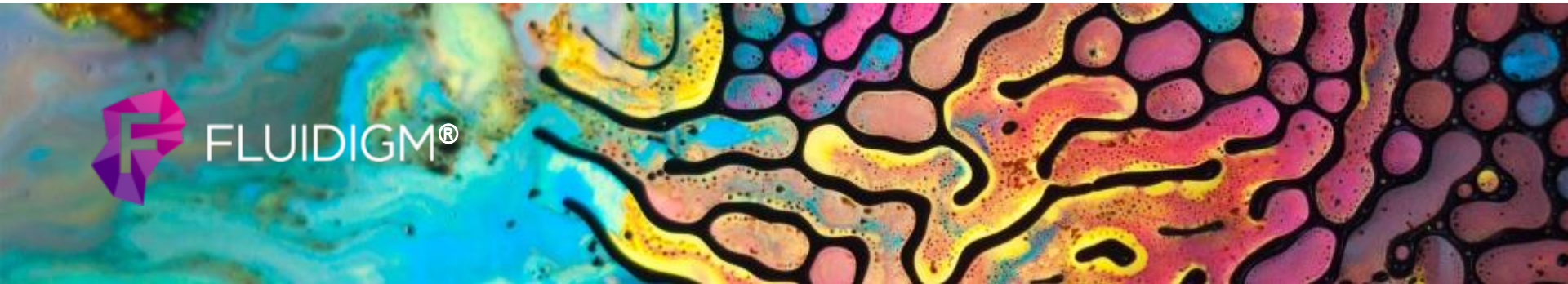
PN 101-8060

Advanta CFTR NGS Library Prep Assay

Analytical Validation



Outline

1. Overview of the Advanta™ CFTR NGS Library Prep Assay
2. Analytical validation experimental design
3. Read performance metrics
4. Variant calling performance metrics
5. Summary and overview of available tools

