Heteroduplex Formation in SMN Gene Dosage Analysis

Shuji Ogino, Debra G. B. Leonard,
Hanna Rennert, Sizhen Gao, and
Robert B. Wilson

From the Department of Pathology and Laboratory Medicine,
University of Pennsylvania Medical Center,
Philadelphia, Pennsylvania

Most spinal muscular atrophy patients lack both copies of SMN1 exon 7 and most carriers have only one copy of SMN1 exon 7. We investigated the effect of SMN1/SMN2 heteroduplex formation on SMN gene dosage analysis, which is an assay to determine copy number of SMN1 exon 7 that utilizes multiplex quantitative polymerase chain reaction (PCR) with *Dra*I digestion to differentiate SMN1 from SMN2. Heteroduplex formation in PCR is a well-described phenomenon. In addition to demonstrating the presence of heteroduplexes by sequence analysis of purified SMN1 bands, we compared the SMN1 signals in various genotype groups (total $n = 260$) to those in a group lacking SMN2 ($n = 13$), and we estimated the relative amounts of SMN1/SMN2 heteroduplexes. The SMN1 signal increased as SMN2 copy number increased despite a constant SMN1 copy number, although not all pairwise comparisons showed a statistically significant difference in the SMN1 signal. In conclusion, SMN1/SMN2 heteroduplexes form in SMN gene dosage analysis, falsely increasing the SMN1 signal. External controls for SMN gene dosage analysis should be chosen carefully with regard to SMN2 copy number. The effect of heteroduplex formation should be considered when performing quantitative multiplex PCR. (*J Mol Diag* 2001, 3:150–157)

Spinal muscular atrophy (SMA) is an autosomal recessive disorder associated with loss of motor neurons in the anterior horn of the spinal cord and caused by mutations in the Survival Motor Neuron 1 gene (*SMN1* or *SMNt*).¹ *SMN1* and its centromeric homologue, *SMN2* (or *SMNc*), lie within the telomeric and centromeric halves, respectively, of a large inverted repeat on chromosome 5q13.¹ *SMN1* and *SMN2* exons differ in only two bases, one in exon 7 and one in exon 8.² A C to T change in a single base of exon 7 of *SMN2* affects the activity of an exonic splice enhancer and changes the splicing pattern of *SMN2* transcripts,³ resulting in a lower level of full-length SMN protein from *SMN2* than from *SMN1*.⁴

Approximately 94% of affected patients lack both copies of *SMN1* exon 7 (reviewed by Wirth⁵). To allow identification of SMA carriers, quantitative polymerase chain reaction (PCR) methods to determine the copy number of *SMN1* exon 7 have been developed.^{6,7} These methods, herein referred to as *SMN* gene dosage analyses, use PCR amplification of *SMN1* and *SMN2* exon 7 segments, followed by *Dra*I restriction enzyme digestion and quantitation of PCR products.^{6,7} *Dra*I digests PCR products derived from *SMN2* but not those derived from *SMN1*. A single copy of *SMN1* by gene dosage analysis confirms carrier status (heterozygous *SMN1* deletion). In addition, up to 6% of affected individuals lack only one copy of *SMN1*, with a presumed non-deletion mutation on the other allele. In an individual with symptoms of SMA, heterozygous *SMN1* deletion supports the diagnosis of SMA. Therefore, *SMN* gene dosage analysis is of clinical importance.

Heteroduplex formation during PCR has been described and shown to affect many genetic tests.^{8–19} We hypothesized that *SMN1*/*SMN2* heteroduplexes form in our *SMN* gene dosage analysis, falsely increasing the *SMN1* signal, and that with a constant *SMN1* copy number more *SMN1*/*SMN2* heteroduplexes form as the *SMN2* copy number increases. The rationale for this hypothesis is as follows. In the annealing step of the last PCR cycle, most single-strand *SMN1* PCR products anneal with *SMN* PCR primers, in which case *SMN1*/*SMN1* homoduplexes form after successful extension. Some single-strand *SMN1* PCR products may anneal with complementary single-strand *SMN2* PCR products as well as complementary single-strand *SMN2* PCR products, forming *SMN1*/*SMN2* heteroduplexes. As PCR proceeds, the relative amount of heteroduplex formation may increase, because the concentrations of *SMN1* and *SMN2* PCR products increase and those of *SMN* primers decrease. *SMN1*/*SMN2* heteroduplexes would be resistant to restriction enzyme digestion because of the single nucleotide mismatch present at the *Dra*I restriction site in the *SMN2* PCR products. After *Dra*I digestion, *SMN1* and *SMN2* PCR products, which are distinguished by their fragment sizes, are denatured and quantitated.^{6,7} *SMN2*

Supported by the Department of Pathology and Laboratory Medicine, University of Pennsylvania Medical Center, Philadelphia, Pennsylvania.

Accepted for publication July 2, 2001.

Address reprint requests to Robert B. Wilson, M.D., Ph.D., Department of Pathology and Laboratory Medicine, University of Pennsylvania Medical Center, Stellar-Chance Laboratories Room 509A, 422 Curie Blvd., Philadelphia, PA 19104. E-mail: wilsonr@mail.med.upenn.edu.

single-strand PCR products denatured from the undigested *SMN1/SMN2* heteroduplexes would be measured as *SMN1* PCR products. In experiments presented here, we tested the heteroduplex hypothesis. We attempted to detect *SMN1/SMN2* heteroduplexes directly by sequence analysis, and to quantitate *SMN1/SMN2* heteroduplex formation in our *SMN* gene dosage analysis.

