Evaluation of a high-throughput NGS SLIMamp™ CFTR Assay

Melanie Prasol; Akuah Kontor; Sean Polvino; Yue Ke; Nicholas Lodato; Dale Yuzuki; Zhaohui Wang

Pillar Biosciences Inc., Natick MA USA, yuzukid@pillar-biosciences.com; wangzh@pillar-biosciences.com

Abstract

Introduction: Next-generation sequencing (NGS) methods have been adopted for clinical genetic testing for targeted variant analysis and full coding-region analysis for the cystic fibrosis transmembrane conductance regulator (CFTR) gene. To improve the efficiency, robustness and cost-effectiveness of CFTR assays, we have applied our SLIMamp™ technology and developed a single-tube multiplex-PCR-based library-prep method allowing high-throughput (up to 384 samples) sequencing of the full-coding region on one MiSeq run. In this study, we evaluated the performance of the SLIMamp™ CFTR NGS assay.

Methods: The SLIMamp amplicons were designed to cover the entire coding regions with splice sites, selected deep intronic regions, and selected structure variants.

A total of 100 known positive samples, containing 69 unique variants, and 5 GIAB samples were used in this study. Forty-seven were clinical samples and 53 were commercially available samples purchased from Coriell. A wide range of DNA input amounts, ranging from 10 – 500 ng, were used for library preparation. Up to 384 libraries were normalized, pooled and sequenced with MiSeq v2 or v3 chemistry. The Pillar Variant Analysis Toolkit (PiVAT ™) v2.0 was used for the NGS data analysis and variant calling.

Results: All 148 known variants were detected from all samples without any false negatives. No false positives were detected in any of the GIAB samples. All replicates from more than 10 unique samples showed 100% concordance in all inter- and intra-runs. The on-target and mapping rates were 99.9% ± 0.06% and 99.8% ± 0.3% respectively. The average minimal base coverage for 384 samples was 307 and 504, on a V2 and v3 kit respectively.

Conclusion: This study demonstrates that the Pillar SLIMamp CFTR sequencing assay and the analysis pipeline offer a simple, robust, scalable and rapid approach for comprehensive and accurate interrogation of genetic variants of the CFTR gene.

Background and Significance

Cystic fibrosis (CF) is caused by biallelic germline mutations of a single gene, the cystic fibrosis transmembrane conductance regulator (CFTR). Over 2000 variants have been described, with greater than 300 currently characterized as pathogenic, and many with uncertain clinical significance. Clinical presentation can vary widely and depends on the specific variant combinations present. Recently NGS has been used for CFTR screening and exome sequencing in clinical settings. Illumina (ILMN) offers two separate FDA approved CF assay kits. The ILMN CF 139-Variant Assay (CF-139) designed for carrier screening allows 46 samples (plus 2 controls) per run while the ILMN CF Clinical Sequencing Assay for full exome sequencing allows for only 6 samples per run. We have developed a user-friendly, high-throughput CFTR NGS assay for MiSeq using patented Pillar SLIMamp tiling technology suitable for both screening and full exome sequencing.

Methods and Assay Design

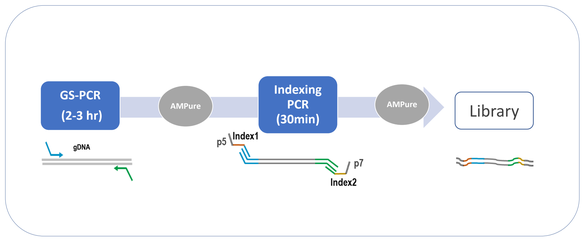
Product Specifications	
Sample type	gDNA
DNA input	5 - 500 ng
Amplicon size	324 - 405 bp with the average size at 361bp
Primer pool	Single pool
Amplicon#	34
Primer Design	No SNPs with minor allele frequency >= 0.1%
Mapping rate	99.8% ± 0.3%
On-target rate	99.9% ± 0.06%
Coverage uniformity	%>0.20 x Mean Coverage: 100%
Variant Frequency	20%
Multiplex level	384 samples on MiSeq (2 x 250 sequencing)

Table 1 – Product specifications for the Pillar CFTR NGS assay.