Advanta CFTR NGS Library Prep Assay

Advanta CFTR NGS Library Prep Assay workflow

1 Prepare reagents.

0.5–1 hr

2 Load.

3 Amplify and barcode.

4 Harvest.

Automated prep on Juno™

4–5.5 hr or overnight harvest

5 Pool and clean up.

6 Run adapter PCR.

7 Clean up.

8 Assess quality.

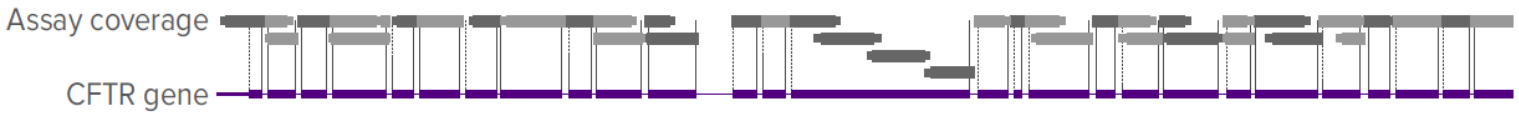
Single-tube reaction

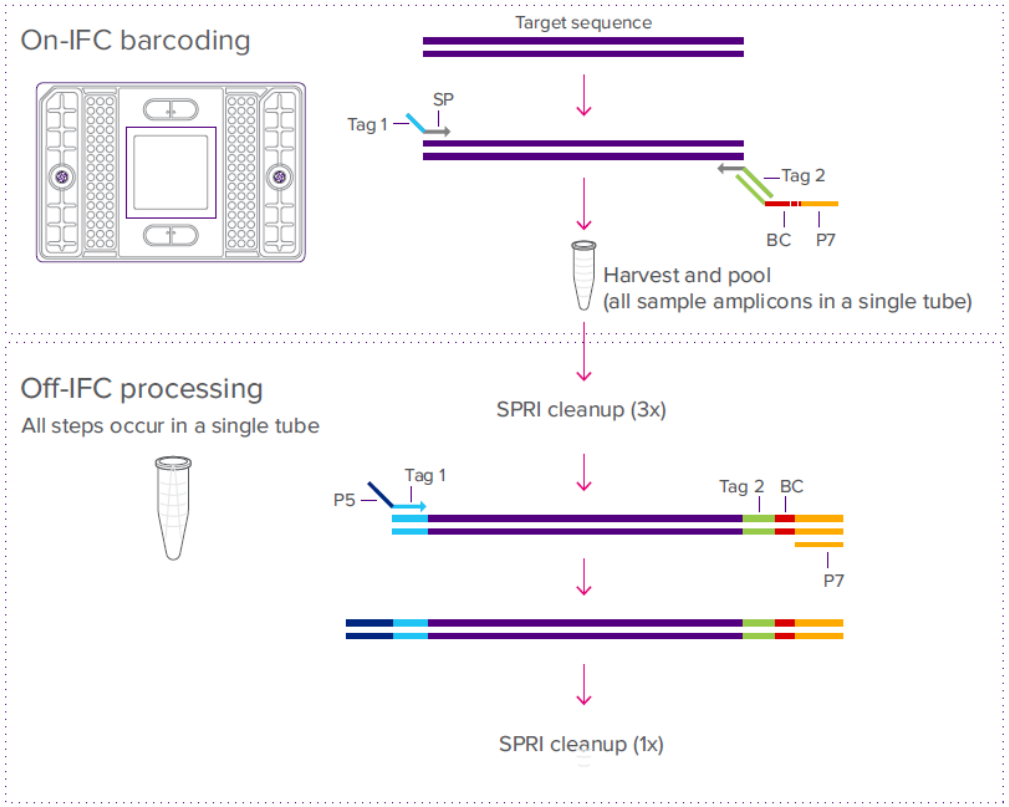
4.5 hr

NGS-ready, barcoded libraries

up to 384/lane (single index)
up to 1536/lane (dual index)

The Advanta CFTR NGS Library Prep Assay provides excellent coverage of the CFTR gene





Analytical validation experimental design

Analytical validation study

The analytical validation by an external party assessed the performance and reproducibility of the Advanta CFTR NGS Library Prep Assay.

Performed by Q² Solutions® | EA Genomics

Q² Solutions
a Quintiles Quest Joint Venture

EA
Genomics

Analytical validation

Using a variety of sample types, performance was assessed for:

- Read performance metrics
- Variant call metrics
  - Accuracy
  - Sensitivity
  - Specificity
- Reproducibility

Variants refer to single-nucleotide variants (SNVs) and insertions and deletions (indels).

Analytical validation run design addressed several variables

Run	IFC	Operator	Reagent Lot
Run 1	LP 48.48 IFC 1 LP 48.48 IFC 2 LP 48.48 IFC 3	Operator 1	Reagent Lot 1
Run 2	LP 48.48 IFC 4 LP 48.48 IFC 5 LP 48.48 IFC 6	Operator 2	Reagent Lot 1
Run 3	LP 48.48 IFC 7 LP 48.48 IFC 8 LP 48.48 IFC 9	Operator 1	Reagent Lot 1
Run 4	LP 48.48 IFC 10 LP 48.48 IFC 11 LP 48.48 IFC 12	Operator 1	Reagent Lot 1
Run 5	LP 48.48 IFC 13 LP 48.48 IFC 14 LP 48.48 IFC 15	Operator 2	Reagent Lot 2
Run 6	LP 192.24 IFC 1	Operator 1	Reagent Lot 2
Run 7	LP 192.24 IFC 2	Operator 2	Reagent Lot 2
Run 8	LP 192.24 IFC 3	Operator 1	Reagent Lot 1

Samples were sequenced across eight flow cells using three MiSeq™ instruments.

Analytical validation sample types

Sample Type	Description
Well-characterized (NIST) genomic reference samples (Coriell)	Purified genomic DNA that does not contain CF-causing variants.
CFTR variant-containing genomic samples (Coriell)	Purified genomic DNA that contains CF-causing variants.
Synthetic control DNA samples (MMQCI)	Plasmid DNA that contains CF-causing variants, including rare variants. (Advanta CFTR assays were not designed for synthetic constructs.)
Buccal swab, saliva, and whole blood	DNA extracted from these sample types that contain CF-causing variants (SNVs only). Samples are unmatched and not from the same donor.
No template control (NTC)	A buffer run in singlet for each integrated fluidic circuit (IFC). NTCs are expected to have no or very few reads.

NIST: National Institute of Standards and Technology
Coriell: Coriell Institute for Medical Research (from either the NIGMS Human Genetic Cell Repository, NHGRI Sample Repository for Human Genetic Research, or CDC Cell and DNA Repository)
MMQCI: Maine Molecular Quality Controls, Inc.

Sample number breakdown

Sample Type	Unique Samples	Total No. of Samples
Coriell NIST reference samples	3	132
Coriell CFTR variant-containing samples	112	732
MMQCI synthetic samples	6	234
Buccal	5	65
Saliva	5	58
Whole blood	5	57
NTC (run in singlet for each IFC)	n/a	18
Grand total	**136**	**1,296**