Materials and Methods

Sample Collection and DNA Extraction

Peripheral blood specimens were received by the Molecular Pathology Laboratory, University of Pennsylvania, for *SMN1* copy number determination. A sequential series of samples (total $n = 273$) were retrospectively selected and anonymized for this study. Genomic DNA was extracted from peripheral blood leukocytes using PUREGENE reagents (Gentra Systems, Inc., Minneapolis, MN) according to the manufacturer's protocol.

SMN1 and *SMN2* Copy Number Assay (*SMN* Gene Dosage Analysis)

The method by Chen et al⁷ was modified for *SMN1* and *SMN2* copy number assay, using 23 cycles of PCR instead of 22 cycles as in the original method. The assay takes advantage of the single nucleotide difference in exon 7 to distinguish *SMN1* from *SMN2* by PCR product size after *DraI* digestion. All assays were run in duplicate. Copy number of *SMN1* per cell (or more precisely, per diploid genome) was determined by quantitation of PCR products and was calculated using a genomic standard (*CFTR* exon 4), internal standards for *SMN1* and *CFTR* (*SMNis* and *CFTRis*, respectively) and five two-copy-*SMN1* controls. These control samples were validated as two-copy *SMN1* controls as described.⁷ Briefly, during the validation step, we observed small sample-to-sample and run-to-run coefficient of validations (CVs) of *SMN1* gene dosage (as described in the Appendix) in arbitrarily-selected controls from individuals without a family history of SMA. We also observed that a vast majority of samples from individuals without a family history showed a tightly clustered *SMN1* gene dosage, approximating the dosage of the controls, which was twice the dosage of a vast majority of samples from known SMA carriers who presumably had only one copy of *SMN1*. The internal standards, *SMNis* and *CFTRis*, have similar PCR amplification characteristics relative to their genomic counterparts. The *SMN1* gene dosage is normalized to five controls with two copies of *SMN1*. The normalized *SMN1* gene dosage is expressed as $C(SMN1)$, which stands for calculated *SMN1* signal (symbols are defined in Table 1.). $C(SMN1)$ is calculated as described in the Appendix. *SMN2* copy number was designated by comparison of the *SMN2* signal to the *SMN1* signal, assuming that the *SMN2* signal is 70 to 80% of the *SMN1* signal for equivalent copy numbers (See Results, Presence of *SMN1/SMN2* Heteroduplexes).

Table 1. Definitions of Symbols

Symbol	Definition
<i>SMNis</i>	<i>SMN</i> internal standard
<i>CFTRis</i>	<i>CFTR</i> internal standard
X:Y	(<i>SMN1</i> copy number):(<i>SMN2</i> copy number)
$C(SMN1)$	Calculated <i>SMN1</i> signal
$MC(SMN1)$	Mean of $C(SMN1)$
$MC(SMN1)_{X:Y}$	Mean of $C(SMN1)$ in a X:Y genotype
% <i>SMN2</i>	Fractional contribution of <i>SMN2</i> to total $C(SMN1)$

Sequence Analysis of Undigestable 188-bp *SMN* PCR Products

To test whether heteroduplexes are present after PCR, we amplified genomic DNA from two samples with one copy of *SMN1* and either three or four copies of *SMN2* as well as two samples with two copies of *SMN1* and lacking *SMN2* as negative controls. The PCR method was modified from the *SMN* gene dosage analysis PCR as follows: *SMNis*, *CFTRis*, and *CFTR* primers were excluded from the PCR master mix. The *SMN* primers were not fluorescently labeled. The PCR cycles were increased to 35 and 50 cycles, instead of 23 cycles, to increase the sensitivity of detection. The PCR cycle number was increased solely for this sequence analysis, but not for *SMN* copy number assay or for quantitation of heteroduplexes. After *DraI* digestion, PCR samples were electrophoresed on an ethidium-bromide-impregnated agarose gel [2% Ultra-Pure agarose (Life Technologies, Rockville, MD) and 1% NuSieve GTG agarose (FMC BioProducts, Rockland, ME)]. The undigestable 188 bp *SMN* PCR band was visualized under long-wave ultraviolet light and cut from the gel. The DNA was extracted from the agarose gel slice using the QIAquick Gel Extraction Kit (QIAGEN Inc., Valencia, CA). The DNA fragments were sequenced with the forward primer *SMN7F* (R111)¹ using the BigDye Terminator Reaction Kit (Applied Biosystems, Inc., Foster City, CA) according to the manufacturer's instructions. The products obtained after cycle sequencing were purified through Centri-Sep columns (Princeton Separations, Adelphia, NJ) and analyzed on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Inc.), according to the manufacturer's instructions.

