Invitae's unique approach to testing *SMN1* and *SMN2* for spinal muscular atrophy

Complete loss of *SMN1* gene function results in spinal muscular atrophy (SMA), an early-onset debilitating neuromuscular disorder characterized by loss of motor neurons in the spinal cord. *SMN1* has a near-identical gene copy named *SMN2* also located on chromosome 5, approximately 800 kilobases from *SMN1*. The coding regions of *SMN2* and *SMN1* differ from one another by a single nucleotide in exon 7*, which we term the gene-determining variant (GDV). This difference adversely affects splicing of the exon and leads to very little full length protein production from the *SMN2* gene.

The majority of pathogenic changes in SMA are deletions of *SMN1* or gene conversion of *SMN1* to *SMN2*. In addition, rare inactivating sequence variants can occur in *SMN1*. About 95%–98% of individuals with SMA have zero copies of *SMN1* and about 2%–5% are compound heterozygotes, with a deletion of *SMN1* on one chromosome and a pathogenic sequence variant in *SMN1* on the other chromosome. Notably, the number of *SMN2* copies is highly variable among individuals. This number influences the SMA phenotype in patients with *SMN1* loss, with severity decreasing and age of onset increasing as the number of *SMN2* copies increases.[1,2] (See appendix for detailed background on SMA.)

CHALLENGES IN SMA TESTING AND INVITAE'S NGS-BASED APPROACH

Most laboratories traditionally diagnose SMA by performing multiplex ligation-dependent probe amplification (MLPA) or quantitative PCR (qPCR) to identify loss of *SMN1* exon 7*. These approaches have significant technical limitations and are difficult to efficiently integrate into broader testing.

To address these limitations we developed a comprehensive next-generation sequencing (NGS)-based approach with a customized bioinformatics solution to offer simultaneous sequencing and copy number analysis of these difficult genes while maintaining our commitment to quality and affordability.

Challenges in SMA genetic testing	Traditional methods	Invitae's NGS approach	Clinical significance
***SMN1* copy number**	MLPA and qPCR methods are not amenable to efficient integration into a comprehensive panel that covers the very broad differential diagnosis for SMA	NGS approach allows inclusion of *SMN1/2* in comprehensive gene panels for neuromuscular disorders and neuropathies	Removes barriers to SMA testing; may identify more cases with atypical presentations
***SMN2* copy number**	Traditional methods lack sufficient resolution to accurately determine increased *SMN2* copy number	Robust customized bioinformatic solution identifies *SMN2* copy number.	*SMN2* copy number can provide critical information for disease progression as well as assist in possible clinical trial enrollment or treatment
***SMN1* sequence variants**	*SMN1* exon 7* sequence variants are not routinely determined by Sanger sequencing due to the inability to locate the variant to *SMN1* vs. *SMN2*; sequence variants in exons 1–6 are also not routinely reported	Invitae's methods can disambiguate and report *SMN1* exon 7* sequence variants; we also report exon 1–6 variants although they are not disambiguated	May identify more cases of *SMN1*-related SMA that would otherwise be missed
Cost	Resource intensive assays prevent testing from being offered affordably	NGS-based approach enables Invitae to offer this test with high quality and affordability	At Invitae, we believe offering high-quality, affordable testing will improve patient care

NGS-BASED METHODOLOGY

Invitae has developed a sophisticated assay and bioinformatics solution to accurately detect pathogenic changes in *SMN1* and determine *SMN2* copy number. First, we align sequencing reads derived from both *SMN1* and *SMN2* to an *SMN1* reference sequence. We then measure total *SMN1* + *SMN2* copy number using a modified version of CNVitae, our custom-built copy number variant detection algorithm that utilizes NGS read counts. Once we have the total *SMN1/2* copy number, individual *SMN1* and *SMN2* exon 7* copy numbers are determined using the exon 7* GDV. This simultaneous determination of *SMN1* and *SMN2* exon 7* copy numbers enables high confidence calls for both *SMN1* and *SMN2*** (Figure 1).

We also use the exon 7* GDV to unambiguously place sequence variants in exon 7* of *SMN1* and *SMN2*. The remaining exons (1–6) of *SMN1* and *SMN2* are identical in sequence, and therefore while we can accurately identify sequence and copy number variants in these exons, their true location within *SMN1* or *SMN2* cannot be determined. Even though disambiguation is not possible for variants in exons 1–6, their identification can inform the diagnosis of rare compound heterozygous affected individuals.

SMN1/2 exon 7* copy number variants are confirmed by ligation-dependent sequencing, an Invitae innovation that transforms traditional MLPA into a highly scalable NGS method. Sequence variants in exon 7* are confirmed using single-molecule PacBio sequencing, which enables the phasing of the variant with the GDV to unambiguously place the variant in either *SMN1* or *SMN2*.