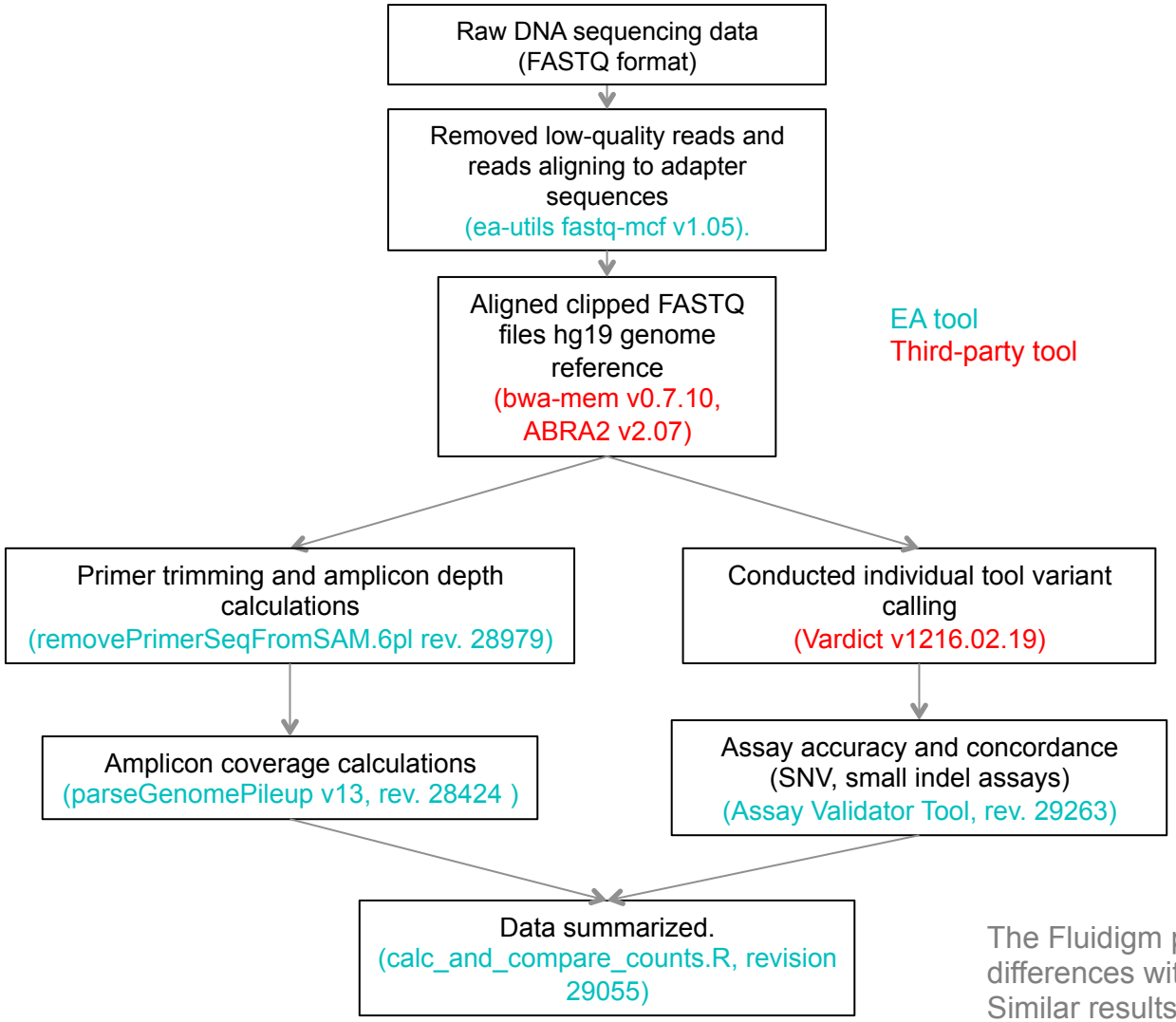
Number of samples by IFC:

- 15 x 48.48 IFCs (n=720)
- 3 x 192.24 IFCs (n=576)

Buccal, saliva, whole blood samples:
Run at two amounts (100 ng, n=117; 60 ng, n=56)

All other samples, except NTC, were run at 100 ng.

Analytical validation bioinformatics pipeline



The Fluidigm pipeline has some differences with the EA pipeline. Similar results were achieved. An overview of the Fluidigm pipeline is available in the protocol appendix.

