The Journal of Molecular Diagnostics, Vol. 21, No. 4, July 2019

the Journal of Molecular Diagnostics

jmd.amjpathol.org

Development of a Novel Next-Generation Sequencing Assay for Carrier Screening in Old Order Amish and Mennonite Populations of Pennsylvania

Erin L. Crowgey,* Michael C. Washburn,† E. Anders Kolb,* and Erik G. Puffenberger‡

From Nemours Biomedical Research Department, Nemours Alfred I. duPont Hospital for Children, Wilmington, Delaware; Assay Development,† ArcherDX Inc., Boulder, Colorado; and the Clinical Laboratory,‡ Clinic for Special Children, Strasburg, Pennsylvania*

Accepted for publication March 13, 2019.

Address correspondence to Erik G. Puffenberger, Ph.D., Clinic for Special Children, 535 Bunker Hill Rd., Strasburg, PA 17579. E-mail: epuffenberger@clinicforspecialchildren.org.

Genetically isolated populations, such as the Old Order Amish and Old Order Mennonite communities, have an increased incidence of specific autosomal recessive disorders caused by the founder effect. In these populations, robust expanded carrier screening and diagnostic testing have the potential to reduce overall medical costs and improve patient outcomes. A novel next-generation sequencing assay was developed using anchored multiplex PCR technology (ArcherDX) for 162 different genetic syndromes caused by 202 pathogenic variants consisting of 150 single-nucleotide changes, 43 small insertion/deletions, and 9 large deletions (>20 nucleotides). To assess the accuracy of the screening panel results, 48 samples were selected on the basis of prior whole exome sequencing results. An additional 15 samples were chosen specifically to validate *SMN1* and *SMN2* copy number analyses. Collectively, the screening panel detected 273 pathogenic single-nucleotide or small insertion/deletion variants, 35 copy number variations, and 1 chromosomal abnormality (Klinefelter syndrome). Concordance with prior whole exome sequencing was 100%. By using a novel next-generation sequencing workflow, a successful targeted gene variant panel was developed for the Old Order Amish and Old Order Mennonite populations of Lancaster County, Pennsylvania. Population-wide carrier screening may help decrease the morbidity and mortality of these conditions in the high-risk populations. *(J Mol Diagn 2019, 21: 687–694; https://doi.org/10.1016/j.jmoldx.2019.03.004)*

The Clinic for Special Children (CSC; Strasburg, PA) was founded in 1989 as a medical home for uninsured Old Order Amish and Old Order Mennonite (Plain) children who experience genetic disease. The mission was to provide accessible and affordable health care in a local setting to ensure improved outcomes for children with complex genetic conditions, such as glutaric aciduria type I and maple syrup urine disease (MSUD). By bringing specialized biochemical and genetic knowledge into a local pediatric service, the clinic lowered health care costs for the community and improved patient outcomes. From the beginning, the clinic invested in an on-site Clinical Laboratory Improvement Amendments–certified analytical laboratory to facilitate research and improve patient care.

Initially, the CSC laboratory provided analytical testing using high-performance liquid chromatography and gas chromatography–mass spectrometry. Laboratory testing protocols were patient centered and designed to address the needs of the local patient population. As such, the laboratory routinely performed amino acid analysis and organic acid analysis and developed a novel assay for measuring amino acids from dried filter paper blood spots, obviating the need for weekly or monthly office visits for venipuncture. In 1998, the clinic hired a geneticist (E.G.P.) to expand the capabilities of the laboratory. At the time, the Human Genome Project was providing unprecedented knowledge and access to genetic data that could be harnessed to improve patient care. The laboratory embarked on a mission to identify the molecular

Supported by the Nemours Center for Cancer and Blood Disorders and the Leukemia Research Foundation of Delaware (E.A.K.).

Disclosures: M.C.W. is an employee of ArcherDx Inc.; however, there are no personal or financial benefits to be gained from this work and publication.

Copyright © 2019 American Society for Investigative Pathology and the Association for Molecular Pathology. Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (*http://creativecommons.org/licenses/by-nc-nd/4.0*).
https://doi.org/10.1016/j.jmoldx.2019.03.004

basis of disease in all clinic patients. Initially, these studies focused on known disease-gene combinations, but they later expanded to include genetic mapping studies and eventually whole exome sequencing. Over the clinic's history, the CSC has identified and published original research articles on 25 novel disease genes.[1–5] In the early days, diagnostic and carrier testing was performed via Sanger sequencing. Although it is considered to be the gold standard, Sanger sequencing is not amenable to multiplexing for detection of multiple variants in one assay. Currently, the CSC has cataloged >270 different molecular lesions that cause genetic disease in the Plain populations of Lancaster County, Pennsylvania. To provide faster and cheaper diagnoses, the laboratory transitioned to high-resolution melt analysis with an unlabeled probe.[6] This allowed providing rapid diagnoses in 2 hours for a modest cost ($50). However, neither technology permits efficient multiplexing or parallel testing. To improve our service, a method was needed to test an individual for all known pathogenic variants in a single assay.

Over the past decade, DNA sequencing techniques have significantly improved to enable the massively parallel sequencing of many DNA molecules simultaneously. These techniques are often referred to as next-generation sequencing, or NGS, and are quickly advancing our knowledge of genetic variation in humans. Although NGS is a powerful technique, there are still inherent limitations with these technologies in their ability to detect chromosomal abnormalities, structural variants, and copy number variations (CNVs) within a single assay.

