

1 **A dual-mode targeted Nanopore sequencing assay for comprehensive *SMN1***  
2 **and *SMN2* variant analysis**

3 **Running title:** Long read sequencing-based analysis of *SMN1/SMN2*

4 Brad Hall<sup>1\*</sup>, Sawsan Alyafei<sup>2\*</sup>, Sathishkumar Ramaswamy<sup>2</sup>, Shruti Sinha<sup>2</sup>, Maha El  
5 Naofal<sup>2</sup>, Fatma Rabea<sup>2,3</sup>, Bryan J. Killinger<sup>1</sup>, Gary J. Latham<sup>1</sup>, Ahmad Abou  
6 Tayoun<sup>2,4</sup>

7

8 <sup>1</sup>Asuragen, A Bio-Techne Brand, Texas, USA

9 <sup>2</sup>Genomics Center of Excellence, Al Jalila Children's Specialty Hospital, Dubai  
10 Health, Dubai, UAE

11 <sup>3</sup>Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai Health,  
12 Dubai, UAE

13 <sup>4</sup>Center for Genomic Discovery, Mohammed Bin Rashid University of Medicine and  
14 Health Sciences, Dubai Health, Dubai, UAE

15 \*Equally contributed to this work

16

17 **Corresponding Author:** Ahmad Abou Tayoun; [Ahmad.Tayoun@dubaihealth.ae](mailto:Ahmad.Tayoun@dubaihealth.ae)

18

19 **Keywords:** Enrichment, Nanopore Sequencing, *SMN1/2*, SMA, Carrier Screening

20

21 **Previously presented at Oxford Nanopore Technologies London Calling 2023**

22

23 **Word Count:** 3523

24

25 **5 Figures and Tables**

26

27 **Abbreviations:**

28 SMA, Spinal Muscular Atrophy

- 29 SNVs, Single nucleotide variants
- 30 Indels, Insertions and deletions
- 31 CNVs, Copy number variants
- 32 NGS, Next generation sequencing
- 33 US, United States
- 34 FDA, Food and Drug Administration
- 35 HRM, High-resolution melt
- 36 MLPA, Multiplex ligation-dependent probe amplification
- 37 ONT, Oxford Nanopore Technologies
- 38 PCR, Polymerase chain reaction
- 39 gDNA, Genomic DNA
- 40 NHGRI, National Human Genome Research Institute
- 41 NIGMS, National Institute of General Medical Sciences
- 42 ddPCR, droplet digital PCR
- 43 CE, Capillary electrophoresis
- 44 QC, Quality control
- 45
- 46 **Human Genes:**
- 47 *SMN1* survival of motor neuron 1, telomeric HGNC:11117
- 48 *SMN2* survival of motor neuron 2, centromeric HGNC:11118
- 49 *CFTR* CF transmembrane conductance regulator HGNC:1884

50 **Abstract**

51 **Background**

52 Spinal Muscular Atrophy (SMA) is one of the most common recessive disorders for  
53 which several life-saving treatment options are currently available. It is essential to  
54 establish universal SMA screening and diagnostic programs using scalable, cost-  
55 effective and accessible platforms to accurately identify all variation types, which is  
56 complicated by homologous *SMN1* and *SMN2* genes.

57 **Methods**

58 We developed a dual-mode PCR-based target enrichment that generates 2.7 to 11.2  
59 kb amplicons spanning *SMN1* and *SMN2* genes for any-length nanopore  
60 sequencing. We trained a variant calling model that utilizes paralog-specific  
61 sequences and read-depth data to accurately detect sequence and copy number  
62 variants specific to each gene.

63 **Results**

64 We present results from the development, optimization, and external evaluation of  
65 this assay using over 750 samples, including cell lines, residual presumed normal  
66 blood donors, and patients with known *SMN1* and *SMN2* genotypes. The assay  
67 detects SNVs, indels, and CNVs with >98% accuracy across all sample sets, with a  
68 highly dynamic throughput range, relatively fast turnaround time, and limited hands-  
69 on-time. Together with the modest capital investment and consumable costs per  
70 sample, this assay can help increase access to SMA testing in low- and middle-  
71 income settings.

72 **Conclusion**

73 We describe a PCR/Nanopore sequencing assay and a customized analysis pipeline  
74 for the comprehensive and accurate detection of variation at the SMA locus and

75 demonstrate its scalability, cost-effectiveness, and potential for the universal  
76 implementation of SMA screening and diagnostic programs.

77

78 **Introduction**

79 Spinal Muscular Atrophy (SMA) is the second most common autosomal recessive  
80 disease and the most common genetic cause of infant death (1). In around 95% of  
81 cases, SMA is caused by a homozygous deletion of the *SMN1* gene, which encodes  
82 for the survival motor neuron 1 protein (2, 3, 4); loss of this protein leads to  
83 degeneration of the motor neurons in the spinal cord and progressive muscle  
84 weakness, paralysis and, if untreated, premature death (5). Additional intragenic and  
85 structural variants have also been identified (6).

86

87 Both *SMN1* and its paralog, *SMN2* encode the same protein, SMN, and copy  
88 numbers of each within the genome can range from zero to four or more. SMA  
89 disease is modified by copy number status of the *SMN2* gene, which has >99.9%  
90 sequence identity to *SMN1* (7), though it does not produce a functional product  
91 because of a single nucleotide change in exon 7 that causes exon skipping and  
92 produces a truncated, nonfunctional protein (8). However, due to leaky expression of  
93 the full-length protein from the *SMN2* locus, its copy number status is inversely  
94 correlated with SMA clinical phenotype and disease severity (9, 8, 10). Patients at  
95 the severe end of the spectrum (SMA type I) often have one copy of *SMN2*, while  
96 those with a milder phenotype (SMA type IV) have greater than or equal to 4 copies  
97 of this gene (9, 11, 5).