Quantitation of *SMN1/SMN2* Heteroduplex Formation

SMN1 and *SMN2* copy numbers were determined in samples from 273 individuals. Then $C(SMN1)$ values were compared among various *SMN* genotype groups, to test the hypothesis that the *SMN1* signal, $C(SMN1)$, increases as *SMN2* copy number increases due to *SMN1/SMN2* heteroduplex formation, despite a constant *SMN1* copy number. The genotypes tested are shown in Table 2. A genotype was designated by a code consisting of two integers separated by a colon: X:Y, where X represents *SMN1* copy number and Y represents *SMN2* copy number. Each sample was tested in duplicate and the

Table 2. *SMN1* and *SMN2* Copy Number Genotype and Heteroduplex Quantitation

(<i>SMN1</i> copy number): (<i>SMN2</i> copy number)	2:0	2:1	2:2	2:3	1:1	1:2	1:3	1:4	3:1	3:2
Number of samples	13	53	81	4	27	57	18	2	11	6
Mean C(<i>SMN1</i>) [‡]	1.902	1.973	2.033	2.211	1.018	1.034	1.063	1.082	2.938	2.999
Variance of C(<i>SMN1</i>) [‡]	0.00565	0.00562	0.00872	0.00983	0.00442	0.00360	0.00462	0.00414	0.00929	0.00549
Standard deviation of C(<i>SMN1</i>) [‡]	0.075	0.075	0.093	0.099	0.066	0.060	0.068	0.064	0.096	0.074
% <i>SMN2</i> in C(<i>SMN1</i>) [‡] with 95% CI	0	3.6 ± 2.3	6.4 ± 2.3	14.0 ± 4.3	6.6 ± 3.1	8.0 ± 2.5	10.5 ± 3.3	12.1 ± 7.6	2.9 ± 2.9	4.9 ± 2.8

[‡]Calculated *SMN1* signal.

average of the two C(*SMN1*) values was used. The coefficient of variation (CV) of the two C(*SMN1*) values for each sample in the duplicate testing of the 273 samples ranged from 0% to 12.5% with a mean of 2.48%, a median of 1.85%, and an 80th percentile point of 4.00%.

The presence of *SMN2*/*SMN1*s heteroduplexes was negligible because at least one control sample lacking *SMN1* in each run consistently lacked the 188-bp peak that would have been present due to undigestable *SMN2* if a significant amount of *SMN2*/*SMN1*s heteroduplexes formed. Therefore, the presence of *SMN1*/*SMN1*s heteroduplexes can also be considered negligible and virtually all undigested *SMN2* PCR products can be considered as derived from *SMN1*/*SMN2* heteroduplexes.

The extent of heteroduplex formation was calculated as the percentage of C(*SMN1*) that was contributed by *SMN2*, by comparing the mean C(*SMN1*) of a particular genotype to that of the 2:0 group. The 2:0 group lacks *SMN1*/*SMN2* heteroduplexes due to an absence of *SMN2* and was used as a standard for quantitation of heteroduplex formation in other genotypes. The mean C(*SMN1*) in all samples within each genotype group is designated as MC(*SMN1*) followed without space by a subscript of *SMN1* and *SMN2* copy number genotype, designated as X:Y. For example, MC(*SMN1*)_{2:2} represents the mean C(*SMN1*) of the 2:2 genotype group. For statistical analyses of the differences in MC(*SMN1*)_{X:Y} between the genotype groups, an F test was first used to determine whether a significant difference in variances was present. If there was no significant difference in variances, then the Student's *t*-test (two-tailed) was used for comparison of MC(*SMN1*)_{X:Y} between the two genotypes. If there was a significant difference in variance, then the Welch *t*-test (two-tailed) was used. MC(*SMN1*)_{1:1}, MC(*SMN1*)_{1:2} and MC(*SMN1*)_{1:3} were compared to MC(*SMN1*)_{2:0} × 1/2 (which serves as a standard for the 1:0 genotype). Similarly, MC(*SMN1*)_{3:1} and MC(*SMN1*)_{3:2} were compared to MC(*SMN1*)_{2:0} × 3/2 (which serves as a standard for the 3:0 genotype). The relative contribution of *SMN2* derived from the *SMN1*/*SMN2* heteroduplexes among the total *SMN1* signal is represented as %*SMN2* and is calculated as described in the Appendix.

Results

Presence of *SMN1*/*SMN2* Heteroduplexes

In our experience using the modified method of Chen et al,⁷ there seemed to be less *SMN2* PCR product than

SMN1 PCR product in most samples. A substantial number of samples with one copy of *SMN1* showed an *SMN2* signal intensity approximately 70 to 80% of the *SMN1* signal. We presumed that such samples had one copy of *SMN2* because the *SMN2* signal was approximately equal to the one-copy *SMN1* signal, and because samples with zero copies of *SMN2* showed a complete absence of *SMN2* signal. Other samples with one copy of *SMN1* frequently showed an *SMN2* signal approximately 1.5 times stronger than that of *SMN1*, implying the presence of two copies of *SMN2*. Although a majority of samples with two copies of *SMN1* showed an *SMN2* signal intensity approximately 70 to 80% of that of *SMN1* (implying the presence of two copies of *SMN2*), more than one third of samples with two copies of *SMN1* showed an *SMN2* signal strength close to one third to one fourth of the *SMN1* signal (implying the presence of one copy of *SMN2*). One hypothesis to explain these discrepancies between copy number and signal intensity was that the PCR amplification efficiencies of *SMN1* and *SMN2* differ. However, this seemed unlikely because *SMN1* and *SMN2* sequences in this PCR differ by only one nucleotide, and this nucleotide difference is not in the primer sites. Moreover, it could not explain why samples lacking *SMN2* generally exhibited weaker *SMN1* signals relative to samples with the same *SMN1* copy number but *SMN2* present. An alternative hypothesis was that incomplete *DraI* digestion falsely increased *SMN1* signal intensity and decreased *SMN2* signal intensity. This also seemed unlikely because of assay controls that lacked *SMN1* in which no undigested *SMN2* signal was detected (see Results, Validation of *SMN* Gene Dosage Analysis). We hypothesized that *SMN1*/*SMN2* heteroduplexes, which formed during PCR and could not be digested by *DraI*, falsely increased the *SMN1* signal intensity (Figure 1). To test this hypothesis, we performed sequence analyses to look for the presence of the *SMN2* sequence in *DraI* undigestable PCR products.

