Determining copy number variation in *SMN1* using a highly multiplexed next-generation sequencing panel

Adam N. Harris[1], Guy Del Mistro[1], Tom Fahland[1], Huimin Xie[1], John D. Bishop[1], James Veitch[2], and Vadim Mozhayskiy[1]
[1]Thermo Fisher Scientific, 5781 Van Allen Way, Carlsbad, CA, USA, 92009
[2]Thermo Fisher Scientific, 180 Oyster Point Blvd, South San Francisco, CA, USA, 94080

STUDY QUESTION

Can a next-generation sequencing (NGS) panel be used to accurately assess samples for copy number variation (CNV) in genes that have high homology, such as *SMN1* and *SMN2*?

SUMMARY ANSWER

We were able to characterize test samples for functional *SMN1* copy number with 97.8% accuracy, and *SMN2* copy number with 95.6% accuracy (n = 91).

WHAT IS ALREADY KNOWN

Loss of *SMN1* activity is a cause of spinal muscular atrophy (SMA). *SMN1* and *SMN2* are highly homologous genes that are difficult to distinguish on NGS platforms due to relatively short read lengths. Typically the copy number determination of these genes is assessed with other technologies such as multiplexed ligation probe assays (MLPA™) and microarray.

STUDY DESIGN

We tested 91 gDNA samples from Coriell cell lines with functional *SMN1/2* copy number status known from microarray analysis. The samples were tested as part of a characterization of the NGS panel in development. Results needed to meet predefined quality control criteria to be included in the assessment.

MATERIALS AND METHODS

Samples were amplified using a two-pool primer panel comprising approximately 14,000 PCR primer pairs targeting coding sequences and known intronic variants in 420 genes implicated in recessive autosomal and X-linked diseases. The resulting DNA barcoded libraries were sequenced in a multiplex of 16 per Ion 540™ Chip using the Ion S5™ System.

MAIN FINDINGS

We used an *in silico* baseline derived from amplicon representation of normal samples and a custom CNV-calling pipeline to determine both (i) the mean copy numbers from amplicons with 100% identity between homologous portions of the genes of interest, and (ii) the homolog-specific copy numbers for amplicons with distinguishing sequence variants between the primers or in the primer-binding sites that enriched amplification in one member of a gene pair. Using this information and orthogonal microarray results taken as "truth," we were able to obtain *SMN1* carrier status with 100% sensitivity (n = 4) and 98.9% specificity (n = 87). Follow-up studies will be needed by individual labs to determine reproducibility of our test with their own samples of interest.

LIMITATIONS AND REASONS FOR CAUTION

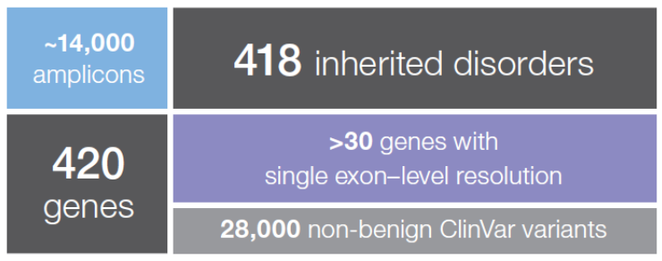
We tested our method on a limited set of gDNA samples isolated from cell lines. Factors that could compromise performance for other samples include amplification and sequencing variation not accommodated by our CNV baseline, and SNPs falling under primers resulting in decreased amplification efficiency that mimics a copy number loss.

WIDER IMPLICATIONS OF THE FINDINGS

Our panel design and analysis methods enable the determination of *SMN1/2* copy number as part of a much broader NGS research assay of hundreds of genes responsible for inherited diseases, enabling the development of more convenient tools in the future.