The CSC serves two large founder populations in southeastern Pennsylvania, the Lancaster Old Order Amish and the Old Order Mennonites (composed primarily of the Weaverland, Groffdale, and Stauffer groups). Because of the small number of founders and a phenomenon called random genetic drift, these Plain populations exhibit relatively high carrier rates for a small subset of genetic diseases compared with the general population. As such, they serve as an ideal study population for the introduction of an NGS carrier screening assay to help minimize the morbidity and mortality associated with their unique genetic risk factors. A multiplexed NGS carrier screening assay has the potential to impact patient care because at-risk couples can be identified and informed of their carrier status before affected children are born. The precious resources of the community can then be used to perform diagnostic testing on high-risk newborns. These interventions then provide an unusual opportunity for presymptomatic treatment, an especially tantalizing prospect in the age of gene therapy.

Because of the need to detect a panel of known pathogenic variants in a single assay, anchored-multiplex PCR technology was coupled with its use of a unique molecular index, to simultaneously target numerous classes of variants with a wide range of allele frequencies. A multidisciplinary team was generated, across patient care sites (Nemours Alfred I. duPont Hospital for Children and Clinic for Special Children) and industry (ArcherDX Inc.) to design and generate a rapid and economical carrier screening assay for the Plain populations of southeastern Pennsylvania.

Materials and Methods

Study Protocol and Subject Enrollment

Sixty-three subjects from the Clinic for Special Children were consented using an approved institutional review board clinical registry and specimen banking protocol. Blood was collected, and DNA was extracted from 1 to 2 mL of blood using a QuickGene-610L Nucleic Acid Isolation System (Autogen, Holliston, MA), according to the manufacturer's instructions. At this time, this targeted gene

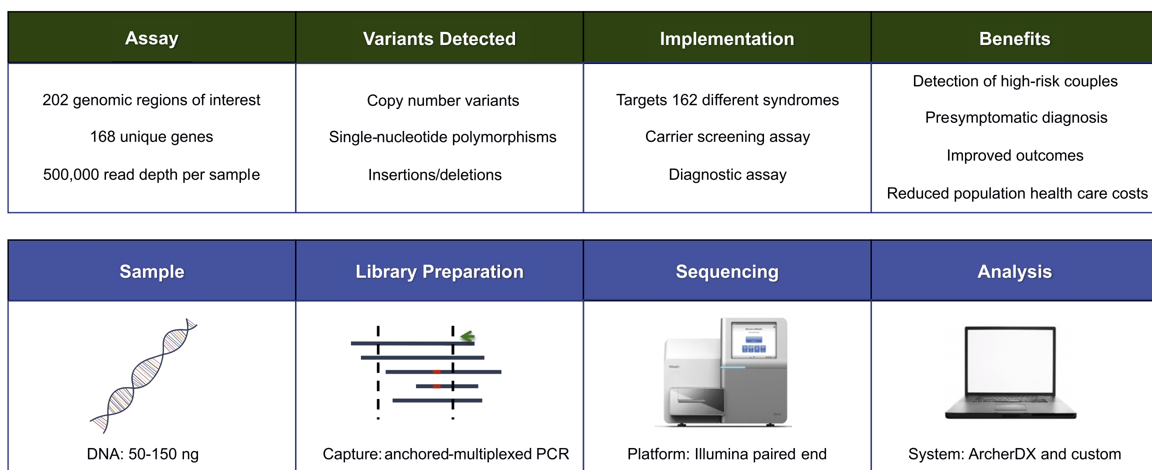


Figure 1 Overview Custom Gene Panel assay for the Lancaster Amish and Mennonite Communities. **A:** A custom anchored-multiplexed PCR (AMP) assay was developed to target 202 genomic regions of interest across 168 unique genes. Collectively, 162 different syndromes can be screened. **B:** The next-generation sequencing (NGS) assay requires DNA as input and uses gene-specific primers (GSPs) as the library preparation method. Library preparation leveraged AMP with a single GSP (**green arrow**). True mutations (**red regions**) are present on different molecules captured and enable robust variant detection. Sequencing is performed on an Illumina platform, and raw NGS data are processed via comprehensive bioinformatic workflows.

panel has only been validated using blood; however, the library preparation technique has been demonstrated to work with dried filter paper and saliva.

Library Preparation and Next-Generation Sequencing

A custom NGS gene panel kit was developed using anchored-multiplex PCR technology (ArcherDx Inc., Boulder, CO). A total of 168 unique genes were targeted, with a focus on 202 alleles associated with 162 different syndromes (Supplemental Table S1). Additional primers were added for sex identification, two gene-specific primer (GSP) pairs that bind on regions that are the same for both X and Y chromosomes; however, on amplification, unique sequences can differentiate them. For male samples, each call on the X chromosome should be homozygous alternative or reference. If a male sequence is not produced, the sample is designated as female. An additional 29 single-nucleotide polymorphisms (SNPs) are also sequenced for sample tracking and identification, but they are not of interest in downstream analysis. A primer uniformity of ≥0.95 was used as a quality threshold for library preparation and complexity. The first 48 samples analyzed with the custom panel did not include the *SMN2* primers. Library preparation was performed following the manufacturer's protocol. Briefly, 50 ng of DNA, quantified by Qubit (Thermo Fischer Scientific, Waltham, MA), was used as input for library preparation for each sample. Libraries were prepared using the Archer VariantPlex protocol (ArcherDx Inc.) for Illumina (San Diego, CA), with 18 cycles for PCR1 and 20 cycles for PCR2, with both using a 3-minute extension at 65°C.