98

99 Given its prevalence, life-threatening outcomes and, most recently, the availability of  
100 three life-saving medications approved by the US FDA (Zolgensma, Spinraza, and  
101 Evyrsdi), genetic diagnostic and screening programs have become essential for risk  
102 assessment, early detection, and timely patient treatment (12, 13). Such programs,

103 however, require rapid, accurate, and comprehensive testing platforms to  
104 unambiguously detect and resolve single nucleotide (SNVs) and copy number  
105 variants (CNVs) in the *SMN1* and *SMN2* genes, as well as complex rearrangements  
106 leading to gene conversions (14) and silent carriers (15).

107

108 SMA testing strategies should also be highly scalable, cost-effective, and not require  
109 significant capital investment or complex infrastructure to support the deployment of  
110 universal screening programs in low-resource settings (16). These qualities are not  
111 only essential for equitable global access to screening but also important for  
112 characterizing disease epidemiology (incidence, prevalence and carrier frequencies)  
113 across populations. Such information can then inform the most appropriate  
114 screening and prevention strategies (17).

115

116 Since *SMN2* has high homology to *SMN1*, current screening methods such as NGS  
117 often require complex analysis methods that are difficult to interpret and may not  
118 accurately resolve *SMN1* copy number variants. Targeted copy number methods  
119 such as high-resolution melt (HRM) and multiplex ligation-dependent probe  
120 amplification (MLPA) are not designed to detect many pathogenic variants (18, 6).

121 Recently, long-read sequencing has emerged as an option to identify copy number  
122 and pathogenic variants in a single workflow (19; 20). Recent advancements in  
123 sequencing technologies may help overcome these challenges by incorporating  
124 much longer unique reads that can differentiate *SMN1* from *SMN2* via paralog-  
125 specific variants. However, such assays are limited by accessibility, capital costs,  
126 and costly and complex operational infrastructure.

127

128 To address these limitations, we developed a complete assay and workflow for  
129 *SMN1* and *SMN2* genetic analysis using targeted PCR amplification, Oxford  
130 Nanopore (ONT) any-length sequencing, and a customized analytical algorithm that  
131 resolves multiple variant classes. We report results with more than 750 samples  
132 demonstrating high accuracy across diverse and complex genotypes. We also  
133 discuss operational and real-world advantages for carrier screening and diagnostic  
134 applications, including simple library preparation, flexible sample throughput, low  
135 capital investment, small instrument footprint, and modest consumable cost per  
136 sample.

137

## 138 **Methods**

### 139 **Assay Prototyping cohort**

140 The prototype assay was developed using genomic DNA (gDNA) from cell lines  
141 (N=97) obtained from the Centers for Disease Control and Prevention Repository  
142 (N=4), National Human Genome Research Institute (NHGRI) Sample Repository for  
143 Human Genetic Research (N=12), and National Institute of General Medical  
144 Sciences (NIGMS) Human Genetic Cell Repository (N=73) at the Coriell Institute for  
145 Medical Research (Camden, NJ; **Supplemental Table 1**). Additionally, cell lines  
146 (N=8; described in [21](#)) were procured from ATCC and isolated using a precipitation-  
147 based method (Qiagen, Hilden, Germany).

148

### 149 **Assay Optimization cohort**

150 Human-derived, presumed normal, de-identified residual whole blood specimens  
151 (N=227) were obtained from We Are Blood (Austin, TX) under the required  
152 regulatory approvals for evaluating clinical specimens. Whole blood was purified

153 using either silica resin/column-based method (Qiagen, Hilden, Germany) or  
154 functionalized magnetic bead (Applied Biosystems, Waltham, MA). Genomic DNA  
155 quantity (ng/uL) and quality ( $A_{260}/A_{280}$ ) were assessed using spectrophotometry.  
156 Samples were diluted in nuclease-free water to the target concentration for analysis  
157 using the assay.

158

### 159 **Assay Evaluation cohort**

160 De-identified samples with known SMA copy number status as determined by clinical  
161 testing using a droplet digital PCR assay (see below) at the CAP-accredited  
162 genomics center, at Al Jalila Children's Specialty Hospital (Dubai, United Arab  
163 Emirates), were used for assay evaluation and test performance characterization.  
164 This study was approved by the Dubai Health Authority Research Ethics Committee  
165 (DSREC-07/2023\_06 and DSREC-SR-03/2023\_08).

166

### 167 **Droplet digital PCR**

168 Genomic DNA extracted from peripheral blood (N = 70) were tested for *SMN1* and  
169 *SMN2* copy number status using a proprietary droplet digital PCR (ddPCR) assay  
170 (Bio-Rad, USA) following manufacturer's instructions. This assay was clinically  
171 validated by the Al Jalila Children's genomics laboratory (22).

172

### 173 **PCR and capillary electrophoresis (PCR/CE)**

174 Only samples determined to have a homozygous deletion of the *SMN1* gene by  
175 ddPCR (N = 32) were clinically tested for *SMN2* copy number status by ddPCR. For  
176 the remaining samples with  $\geq 1$  *SMN1* copies (N = 44), *SMN2* dosage status was  
177 determined by the AmplideX® PCR/CE *SMN1/2* Plus Kit as previously described



178 (23). In addition, this method was utilized to assess both *SMN1* and *SMN2* copy  
179 number for all cell line and presumed normal whole blood samples during assay  
180 development and optimization.

181

## 182 **PCR and Nanopore Sequencing (PCR/Nanopore)**

183 The prototype assay workflow includes PCR master mix setup, gene-specific  
184 amplification, sample specific barcoding, paramagnetic bead size selection and  
185 concentration, library adapter ligation, Nanopore sequencing, and assay-specific  
186 analysis pipelines (**Figure 1**). Roughly 40ng gDNA was amplified by multiplex PCR  
187 targeting 2.7kb regions of *SMN1* or *SMN2* exons 7-8 and copy number neutral  
188 endogenous control amplicons from *CFTR* in one mix or a larger 11.2kb amplicon  
189 encompassing exons 3-8 in a separate mix. Reactions were cleaned up using 0.6X  
190 AMPure XP bead ratio (Beckman Coulter, Indianapolis, IN). Samples were tagged  
191 with unique barcodes during a second PCR reaction, normalized by mass after Qubit  
192 quantitation (Invitrogen, Carlsbad CA), pooled into a single reaction, and  
193 concentrated using 0.6X AMPure XP bead ratio.