Assay Design – The Pillar CFTR assay is a single-tube multiplexed PCR based assay. **SLIMamp** technology allows for the simultaneous enrichment of overlapping amplicons. The work-flow consists of only two PCR reactions, each followed by AMPure bead purification. Libraries are then sequenced with MiSeq v2 or v3 chemistry using a 2x250 paired-end read length. Up to 384 samples can be sequenced simultaneously.

A total of 31 amplicons (214-354 bp) cover the full sequencing regions (CDS±10bp + intron12/22), including all SNV/indel variants annotated as CF-causing or varying-clinical consequence described in CFTR2 (Dec2017). Additionally, three primer sets flanking the junctions to capture three large deletions (10 Kb-21Kb).

Figure 1 – The Pillar CFTR Assay Work-flow



Data processing and analysis – The Pillar Variant Analysis Toolkit (PiVAT ™) v2.0 was used for the NGS data analysis and variant calling. The software analyzes the sequence data and maps it against the hg19 reference genome. It calculates and reports general sequencing quality metrics, including coverage, mapping rate, on-target rate, and uniformity. It also produces variant calls. Any variation from the hg19 genome sequence is reported by the variant caller, as well as the zygosity (20-80% of reads is called as heterozygous, greater than 80% as homozygous). Additionally, the type of variant (such as SNV or deletion), allele frequency, and the clinical significance are also reported.

Results and Conclusions

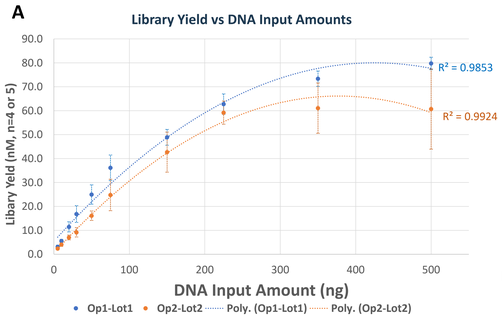


Figure 2 – DNA Input Study – Five samples were used for an input study to evaluate the performance of the assay with different starting amounts of DNA: 4 Coriell cell line DNAs with 10 different input amounts, from 5 – 500 ng, and one clinical sample with 6 different input amounts, from 5 – 75 ng. Two operators each used two different reagent lots. **A)** The final library yield was dependent on the initial sample input. However, overall sequencing results were unchanged. **B)** The mapping rate, on-target rate and variant calls were unaffected. Both operators achieved mapping and on-target rates greater than 99%.

A 14 Variants detected by the **Pillar CFTR Assay** but **not** covered by ILMN CF-139

HGVSC	Legacy	CFTR2-Dec2017: Allele frequency
c.3454G>C	D1152H	0.40%
c.443T>C*	I148T	0.10%
c.1721C>A	P574H	0.02%
c.4046G>A	G1349D	0.02%
c.3485G>T*	R1162L	0.01%
c.313delA	444delA	0.01%
c.3368-2A>T	3500-2A->T	0.01%
c.509G>A*	R170H	0.01%
c.1523T>G*	F508C	0.01%
c.1519A>G	I507V	
c.429delT		
c.935_937delTCT		
c.1820_1903del84		
c.-4G>C		

B All 23 ACMG23 variants detected

HGVSC	Legacy	CFTR2-Dec2017: Allele frequency
c.1521_1523delCTT	F508del	69.86%
c.1624G>T	G542X	2.54%
c.1652G>A	G551D	2.11%
c.3909C>G	N1303K	1.58%
c.350G>A	R117H	1.31%
c.3846G>A	W1282X	1.22%
c.1657C>T	R553X	0.94%
c.489+1G>T	621+1G->T	0.93%
c.1585-1G>A	1717-1G->A	0.86%
c.3718-2477C>T	3849+10kbC->T	0.82%
c.2657+5G>A	2789+5G->A	0.71%
c.3484C>T	R1162X	0.46%
c.1519_1521delATC	I507del	0.46%
c.254G>A	G85E	0.43%
c.3528delC	3659delC	0.38%
c.1040G>C	R347P	0.37%
c.1364C>A	A455E	0.35%
c.2988+1G>A	3120+1G->A	0.33%
c.1000C>T	R334W	0.30%
c.1766+1G>A	1898+1G->A	0.30%
c.1679G>C	R560T	0.24%
c.579+1G>T	711+1G->T	0.19%
c.2052delA	2184delA	0.18%