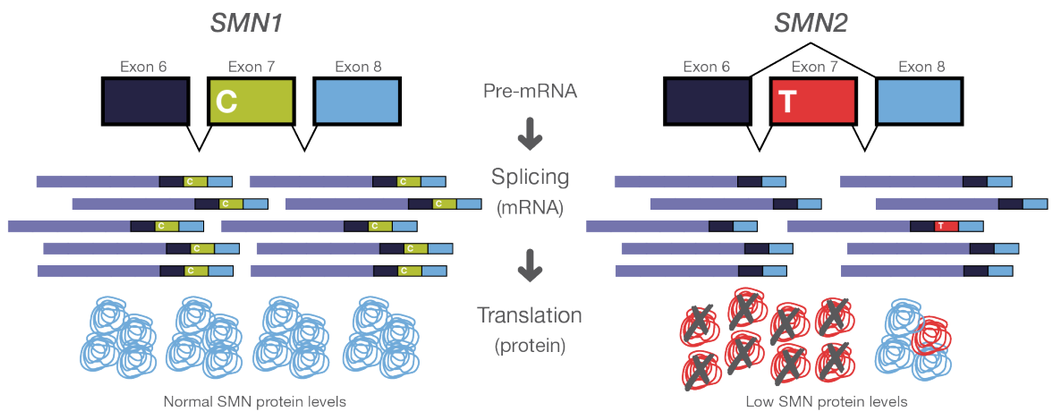
RESULTS

Figure 1. Ion AmpliSeq CarrierSeq Expanded Carrier Screening (ECS) Panel.



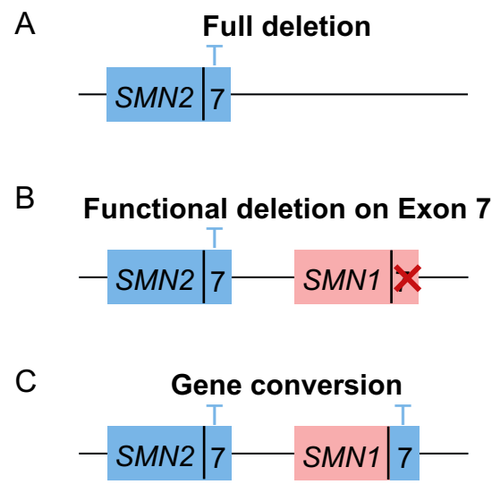
The Ion AmpliSeq™ CarrierSeq™ ECS Panel is a ready-to-implement, end-to-end solution that analyzes SNVs, indels, and CNVs across the full coding sequences of 420 genes covering >28,000 non-benign ClinVar variants in a single assay. The panel allows researchers to consolidate stand-alone assays into a single NGS assay for difficult-to-sequence genes like *SMN1*/*SMN2* and *HBA1*/*HBA2* for SMA and α-thalassemia research, respectively.

Figure 2. Structural and functional differences in *SMN1* and *SMN2*.



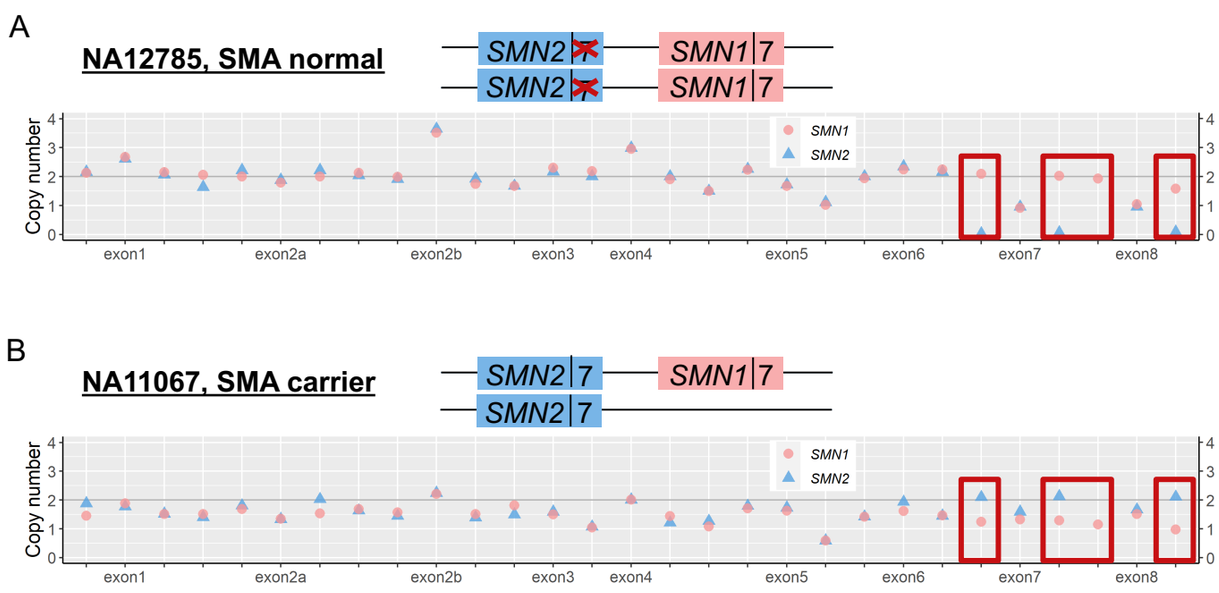
SMN1 and *SMN2* are highly homologous genes in a ~500 kb duplicated block on human chromosome 5q. The two genes differ in their coding sequence by a single base (C in *SMN1*, T in *SMN2*) in exon 7. There are additional differences in intron 7 and noncoding exon 8. The base change causes exon 7 to be skipped in the majority of transcripts for *SMN2*, producing a nonfunctional truncated protein. SMA results when there are defects in the *SMN1* gene that cannot be compensated by the low levels of functional protein produced by *SMN2*. However, increased copies of *SMN2* can reduce the severity of the disease. Image adapted from Butchbach 2016 [1].

