step of 95°C for 2 min., followed by 26 cycles of 95°C for 15 s, 56°C for 15s and 72°C for 7s, with a final extension step at 72°C for 30s. According to the manufacturer's instructions, 3 μ l of PCR products were incubated with Exo-SAP IT (Amersham Biosciences) prior to the primer extension reaction. Primer extension was carried out with the SNaP shot Multiplex Ready Reaction Mix (AppleraBiosystem). The reaction was performed in a total volume of 10 μ l, containing 3 μ l of cleaned PCR products, 1 μ l of SNaP shot premix, 1 μ l each of 0.2 μ M SBE primers and 2 μ l of nuclease-free H₂O. The primer extension thermo-cycling conditions consisted of 26 cycles of 96°C for 10s, 50°C for 15s and 60°C for 30s.

Following the primer extension, the reactions products were purified by SAP (Amersham Biosciences), according to the manu-

facturer's instructions. The cleaned products were combined with 0.2 μ l of GeneScan-120 LIZ Size Standard Mix and 9.8 μ l of form amide and run on Applied Bio systems 3730 DNA Analyzer (Applied Bio systems). The peaks of dye intensities corresponding to extensions of the SBE primers were determined by inspecting the output from the ABI 3730 DNA Analyzer.

Data Analysis

The quantification of the SMN1 and SMN2 copy numbers was performed independently for exon 7 and exon 8 by dividing the height of each gene-specific peak by the sum of the 2 endogenous control peaks (F8 and ALB). The ALB gene on chromosome 4 is present in 2 copies per diploid genome. F8 on chromosome X is present as a single copy gene per male diploid genome. The ratio was then compared, separately for males and females, to the ratios obtained from the three control samples with known SMN1 and SMN2 copy numbers to obtain a dosage quotient that indicated the copy number of each SMN gene.

Results

We used the PCR co-amplification and SBE technologies to determine the copy number of a target sequence of an unknown genomic copy number relative to a reference sequence of a known genomic copy number. Briefly, upon target and reference sequence co-amplification, the excess dNTPs and primers were removed by Shrimp Alkaline Phosphatase (SAP) and exonuclease treatment. The SBE of a primer that was specific for the target sequence as well as of a primer that was specific for the reference sequence was carried out in the presence of fluorescently labelled ddNTPs. The unincorporated fluorescent ddNTPs were then removed by SAP, and the extension products were readily identified using a DNA sequence detector. The quantification of the target sequence is achieved by the comparison of the target fluorescent signal versus the reference signal, exploiting the fundamental concept that the accumulation of fluorescence is proportional to the amplification of genomic regions. We utilised qualitative and quantitative capacities of this method by applying it as a diagnostic protocol for the identification of SMA carriers and for the characterisation of affected individuals, concurrently establishing the copy number of the primary SMA-determining gene SMN1 and the potential disease modulator gene SMN2. The assay was applied to a cohort of 160 individuals with different SMN1 and SMN2 copy numbers, including 88 apparently healthy individuals from the general population, 14 SMA-affected patients and 58 SMA carriers (Table 2).

Number of subjects	Genotype (SMN1:SMN2)	Expected total copy number (SMN1+SMN2)	Measured total copy number (mean \pm SD)	exon7 measured SMN1/SMN2 ratio (mean \pm SD)	exon8 measured SMN1/SMN2 ratio (mean \pm SD)	Expected SMN1/SMN2 ratio
5	2:0	2	2.28 \pm 0.22			
4	1:1	2	2.11 \pm 0.17	1.07 \pm 0.03	1.10 \pm 0.02	1.00
20	1:2	3	3.08 \pm 0.29	0.52 \pm 0.03	0.54 \pm 0.06	0.50
40	2:1	3	3.02 \pm 0.32	2.05 \pm 0.06	2.00 \pm 0.02	2.00
32	1:3	4	3.87 \pm 0.34	0.34 \pm 0.03	0.35 \pm 0.02	0.33
36	2:2	4	3.97 \pm 0.31	1.09 \pm 0.04	1.11 \pm 0.03	1.00
2	3:1	4	4.15 \pm 0.21	3.02 \pm 0.10	2.95 \pm 0.05	3.00
5	2:3	5	5.03 \pm 0.21	0.62 \pm 0.03	0.67 \pm 0.06	0.66
2	1:4	5	4.85 \pm 0.39	0.26 \pm 0.01	0.22 \pm 0.02	0.25
9	0:3	3	3.12 \pm 0.14			
4	0:4	4	4.45 \pm 0.22			
1	0:5	5	5.22 \pm 0.20			