194

195 Sequencing libraries were prepared using the Nanopore Ligation Sequencing Kit  
196 (LSK-110 or LSK-114; ONT). Sequencing was conducted using MinKNOW software  
197 (22.10.10) on MinION flow cells (R9.4.1 or R10.4.1) with a Mk1B sequencer  
198 connected to a computer. Sequencing proceeded for at least 5 hours to obtain a  
199 minimum of 150 fully-spanning reads per copy per region. Super-accurate base  
200 calling was performed using Guppy (6.3.9) after sequencing and fully spanning reads  
201 were aligned to GRCh38 using minimap2 (2.15+dfsg-1).

202

203 To determine copy number, we trained a gradient boosting tree model to predict  
204 copy number using the ratio of Cs and Ts at the paralog specific variant c.840C>T in  
205 exon 7 (NM\_000344.3c.840C>T; Single Nucleotide Polymorphism database,  
206 <https://www.ncbi.nlm.nih.gov/snp>, accession number rs1164325688, build 155, last  
207 accessed August 15<sup>th</sup>, 2023) (**Supplemental Figure 1**). The model incorporates  
208 read depth of fully spanning 2.7kb fragments aligned to *SMN1* (c.840C) or *SMN2*  
209 (c.840T) and the geometric mean of endogenous control amplicons (*CFTR*) to infer  
210 fold change based on normalized read depth in 2-copy calibrator samples  
211 (**Supplemental Figure 2**). The machine learning algorithm was trained on a subset  
212 of cell lines (N=79) and an independent set of residual whole blood samples  
213 (N=352). Hyperparameters for the decision tree model were selected using an 80:20  
214 train:test split in a stratified randomly selected five-fold cross validation scheme.

215

216 Additionally, *SMN1* and *SMN2* copy number was informed with the longer 11.2kb  
217 amplicon flagged as edge cases where the machine learning algorithm predicted a  
218 copy number of 2.25 to 2.75. First, unique haplotype groups were identified and  
219 differentiated by sequence variability (e.g. SNVs) corresponding to unique copies  
220 aligning to the same region of the genome (**Supplemental Figure 3**). This allows for  
221 the identification of all unique copies for a gene target, and thus can be used to  
222 inform copy number calls and provide proper input to variant calling software that  
223 often assumes no more than two copies of the gene are present. Group sizes were  
224 normalized according to read depth and copy number was inferred for each group.  
225 The number of normalized haplotype groups were reported as the predicted copy  
226 number for samples that were flagged as edge cases when more than one group  
227 was identified.

228 Predicted copy number from the PCR/Nanopore assay was compared to data  
229 collected by orthogonal methods (ddPCR and/or PCR/CE). Small nucleotide variants  
230 and insertion/deletions were identified using Clair3 (24).

231

## 232 **Results**

### 233 **Assay Development and Optimization**

234 We developed a complementary, two-in-one MinION-based assay utilizing amplicons  
235 ranging more than 10-fold in length, including reference loci, to identify *SMN1* and  
236 *SMN2* variants. Assay designs and analysis methods were optimized for copy  
237 number changes and phased SNVs and indels (**Figure 1**). Primers were developed  
238 to amplify ~800-3000 base fragments (“short” amplicons) encompassing both *SMN1*  
239 and *SMN2* exon 7-8 and endogenous controls (**Supplemental Figure 2**). The assay  
240 assesses *SMN1* or *SMN2* copy number by aligning reads spanning exons 7 and 8 to  
241 either *SMN1* or *SMN2*, then determining the raw read depth and ratio of reads  
242 associated with c.840C (*SMN1*) or c.840T (*SMN2*). A machine learning method  
243 infers copy number by comparing the C:T ratio in exon 7 and corresponding read  
244 depths to endogenous control amplicons in calibrator samples.

245

246 We also designed primers to amplify an 11.2kb fragment (“long” amplicon)  
247 encompassing exons 3-8 to enhance phasing SNVs in *SMN1* or *SMN2*  
248 (**Supplemental Figure 3**). This long amplicon typically captures several paralog-  
249 specific variants and facilitates sequence deconvolution of reads into haplotype  
250 groups to infer copy number. In some samples, however, there are no copy  
251 distinguishing sequence variants within one or both paralogs to independently  
252 resolve copy number.

253

254 The assay was developed and evaluated in three phases (**Figure 2**). In the  
255 prototyping phase, gene-specific primers, protocols, and reagents were designed  
256 and tested on a diverse set of cell line (N=79) and presumed normal whole blood  
257 (N=352) samples. This set included independently analyzed *SMN1* and *SMN2*  
258 genotypes ranging from 0 to  $\geq 4$  copies along with SNVs and Indels described below.  
259 Sequencing data were collected with short and long amplicons. Algorithms were  
260 trained using amplicon sequence and read depth within *SMN1* and *SMN2* compared  
261 to known genotypes for each sample. We also identified endogenous control  
262 amplicons and calibrator samples to convert *SMN1* and *SMN2* read-depth ratios into  
263 copy numbers.

264

265 In the assay optimization phase, performance was first evaluated with the short,  
266 2.7kb amplicon design utilizing a sample cohort of 18 additional cell lines (N=97  
267 total) and an independent set of whole-blood samples (N=227) across 0 to  $\geq 3$  *SMN1*  
268 and *SMN2* copies (**Figure 33**). Assay copy number predictions were compared to  
269 PCR/CE data and segregated by sample type.