All purifications during library preparation were performed with Agencourt AMPure XP (catalog number A63882; Beckman Coulter, Pasadena, CA). Libraries were quantified using the KAPA Library Quantification Kit (catalog number KK4824; KAPA Biosystems, Wilmington, MA) and pooled to an equimolar concentration (10 nmol/L). The concentration of the final pool of libraries was confirmed with the KAPA Library Quantification Kit. Libraries were sequenced on the Illumina NextSeq 500 platform with 2 × 150-bp sequencing using the NextSeq 500/550 Mid Output Kit version 2 (300 cycles). All libraries were loaded onto the

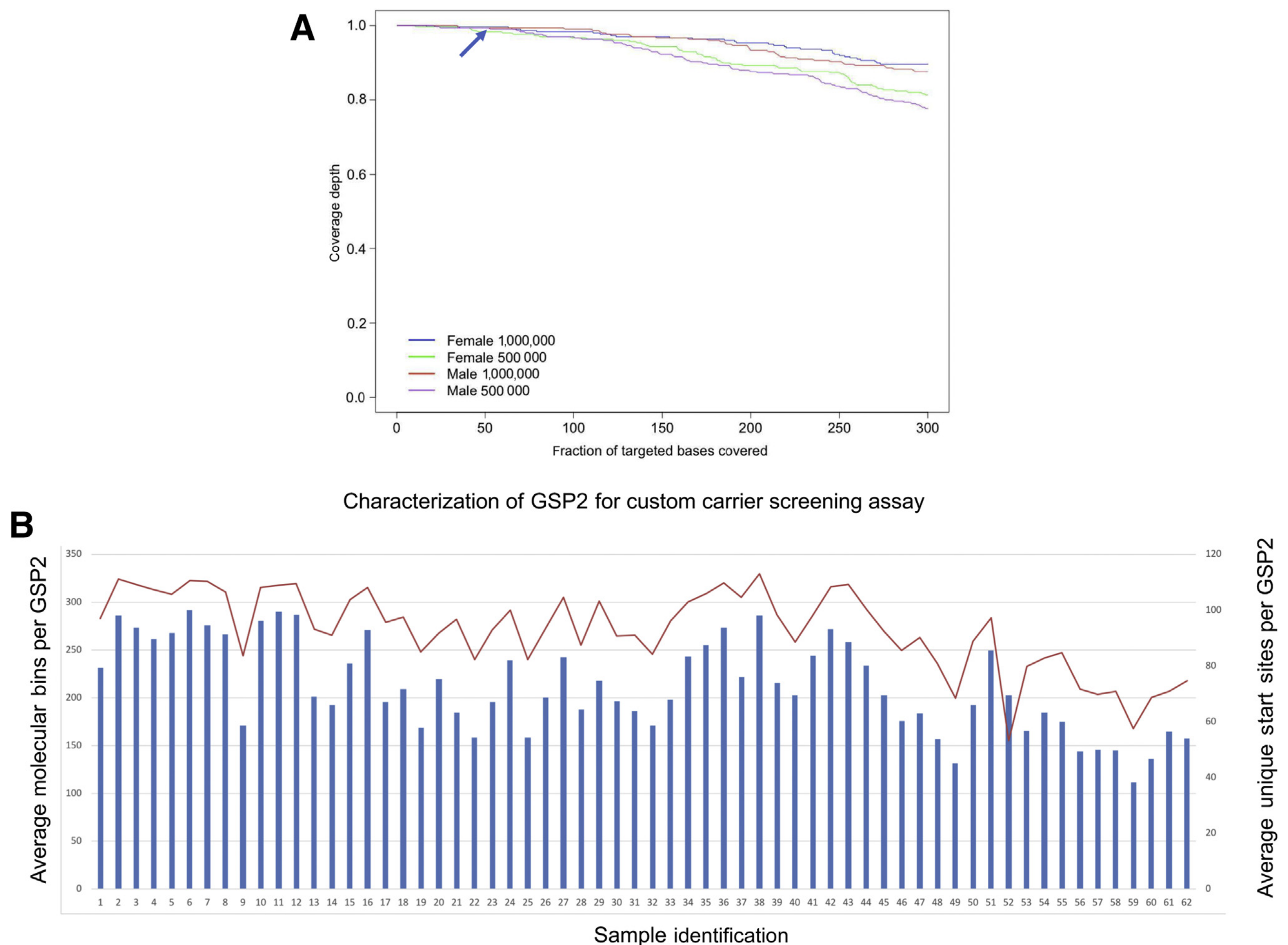
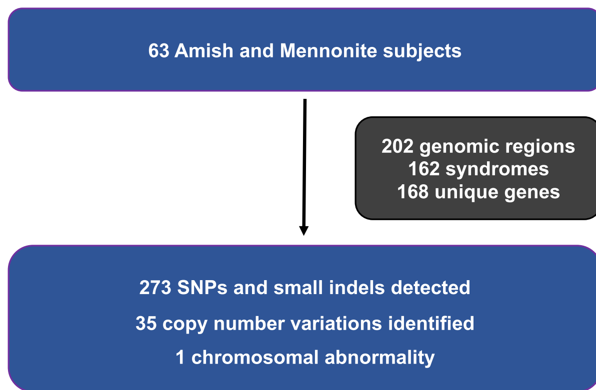


Figure 2 Analysis of assay performance. **A:** Genome-in-a-Bottle samples were sequenced at 500,000 (**blue** and **purple lines**) and 1,000,000 (**red** and **green lines**) read depths, and coverage depth per bases sequenced was plotted. **Blue arrow** indicates sequencing depth requirements for single-nucleotide polymorphisms. **B:** Characterization of gene-specific primers (GSPs) for custom carrier screening assay. For each sample, per GSP2 primer across all genes targeted, the average number of molecular bins per GSP2 (**blue bars**) and average unique start sites per GSP2 (**red line**) were calculated per subject sequenced.