The *SMN2* sequence derived from *SMN1*/*SMN2* heteroduplexes was identified by sequencing the *DraI* undigestable 188 bp band from agarose gels (Figure 2). In two samples lacking *SMN2* (genotype 2:0), the *SMN2*-specific nucleotide T in position 27141 was absent, as expected (Figure 2A). By contrast, in two samples with one *SMN1* copy and three or four *SMN2* copies (genotype 1:3 and 1:4), respectively, a significant amount of the *SMN2*-specific nucleotide T in position 27141 was present under the expected C peak of *SMN1* (Figure 2, B

Figure 1. SMN1/SMN2 heteroduplexes contribute to the *SMN1* signal in *SMN* gene dosage analysis. *SMN1* (or *SMN2*) homoduplexes form either when complementary *SMN1* (or *SMN2*) single strand PCR products anneal with each other or when successful extension occurs after *SMN1* (or *SMN2*) single strands anneal with their corresponding PCR primers. *SMN1*/SMN2 heteroduplexes form when *SMN1* single strands anneal with their near-complementary *SMN2* single strands. Since only *SMN2* homoduplexes can be digested by *DraI* to form 165-bp DNA fragments, *SMN2* strands in *SMN1*/SMN2 heteroduplexes cannot be digested and remain as intact 188-bp PCR products, which are measured as *SMN1*.

and C). The amount of the *SMN2*-specific nucleotide T increased as the PCR cycle number was increased from 35 to 50, as predicted by the heteroduplex hypothesis.

Validation of *SMN* Gene Dosage Analysis

In *SMN* gene dosage analysis, the calculated *SMN1* signal, $C(SMN1)$, should always be close to an integer, which is the *SMN1* copy number of a sample, unless there is mosaicism. When individual samples were classified into *SMN1* and *SMN2* copy number genotypes, $MC(SMN1)_{x,y} \pm 2$ SD was within the ranges 0.89 to 1.21 for one copy of *SMN1* (the 1:1, 1:2, 1:3, and 1:4 genotypes), 1.75 to 2.41 for two copies of *SMN1* (the 2:0, 2:1, 2:2, and 2:3 genotypes) and 2.74 to 3.15 for three copies of *SMN1* (the 3:1 and 3:2 genotypes) and there was no overlap between these ranges (Table 2 and Figure 3). Incomplete restriction-enzyme digestion could result in a falsely high signal intensity for *SMN1*. Therefore, at least one sample lacking *SMN1* but possessing *SMN2* served in every run as a control to verify complete *DraI* digestion.

Quantitation of *SMN1*/SMN2 Heteroduplex Formation

One sample with a $C(SMN1)$ of 2.218 was initially included in the 2:0 genotype group but was shown to be an outlier by the Grubbs-Smirnov test ($P < 0.01$). Excluding this sample, the 2:0 genotype group ($n = 13$) demonstrated that a distribution of $C(SMN1)$ was close to a normal distribution with the mean $C(SMN1) = 1.902$ (95% confidence interval of 1.857 to 1.947), a SD of 0.075 and a variance of 0.00565. When the outlier sample was included in the 2:0 group, the mean $C(SMN1) = 1.924$ (95% confidence interval of 1.860 to 1.989) with an SD of 0.111, reflecting a larger variance of 0.0124. We used the mean $C(SMN1)$ of the other 13 samples in the 2:0 genotype as a standard. There are several possible reasons why the $C(SMN1)$ of 2.218 for the one outlier sample was

Figure 2. Sequence analysis demonstrated the presence of *SMN2* sequence derived from *SMN1*/SMN2 heteroduplexes in the *DraI* undigestible 188 bp bands on agarose gels. Genotypes are described as (*SMN1* copy number): (*SMN2* copy number) as in the text. **A:** The absence of a nucleotide T peak at position 2714C is shown (arrow) in one of the two 2:0 samples after 35 cycles of PCR. The absence of the nucleotide T peak was observed in both 2:0 samples after both 35 cycles and 50 cycles of PCR. **B:** The presence of the nucleotide T peak at position 2714C is indicated (arrow) under the C peak in the 1:4 genotype sample after 35 cycles of PCR. A similar result was also obtained in the 1:3 genotype sample after 35 cycles of PCR. **C:** The increase in the nucleotide T peak relative to the C peak is shown (arrow) as PCR cycle number was increased to 50. A similar result was also obtained in the 1:3 genotype sample after 50 cycles of PCR.

Figure 3. The calculated *SMN1* signal, $C(SMN1)$, in various *SMN* genotypes. The x axis represents *SMN* genotypes described as (*SMN1* copy number):(*SMN2* copy number) as in the text. The y axis represents the calculated *SMN1* signal, $C(SMN1)$. The mean $C(SMN1)$ is represented by a column. The solid vertical line across the top of each column represents ± 1 SD. The mean $C(SMN1)$ increased as *SMN2* copy number increased despite a constant *SMN1* copy number.

greater than 4 SD above the mean $C(SMN1)$ of the 2:0 genotype excluding this sample. There might be a polymorphism affecting one of the primer sites on the *CFTR* exon 4 genomic reference sequence, causing less efficient PCR amplification. Other possibilities include mosaicism and partial DNA degradation in the sample, causing less efficient PCR amplification of the longer *CFTR* genomic sequence (239 bp) compared to the *SMN1* sequence (188 bp).