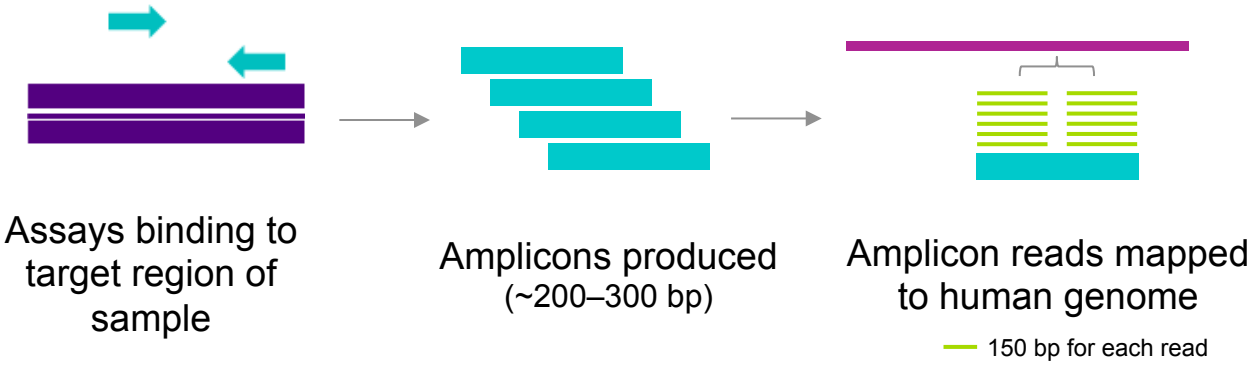
Sample filters during analysis

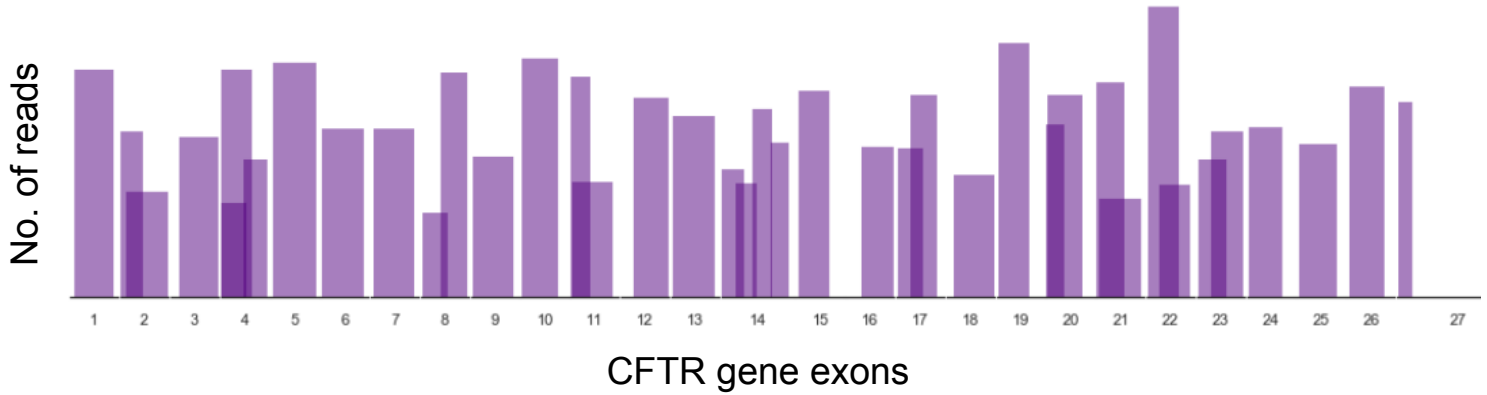
- 14 samples did not meet threshold for reads during demultiplexing (1.1% of 1,296 samples), and these samples were excluded for analysis.

- Seven samples were excluded for analysis due to operator error during the library preparation process.

Analytical validation read performance metrics

The 44 assays for SNVs/small indels were used to assess read performance metrics





All sample types were used in read performance assessments

- Specimens included gDNA (Coriell cell lines, buccal swabs, saliva, whole blood) and MMQCI synthetic samples.
- Design of MMQCI synthetic samples prevented proper annealing of some primer sets, leading to modification of metric calculations for MMQCI samples.
- Assays were designed for gDNA from natural, biologically derived samples and not for MMQCI samples.

Average assay read depth was at least 1,150x for all samples

Sample Type	Mean Depth across Amplicons
All Coriell genomic samples	2,576
MMQCI synthetic samples	1,469
Buccal	1,155
Saliva	1,389
Whole blood	3,105

139 to 188 samples per MiSeq run using MiSeq Reagent Kit v2.

Performance summary

Analytical validation read performance metrics

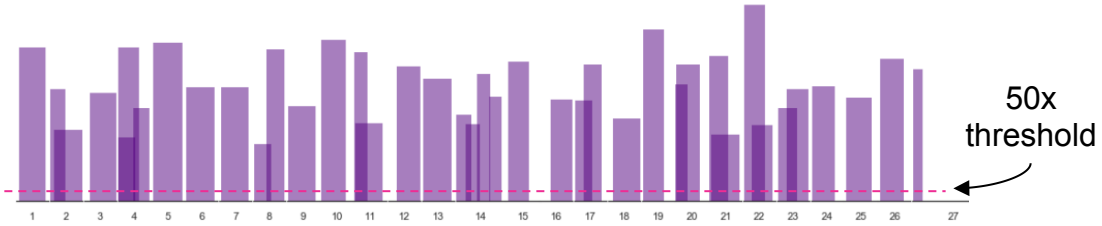
Attribute	Observed Performance
Assay pass rate	99.3% (gDNA) 97.8% (MMQCI)
Amplicon uniformity	99.4% (gDNA) 98.2% (MMQCI)
Reads mapped to genome	≥99.9% (all sample types)
Reads mapped to target	≥99.99% (all sample types)
No template control (NTC) reads	NTC undetectable • 15/18 IFCs (undetected) • 3/18 IFCs (<0.02% total seq reads)

Assay pass rate

Analytical validation read performance metrics

Experimental Approach	Observed Performance
Proportion of samples with at least 95% of the target region covered by at least 50x read depth	99.3% (gDNA) 97.8% (MMQCI)