Table 2: A cohort of 160 individuals with different SMN1 and SMN2 copy numbers.

Base pair exchanges, C-to-T at position +6 in exon7 (c.840C>T) and G-to-A in the un translated region of exon8 (Burglen et al. 1996) [32] were used to distinguish between the two SMN genes (Figure 1A), while the regions of the ALB and F8 genes were used as internal references (Figure 1A and Materials and Methods). We unambiguously identified the SMN1 and SMN2 peaks as well as the reference peaks from the genotyping outputs (Figure 1B).

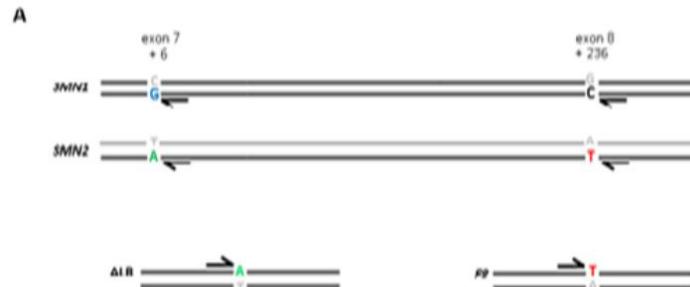


Figure 1A: SBE genotyping performed for target SMN genes and the reference Albumin and Factor 8 genes. For target SMN genes genotyped residues were C/T at position +6 in exon 7 and G/A at position +236 in exon 8. For reference Albumin (ALB) and Factor8 (F8) genes, genotyped residues were A at position +102 in exon 12 and T at position +52 in exon 8, respectively. Arrows represent SBE primers.

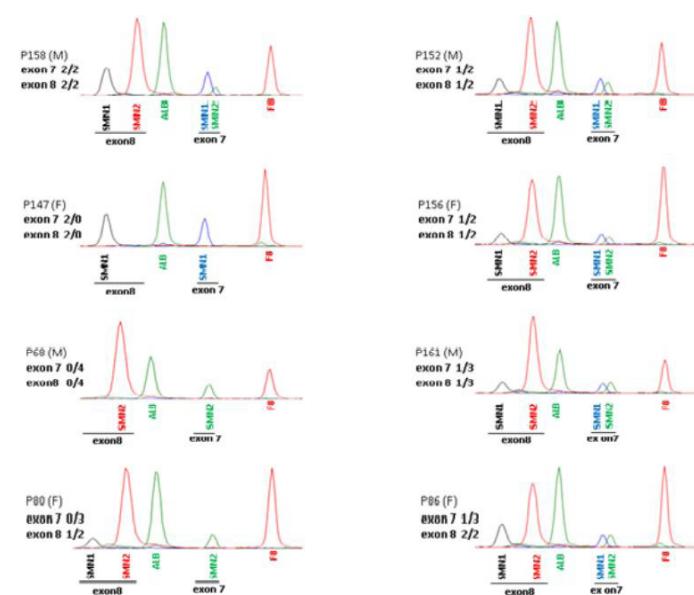


Figure 1B: SBE genotyping of multiplex PCR analysis from healthy controls, SMA patients and carriers with different SMN1/SMN2 ratios. SMN1 and SMN2 in exons 7 and 8 as well as ALB and F8 are indicated in colour code under respective peaks in sequencing outputs. Sample name, gender and SMN1/SMN2 copy numbers, separately reported for exon 7 and exon 8, and are indicated on the left side of each output. F = female; M = male; Individuals P80 (affected subject) and P86 (healthy carrier) experienced gene conversion of SMN1 to SMN2 in exon 7. Males have only one copy of the F8 gene since it is on chromosome X.