270

271 Copy number agreement was 97.9% (95/97) between PCR/Nanopore and  
272 comparator method for both *SMN1* and *SMN2* in the cell-line set. One sample  
273 (NA20232) was discordant for both *SMN1* and *SMN2*. This cell line was expected to  
274 be 3 copy *SMN1* and 2 copy *SMN2* (3/2) by PCR/CE yet reported as 2/1 by the  
275 PCR/Nanopore assay. Two other samples had either *SMN1* or *SMN2* discordant  
276 copy number data where a 3 copy was expected, yet a 2 copy was predicted by the

277 algorithm. In each case, haplotype deconvolution with the long amplicon supported  
278 the expected copy number.

279

280 In the whole blood sample cohort, 218/227 (96.0%) *SMN1* and 224/227 (98.7%)  
281 *SMN2* copy number calls agreed between PCR/Nanopore and comparator method.

282 Of the 9 discordant samples for *SMN1*, 7 discordant samples had more than one  
283 haplotype. We used the long amplicon design to investigate these discordances,  
284 consistent with methods described. Six of the seven samples agreed with  
285 comparator data when both amplicon designs were utilized. Four of 9 discordant  
286 samples were flagged as edge cases at or near the call threshold with the short  
287 amplicon and 3 were resolved by haplotype deconvolution with the long amplicon  
288 design.

289

290 Consequently, we applied the flagging QC across the entire dataset to determine the  
291 effect on accuracy (**Figure 33**). One cell line and 12 whole blood samples were  
292 flagged as edge cases at or near the call threshold with the 2.7kb amplicon for  
293 *SMN1*. Of these, 4 were discordant using the short amplicon alone, but 3 were  
294 corrected by calls using the long amplicon. Similarly, of the 3 *SMN2* calls flagged  
295 only 1 was discordant and long amplicon design supported the expected copy  
296 number. In no instance was a sample flagged where the long amplicon call  
297 disagreed with comparator data when at least 2 haplotype groups were identified. As  
298 a result, both amplicon designs were used for all subsequent analyses where call  
299 threshold flagging prioritized the long amplicon copy number call.

300

301 In addition to copy number evaluation, the assay was designed to identify and phase  
302 SNVs and indels without reflexing to other assays. For example, the *SMN2*,  
303 NM\_017411.3: c.859G>C variant (dbSNP, rs121909192, build 155, last accessed  
304 August 15<sup>th</sup> 2023) is associated with a less severe SMA phenotype (25, 26, 27, 28).  
305 Two variants linked with the *SMN1* duplication haplotype have been shown to flag  
306 silent carriers and increase carrier detection rates when 2 *SMN1* copies occur on the  
307 same chromosome. These variants are NM\_000344.3:c.\*3+80T>G (alias  
308 g.27134T>G; dbSNP, rs143838139, build 155, last accessed August 15<sup>th</sup> 2023) and  
309 NM\_000344.3: c.\*211\_\*212del (alias g.27706\_27707delAT; dbSNP, rs200800214,  
310 build 155, last accessed August 15<sup>th</sup> 2023).

311

312 PCR/Nanopore genotyping of these variants agreed with the PCR/CE comparator  
313 method in 324/324 samples for c.\*3+80T>G and c.859G>C, and in 322/324 (99.4%)  
314 for c.\*211\_212del (**Figure 4**). Investigation of the two discordant samples suggested  
315 erroneous comparator data. In HG00691, a rare non-pathogenic deletion in *SMN2*  
316 (rs576032516) shifted a CE peak into the mutant CE bin, resulting in a false positive.  
317 In the second sample, the automated peak caller in the CE assay incorrectly called  
318 noise along the baseline within the c.\*211\_212del bin, resulting in another false  
319 positive call.

320

### 321 **Independent Evaluation of Residual Clinical Samples**

322 Next, we evaluated assay performance across a diverse, independent cohort of  
323 samples from 70 individuals (57% females, average age 2.17 years, range 1 day –  
324 16 years) with known *SMN1* and *SMN2* copy number status originally tested at AI

325 Jalila Children's Specialty Hospital, Dubai, UAE. Those individuals represented 14  
326 countries, primarily from the Middle East and North Africa (**Supplemental Table 2**).

327

328 Of all individuals, 30 (42.9%) had homozygous deletions in the *SMN1* gene, while  
329 the remaining 40 (57.1%) had 1 to 4 copies of this gene. Sixty-seven of the 70  
330 samples were also evaluated for *SMN2* copy number. Most individuals (58%) had 2  
331 *SMN2* copies, 10% had 0 copies and 42% had 1, 3, or  $\geq 4$  copies of this gene  
332 (**Supplemental Table 3**).

333

334 The PCR/Nanopore assay correctly identified 70/70 and 65/67 of *SMN1* and *SMN2*  
335 copy number states for a concordance rate of 100% (95% CI, 94.80% - 100.00 %)   
336 and 97.0% (95% CI, 89.75% - 99.18%), respectively (**Table 1**).

337

### 338 **Cost Effectiveness and Hand-On Time**

339 The accessibility and broad use of an *SMN1/2* screening and genotyping assay  
340 depends on numerous criteria, including performance, operational factors and cost.  
341 Assay workflow, analysis and economic considerations are especially important for  
342 laboratories in lower resource environments. We assessed these elements by  
343 performing a time-motion analysis, quantifying the scalability of the assay across  
344 different sample batch volumes, and calculating estimated costs per sample.

345

346 Time-motion analysis was estimated across 6 operators based on experience for  
347 both a 24- and 96-sample batch (**Supplemental Table 4**). The workflow from sample  
348 to answer requires less than 48 hours for a 24-sample batch on a Mk1B connected  
349 to a computer with a recommended GPU. Gene-specific and barcoding PCR can be

350 completed in a single work shift. Subsequently, samples can be pooled, the library  
351 prepared, and sequencing initiated with sequencing and analysis completed  
352 overnight for review the next day. A 96-sample batch required additional sequencing  
353 and data processing time but was still completed within 72 hours.