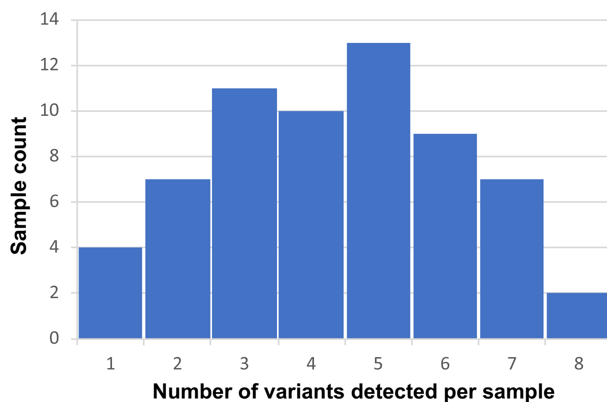


Figure 3 Summary of variants identified in the Amish and Mennonite cohort screened. **A:** Collectively, across all 63 subjects analyzed, 273 single-nucleotide polymorphisms (SNPs) and small insertions/deletions (indels), 35 copy number variants, and 1 chromosomal abnormality were detected. **B:** Distribution of single-nucleotide polymorphisms and small insertions and deletions for the 63 subjects analyzed.

sequenced 1.8 pmol/L with a target PhiX concentration of 20%. The assay was initially designed to detect CNVs in five genes (Supplemental Table S1) and was tested on an initial cohort of 48 samples as well as 2 controls: Genome-in-a-Bottle samples [NA12877 (male) and NA12878 (female); Coriell Institute, Camden, NJ]. An additional module was added later to specifically detect copy number variation in *SMN1* and *SMN2*, and the remaining 16 subjects were sequenced with this updated panel. All FASTQ data were deposited in the European Nucleotide Archive (*https://www.ebi.ac.uk/ena*, accession number PRJEB31811; Carrier Screening Assay for Plain Populations).

Data Analysis

The Amazon Cloud was used for processing the data via the ArcherDX analysis platform version 5.1.3 (Archer DX Inc.), and the process included FASTQ trimming, read deduplication, genome alignment, and variant detection and annotation. The analysis pipeline contains the following applications: an assembly-based realigner (ABRA),[7] bamaddrg, bcftools, bedtools, blast,[8] bowtie2,[9] bwa, EMBOSS, fastqc, freeBayes,[10] Lofreq,[11] Mixcr,[12] Muscle,[13] samtools, variant effect predictor (VEP),[14] Velvet,[15] HTSeq,[16] complete-striped-smith-waterman-library, JBrowse,[17] JQuery DataTables, Django Solo, and plot.js.

FASTQ files are analyzed via fastqc for quality and aligned to the reference genome hg19 using bwa and bowtie2, and alignment files are processed via Genome Analysis Toolkit (GATK) best practices. SNPs and small insertions/deletions (indels) of <25 bp are called using FreeBayes and Lofreq. To aid in detection of variants of interest, the ArcherDX variant caller Vision (ArcherDX Inc.) focused on detecting SNPs and small indels of interest by using a targeted variant call file. This restricted the analysis to the pathogenic variants of interest and excluded most off-target variants in the same region. A minimum base quality of $\geq$20 for variant detection was implemented. SNPs and small indels required a minimum of 30$\times$ coverage for detection; and for CNVs, >50$\times$ coverage was required. For autosomal dominant and recessive SNPs, a threshold of 40% to 100% (0.4 to 1.0 variant allele frequency) was used for calling variants. The de novo variants, CNVs, and indels have a threshold of 0.25 to 1.

CNV filtering was based on the following: i) a *P*-value calculation, based on a two-tailed Wilcoxon rank sum test, with a null hypothesis that the median value of the copy number called for each probe of the given target was equal to the median value of the copy number called for all primers identified as being in the baseline, across all samples analyzed; ii) SD of the called copy number for all probes of the target; iii) total number of adjacent GSPs supporting the CNV event; and iv) variants deemed likely to be a chemistry or sequencing artifact were filtered out. Variants were annotated with ClinVar annotations or an internal database of variants detected in >3000 research exomes, which included calculated allele frequencies for the Lancaster Amish and Mennonite communities. Binary alignment map (BAM) files were visualized for all results, and variants were confirmed via whole exome sequencing and/or Sanger sequencing.

Only variants targeted by this gene panel were analyzed, and incidental variants, including variants of unknown significance, were not analyzed. In the future, the clinical implementation of a targeted gene panel will not report incidental variants. Furthermore, the methyl-CpG-binding protein 2 region does not yield high sequencing coverage because of high GC content, and currently this region is not included as part of the results.