Two independent references were used simultaneously to avoid biases due to possible individual variations in the copy number of one reference sequence. Because ALB on chromosome 4 is present in two copies per diploid genome and F8 on chromosome X is present in two copies in females and in a single copy in males, we examined the reliability of our system by looking at the relative ratios of the two reference signals across males and females. The ALB/F8 ratios found in males (1.85 ± 0.10) were consistently doubled with respect to those found in females (0.98 ± 0.12), underlining the reliability and the quantitative nature of the assay. In addition, the fixed relative ratios between the two reference signals across individuals indicated stability in their copy numbers, confirming them as suitable references in all of the samples. It is

important to note that the reference sequence can be any genomic region with stable copy numbers. Control samples with known SMN1/SMN2 copy numbers (1/1, 1/2, 1/3, 2/2, 2/1 and 3/1) were run alongside the test samples. First, the controls were used to determine the proportionality between SMN1 and SMN2 relative signal intensities, separately for exon 7 and exon 8, showing the linearity of the assay in response to the varying copy numbers of the two SMN genes (exon7 R2=0.9915 and exon8 R2=0.9653) and further confirming the quantitative nature of the assay. Second, the target-to-reference ratios in the control subjects allowed for the assessment of the SMN1 and SMN2 copy numbers in the test samples with an unknown SMN dosage. The copy numbers of the SMN1 and SMN2 genes in the Test samples (T) compared to those of the Control samples (C) were calculated separately for males and females by the following equations:

$$\text{Peak heights SMN1 (T)}/[\text{ALB+F8(T)}] \text{ and } \text{Peak heights SMN2(T)}/[\text{ALB+F8(T)}]$$

$$\text{Peak heights SMN1(C)}/[\text{ALB+F8(C)}] \text{ Peak heights SMN2(C)}/[\text{ALB+F8(C)}]$$

This comparison resulted in a dosage quotient that indicated the copy number of each SMN gene in the test sample. For each individual, we quantified the SMN1 and SMN2 dosage independently for exon7 and exon 8, enabling the identification of the gene-conversion events that result in the creation of a hybrid SMN gene. The gene conversion of SMN1 to SMN2 in exon 7 is one cause of SMA. The +6 C-to-T substitution in SMN2 exon 7 decreases the activity of an exotic splice enhancer and alters the splicing pattern so that the SMN2 mRNA excludes the exon 7 sequences. Consequently, SMN2 produces insufficient amounts of the full-length SMN transcript and protein to rescue the SMA phenotype [26,28]. However, the conversion of SMN2 to SMN1 in exon 7 could be used as a therapeutic approach for SMA [34]. By identifying individuals with unequal SMN1/SMN2 ratios in the two exons we were able to distinguish between the hybrids genes derived from conversion events. Individual P86 had an SMN1/SMN2 ratio of 1:3 for exon7 and 2:2 for exon8, while individual P80 had a ratio of 0:3 in exon 7 and 1:2 in exon 8 (Figure 1B), indicating a conversion from SMN1 to SMN2 in exon7. Due to this conversion, individuals P86 and P80 became an SMA carrier and an SMA-affected individual, respectively.