95% of the target region covered by assays with at least 50x read depth



Amplicon uniformity

Analytical validation read performance metrics

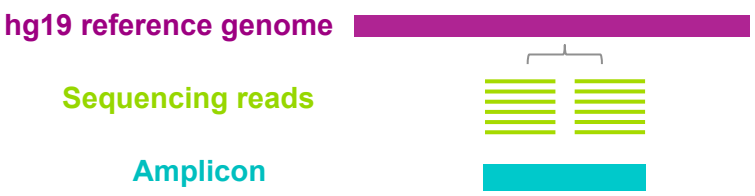
Experimental Approach	Observed Performance
Proportion of amplicons designed for SNV and indel detection within 1/5 to 5x mean amplicon depth within the sample	99.4% (gDNA) 98.2% (MMQCI)

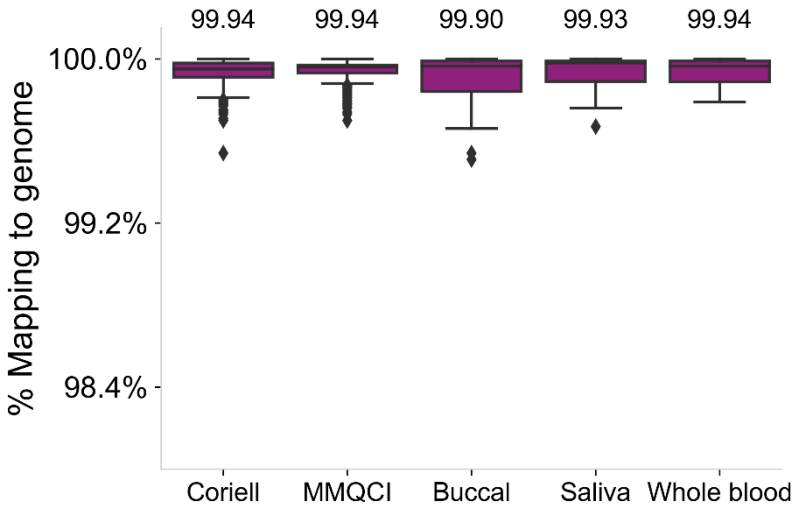
Sample Type	% Amplicons within Uniformity 1/5 to 5x of Mean
All Coriell genomic samples	99.27%
Buccal	99.24%
Saliva	99.48%
Whole blood	99.47%

Reads mapped to genome (%)

Analytical validation read performance metrics

Experimental Approach	Observed Performance
% sequencing reads that map to genome hg19	≥99.90% (all sample types)

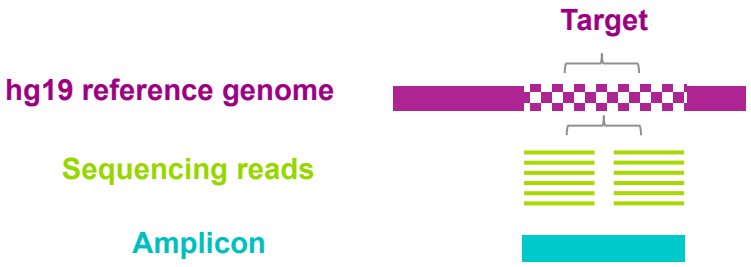


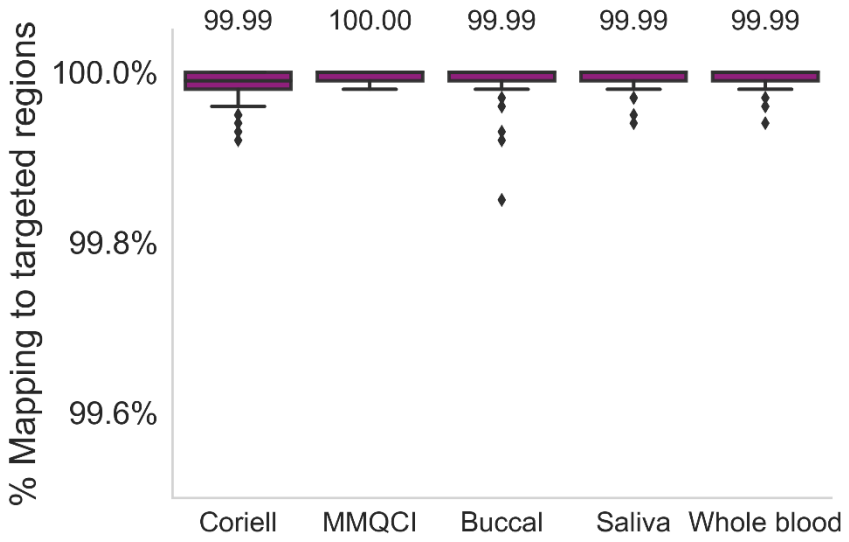


Reads mapped to target (%)