Results

Assessment of the Targeted Gene Panel

A total of 202 genomic regions of interest were targeted using a custom anchored-multiplex PCR and Illumina

Table 1 Results for CNV Read Depth Requirements

Gene	Subject ID	Read depth	Unique fragment	Reads		CNV		Copy no.		*P* value	Call
				Mean	SD of mean	Mean reads	SD of mean	Mean	SD		
SMN1	32,649	5000	3907	7.57	11.65	6.75	5.60	0.54	0.20	0.7277	Below
		10,000	7701	13.14	16.33	13.56	9.92	0.52	0.17	0.1642	Below
		25,000	18,681	29.29	36.61	32.91	23.09	0.54	0.13	0.0214	Below
		50,000	34,459	48.14	57.48	60.69	39.48	0.51	0.16	0.0192	Below
		75,000	49,053	70.43	85.34	87.69	53.23	0.49	0.17	0.0141	Below
		100,000	61,951	83.57	97.19	111.67	64.07	0.52	0.14	0.0087	High quality
ADAMTS10	33,908	5000	3637	5.09	4.16	6.39	5.92	0.74	0.21	0.6805	Below
		10,000	7152	8.64	5.70	12.21	10.12	0.49	0.12	0.0050	High quality
		25,000	16,283	18.73	10.01	28.26	22.18	0.50	0.06	7.484×10^{-5}	High quality
		50,000	29,166	34.18	18.25	52.43	36.75	0.50	0.06	3.204×10^{-5}	High quality
		75,000	39,936	45.64	19.84	72.01	46.65	0.49	0.05	3.928×10^{-6}	High quality
		100,000	48,902	56.73	27.12	89.57	55.72	0.51	0.05	2.189×10^{-5}	High quality
ANAPC7	34,683	5000	3630	2.56	2.40	6.20	5.15	0.70	0.47	0.6785	Below
		10,000	7138	6.22	3.77	12.63	10.37	0.58	0.17	0.3463	Below
		25,000	16,172	13.44	7.49	28.59	20.02	0.61	0.20	0.4996	Below
		50,000	28,745	22.78	11.00	50.95	32.52	0.58	0.15	0.0413	Below
		75,000	38,448	30.78	9.58	69.70	40.71	0.55	0.09	0.0003	High quality
		100,000	47,048	37.56	14.72	86.63	49.66	0.52	0.11	0.0034	High quality
STRADA	34,691	5000	3705	3.40	4.67	6.22	5.59	0.42	0.16	0.0125	Below
		10,000	7379	8.20	8.83	12.41	9.68	0.54	0.14	0.0265	below
		25,000	17,829	19.90	24.39	30.06	22.19	0.45	0.07	0.0002	High quality
		50,000	33,906	40.50	44.91	58.21	39.62	0.46	0.08	0.0001	High quality
		75,000	48,610	55.30	51.26	83.54	53.25	0.45	0.04	6.81×10^{-5}	High quality
		100,000	62,111	74.60	70.91	109.29	68.47	0.48	0.07	5.86×10^{-5}	High quality
SMN1	35,009	5000	3696	4.43	5.16	6.57	5.22	0.43	0.05	0.1538	Below
		10,000	7360	10.43	13.24	12.81	9.98	0.49	0.12	0.1011	Below
		25,000	17,533	28.57	35.57	31.01	21.93	0.56	0.10	0.0409	Below
		50,000	32,662	45.00	50.74	57.56	35.85	0.5753	0.0707	0.0014	High quality
		75,000	46,111	59.29	64.11	82.43	48.28	0.4903	0.1137	0.0053	High quality
		100,000	57,857	70.00	72.71	104.78	57.84	0.5229	0.0952	0.0026	High quality

CNV, copy number variation; ID, identification.

(paired-end) sequencing protocol (ArcherDX Inc.) (Figure 1A). A novel and robust workflow was implemented from DNA isolation to data interpretation (Figure 1B). The sequencing characteristics of the custom kit were assessed using two Genome-in-a-Bottle samples [NA12877 (male) and NA12878 (female)]. The coverage of targeted bases was analyzed (Figure 2A) to determine read depth requirements, and at sequencing depths of 500,000, the samples had >95% of bases covered at a depth of 60×. The 500,000 read depths had similar coverage as the 1,000,000 at 50×, 100×, and 150×; however, at read depth coverage >200, the number of bases covered at those depths declined compared with the 1,000,000 depth of coverage.

The library technique used a unique molecular index approach.[18,19] During library preparation, each DNA fragment was tagged with a unique molecular index, enabling error-corrected sequencing and the calculation of unique molecules per targeted locus. The custom assay was characterized for each GSP by evaluating the number of total molecules mapping to the target, the number of unique molecules mapping to the target, and primer uniformity. For each subject sequenced, across all GSP2 primers, the average number of DNA molecules per GSP2 and average unique start sites per GSP2, or rather unique molecules, are plotted in Figure 2B. The average molecular bins per GSP2 (Figure 2B) indicates the average number of fragments in which the aligner was able to detect a DNA molecule per GSP2 per sample, and a range of 111.65 to 291.05 was displayed across the 63 subjects, indicating high-complexity libraries. The average unique start sites per GSP2 (Figure 2B) indicates the average number of start sites at which the aligner was able to detect a DNA molecule per GSP2 per sample, and a range of 53.05 to 112.85 was displayed across the 63 subjects, indicating high-complexity libraries. The goal is to have a high-complexity library as it enhances variant detection.