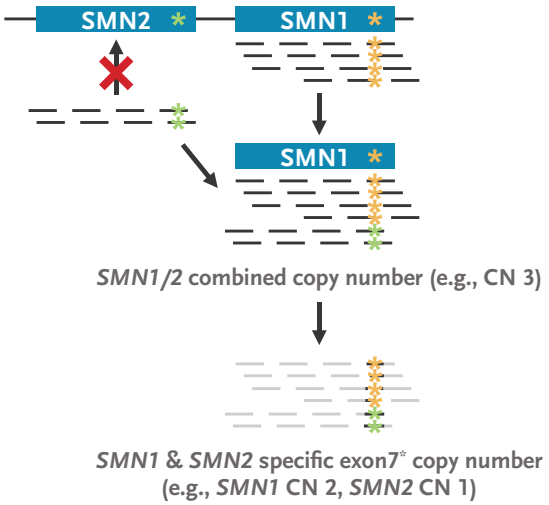

Figure 1: *SMN1/2* bioinformatics method. Reads derived from both *SMN1* and *SMN2* are aligned to *SMN1*, and combined *SMN1/2* copy number is determined using Invitae's read count-based copy number variant detection algorithm, CNVitae. *SMN1*- and *SMN2*-specific exon 7* copy number is resolved by counting reads with the gene determining variant in exon 7*.

VALIDATION

Our *SMN1/2* approach was validated on a set of nine samples available from an external commercial repository of biological samples. *SMN1* exon 7* copy number information was previously determined through traditional methods, and *SMN2* copy number was known for a subset of these samples.[3] Our method showed 100% sensitivity and specificity for *SMN1* and *SMN2* copy number, and notably its higher resolution for determining *SMN2* copy number enabled us to obtain accurate results for three samples for which copy number had been imprecisely determined with traditional methods previously (see appendix validation table for details).[3]

SMN1 AND *SMN2* POPULATION FREQUENCY

We applied Invitae's novel *SMN1/2* bioinformatics method to examine *SMN1/2* genotypes in the internal Invitae database of sequencing results. In 2,141 de-identified individuals, we identified 34 individuals carrying a single copy of *SMN1*. This observed carrier rate of 1:63 is in accordance with previously reported SMA carrier rates,[4] and provides additional evidence of the sensitivity of our method. We also investigated the distribution of *SMN2* copy number in this population and found significant variability.

Copy number	0	1	2	3	4	5
SMN1	0.0%	1.6%	92.9%	5.9%	0.4%	0.1%
SMN2	7.1%	29.2%	61.8%	1.9%	0.2%	0.0%

FOOTNOTES

*Reference sequence NM_000344.3, which is used to describe *SMN1* sequence variants, contains 8 protein-coding exons. Due to historical reasons, the second and third exons are conventionally referred to as exons 2a and 2b, and the subsequent exons are referred to as exons 3–7 (PMID: 8838816). At Invitae, systematic exon numbering is used for all genes, including *SMN1* and *SMN2*. For this reason, the gene-differentiating exon conventionally referred to as exon 7 in the literature and in this whitepaper is referred to as exon 8 in our clinical reports.

**Copy number of *SMN2* exon 7* is expected to represent copy number for the entire *SMN2* gene, and will only be reported for individuals with a positive result in *SMN1*. CNVs limited to exons 1–6 of *SMN1* or *SMN2* will not be reported. This assay cannot detect silent carriers (individuals that have 2 functional copies of *SMN1* on one chromosome and zero copies on the other). Therefore a negative result for carrier testing greatly reduces but does not eliminate the chance that a person is a carrier.

REFERENCES

1. Mailman MD, *et al*. Molecular analysis of spinal muscular atrophy and modification of the phenotype by *SMN2*. *Genet Med*. 2002;4:20–6. PMID: 11839954
2. Swoboda KJ, *et al*. Natural history of denervation in SMA: relation to age, *SMN2* copy number, and function. *Ann Neurol*. 2005;57:704–12. PMID: 15852397
3. Stabley DL, *et al*. *SMN1* and *SMN2* copy numbers in cell lines derived from patients with spinal muscular atrophy as measured by array digital PCR. *Molecular Genetics & Genomic Medicine*. 2015;3(4):248-257. PMID: 26247043
4. Hendrickson BC, *et al*. Differences in *SMN1* allele frequencies among ethnic groups within North America. *Journal of Medical Genetics*. 2009;46:641-644. PMID: 19625283
5. Wirth B, *et al*. De novo rearrangements found in 2% of index patients with spinal muscular atrophy: mutational mechanisms, parental origin, mutation rate, and implications for genetic counseling. *American Journal of Human Genetics*. 1997;61(5):1102-1111. PMID: 9345102
6. McAndrew PE, *et al*. Identification of proximal spinal muscular atrophy carriers and patients by analysis of *SMNT* and *SMNC* gene copy number. *Am J Hum Genet*. 1997 Jun;60(6):1411-22. PMID: 9199562
7. Chen KL, *et al*. Duplications and de novo deletions of the *SMNt* gene demonstrated by fluorescence-based carrier testing for spinal muscular atrophy. *Am J Med Genet*. 1999 Aug 27;85(5):463-9. PMID: 10405443
8. Prior TW, *et al*. Homozygous *SMN1* deletions in unaffected family members and modification of the phenotype by *SMN2*. *Am J Med Genet A*. 2004 Oct 15;130A(3):307-10. PMID: 15378550
9. Pearn J. Incidence, prevalence, and gene frequency studies of chronic childhood spinal muscular atrophy. *J Med Genet*. 1978 Dec;15(6):409-13. PMID: 745211
10. Mostacciuolo ML, *et al*. Epidemiology of spinal muscular atrophies in a sample of the Italian population. *Neuroepidemiology*. 1992;11(1):34-8. PMID: 1608493
11. Thieme A, *et al*. Epidemiological data on Werdnig-Hoffmann disease in Germany (West-Thüringen). *Hum Genet*. 1993 Apr;91(3):295-7. PMID: 8478016