Analytical validation read performance metrics

Experimental Approach	Observed Performance
% reads that map to target region out of total mapped reads	≥99.99% (all sample types)





No template control (NTC) reads

Analytical validation read performance metrics

Experimental Approach	Observed Performance
Used to assess cross-contamination within the 18 IFC runs	NTC undetectable • 15/18 IFCs (undetected) • 3/18 IFCs (<0.02% total seq reads)

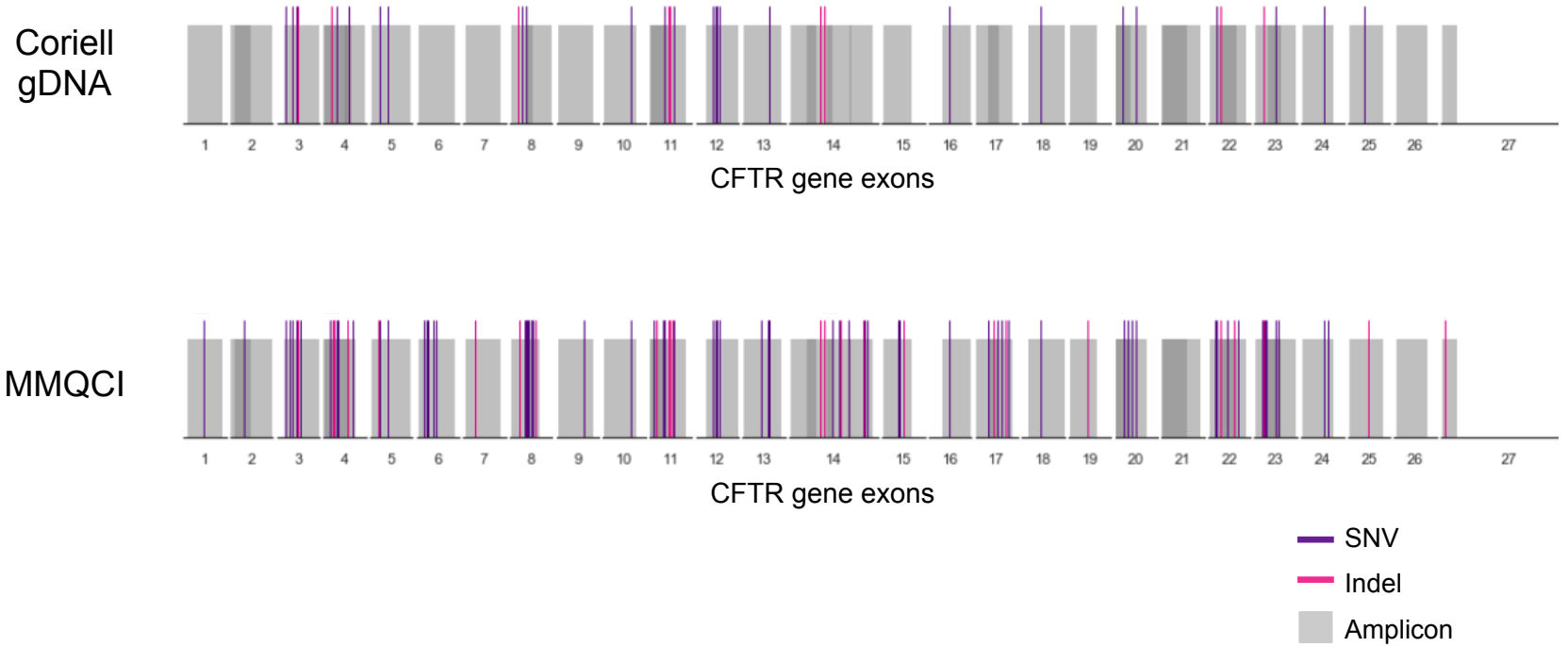
Analytical validation variant call metrics