354

355 The assay supports 12 to 96 samples per batch (**Supplemental Table 5**). Batch  
356 size is currently only limited by the recommended barcodes available from ONT  
357 (N=96) since flow cells can routinely process >10M reads. A single Mk1B running 12  
358 samples once per week could screen 624 samples per year. By comparison, 480  
359 samples could be processed per batch on one GridION with 5 independently  
360 accessible flow cells, real-time base calling, demultiplexing, and alignment. Utilizing  
361 3 staggered run batches per week a lab could process 75,000 samples per year.  
362 This throughput can scale linearly by adding additional GridION instruments. In  
363 addition, the assay supports automated reaction setup, bead isolation, quantitation,  
364 and data analysis to further improve workflow efficiency.

365

366 We estimated less than \$20 USD per sample for materials costs excluding PCR  
367 (**Supplemental Table 6**). Since the assay is still in development, PCR enrichment  
368 costs cannot be accurately determined. For example, the assay design must be  
369 finalized and verified, reagents and kits manufactured, and quality control and  
370 release testing established. However, PCR is well documented to be highly cost-  
371 effective at scale, and we expect that the total per-sample costs will be comparable  
372 to other *SMN1/2* diagnostic kits that provide far less genotyping information and  
373 insight. These costs also do not include instrument-related expenses. Importantly,  
374 ONT instruments (e.g. Mk1B and GridION) have low capital requirements and



375 benchtop footprints, which creates flexibility for laboratories in how they implement  
376 the technology. For either instrument, a computer is required for data analysis. For  
377 Mk1B instruments, the computer must be equipped with a high-performance GPU to  
378 utilize live basecalling, demultiplexing, and alignment during sequencing.

379

## 380 **Discussion**

381 We developed a novel dual-mode PCR/Nanopore sequencing assay for  
382 comprehensive, scalable and cost-effective *SMN1/2* genotyping. The assay design  
383 utilizes a PCR target enrichment approach to generate 2.7kb “short” and 11.2kb  
384 “long” amplicons spanning regions of the *SMN1* and *SMN2* genes, as well as a  
385 machine learning-based analytical pipeline haplotype phasing and read depth data to  
386 decipher sequence and copy number variants specific to both highly homologous  
387 genes.

388

389 Data from both amplicons were combined to achieve the highest accuracy for the  
390 prototype assay (>97%). Other long-read sequencing designs utilize only phased  
391 haplotype analysis either by hybrid capture (20) or multiple ultra-long (>26kb)  
392 amplicons on the PacBio SMRT platform (19). Robustness may be adversely  
393 affected by identical haplotypes without read depth normalized to endogenous and  
394 exogenous control amplicons. Indeed, we observed identical haplotype groups in our  
395 sample set when only the long amplicon was used for copy number evaluation. In  
396 contrast, the PCR/Nanopore assay utilizes a combination of read-depth  
397 normalization (short amplicon) and haplotype phasing (long amplicon) to resolve  
398 *SMN1* and *SMN2* copy number and phase pathogenic SNV/introns. In addition, the  
399 assay scales from tens to tens of thousands of samples per year with reduced

400 capital and per-sample costs compared to other sequencing-based assays.  
401 Reducing costs is critical for many existing screening labs or those looking to adopt  
402 such screening assays, as reimbursement costs are not always able to cover test  
403 costs with NGS workflows. Lastly, the assay includes analysis and reporting software  
404 to reduce interpretation expertise and overhead, though the full suite of push-button  
405 automated analysis software is still in development.

406

407 We acknowledge a few limitations of the current study. The optimized prototype was  
408 designed to phase variants across *SMN1* or *SMN2* exons 3-8. Although we  
409 demonstrated variant phasing with non-pathogenic silent carrier and disease  
410 modifier SNVs, full exon coverage of *SMN1* and *SMN2* is preferred to assess all  
411 potential SNV/indel variants. In ongoing work, we have extended the assay design to  
412 cover exons 1, 2a, and 2b, though performance has not yet been evaluated. Further,  
413 amplification efficiency must be optimized to reduce read depth variability between  
414 samples by optimizing primers and cycling conditions. These modifications are  
415 expected to improve the resolution and differentiation of 3- and 4-copy genotypes,  
416 especially important for treatment decisions that may rely on accurate high copy  
417 number *SMN2* calls when *SMN1* is not detected. We have also begun evaluating  
418 additional endogenous control amplicons to improve accuracy. Finally, we  
419 recognized that amplicon-based methods can be affected by sample-specific SNVs  
420 in primer-binding regions. This risk was reduced by utilizing two different primer sets  
421 across different amplicon sizes to resolve copy number.

422

423 In summary, we present results from the development, optimization, and external  
424 evaluation of a novel PCR/Nanopore assay using over 750 samples, including cell

425 lines, residual presumed normal blood donors, and patient specimens with known  
426 *SMN1* and *SMN2* genotypes. The results reveal accurate detection of multiple  
427 categories of clinically informative variants, including SNVs, indels and CNVs. SNV  
428 phasing was demonstrated through known silent carrier (2+0) and disease-modifier  
429 variants. Importantly, the assay is cost-effective and scalable, showing potential for  
430 broad implementation in diagnostic and screening programs. Last, we note a key  
431 benefit of the assay chemistry, platform and workflow is its flexibility to include  
432 additional, variants in *CFTR*, *FMR1*, and *HBA1/2* and other genes associated with  
433 commonly screened genetic disorders. This extensibility may further expand the  
434 utility of the approach and application and represents an important future direction  
435 for technology development.

436

### 437 **Funding**

438 This work received funding support from Asuragen and Oxford Nanopore  
439 Technologies in the form of reagents and consumables.

440

### 441 **Conflict of Interest**

442 BH, BK, and GL are employees of Bio-Techne with stock and stock options in this  
443 company.

444

### 445 **Acknowledgements**

446 We would like to thank all members of Al Jalila Children's Genomics Center,  
447 Asuragen, and Oxford Nanopore Technologies, specifically Frédérique Lerêteux,  
448 Rita Aoun, and Hannah Lucio, for their valuable input on this work.