We obtained data for the quantitation of %*SMN2* from a total of 260 samples with various genotypes with *SMN2* copies and the 13 samples with the 2:0 genotype (a standard) (Table 2 and Figure 3). The $C(SMN1)$ for each genotype exhibited a distribution that was close to a normal distribution, allowing F-tests, Student's t -tests, and Welch t -tests to be applied to the data. Due to *SMN1/SMN2* heteroduplex formation, the *SMN1* signal of the genotypes with two copies of *SMN1* increased as *SMN2* copy number increased despite the constant *SMN1* copy number ($P < 0.005$). The genotypes 1:Y with at least one copy of *SMN2* also showed significantly higher $C(SMN1)$ than $[C(SMN1)_{2:0} \times 1/2]$ ($P < 0.0005$) and the genotypes 3:1 and 3:2 showed higher $C(SMN1)$ than $[C(SMN1)_{2:0} \times 3/2]$ ($P = 0.01$ and $P = 0.062$, respectively). However, all pairwise comparisons of $C(SMN1)$ within the genotypes with one copy of *SMN1*, and comparison of $C(SMN1)$ between the 3:1 and 3:2 genotypes, showed only one pair (the 1:1 vs. 1:3 genotypes) with a significant difference ($P < 0.032$), although there was an overall tendency that $C(SMN1)$ increased as *SMN2* copy number increased despite a constant *SMN1* copy number (Table 2 and Figure 3).

Discussion

Heteroduplex formation in PCR has been described.⁸⁻¹⁹ Theoretically, heteroduplexes form whenever different templates with similar sequences are annealed. In non-quantitative PCR-Restriction Fragment Length Polymorphism (RFLP) assays, which are used in many genetic tests, heteroduplexes usually do not affect the interpre-

tation of results, because one can determine the presence or absence of a mutant allele (associated with the presence or absence of a restriction site) regardless of the presence of a low level of heteroduplex formation. Heteroduplexes rarely cause diagnostic problems in non-quantitative PCR analyses, exemplified by PCR analysis for the GAA trinucleotide expansion in the *FRDA* gene Friedreich ataxia¹⁷ and by PCR-RFLP analysis for the detection of two different mutations in the steroid 21-hydroxylase gene for congenital adrenal hyperplasia.¹⁴

On the other hand, heteroduplexes may affect the interpretation of results for quantitative PCR assays, such as in methods for *SMN1* quantitation by McAndrew et al.,⁶ Chen et al.,⁷ and Wirth et al.²⁰ Methods that circumvent problems of heteroduplex formation in quantitative PCR assays have been reported.²¹⁻²⁴ Most of these methods are complex and require individual optimization. A relatively simple method, hot-stop PCR, was recently reported by Uejima et al.,²⁵ but required utilization of radioactivity and an extra manipulation before the final cycle of PCR. Becker-Andre and Hahlbrock,¹¹ who first mentioned the possibility of heteroduplex formation affecting the results of DNA quantitation by PCR, considered that essentially only homodimeric DNA species formed in the exponential phase of PCR. Uejima et al.²⁵ found that the quantitated allele ratio in regular 25-cycle PCR was equal to the theoretical ratio, implying no significant heteroduplex formation after 25 cycles of PCR, whereas there was a significant distortion of the quantitated allele ratio in regular 35-cycle PCR presumably due to heteroduplex formation. By contrast, our results indicate that there is a significant amount of heteroduplex formation in our *SMN* gene dosage analysis even using 23 cycles of PCR. Reducing the PCR cycle number will likely decrease heteroduplex formation. The method of McAndrew et al.⁶ used 16 cycles of PCR and clearly showed a smaller effect of heteroduplexes, as well as good accuracy and precision. Our method has the disadvantage of having greater heteroduplex formation, but not to the point of affecting test accuracy or precision, while having the advantage of using a non-radioisotopic detection system.

This study is the first to evaluate the effect of heteroduplex formation on the quantitation of *SMN1*. We demonstrated that the calculated *SMN1* signal for individuals with two *SMN1* copies increased significantly as *SMN2* copy number increased. Although not statistically significant, there was a consistent tendency for the calculated *SMN1* signal to increase for individuals with one or three *SMN1* copies, as *SMN2* copy number increased. We speculate that our assay was insufficiently accurate to detect statistically significant differences with these relatively small sample sizes. However, for all practical purposes, the accuracy and precision of our assay to determine copy number of *SMN1* was quite high. The mean $C(SMN1) \pm 2$ SD in each *SMN1* and *SMN2* genotype group was still within the ranges 0.89 to 1.21 for one copy of *SMN1*, 1.75 to 2.41 for two copies of *SMN1*, and 2.74 to 3.15 for three copies of *SMN1*, and there was no overlap between these ranges. It should be noted that we obtained the results of this study using two-copy *SMN1*