Primer uniformity, as measured by the ratio of the number of unique molecules generated for each strand-specific GSP primer per locus, was also analyzed, and uniform amplification across all samples and primers was detected in both sequencing directions (>0.96 for all samples). The reproducibility of the assay performance was measured using samples NA12877 (male) and NA12878 (female). The number of reads per GSP between technical replicates is plotted in Supplemental Figure S1, and both samples had a Pearson correlation

coefficient ≥0.94. The number of reads per GSP was also plotted between two different runs and two different samples (Supplemental Figure S1) and had a Pearson correlation coefficient = 0.96. The distribution of reads per GSP across six samples was also plotted (Supplemental Figure S1), and no significant differences were identified. Furthermore, both of the control samples have known and published polymorphisms (Supplemental Table S2), and when compared with regions targeted on this panel, eight polymorphisms across the two samples were identified. All eight polymorphisms were identified via the presented panel (Supplemental Table S2), no additional variants were identified, and no false positives or false negatives were detected across the samples.

Variant Analysis

In total, 63 subjects were sequenced via the custom gene panel, focusing on the top 202 variant alleles associated with 162 syndromes within the Lancaster Amish and Mennonite communities (Figure 3A). In total, 309 variants were detected (Supplemental Table S3): 273 pathogenic SNPs and small indels, 35 CNVs, of which 33 were associated with a known disease (Table 1), and 1 chromosomal abnormality (Figure 3B).

Interestingly, the assay detected a CNV gain in *CLCNKB*, which is a variant of unknown significance, even though the assay was developed to target Bartter syndrome (a CNV loss in *CLCNKB*). An XXY chromosomal abnormality, Klinefelter syndrome, was detected in one of the subjects (Figure 3B). All variants identified were validated via an alternative method. The distribution of SNPs and indels per subject was plotted (Figure 3C) and indicates a normal distribution of variants called, with an average of five per subject.

An individual subject is plotted in Figure 4, and highlights the ability to detect different types of variants, CNVs (Figure 4), and SNPs (Table 2), in a single assay in a single subject. The subject highlighted had two CNVs (Figure 4), a copy number loss of *SMN1* and a copy number gain (n = 3) of *SMN2* (Table 2). Furthermore, the phenotype, spinal muscular atrophy, was validated by our clinical collaborators. Of interest, this subject was also genotyped as heterozygous for both *ATM* (c.1229T>C) and *MKKS* (c.724G>T) variants, indicating a carrier status for both variant ataxia-telangiectasia (atypical) and McKusick-Kauffman syndrome, respectively.

To determine the threshold for calling CNVs, five subjects with confirmed CNVs were subsampled to read depths of 5000, 10,000, 25,000, 50,000, 75,000, and 100,000 (the original read depth was 500,000), to determine the read depth threshold for accurately calling a CNV. All of the CNVs tested, *SMN1*, *ADAMTS10*, *ANAPC7*, and *STRADA*, were accurately called at a read depth of 100,000 (Table 3). *ADAMTS10* was the only CNV accurately called at a read depth of 10,000, and *STRADA* threshold for accuracy was 25,000 read depth (Table 3).

Discussion

Using a novel NGS approach, a custom gene panel to screen 202 known pathogenic and/or actionable variants was developed for the southeastern Pennsylvania Plain populations. This screening assay is capable of determining carrier status for 168 syndromes, which have significant health implications for future generations, family planning, and community public health. In the case of MSUD, diagnosis and initiation of therapy for newborns within the first day or two of life, before the results of standard newborn screening are typically available, lead to prevention of acute intoxication in the newborn period and accompanying decreased health care costs.[20] Indeed, there are many disorders present in the Plain populations (eg, Crigler-Najjar syndrome, glutaric aciduria, familial hypercholanemia, spinal muscular atrophy, and 3-β-hydroxysteroid dehydrogenase deficiency), in whom detection of asymptomatic infants is beneficial. Although some conditions are detected via Pennsylvania newborn screening, most disorders on our panel require molecular genetic testing.

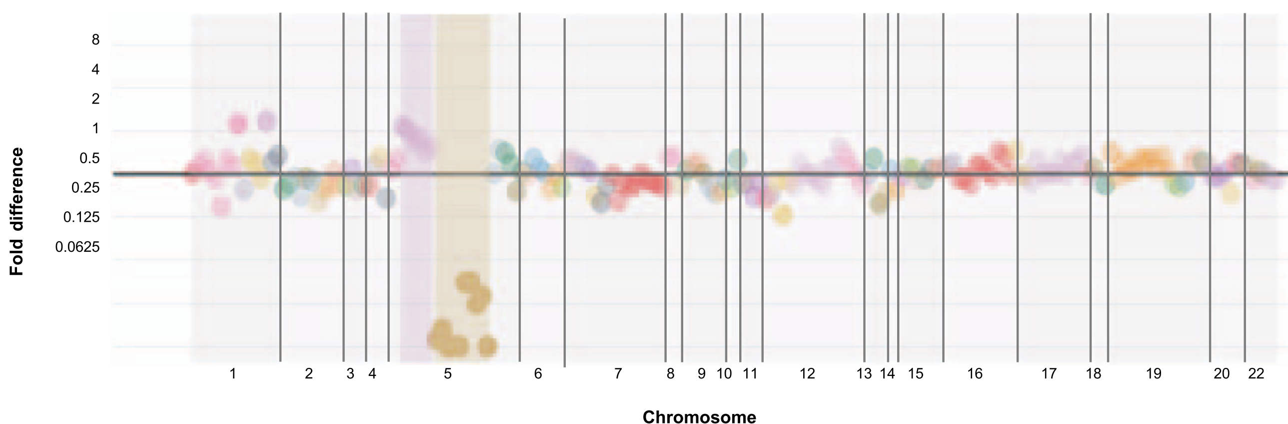


Figure 4 Single patient overview and summary. Copy number variation (CNV) analysis. The subject analyzed has a CNV loss of *SMN1* and a CNV gain of *SMN2*.