Evidence-based CFTR variant coverage

1078delT (d)	**1898+3A->G**	306delTAGA	4040delA	C524X	G550X	**Q2X**	**S1251N**
1119delA	1924del7	306insA	405+1G->A	CFTR50kbdel	**G551D** (a) (d)	**Q39X**	S1255P
1138insG	**2055del9->A**	**3120+1G->A** (a)	**405+3A->C**	CFTRdele1	G551S	Q414X	**S1255X**
1154insTC	2105-2117del13ins AGAAA	**3120G->A**	**406-1G->A**	CFTRdele14b-17b	G673X	**Q493X**	**S341P**
1161delC	2118del4	**3121-1G->A**	**4209TGTT->AA**	CFTRdele17a-18	**G85E** (a) (d)	Q525X	**S466X**
1213delT	**2143delT** (d)	3121-2A->G	4218insT	CFTRdele17a,17b	**G970R**	**Q552X**	**S489X**
1248+1G->A	**2183AA->G or 2183delAA->G** (d)	3132delTG	4279insA	CFTRdele2	H1054D	Q685X	**S492F**
1249-1G->A	**2184delA** (a) (d)	**3171delC**	4326delTC	CFTRdele2-4	**H199Y**	Q715X	S4X
124del23bp	2184insA (d)	3171insC	4374+1G->A	CFTRdele2,3 (d)	**I1234V**	**Q890X**	**S549N**
1259insA	2185insC	3271delGG	4374+1G->T	CFTRdele22-24	**I336K** (d)	**Q98X**	**S549R**
1288insTA	2307insA	**3272-26A->G** (d)	**4382delA**	CFTRdele22,23	**I507del** (a) (d)	**R1066C**	S912X
1341+1G->A	**2347delG**	3500-2A->G	4428insGA	CFTRdele4-7	**K710X**	**R1066H**	**S945L**
1461ins4	2372del8	365-366insT	442delA	CFTRdup6b-10	**L1065P**	R1102X	**T338I**
1471delA	2556insAT	**3659delC** (a) (d)	**444delA**	**D110H**	**L1077P**	**R1158X**	**V520F**
1497delGG	2585delT	3667ins4	**457TAT->G**	E1104X	L1254X	**R1162X** (a) (d)	**W1089X**
1525-1G->A	2594delGT	3737delA	541delC	E1371X	**L206W**	**R117C**	W1098X
1525-2A->G	**2622+1G->A**	**3791delC**	**574delA**	E193X	**L227R**	**R117H + 5T** (a)	W1145X
1548delG	**2711delT**	3821delT	**621+1G->T** (a) (d)	E56K	**L467P**	**R334W** (a) (d)	W1204X
1609delCA	2721del11	3849+10kbC->T (a) (d)	**663delT**	**E585X**	**L732X**	**R347H**	**W1282X** (a) (d)
1677delTA (d)	2732insA	3849+4A->G	675del4	**E60X** (d)	L88X	**R347P** (a) (d)	W19X
1717-1G->A (a) (d)	**2789+5G->A** (a) (d)	3850-1G->A	**711+1G->T** (a)	**E822X**	**L927P**	**R352Q**	W216X
1717-8G->A	2790-1G->C	3850-3T->G	**711+3A->G**	**E831X**	**M1101K** (d)	**R553X** (a) (d)	**W401X**
1782delA	**2869insG**	**3876delA**	**711+5G->A**	**E92K**	**M1V**	**R560K**	W496X
1811+1.6kbA->G	2896insAG	3878delG	**712-1G->T**	**E92X** (d)	**N1303K** (a) (d)	R560S	W57X
1811+1G->C	2942insT	3905insT (d)	**852del22**	**F508del** (a) (d)	**P205S**	**R560T** (a)	**W846X**
1812-1G->A	2957delT	**394delTT**	**935delA**	**G1061R**	**P67L**	**R709X**	W882X
1824delA	296+1G->A	**4005+1G->A**	991del5	**G1244E**	Q1042X	**R75X**	**Y1092X** (d)
182delT	296+1G->T	4010del4	**A455E** (a) (d)	**G1349D**	**Q1313X**	**R764X**	**Y122X**
1833delT	297-1G->A	4015delA	A46D	**G178R**	Q1382X	R785X	Y275X
185+1G->T	2991del32	4016insT	**A559T**	G27X	Q1411X	R792X	**Y569D**
1898+1G->A (a)	**3007delG**	**4021dupT**	A561E	**G330X**	Q1412X	**R851X**	Y849X
1898+1G->C	3028delA	4022insT	C276X	**G542X** (a) (d)	**Q220X**	**S1196X**	Y913X

Panel of 256 CFTR variants, from point mutations to large exonic del/dups, derived from the Clinical and Functional Translation of CFTR (CFTR2); available at cftr2.org; CF-causing variants from CFTR2_8August2016.xlsx
(a) indicates variants (23) within the ACMG/ACOG recommendations. (d) indicates the minimum variants (31) required by German guidelines. **Bold** indicates the inclusion of at least one variant-bearing sample in the analytical validation.

CFTR variants tested in CorielI gDNA and MMQCI synthetic samples



Sample types used in variant call assessments

Sample Type	Sensitivity	Specificity	Accuracy	Reproducibility
*Coriell NIST reference samples	X	X	X	X
Coriell CFTR variant-containing samples	X			X
*MMQCI synthetic samples	X	X	X	X
Buccal, saliva, whole blood				X

* NIST samples and synthetics had high confidence in known negative data.
Variant calls were only considered with minimum allele frequency >10% and read depth >50x.