controls consistently composed of three samples with two copies of *SMN2* (genotype 2:2) and two samples with one copy of *SMN2* (genotype 2:1). Based on our data, two-copy *SMN1* controls for *SMN* gene dosage analysis should be chosen carefully with regard to the *SMN2* copy number. For example, if one uses two-copy *SMN1* samples lacking *SMN2*, as two-copy *SMN1* controls, the calculated *SMN1* signal will be overestimated for all samples that have at least one copy of *SMN2*. On the other hand, if one uses two-copy *SMN1* samples with three or more copies of *SMN2* as controls, the calculated *SMN1* signal will be underestimated for all samples that have one or two copies of *SMN2*, or lack *SMN2* entirely. Individuals who lack *SMN2* may have a calculated *SMN1* signal between the one-copy range (consistent with carrier status) and the two-copy range (more typical of non-carrier status), causing a significant diagnostic problem. One possible solution to this problem is to modify the calculated *SMN1* signal, $C(SMN1)$, for each *SMN1* and *SMN2* genotype. For example, if a sample lacking *SMN2* shows a $C(SMN1)$ of 1.62, this $C(SMN1)$ can be modified to 1.71, which is closer to two-copy *SMN1* samples (by dividing 1.62 by the factor $MC(SMN1)_{2:0} \times 1/2 = 0.95$ from Table 2). For two-copy *SMN1* controls for clinical assays, we recommend using three samples from the 2:2 genotype and two samples from the 2:1 genotype. This composition of genotypes roughly reflects the frequencies of *SMN* genotypes in the general population (unpublished data). One should be aware of the effect of *SMN1/SMN2* heteroduplexes when interpreting results of *SMN* gene dosage analysis.

Intuitively, one might expect a higher %*SMN2* in samples with more total *SMN* gene copies because more overall PCR products and fewer primers favor heteroduplex formation rather than primer binding. However, we estimated nearly identical %*SMN2* (6.6% and 6.4%) in the 1:1 and 2:2 groups. We speculate that the 23-cycle PCR in our assay is still within the exponential phase and that the primers are still in relative excess. One might also predict a higher %*SMN2* where there is a higher fraction of *SMN2*, i.e., $(SMN2 \text{ copy number})/(\text{total } SMN \text{ copy number})$, in the specimen. Our results were generally concordant with this prediction. The observed values of %*SMN2* in each genotype (followed by the fraction of *SMN2* in the specimen and number of samples) from the lowest to highest are: 2.9% in the 3:1 (0.25, $n = 11$), 3.6% in the 2:1 (0.33, $n = 53$), 4.9% in the 3:2 (0.40, $n = 6$), 6.4% in the 2:2 (0.50, $n = 81$), 6.6% in the 1:1 (0.50, $n = 27$), 8.0% in the 1:2 (0.67, $n = 57$), 10.5% in the 1:3 (0.75, $n = 18$), 12.1% in the 1:4 (0.80, $n = 2$), and 14.0% in the 2:3 (0.60, $n = 4$). Therefore, in general, %*SMN2* tends to increase as the fraction of *SMN2* in the specimen increases. The only pairwise comparisons that contradict the prediction involve the 2:3 genotype. This may be due to the small sample size ($n = 4$). More samples are necessary to evaluate this issue.

In conclusion, we have demonstrated that *SMN1/SMN2* heteroduplexes form during the final annealing step of PCR in our *SMN* gene dosage analysis assay, resulting in an increased *SMN1* signal as *SMN2* copy number increases. Therefore, control samples with two

copies of *SMN1* for *SMN1* quantitation should be selected judiciously with regard to *SMN2* copy number. The effect of heteroduplex formation should be considered whenever performing multiplex PCR amplification of templates with similar sequences.

Acknowledgments

The authors thank Patricia E. McAndrew and Thomas W. Prior for generously providing us with the plasmids containing *SMNis* and *CFTRis* inserts, Jean B. Patel for assisting us in sequence analysis, Cynthia Turino and Treasa Smith for performing *SMN* gene dosage analysis and their technical support, and Warren J. Ewens for assisting us in calculating 95% CI for %*SMN2*.

Appendix

1. Calculation of *SMN1* Copy Number

To quantitate *SMN1* copy number, we defined signal intensity as the relevant peak area on an ABI PRISM 310 capillary electropherogram. The signal intensities of the relevant peaks are defined as follows: A_1 , *SMNis* PCR product; A_2 , *SMN2* PCR product; A_3 , *SMN1* PCR product; A_4 , *CFTRis* PCR product; A_5 , *CFTR* PCR product.

Because A_3 and A_1 are proportional to initial copy numbers of *SMN1* and *SMNis*, respectively, in a sample tube,

$$\begin{aligned} A_3 &= a(\text{initial copy number of } SMN1) \\ &= a(SMN1 \text{ copy number per cell}) \\ &\quad (\text{number of cells}) \end{aligned} \quad (1)$$

where a equals the PCR efficiency constant for *SMN*. In addition, because the initial copy number of *SMNis* equals the PCR master mix volume added into a sample tube, multiplied by the *SMNis* concentration in the PCR master mix,

$$A_1 = ak_1(\text{initial copy number of } SMNis) = ak_1cv \quad (2)$$

where k_1 equals the PCR efficiency difference factor for *SMNis* compared to *SMN1*, c equals the *SMNis* concentration in the PCR master mix, and v equals the PCR master mix volume added into a sample tube.