Figure 3. Modes of *SMN1* copy loss.



SMA carrier status results from defects in or copy loss of *SMN1*. This can occur as **(A)** a full deletion of the *SMN1* locus, **(B)** a reduction in the functional copy number (FCN) of *SMN1,* often by deletion of exons 7 and 8, or **(C)** gene conversion of exon 7 in *SMN1* to *SMN2*, which brings the T responsible for mis-splicing into the *SMN1* locus. These gene conversion events are distinguishable from point mutations due to the inclusion of additional downstream variations associated with *SMN2*.

Figure 4. Copy number determination for *SMN1* and *SMN2*.



Ion Reporter™ Software and Carrier Reporter Software interpret amplicon representation to call copy number variation in *SMN1*/*SMN2* and annotate the significance of variants, respectively. Normalized read counts for each amplicon are compared to a normal baseline. An algorithm is then used to call copy number and FCN for both genes. Most amplicons and their primers are completely identical between both *SMN1* (red circles) and *SMN2* (blue triangles) and thus represent the average copy number of both genes. However, several amplicons in and between exons 7 and 8 (red boxes) distinguish between the two genes or are exclusively amplified from *SMN1*. **(A)** An SMA normal sample from Coriell with 2 intact *SMN1* copies but a double deletion of exon 7 in *SMN2*. This is reflected by most amplicons showing copy number of two across exons 1–6 (some noise is expected). However, identical amplicons in exon 7 and exon 8 (between red boxes) show average copy number of one, while distinguishing amplicons show two copies of *SMN1* and no *SMN2*. **(B)** An SMA carrier sample from Coriell with a deletion of one copy of the *SMN1* gene. Most amplicons now support an average copy number of ~1.5, whereas distinguishing amplicons (red boxes) resolve this as two copies of *SMN2* and one copy of *SMN1*.

Table 1. Test sample copy number.

SMN1/SMN2 FCN	No. of samples
1/1	2
1/2	2
2/0	2
2/1	33
2/2	36
2/3	5
2/4	1
3/0	2
3/1	6
3/2	1
3/3	1
Total	91

SMN1/*SMN2* functional copy number in the 91 samples used to test the Ion AmpliSeq CarrierSeq ECS Panel and software. Four samples were *SMN1* carriers (red outline). 57% had copy number losses or gains in *SMN2*.

Table 2. FCN accuracy.

	SMN1 FCN	*SMN2* FCN
Correct	89	87
Incorrect	2	4
Accuracy	97.8%	95.6%

The number of samples called correctly and incorrectly for *SMN1* or *SMN2* FCN (intact exon 7).

Table 3. *SMN1* carrier status.

		True condition	
		Pos.	Neg.
Test result	Pos.	4	1
	Neg.	0	86
		Sensitivity	Specificity
		100%	98.9%

The number of positive and negative *SMN1* copy number calls (columns) that were independently determined to be positive or negative by microarray (rows). The sensitivity and specificity are shown.

CONCLUSIONS

The Ion AmpliSeq CarrierSeq ECS Panel and accompanying software enable accurate FCN calling for *SMN1* with high sensitivity and specificity, as part of a broader panel interrogating hundreds of genes and thousands of variants.

ACKNOWLEDGMENTS

The authors thank Colin Davidson for extensive guidance during the development of this product and Kevin Tsai and Tamra Hill for help with graphics.

REFERENCES

1. Butchbach, ME (2016) *Front. Mol. Biosci.* https://doi.org/10.3389/fmolb.2016.00007.

PRODUCT INFORMATION

Product	Cat. No.
Ion AmpliSeq CarrierSeq ECS Panel	A48036
CarrierSeq ECS Kit with Ion 530 Chips (4 samples/chip)	A43585
CarrierSeq ECS Kit with Ion 540 Chips (16 samples/chip)	A43586

TRADEMARKS/LICENSING

ThermoFisher SCIENTIFIC