C Additional variants detected

HGVSC	Legacy	CFTR2-Dec2017: Allele frequency
c.2051_2052delAAinsG	2183AA->G	0.37%
c.3140-26A>G	3272-26A->G	0.33%
c.54-5940_273+10250del21kb**	CFTRdele2,3	0.30%
c.617T>G	L206W	0.24%
c.2052dupA	2184insA	0.23%
c.262_263delTT	394delTT	0.22%
c.178G>T	E60X	0.21%
c.1477C>T	Q493X	0.21%
c.3302T>A	M1101K	0.17%
c.3196C>T	R1066C	0.16%
c.1021_1022dupTC	1154insTC	0.15%
c.3773dupT	3905insT	0.15%
c.3276C>A	Y1092X	0.14%
c.1646G>A	S549N	0.14%
c.1040G>A	R347H	0.14%
c.948delT	1078delT	0.13%
c.3472C>T	R1158X	0.13%
c.1558G>T	V520F	0.11%
c.2012delT	2143delT	0.07%
c.223C>T	R75X	0.07%
c.3744delA	3876delA	0.06%
c.366T>A	Y122X	0.06%
c.532G>A	G178R	0.06%
c.1647T>G	S549R	0.05%
c.579+3A>G	711+3A->G	0.04%
c.2125C>T	R709X	0.04%
c.2128A>T	K710X	0.04%
c.2668C>T	Q890X	0.03%
c.274-1G>A	406-1G->A	0.03%
c.3964-78_4242+577del**	CFTRdele22,23	0.03%
c.274G>T	E92X	0.03%
c.1680-1G>A	1812-1G->A	0.02%
c.4243-35delT	c.4375-36delT	

Table 2 – Variant detection study – A total of 100 known positive samples, (47 clinical samples and 53 commercial) containing 69 unique variants were examined. Additionally, 5 GIAB samples were used in this study. Each sample was tested at least twice. **A)** The Pillar CFTR assay detects an additional 14 variants that are not included in the Illumina 139 variant assay, at least 5 of which are pathogenic. **B)** All 23 of the ACMG23 variants were represented by the sample selection. **C)** 33 additional clinically significant variants were detected in our assay. Variants in B and C are covered by both the Pillar CFTR assay and ILMN CF-139 assay. There were no false positives in the GIAB samples. Although not present in the sample group, there was sufficient coverage to detect the remaining ILMN 139 variants not listed in B and C.

The intron 9 poly-T/TG repeat region is included in panel for sequencing, but is currently excluded from the variant caller analysis.

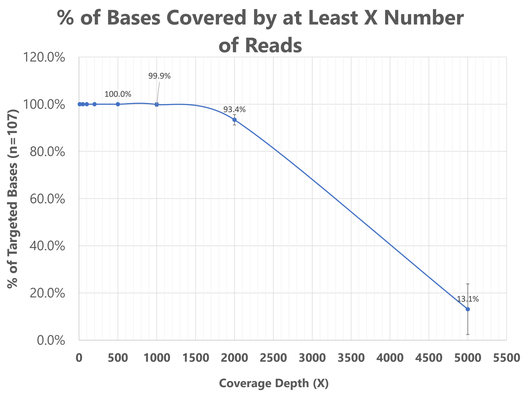


Figure 3 – 107 samples were sequenced on a MiSeq v2 flow cell. Greater than 90% of all amplicons achieved a coverage depth of at least 2000 reads. The mean base coverage was 3660 ± 338. Up to 384 samples can be run on a v2. Higher coverage can be obtained by running a v3 kit. The average minimal base coverage for 384 samples was 307 and 504, on a v2 and v3 kit respectively.

Conclusions:

1. Due to the high multiplexing level (up to 384 samples on one MiSeq run), the Pillar CFTR Assay could be used for both carrier screening and whole exome sequencing.

2. The assay is robust, accurate, reproducible and cost-effective. The assay can take a wide range of DNA input amounts from 5 to 500 ng, eliminating the need for DNA input normalization.

2. The Pillar Assay is designed to be automation friendly to support high throughput screening. The PCR master mixes contain all PCR components except for DNA input and primers. The minimal volume pipetted per reaction during the whole process is 4 µL.

3. The Pillar CFTR Assay library preparation process is stream-lined and rapid with only two PCR reactions. The total process from DNA to sequencer takes less than 1 day.