Similar formulas can be generated for *CFTR* and *CFTRis*, where b equals the PCR efficiency constant for *CFTR*, k_2 equals the PCR efficiency difference factor for *CFTRis* compared to *CFTR*, d equals the *CFTRis* concentration in the PCR master mix, and all cells are assumed to have 2 copies of *CFTR* exon 4.

$$\begin{aligned} A_5 &= b(\text{initial copy number of } CFTR) \\ &= b(CFTR \text{ copy number per cell})(\text{number of cells}) \\ &= 2b(\text{number of cells}) \end{aligned} \quad (3)$$

$$A_4 = bk_2(\text{initial copy number of } CFTRis) = bk_2dv \quad (4)$$

$$A_3/A_1 = (\text{SMN1 copy number per cell}) \cdot (\text{number of cells})/(k_1cv) \quad (5)$$

$$A_5/A_4 = 2(\text{number of cells})/(k_2dv) \quad (6)$$

$$(A_3/A_1)/(A_5/A_4) = [k_2d/(2k_1c)] (\text{SMN1 copy number per cell}) \quad (7)$$

$$k_3 = (k_2d)/(2k_1c) \quad (8)$$

$$(A_3/A_1)/(A_5/A_4) = k_3(\text{SMN1 copy number per cell}) \quad (9)$$

$[(A_3/A_1)/(A_5/A_4)]$ is described as *SMN1* gene dosage, which is normalized in the next step.

$[(A_3/A_1)/(A_5/A_4)]$ is described as *SMN1* gene dosage, which is normalized in the next step.

$$k_3 = [(A_3/A_1)/(A_5/A_4)]/(\text{SMN1 copy number per cell}) \quad (10)$$

We obtained a mean of k_3 , designated as k_3^* , from (10) in five samples with two copies of *SMN1*, composed of three samples with two copies of *SMN2* and two samples with one copy of *SMN2* (11)

$$(\text{SMN1 copy number per cell}) = [(A_3/A_1)/(A_5/A_4)]/k_3^* \quad (12)$$

The formula (12) is designated herein as the calculated *SMN1* signal and described as $C(\text{SMN1})$.

2. Calculation of %SMN2

To quantitate *SMN1/SMN2* heteroduplexes, let the mean $C(\text{SMN1})$ in the 2:0 genotype, $\text{MC}(\text{SMN1})_{2:0}$ equal to 2α , where α represents the contribution of single-stranded *SMN1* PCR products, all of which were derived from *SMN1/SMN1* homoduplexes in this genotype.

$$\text{MC}(\text{SMN1})_{2:0} = 2\alpha \quad (13)$$

We assume that the absolute value among the mean $C(\text{SMN1})$ contributed by single-stranded *SMN1* PCR products to the mean $C(\text{SMN1})$ is constant ($= 2\alpha$) regardless of the *SMN2* copy number when the *SMN1* copy number remains constant. When we define 2β as the contribution of total single-stranded PCR products from *SMN1/SMN2* heteroduplexes to $\text{MC}(\text{SMN1})_{2:Y}$ in a sample from the 2:Y genotype (where Y is a positive integer; 1, 2, 3, or 4 in this study), β represents a contribution of either single-stranded *SMN1* PCR products or single-stranded *SMN2* PCR products, both of which were derived from *SMN1/SMN2* heteroduplexes. Then, the contribution of *SMN1* strands from *SMN1/SMN1* homoduplexes is equal to $2\alpha - \beta$, because total contribution of single-stranded *SMN1* PCR products is constant ($= 2\alpha$) and some contribution of *SMN1* strands, equivalent to β , were derived from *SMN1/SMN2* heteroduplexes. Therefore, the total contribution of *SMN1* and

SMN2 single strands to $\text{MC}(\text{SMN1})_{2:Y}$ equals $(2\alpha - \beta) + 2\beta = 2\alpha + \beta$.

$$\text{MC}(\text{SMN1})_{2:Y} = 2\alpha + \beta \quad (14)$$

$$\text{MC}(\text{SMN1})_{2:Y} - \text{MC}(\text{SMN1})_{2:0} = \beta \quad (15)$$

From (14) and (15) the fraction of *SMN2* contribution to the total $\text{MC}(\text{SMN1})_{2:Y}$ (represented as %*SMN2*) equals:

$$\begin{aligned} \% \text{SMN2} &= \beta/(2\alpha + \beta) \\ &= [\text{MC}(\text{SMN1})_{2:Y} \\ &\quad - \text{MC}(\text{SMN1})_{2:0}]/\text{MC}(\text{SMN1})_{2:Y} \end{aligned} \quad (16)$$

For genotypes 1:Y and 3:Y, $[\text{MC}(\text{SMN1})_{2:0} \times 1/2]$ and $[\text{MC}(\text{SMN1})_{2:0} \times 3/2]$, respectively, were used as standards. Therefore, from (16) the fraction of *SMN2* strand contribution to the total $\text{MC}(\text{SMN1})_{X:Y}$ (%*SMN2*) equals:

$$\begin{aligned} \% \text{SMN2} &= \beta/(2\alpha + \beta) \\ &= [\text{MC}(\text{SMN1})_{1:Y} - \text{MC}(\text{SMN1})_{2:0} \times 1/2] \\ &\quad / \text{MC}(\text{SMN1})_{1:Y} \end{aligned} \quad (17)$$

$$\begin{aligned} \% \text{SMN2} &= \beta/(2\alpha + \beta) \\ &= [\text{MC}(\text{SMN1})_{3:Y} - \text{MC}(\text{SMN1})_{2:0} \times 3/2] \\ &\quad / \text{MC}(\text{SMN1})_{3:Y} \end{aligned} \quad (18)$$