APPENDIX

BACKGROUND ON SMA

Clinical description

Spinal muscular atrophy (SMA) is a neuromuscular disorder caused by the loss of motor neurons within the spinal cord, which results in progressive muscle weakness and atrophy. Other features of SMA may include muscle fasciculations, tremor, poor weight gain, sleeping difficulties, pneumonia, scoliosis, joint contractures, and congenital heart disease. Four clinical SMA subtypes have been defined: severe infantile acute SMA type I (OMIM 253300, also referred to as Werdnig-Hoffman disease), infantile chronic SMA type II (OMIM 253550), juvenile SMA type III (OMIM 253400, also referred to as Wohlfard-Kugelberg-Welander disease), and adult-onset SMA type IV (OMIM 271150).

Molecular genetics

SMA is caused by biallelic loss of *SMN1* gene function. Loss of *SMN1* gene function can occur through at least 3 different mechanisms: 1) deletion of *SMN1*, 2) inactivating sequence variants in *SMN1*, or 3) gene conversion of *SMN1* to *SMN2*.

SMA testing is complicated by the presence of the nearby *SMN2* gene, which does not itself cause SMA if mutated, but it has nearly identical sequence homology to *SMN1*. Among the 882 nucleotides that code for amino acids, only one nucleotide differs between *SMN1* and *SMN2*: in the terminal coding exon at position c.840, a C nucleotide occurs in the *SMN1* gene while a T nucleotide occurs in the *SMN2* gene. The variant at position c.840 does not result in a different amino acid being encoded, but causes improper processing of the *SMN2* mRNA and consequently reduced protein production.

Although the *SMN2* gene produces minimal functional protein, it can significantly modify the severity of SMA, with more copies of *SMN2* correlated with decreased SMA severity.[1,2] The c.840 position lies in exon 8 (conventionally referred to in the literature as exon 7*), and this difference is routinely used to differentiate *SMN1* and *SMN2* copy number. Therefore, we call the c.840 difference the *SMN1/2* gene determining variant (GDV).

Inheritance

SMA is inherited in an autosomal recessive manner, with a new mutation rate of up to 2%.[5] Some individuals (an estimated 8% of Ashkenazi Jewish carriers) appear to have normal *SMN1* results due to two functional copies of *SMN1* on one chromosome and zero functional copies on the other chromosome, an allelic combination termed "silent carrier."[6,7]

Penetrance

On its own, biallelic loss of *SMN1* has high penetrance. However, penetrance is complicated by modifying effects of *SMN2* copy number (see above). Unaffected individuals with five copies of *SMN2* and a homozygous deletion in *SMN1* have been reported, suggesting that five copies of *SMN2* may compensate for the lack of *SMN1* expression.[8]

Prevalence

The prevalence of SMA I, II and III is estimated to be <2 per 100,000 of the general population,[9] and the incidence of the severe infantile acute SMA type I is estimated to be 4–10 individuals per 100,000 live births.[9–11]

About 95%–98% of individuals with SMA have zero copies of *SMN1* and about 2%–5% are compound heterozygotes, who have a deletion of *SMN1* along with one copy of *SMN1* that contains a pathogenic sequence variant. Approximately 1 in 50 individuals are carriers of a single *SMN1* deletion or pathogenic variant, although this frequency has been found to vary between ethnic groups.[4]

VALIDATION TABLE

Sample ID	External lab *SMN1* copy number	External lab *SMN2* copy number	Published *SMN2* copy number[3]	Invitae *SMN1* copy number	Invitae *SMN2* copy number	Orthogonal confirmation status
NA22592	0	3		0	3	confirmed
NA00232	0	1→2	2	0	2	confirmed
NA09677	0	2→3	3	0	3	confirmed
NA10684	0	2	2	0	2	confirmed
NA03815	1	1		1	1	confirmed
NA23687	1			1	2	confirmed
NA23688	1			1	2	confirmed
NA03813†	0	2→3	3	0	3	confirmed
NA03814†	1	3 or more	5	1	5	confirmed

***SMN1/2* copy number validation was conducted with 9 samples from an external reference laboratory. Samples NA00232, NA03813 and NA09677 had incorrect external *SMN2* copy number annotations and were reclassified during our validation process to reflect copy number confirmed by Stabley *et al.*[3] and Invitae.**

†Samples were run in five replicates and showed concordant results.