449

## 450 **References**

- 451 1. D'Amico A, Mercuri E, Tiziano FD, Bertini E. Spinal muscular atrophy.  
452 Orphanet J Rare Dis 2011;6:71. <https://pubmed.ncbi.nlm.nih.gov/7813012/>  
453
- 454 2. Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, et al.  
455 Identification and characterization of a spinal muscular atrophy-determining gene.  
456 Cell 1995;80:155-65. <https://www.ncbi.nlm.nih.gov/pubmed/7813012/>  
457
- 458 3. Roy N, Mahadevan MS, McLean M, Shutter G, Zahra Yaraghi, Reza Zanjirani  
459 Farahani, et al. The gene for neuronal apoptosis inhibitory protein is partially  
460 deleted in individuals with spinal muscular atrophy. Cell 1995;80:167–78.  
461
- 462 4. Thompson TG, DiDonato CJ, Simard LR, Ingraham SE, Burghes AH, Crawford  
463 TO, et al. A novel cDNA detects homozygous microdeletions in greater than 50%  
464 of type I spinal muscular atrophy patients. Nat Genet 1995;9:56-62.  
465 <https://pubmed.ncbi.nlm.nih.gov/7704025/>  
466
- 467 5. Prior TW, Leach ME, Finanger E. Spinal Muscular Atrophy [Internet]. Adam  
468 MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Mirzaa G, et al., editors.  
469 PubMed. Seattle (WA): University of Washington, Seattle; 1993.  
470 <https://pubmed.ncbi.nlm.nih.gov/20301526/>  
471
- 472 6. Wadman RI, Jansen MD, Stam M, Wijngaarde CA, Curial CAD, Medic J, et al.  
473 Intragenic and structural variation in the SMN locus and clinical variability in

474 spinal muscular atrophy [Internet]. Brain Commun 2020;2:fcaa075

475 <https://pubmed.ncbi.nlm.nih.gov/32954327/>

476

477 7. Chen X, Sanchis-Juan A, French CE, Connell AJ, Delon I, Kingsbury Z, et al.

478 Spinal muscular atrophy diagnosis and carrier screening from genome

479 sequencing data. Genetics in Medicine 2020;22:945–53.

480 <https://www.nature.com/articles/s41436-020-0754-0#ref-CR6>

481

482 8. Lorson CL, Hahnen E, Androphy EJ, Wirth B. A single nucleotide in the SMN

483 gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl

484 Acad Sci 1999;96:6307–11

485 <https://www.pnas.org/content/96/11/6307.short>

486

487 9. Coovert DD, Le TT, McAndrew PE, Strasswimmer J, Crawford TO, Mendell

488 JR, et al. The Survival Motor Neuron Protein in Spinal Muscular Atrophy. Hum

489 Mol Genet 1997;6:1205–14.

490

491 10. Mailman MD, Heinz JW, Papp AC, Snyder PJ, Sedra MS, Wirth B, et al.

492 Molecular analysis of spinal muscular atrophy and modification of the phenotype

493 by SMN2. Genet Med 2002;4:20–6. <https://www.nature.com/articles/gim20024>

494

495 11. Lefebvre S, Burlet P, Liu Q, Bertrand S, Clermont O, Munnich A, et al.

496 Correlation between severity and SMN protein level in spinal muscular atrophy.

497 Nature genetics. 1997;16:265–9.

498 <https://www.ncbi.nlm.nih.gov/pubmed/9207792>

499

500 12. The American College of Obstetricians and Gynecologists. Committee

501 Opinion No. 691. Obstetrics & Gynecology. 2017;129:e41–55.

502

503

504 13. Prior TW, Nagan N, Sugarman EA, Batish SD, Braastad C. Technical

505 standards and guidelines for spinal muscular atrophy testing. Genet Med

506 2011;13:686–94.

507

508 14. Campbell L, Potter A, Ignatius J, Dubowitz V, Davies K. Genomic Variation

509 and Gene Conversion in Spinal Muscular Atrophy: Implications for Disease

510 Process and Clinical Phenotype. The American Journal of Human Genetics

511 1997;61:40–50.

512

513 15. Luo M, Liu L, Peter I, Zhu J, Scott SA, Zhao G, et al. An Ashkenazi Jewish

514 SMN1 haplotype specific to duplication alleles improves pan-ethnic carrier

515 screening for spinal muscular atrophy. Genet Med 2013;16:149–56.

516

517 16. Abou Tayoun AN, Burchard PR, Malik I, Scherer A, Tsongalis GJ.

518 Democratizing Molecular Diagnostics for the Developing World. American Journal

519 of Clinical Pathology. 2014;141:17–24.

520 <https://academic.oup.com/ajcp/article/141/1/17/1766064>

521

- 522 17. Abou Tayoun AN. Unequal global implementation of genomic newborn  
523 screening. *Nature Reviews Genetics* 2023;24:801–2.  
524 <https://pubmed.ncbi.nlm.nih.gov/37723349/>  
525
- 526 18. Sa'adah N, Harahap NIF, Nurputra DK, Rochmah MA, Morikawa S,  
527 Nishimura N, et al. A Rapid, Accurate and Simple Screening Method for Spinal  
528 Muscular Atrophy: High-Resolution Melting Analysis Using Dried Blood Spots on  
529 Filter Paper. *Clinical Laboratory* 2015;61:575–80.  
530 <https://pubmed.ncbi.nlm.nih.gov/26118191/>  
531
- 532 19. Li S, Han X, Xu Y, Chang C, Gao L, Li J, et al. Comprehensive analysis of  
533 spinal muscular atrophy (CASMA): SMN1 copy number, intragenic mutation and  
534 2 + 0 carrier analysis by third-generation sequencing. *J Mol Diagnostics* 2022;  
535 24:1009–20.  
536
- 537 20. Chen X, Harting J, Farrow EG, Thiffault I, Kasperaviciute D, Hoischen A, et  
538 al. Comprehensive SMN1 and SMN2 profiling for spinal muscular atrophy  
539 analysis using long-read PacBio HiFi sequencing. *Am J of Hum Genetics*  
540 2023;110:240–50.  
541
- 542 21. Budimirovic DB, Schlageter A, Filipovic-Sadic S, Protic DD, Bram E, Mahone  
543 EM, et al. A Genotype-Phenotype Study of High-Resolution FMR1 Nucleic Acid  
544 and Protein Analyses in Fragile X Patients with Neurobehavioral Assessments.  
545 *Brain Sci* 2020;10:694.  
546 <https://www.mdpi.com/2076-3425/10/10/694/htm>

547

548 22. Naofal ME, Ramaswamy S, Alsarhan A, Nugud A, Sarfraz F, Janbaz H, et al.