Table 2 Summary CNV Analysis

Gene	CNV description	No. of samples
CLCNKB	Loss	2
SMN1	Loss	17
ANAPC7	Loss	1
ADAMTS10	Loss	1
X chromosome	Gain	1
STRADA	Loss	1
SMN2	Loss	3
SMN2	Gain	7

CNV, copy number variation.

For other genetic conditions, novel therapies are forthcoming or currently available, which promise to forever change the perception and treatment of these disorders. Historically, spinal muscular atrophy type 1 has been clinically diagnosed after the patient becomes symptomatic. New interventions, such as small molecules or gene therapies, have, for the first time, led to meaningful changes in the natural history of this disease.[21,22] Identification of the asymptomatic patient will be crucial when implementing these new therapies. For disorders such as spinal muscular atrophy, the new therapies can halt the progression of the disease, but evidence so far does not indicate reversal or restoration of normal motor neuron function, so detection of the asymptomatic infant is critically important.

The Plain populations of Pennsylvania have historically used carrier testing in a unique way. Unlike other founder groups, such as the Ashkenazi Jews,[6] Amish and Mennonite communities have traditionally pursued carrier testing infrequently and use the resulting information differently. Nearly all carrier testing performed at our center is elective and is requested after a couple is married. There is no attempt to use carrier testing as a means to prevent two carriers from marrying or having children. In this community, carrier testing permits anticipatory health care to prevent the devastating effects of genetic diseases that can affect the newborn.

The Mennonite community is especially receptive to carrier screening for MSUD. An affected newborn can become encephalopathic in the first few days of life. Without rapid diagnosis and implementation of therapy, these children will end up in the intensive care unit and endure costly and protracted efforts to reduce plasma leucine levels to normal. Indeed, our carrier testing efforts in this community have not entirely eliminated this problem. Over the past 10 years, 26 MSUD children were born to Mennonite couples. Because of carrier testing and prior MSUD children born into a family, 19 affected children were diagnosed in the first 24 hours of life, and therapy was instituted promptly. None of these children was hospitalized in the newborn period. Unfortunately, seven MSUD babies were born to parents who were unaware that they were both carriers. All seven were symptomatic and hospitalized in the newborn period because of late diagnosis.

These late diagnoses affect patients, families, and the broader Mennonite community. For example, a subject was diagnosed at the Clinic for Special Children on day of life 5 with a leucine level of 54 mg/dL (normal, 1.5 to 3.5 mg/dL). The patient was promptly admitted to Lancaster General Hospital (Lancaster, PA) for emergent treatment to lower branched-chain amino acid levels. Two days later, the patient was transferred to another institution because of critical cerebral edema. Once this subject was stabilized and breathing without assistance, he was transferred back to Lancaster General Hospital and eventually discharged on day of life 17. Even with a significant self-pay discount, medical expenses topped $110,000. In extraordinary cases, the community will raise funds to alleviate the economic burden of a family, but the neonatal encephalopathy that precipitated the hospitalization can have lasting effects. In this case, the subject has a significant behavioral and mood disorder, a finding common in MSUD patients who were symptomatic neonates.[6]

Implementation of a community-wide carrier screening program would ultimately serve multiple purposes. First, by identifying at-risk couples and performing diagnostic testing on cord blood from their children, it can be ensured that most children affected by a recessive genetic disease in these communities are identified as asymptomatic newborns. For diseases like MSUD, this early detection decreases morbidity and mortality. Second, early diagnoses obviate lengthy and costly hospitalizations, potentially saving these uninsured communities millions of health care dollars. As the MSUD case illustrated above, detection of all affected children at birth could save the community approximately $90,000 per year in avoidable hospital bills with respect to MSUD alone. If these savings are extrapolated to other disorders in which late diagnoses can lead to newborn hospitalizations, the community savings could approach several hundred thousand dollars per year. Third, diagnosis of asymptomatic newborns ensures optimal timing for new therapeutics, such as gene therapy, which are rapidly entering the clinical domain.[21,23]

With a single assay, which costs $115 for library preparation and $50 in sequencing costs (Illumina paired-end, 500,000 reads per sample, batched sequencing), 168

Table 3 Overview of a Single Subject's Results

Gene	Variant	Copy no.	Clinical impact	Phenotype
SMN1	Gene loss	0	Affected	Spinal muscular atrophy
SMN2	Gene gain	3		
ATM	c.1229T>C	1	Heterozygote	Variant ataxia-telangiectasia (atypical)
MKKS	c.724G>T	1	Heterozygote	McKusick-Kauffman syndrome

syndromes can be simultaneously screened. The costs mentioned previously do not include personnel time, including medical sign off. The assay presented herein is economical and allows for parallel detection of multiple types of pathogenic changes, including single-nucleotide, insertion/deletion, and copy number variants. Implementing a screening assay within the Plain population will require extensive community outreach efforts and robust training and education for both health care providers and patients. When effectively implemented, these efforts will have a huge impact on decreasing overall medical costs, while improving patient outcomes. Ultimately, it is essential that bioinformaticians, health care providers, wet-bench laboratory experts, and health informaticians work together to generate an effective screening assay.

Acknowledgments

We thank Drs. Kevin Strauss, Michael Fox, and Matthew Demczko for manuscript editing and scientific discussions; and the Nemours Biomedical Research Department for supporting E.L.C.