We analysed each sample at least three times obtaining analogous results, demonstrating the reproducibility of our detection system for gene dosage determination (Table 2). The measured copy numbers were in accordance with the genotypes that were previously determined by real-time PCR [30] and MLPA [31], proving that we can successfully assign all genotypes with different SMN1/SMN2 gene copy numbers, including SMN1/SMN2 ratios equal and not equal to one. We initially designed SBE oligonucleotide on the (+)- and (-)-strands for to detect nucleotide substitutions in exon 7 and exon 8 of the SMN genes as well as sev-

eral oligonucleotide to genotype different ALB and F8 residues. This oligonucleotide were used in various combinations to obtain the optimal output pattern, i.e., comparable peak heights and clear peak separation to avoid interference due to overlapping signals (data not shown). All of the tested combinations gave distinct, well-separated peaks; however, for some combinations, there was a large discrepancy in the signal intensities favouring one peak (up to 20-fold). We were always able to correctly determine the copy number of both of the SMN genes. Good peak separation was obtained using SBE oligonucleotide that differed in length by at least 3 nucleotides.

Discussion

Here, we report a molecular tool for the targeted detection and characterisation of CNVs. At our knowledge we describe for the first time a PCR co-amplification and SBE-based technology in determining the copy number of a target sequence of unknown genomic copy number relative to a reference sequence of known genomic copy number. The quantification of the target sequence is achieved by comparing the target fluorescent signal versus the reference signal, exploiting the fundamental concept that the accumulation of fluorescence is proportional to the amplification of genomic regions. This method can be used to simultaneously determine the copy numbers of several different targets as well as to discriminate among highly similar/almost identical targets that differ by only one SNP, establishing their copy numbers. To fully exploit the potential of this method, we validated it by genotyping the human SMA locus, one of the most complex loci in terms of determining the correct CNV configuration due to the extreme similarity of the two SMN genes, making simultaneous and reliable SMN1 and SMN2 quantification quite challenging. A previous method based on the principle of primer extension has been developed for genotype determination at the SMN locus [35]. However, this method is more laborious and time-consuming than the method we describe here. Indeed, this previous method is based on two independent multiplex and complementary PCRs before the primer extension reaction, including a competitive PCR requiring a known amount of pre-constructed internal controls. Furthermore, this method identifies the SMN1/SMN2 ratio referring only to exon 7 of both genes; therefore, this method is not able to detect the conversion of SMN1 to SMN2. The method we report allows for the complete genotyping of the SMN locus by the concurrent identification of SNPs in exon 7 and exon 8 of the two SMN genes and of arbitrary residues in ALB and F8, which were selected as reference genes. We were able to accurately establish the copy number of the primary SMA-determining gene SMN1 and the potential disease modulator gene SMN2 in a cohort of 160 individuals with different SMN1 and SMN2 copy numbers. The cohort had already been well-genotyped in our laboratory by other semi-quantitative and reference methods [30,31]. We successfully identified all of the carriers and correctly determined the copy numbers of the SMN2 gene in the SMA patients. The precise

assessment of the SMN2 copy number is particularly important, as it could have a prognostic relevance for affected subjects [29] and as SMN2 is a common target in gene therapy [34]. In our assay, the fluorescence intensities appeared to be dose-dependent, never reaching a plateau, allowing us to unambiguously distinguish even between individuals with 4 copies (Figure 1B) and 5 copies (data not shown) of the SMN2 gene. We were also able to identify the hybrid genes derived from the gene conversion events of SMN1 to SMN2, which is valuable for carrier identification and for diagnosing SMA patients.

Our assay was highly reliable with the potential for high-throughput and low-cost; therefore, this assay is suitable for clinical diagnostics as well as for the screening of large populations in allele-specific and CNV distribution studies. As primer extension followed by sequencing is a classic approach for SNP detection, our method also allows for concurrent SNP genotyping. The assay is based on the recognition of a specific fluorescent signal, circumventing sequence/probe cross-hybridisation that could contaminate the final analytical results, as in MLPA-based and real-time PCR-based methods. In addition, the availability of more than one method for CNV quantification in diagnostic procedures could be useful to avoid erroneous diagnoses due to somatic mosaicism and/or polymorphisms at the primer site.