References

1. Lefebvre S, Burglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M, Le Paslier D, Frezal J, Cohen D, Weissenbach J, Munnich A, Melki J: Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 1995, 80:155-165
2. Burglen L, Lefebvre S, Clermont O, Burlet P, Viollet L, Cruaud C, Munnich A, Melki J: Structure and organization of the human survival motor neurone (SMN) gene. *Genomics* 1996, 32:479-482
3. Pellizzoni L, Kataoka N, Charroux B, Dreyfuss G: A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* 1998, 95:615-624
4. Jong YJ, Chang JG, Lin SP, Yang TY, Wang JC, Chang CP, Lee CC, Li H, Hsieh-Li HM, Tsai CH: Analysis of the mRNA transcripts of the survival motor neuron (SMN) gene in the tissue of an SMA fetus and the peripheral blood mononuclear cells of normals, carriers and SMA patients. *J Neurol Sci* 2000, 173:147-153
5. Wirth B: An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). *Hum Mutat* 2000, 15:228-237
6. McAndrew PE, Parsons DW, Simard LR, Rochette C, Ray PN, Mendell JR, Prior TW, Burghes AH: Identification of proximal spinal muscular atrophy carriers and patients by analysis of SMNT and SMNC gene copy number. *Am J Hum Genet* 1997, 60:1411-1422
7. Chen KL, Wang YL, Rennert H, Joshi I, Mills JK, Leonard DG, Wilson RB: Duplications and de novo deletions of the SMNT gene demonstrated by fluorescence-based carrier testing for spinal muscular atrophy. *Am J Med Genet* 1999, 85:463-469
8. Jensen MA, Hubner RJ: Use of homoduplex ribosomal DNA spacer amplification products and heteroduplex cross-hybridization products in the identification of *Salmonella* serovars. *Appl Environ Microbiol* 1996, 62:2741-2746
9. Jensen MA, Straus N: Effect of PCR conditions on the formation of heteroduplex and single-stranded DNA products in the amplification of bacterial ribosomal DNA spacer regions. *PCR Methods Appl* 1993, 3:186-194

10. Hsieh CH, Griffith JD: Deletions of bases in one strand of duplex DNA, in contrast to single-base mismatches, produce highly kinked molecules: possible relevance to the folding of single-stranded nucleic acids. *Proc Natl Acad Sci USA* 1989, 86:4833–4837
11. Becker-Andre M, Hahlbrock K: Absolute mRNA quantification using the polymerase chain reaction (PCR): a novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Res* 1989, 17:9437–9446
12. Ruano G, Kidd KK: Modeling of heteroduplex formation during PCR from mixtures of DNA templates. *PCR Methods Appl* 1992, 2:112–116
13. Ruano G, Deinard AS, Tishkoff S, Kidd KK: Detection of DNA sequence variation via deliberate heteroduplex formation from genomic DNAs amplified en masse in "population tubes." *PCR Methods Appl* 1994, 3:225–231
14. Bradley JF, Baker D, Schwartz ID, Rothberg PG: The importance of heteroduplexes in interpreting the results of PCR-RED diagnostic assays: application to the analysis of mutations in the steroid 21-hydroxylase gene in a case of congenital adrenal hyperplasia. *Mol Diagn* 1998, 3:119–123
15. Henley WN, Schuebel KE, Nielsen DA: Limitations imposed by heteroduplex formation on quantitative RT-PCR. *Biochem Biophys Res Commun* 1996, 226:113–117
16. Boer PH, Ramamoorthy J: How to correct for errors in mRNA quantitation by competitive PCR due to heteroduplex formation of amplification products. *Cell Mol Biol* 1997, 43:841–850
17. Poirier J, Ohshima K, Pandolfo M: Heteroduplexes may confuse the interpretation of PCR-based molecular tests for the Friedreich ataxia GAA triplet repeat [letter]. *Hum Mutat* 1999, 13:328–330
18. Mashal RD, Koontz J, Sklar J: Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat Genet* 1995, 9:177–183
19. Pushnova EA, Geier M, Zhu YS: An easy and accurate agarose gel assay for quantitation of bacterial plasmid copy numbers [In Process Citation]. *Anal Biochem* 2000, 284:70–76
20. Wirth B, Herz M, Wetter A, Moskau S, Hahnen E, Rudnik-Schoeneborn S, Wienker T, Zerres K: Quantitative analysis of survival motor neuron copies: identification of subtle *SMN1* mutations in patients with spinal muscular atrophy, genotype-phenotype correlation, and implications for genetic counseling. *Am J Hum Genet* 1999, 64:1340–1356
21. Yun K, Soejima H, Merrie AE, McCall JL, Reeve AE: Analysis of *IGF2* gene imprinting in breast and colorectal cancer by allele specific-PCR. *J Pathol* 1999, 187:518–522
22. Takeda S, Ichii S, Nakamura Y: Detection of K-ras mutation in sputum by mutant-allele-specific amplification (MASA). *Hum Mutat* 1993, 2:112–117
23. Greenwood AD, Burke DT: Single nucleotide primer extension: quantitative range, variability, and multiplex analysis. *Genome Res* 1996, 6:336–348
24. Pushnova EA, Zhu YS: Quantitative restriction fragment length polymorphism: a procedure for quantitation of diphtheria toxin gene CRM197 allele. *Anal Biochem* 1998, 260:24–29
25. Uejima H, Lee MP, Cui H, Feinberg AP: Hot-stop PCR: a simple and general assay for linear quantitation of allele ratios. *Nat Genet* 2000, 25:375–376