E.L.C. was responsible for data analysis, writing, and coordinating the manufacturing of the panel; M.C.W. was responsible for leading the manufacturing efforts for the targeted gene panel; E.A.K. was responsible for the funding, helping to develop the concept for the panel, and writing of the manuscript; E.G.P. is the guarantor of this work and had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.jmoldx.2019.03.004*.

References

1. Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, Stephan DA, Morton DH: Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. N Engl J Med 2006, 354:1370−1377
2. Puffenberger EG, Strauss KA, Ramsey KE, Craig DW, Stephan DA, Robinson DL, Hendrickson CL, Gottlieb S, Ramsay DA, Siu VM, Heuer GG, Crino PB, Morton DH: Polyhydramnios, megalencephaly and symptomatic epilepsy caused by a homozygous 7-kilobase deletion in LYK5. Brain 2007, 130(Pt 7):1929−1941
3. Puffenberger EG, Jinks RN, Sougnez C, Cibulskis K, Willert RA, Achilly NP, Caddisy RP, Fiorentini CJ, Heiken KF, Lawrence JJ, Mahoney MH, Miller CJ, Nair DT, Politi KA, Worcester KN, Setton RA, Dipiazza R, Sherman EA, Eastman JT, Francklyn C, Robey-Bond S, Rider NL, Gabriel S, Morton DH, Strauss KA: Genetic mapping and exome sequencing identify variants associated with five novel diseases. PLoS One 2012, 7:e28936
4. Puffenberger EG, Jinks RN, Wang H, Xin B, Fiorentini C, Sherman EA, Degrazio D, Shaw C, Sougnez C, Cibulskis K, Gabriel S, Kelley RI, Morton DH, Strauss KA: A homozygous missense mutation in HERC2 associated with global developmental delay and autism spectrum disorder. Hum Mutat 2012, 33:1639−1646
5. Strauss KA, Jinks RN, Puffenberger EG, Venkatesh S, Singh K, Cheng I, Mikita N, Thilagavathi J, Lee J, Sarafianos S, Benkert A, Koehler A, Zhu A, Trovillion V, McGlincy M, Morlet T, Deardorff M, Innes AM, Prasad C, Chudley AE, Lee IN, Suzuki CK: CODAS syndrome is associated with mutations of LONP1, encoding mitochondrial AAA+ Lon protease. Am J Hum Genet 2015, 96:121−135
6. Zhou L, Myers AN, Vandersteen JG, Wang L, Wittwer CT: Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. Clin Chem 2004, 50:1328−1335
7. Mose LE, Wilkerson MD, Hayes DN, Perou CM, Parker JS: ABRA: improved coding indel detection via assembly-based realignment. Bioinformatics 2014, 30:2813−2815
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990, 215:403−410
9. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. Nat Methods 2012, 9:357−359
10. Garrison E, Marth G: Haplotype-based variant detection from short-read sequencing. Available at: https://arxiv.org/abs/1207.3907; 2012 (accessed May 29, 2019)
11. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Patric R, Hibberd ML, Nagarajan N: LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res 2012, 40:11189−11201
12. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov: MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods 2015, 12:380−381
13. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004, 32:1792−1797
14. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F: The Ensembl Variant Effect Predictor. Genome Biol 2016, 17:122
15. Zerbino DR, Birney E: Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008, 18:821−829
16. Anders S, Pyl PT, Huber W: HTSeq: a Python framework to work with high-throughput sequencing data. Bioinformatics 2015, 31:166−169
17. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH: JBrowse: a next-generation genome browser. Genome Res 2009, 19:1630−1638
18. Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, Chen J, Robinson HE, Shim HS, Chimielecki J, Pao W, Engleman JA, Iafrate AJ, Le LP: Anchored multiplex PCR for targeted next-generation sequencing. Nat Med 2014, 20:1479−1484
19. Young AL, Challen GA, Birmann BM, Druley TE: Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nat Commun 2016, 7:12484
20. Morton DH, Strauss KA, Robinson DL, Puffenberger EG, Kelley RI: Diagnosis and treatment of maple syrup disease: a study of 36 patients. Pediatrics 2002, 109:999−1008
21. Mendell JR, Al-Zaidy S, Shell R, Arnold WD, Rodino-Klapac LR, Prior TW, Lowes L, Alfano L, Berry K, Church K, Kissel JT, Nagendran S, L'Italien J, Sproule DM, Wells C, Cardenas JA, Heitzer MD, Kaspar A, Corcoran S, Braun L, Likhite S, Miranda C, Meyer K, Foust KD, Burghes AHM, Kaspar BK: Single-dose gene-replacement therapy for spinal muscular atrophy. N Engl J Med 2017, 377:1713−1722
22. Finkel RS, Mercuri E, Darras BT, Connolly AM, Kuntz NL, Kirschner J, Chiriboga CA, Saito K, Servais L, Tizzano E, Topaloglu H, Tulinius M, Montes J, Glanzman AM, Bishop K, Zhong ZJ, Gheuens S, Bennett CF, Schneider E, Farwell W, De Vivo DC; ENDEAR Study Group: Nusinersen versus Sham control in infantile-onset spinal muscular atrophy. N Engl J Med 2017, 377:1723−1732
23. Glascock J, Lenz M, Hobby K, Jarecki J: Cure SMA and our patient community celebrate the first approved drug for SMA. Gene Ther 2017, 24:498−500