Our method can be directly used as a diagnostic genotyping test for the quantification of the SMN1 and SMN2 genes in SMA families and in the general population, allowing for the identification of carriers at reproductive risk and providing insight into the frequency of gene conversion events. In general, the simultaneous assessment of SNPs and CNVs increases the resolution of the assay to the nucleotide level, allowing for the detection of very small aberrations. This detection ability could contribute to basic research as well as to routine diagnostics. Therefore, we believe that this method can be applied for the rapid detection of trisomy syndromes and micro deletions as well as of micro duplication syndromes, including Trisomy 13, Down syndrome, Di George syndrome and autism spectrum disorders. The assay described here could be further applied to the validation of CNVs that are discovered by genome-wide screening methods. Furthermore, this methodology could be directly applied in gene expression studies to quantify the steady-state mRNA levels of a target gene(s) or of different alleles of a target gene(s) by expressing these mRNA levels relative to those of an internal control RNA (Radovic, unpublished observations).

The evaluation of both the variation in gene expression and the variation in allelic expression within and among individuals/populations is of fundamental value in biomedical research, where it may lead to the discovery of the causative genes of common hereditary diseases and their mechanism of action [36]. Coupling genotyping to gene expression studies could reveal the existence of direct and/or indirect mechanisms that require further dissection

to appropriately elucidate a particular phenotype. In conclusion, we report a rapid and simple quantitative assay that allows for the identification and allele-specific characterisation of CNVs. The assay has the potential for high-throughput, as 96 samples can be analysed in only a few hours, and it is a cost-effective tool. In addition, the assay we describe here could be easily employed in gene expression studies, allowing for the quantitative and allele-specific expression screening of a large number of genes, providing insight into questions of crucial importance to our understanding of the genomics of gene regulation.

Financial Support and Acknowledgments

The work was supported by European funds for regional development, by the Italian Ministry of Economic Development and by Friuli Venezia Giulia autonomous Region (POR FESR 2007-2013). We thank the patients and relatives for their collaboration on this project and for their voluntary enrolment.

References

1. Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* 12: 363-376.
2. Cook DE, Lee TG, Guo X, Melito S, Wang K, et al. (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* 338: 1206-1209.
3. McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, et al. (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol* 159: 1295-1308.
4. Craciun AR, Meyer CL, Chen J, Roosens N, De Groot R, et al. (2012) Variation in HMA4 gene copy number and expression among *Noccaea caerulescens* populations presenting different levels of Cd tolerance and accumulation. *J Exp Bot* 63: 4179- 4189.
5. Li Y, Xiao J, Wu J, Duan J, Liu Y, et al. (2012) A tandem segmental duplication (TSD) in green revolution gene Rht-D1b region underlies plant height variation. *New Phytol* 196: 282-291.
6. Diaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012) Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS One* 7: e33234.
7. Clop A, Vidal O, Amills M (2012) Copy number variation in the genomes of domestic animals. *Anim Genet* 43: 503-517.
8. Malhotra D, Sebat J (2012) CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* 148: 1223-1241.
9. Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, et al. (2009) High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res* 19: 1682-1690.
10. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, et al. (2009) Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 84: 148-161.
11. Gonzaga-Jauregui C, Lupski JR, Gibbs RA (2012) Human genome sequencing in health and disease. *Annu Rev Med* 63: 35-61.