549 The genomic landscape of rare disorders in the Middle East. *Genome Med*

550 2023;15:5.

551

552 23. Milligan JN, Larson JL, Filipovic-Sadic S, Laosinchai-Wolf W, Huang Y-W, Ko

553 T-M, et al. Multisite Evaluation and Validation of a Sensitive Diagnostic and

554 Screening System for Spinal Muscular Atrophy that Reports SMN1 and SMN2

555 Copy Number, along with Disease Modifier and Gene Duplication Variants. *J Mol*

556 *Diagn* 2021;23:753–64.

557 <https://pubmed.ncbi.nlm.nih.gov/33798739/>

558

559 24. Zheng Z, Li S, Su J, Leung AW, Lam T, Luo R. Symphonizing pileup and full-

560 alignment for deep learning-based long-read variant calling. *Nat Comput Sci*

561 2022;2:797-803.

562

563 25. Vezain M, Saugier-veber P, Goïna E, Touraine R, Manel V, Toutain A, et al.

564 A rare SMN2 variant in a previously unrecognized composite splicing regulatory

565 element induces exon 7 inclusion and reduces the clinical severity of spinal

566 muscular atrophy. *Human Mutation* 2010;31:E1110–25.

567

568 26. Ruhno C, McGovern VL, Avenarius MR, Snyder PJ, Prior TW, Nery FC, et al.

569 Complete sequencing of the SMN2 gene in SMA patients detects SMN gene

570 deletion junctions and variants in SMN2 that modify the SMA phenotype. *Human*



571 genetics 2019;138:241–56.

572 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6503527/>

573

574 27. Finkel RS, Mercuri E, Darras BT, Connolly AM, Kuntz NL, Kirschner J, et al.

575 Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. N

576 Engl J Med 2017;377:1723–32.

577 <https://www.ncbi.nlm.nih.gov/pubmed/29091570>

578

579 28. Pechmann A, Langer T, Schorling D, Stein S, Vogt S, Schara U, et al.

580 Evaluation of Children with SMA Type 1 Under Treatment with Nusinersen within

581 the Expanded Access Program in Germany. J Neuromuscul Dis 2018;5:135–43.

582 <https://www.ncbi.nlm.nih.gov/pubmed/29689734>

583

584

585

586

587

588

589

590

591

592

593 **Table 1.** PCR/Nanopore assay performance based on concordance data using  
594 clinical samples with known SMN1/2 copy number status.

595

<b>COPY NUMBER</b>	<b>NUMBER OF CLINICAL BLOOD SAMPLES (BOTH DESIGNS)</b>
--------------------	--

---

STATE	SMN1 (N=70)	SMN2 (N = 67)
0	30/30	7/7
1	8/8	7/7
2	22/22	37/39
≥ 3	10/10	14/14

596

597

598

### 599 **Figure Legends**

600 **Figure 1.** The prototype PCR/Nanopore Assay design and workflow involves two  
601 separate PCR reactions that amplify either 2.7kb or 11.2kb amplicon products.  
602 Amplicons are barcode tagged for each sample in a second PCR reaction, then  
603 pooled by mass into a single sequencing library. The library is prepared by  
604 appending nanopore-specific adaptors and loaded into a MinION flowcell for  
605 sequencing on a Mk1B connected to a computer running MinKNOW software  
606 (Oxford Nanopore Technologies). Data are analyzed with off-the-shelf and assay-  
607 specific software.

608 **Figure 2.** Study design for assay development consisted of three phases.

609 **Figure 3.** Performance of the assay to identify copy number across cell lines or  
610 presumed normal whole blood samples within the optimization data set for SMN1  
611 (orange) or SMN2 (blue) utilizing the 2.7kb fragment or both designs where call  
612 threshold flagging would prioritize the copy number call from the 11.2kb amplicon.  
613 Genotype agreement was >97% for both SMN1 and SMN2 using both designs, and  
614 100% for the limited number of 0 and 1 copy SMN1 samples in this data set.

615 **Figure 4.** Confusion matrix comparing the expected and predicted SNVs associated  
616 with silent carrier or disease modifier risk. The algorithm accurately identified each  
617 variant type except for two. †Inspection of the underlying trace data for these two  
618 samples in the comparator assay revealed a profile consistent with false-positive  
619 calls and a defined root cause in each case. This correction would result in 100%  
620 agreement across each variant for the PCR/Nanopore assay.

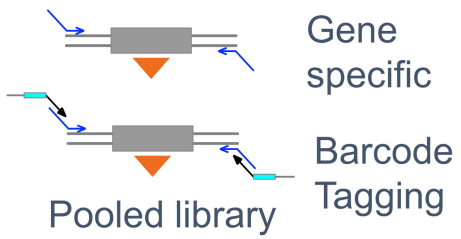
621

**SMN1/2**

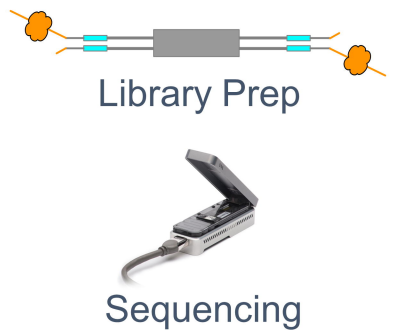
2.7kb Amplicon  
Exon 7-8

11.2kb Amplicon  
Exon 3-8

PCR Enrichment



Any-Length Sequencing



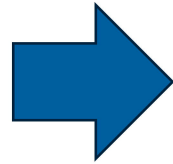
Software



## Prototyping Phase

79 Cell Lines  
352 Whole Blood

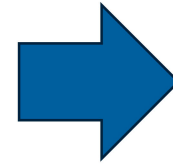
Combine multiplex primer design, reagents, algorithms, and protocol into a proof-of-concept assay with known samples across a collection of diverse *SMN1* and *SMN2* copy number samples.



## Optimization Phase

97 Cell Lines (79 + 18 new)  
227 new Whole Blood

Optimize performance on a partially independent set of samples that exemplifies *SMN1* and *SMN2* copy numbers associated with screening and diagnostics, across operators, instruments, and input quantities.

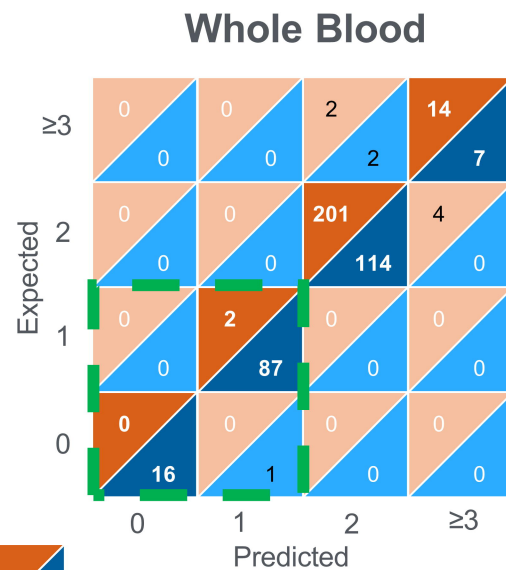
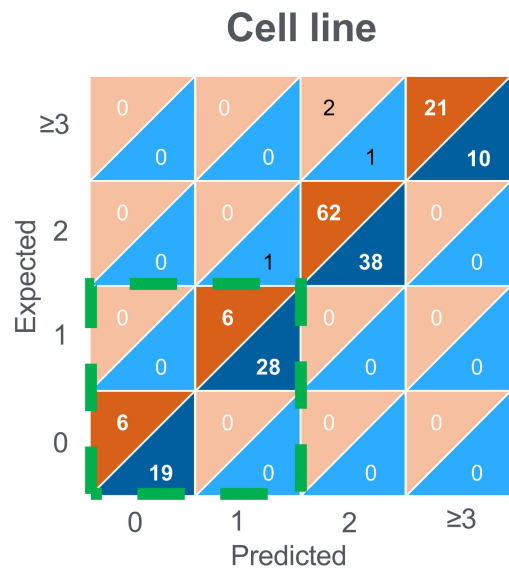


## Evaluation Phase

70 Clinical Samples

Evaluate assay performance across an independent clinical cohort of patients from 15 primarily Middle East and North Africa countries.

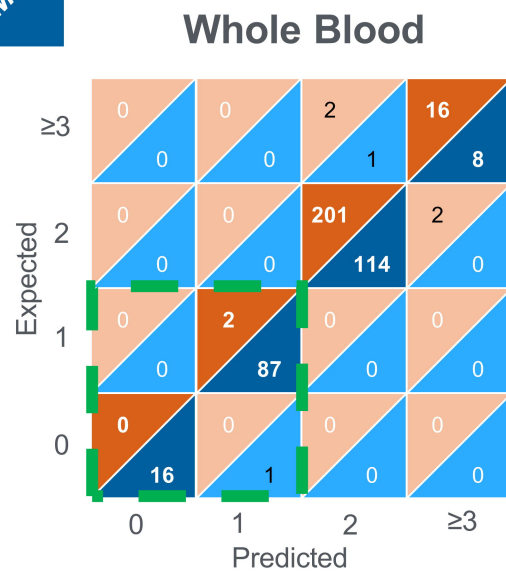
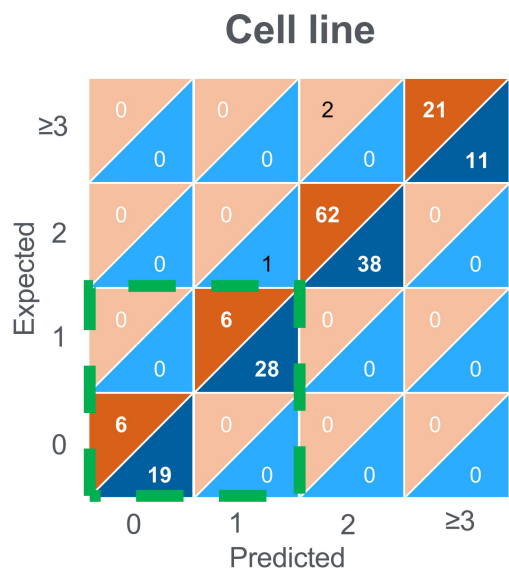
2.7kb design



### Genotype Agreement

	SMN1	SMN2	SMN1 0 or 1 cp
Cell Line	95/97 (97.9%)	95/97 (97.9%)	12/12 (100%)
Whole Blood	218/227 (96.0%)	224/227 (98.7%)	2/2 (100%)

2.7kb & 11.2kb design



### Genotype Agreement

	SMN1	SMN2	SMN1 0 or 1 cp
Cell Line	95/97 (97.9%)	96/97 (99.0%)	12/12 (100%)
Whole Blood	221/227 (97.3%)	224/227 (98.7%)	2/2 (100%)

**c.\*3+80T>G**

<b>Expected</b>	Positive	0	21
	Negative	303	0
		Negative	Positive
		<b>Predicted</b>	

**c.\*211\_212del**

<b>Expected</b>	Positive	2†	20
	Negative	302	0
		Negative	Positive
		<b>Predicted</b>	

**c.859G>C**

<b>Expected</b>	Positive	0	3
	Negative	321	0
		Negative	Positive
		<b>Predicted</b>	