12. Beckmann JS, Sharp AJ, Antonarakis SE (2008) CNVs and genetic medicine (excitement and consequences of a rediscovery). *Cytogenet Genome Res* 123: 7-16.
13. Lee C, Scherer SW (2010) The clinical context of copy number variation in the human genome. *Expert Rev Mol Med* 12: e8.
14. Ruderfer DM, Chambert K, Moran J, Talkowski M, Chen ES, et al. (2013) Mosaic copy number variation in schizophrenia. *Eur J Hum Genet* 21: 1007-1011.
15. D'Angelo CS, Koiffmann CP (2012) Copy number variants in obesity-related syndromes: review and perspectives on novel molecular approaches. *J Obes*: 1-15.
16. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, et al. (2012) Whole-genome sequencing in autism identifies hot spots for *de novo* germline mutation. *Cell* 151: 1431-1442.
17. McCarroll SA (2008) Extending genome-wide association studies to copy-number variation. *Hum Mol Genet* 17: R135-142.
18. Merikangas AK, Corvin AP, Gallagher L (2009) Copy-number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet* 25: 536-544.
19. Vitali T, Sossi V, Tiziano F, Zappata S, Giuli A, et al. (1999) Detection of the survival motor neuron (SMN) genes by FISH: further evidence for a role for SMN2 in the modulation of disease severity in SMA patients. *Hum Mol Genet* 8: 2525-2532.
20. Knuutila S, Bjorkqvist AM, Autio K, Tarkkanen M, Wolf M, et al. (1998) DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am J Pathol* 152: 1107-1123.
21. Albertson DG, Ylstra B, Segraves R, Collins C, Dairkee SH, et al. (2000) Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat Genet* 25: 144-146.
22. Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. *Genome Res* 6: 986-994.
23. Szilagyi A, Blasko B, Ronai Z, Fust G, Sasvari-Szekely M, et al. (2006) Rapid quantification of human complement component C4A and C4B genes by capillary gel electrophoresis. *Electrophoresis* 27: 1437-1443.
24. Schouten JP, Mc Elgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, et al. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 30: e57.
25. Pearn J (1980) Classification of spinal muscular atrophies. *Lancet* 1: 919-922.
26. Lefebvre S, Burglen L, Reboullet S, Clermont O, Burlet P, et al. (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 80: 155-165.
27. Hahnen E, Forkert R, Marke C, Rudnik-Schöneborn S, Schöning J, et al. (1995) Molecular analysis of candidate genes on chromosome 5q13 in autosomal recessive spinal muscular atrophy: evidence of homozygous deletions of the SMN gene in unaffected individuals. *Hum Mol Genet* 4: 1927-1933.
28. Lorson CL, Hahnen E, Androphy EJ, Wirth B (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc Natl Acad Sci USA* 96: 6307- 6311.
29. Wirth B, Brichta L, Hahnen E (2006) Spinal muscular atrophy and therapeutic prospects. *Prog Mol Sub Cell Bio* 144: 109-132.
30. Passon N, Pozzo F, Molinis C, Bregant E, Gellera C, et al. (2009) A simple multiplex real-time PCR methodology for the SMN1 gene copy number quantification. *Genet Test Mol Biomarkers* 13: 37- 42.
31. Passon N, Dubsky de Wittenau G, Jurman I, Radovic S, Bregant E, et al. (2010) Quick MLPA test for quantification of SMN1 and SMN2 copy numbers. *Mol Cell Probes* 24: 310-314.
32. Burglen L, Lefebvre S, Clermont O, Burlet P, Viollet L, et al. (1996) Structure and organization of the human survival motor neurone (SMN) gene. *Genomics* 32: 479- 482.
33. Lefebvre S, Burlet P, Liu Q, Bertrand S, Clermont O, et al. (1997) Correlation between severity and SMN protein level in spinal muscular atrophy. *Nat Genet* 16: 265-269.
34. DiMatteo D, Callahan S, Kmiec EB (2008) Genetic conversion of an SMN2 gene to SMN1: a novel approach to the treatment of spinal muscular atrophy. *Exp Cell Res* 314: 878-886.
35. Gerard B, Ginet N, Matthijs G, Evrard P, Baumann C, et al. (2000) Genotype determination at the survival motor neuron locus in a normal population and SMA carriers using competitive PCR and primer extension. *Human Mutation* 16: 253-263.
36. Yan H, Zhou W (2004) Allelic variations in gene expression. *Curr Opin Oncol* 16: 39-43.