Performance summary

Analytical validation variant call metrics

Attribute	Observed Performance		
Sensitivity	99.1% SNVs (Coriell) 99.7% indels (Coriell) 99.9% SNVs (MMQCI) 96.1% indels (MMQCI)		
Specificity	100% for all samples (SNVs and indels)		
Accuracy	100% (Coriell) 99.99% (MMQCI)		
Reproducibility	**Sample**	**SNVs**	**Indels**
	Coriell	99.9%	99.5%
	MMQCI	98.7%	96.4%
	Blood	100%	N/A
	Buccal	98.5%	N/A
	Saliva	98.4%	N/A

Sensitivity

The proportion of CFTR variant-positive positions correctly identified

Experimental Approach	Observed Performance
$Sensitivity = TP / TP + FN$ where TP = true positive and FN = false negative 112 Coriell samples (≥4 replicates each) + MMQCI samples	99.1% SNVs (Coriell) 99.7% indels (Coriell) 99.9% SNVs (MMQCI) 96.1% indels (MMQCI)

Specificity

The proportion of positions negative for variants correctly identified

Experimental Approach	Observed Performance
$Specificity = TN/TN+FP$ where TN = true negative and FP = false positive 3 Coriell samples with known homozygous reference positions in the NIST database + MMQCI samples	100% for all samples (SNVs and indels)

Accuracy

The proportion of correctly identified positive and negative for variants

Experimental Approach	Observed Performance
$Accuracy = TP + TN / TP + TN + FP + FN$	100% (Coriell)
Multiple replicates (37 to 44) of three Coriell samples + MMQCI samples	99.99% (MMQCI)

Reproducibility

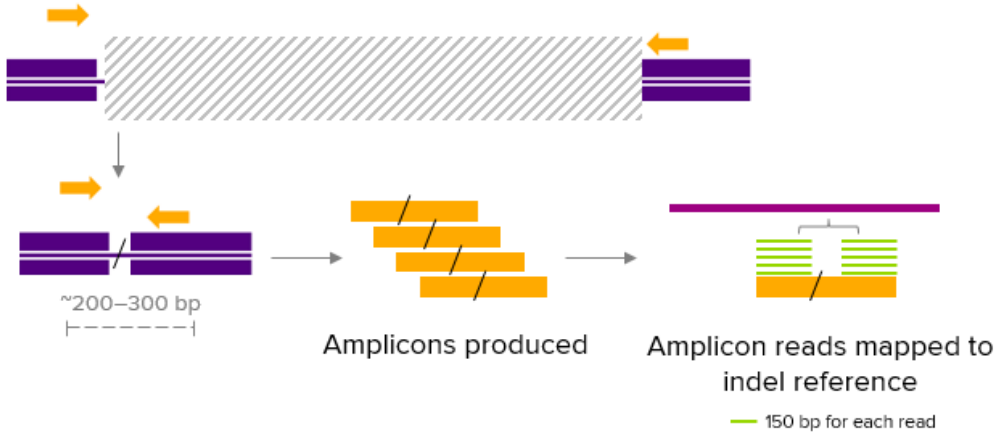
Ability to get the same result upon repeat testing that accounts for sources of variability

Experimental Approach	Observed Performance		
	Sample	**SNVs**	**Indels**
$Reproducibility = \sum(Matches\ across\ all\ comparisons) / \sum(Matches\ and\ Mismatches\ across\ all\ comparisons)$	Coriell	99.3%	99.5%
	MMQCI	98.7%	96.4%
% concordance based on pairwise comparisons of all variables	Blood	100%	N/A
	Buccal	98.5%	N/A
	Saliva	98.4%	N/A

Variables tested for pairwise comparison: operator, reagent lot, library preparation day, IFC type, and input amount.

Performance of large deletions and duplications

- CF-variant Coriell sample NA18688 was expected to have a 21 kb heterozygous deletion resulting in loss of exons 2 and 3.
- Amplification of the CFTR047_Del primer set was expected in samples positive for this deletion.
- This was identified across NA18688 all replicates (100% sensitivity).



Summary

Advanta CFTR NGS Library Prep Assay analytical validation summary

- External evaluator results demonstrate strong performance across a variety of sample types.
- The assay shows excellent read performance metrics and amplicon uniformity.
- High sensitivity, specificity and accuracy of CFTR variant calls is demonstrated across a variety of samples.
- Results were highly reproducible between technical replicates, reagent lots and operators.

Analytical validation white paper, study guidelines and example data are available upon request.

Thank you.

For Research Use Only. Not for use in diagnostic procedures.

Appendix

Estimated sequencing read depth

IFC Type	Number of IFCs	Number of Samples	Average Read Depth per Sample	Average Read Depth per Amplicon per Sample
48.48	1	48	312,000	4,200
48.48	2	96	156,000	2,100
48.48	4	192	78,000	1,000
48.48	8	384	39,000	500
192.24	1	192	78,000	1,000
192.24	2	384	39,000	500
192.24	4	768*	19,500	260
192.24	8	1,536*	9,700	130

* Dual indexing is required when sequencing more than 384 samples on a single lane.

Estimates are conservative. Factors used to determine the estimates:

- Use of the MiSeq Reagent Kit v2 and 15 M total reads passing filter per lane
- No negative control samples are included in these calculations.
- Assumes all 73 amplicons of CFTR panel amplified in all samples
- Average read depth per amplicon includes both ends of a paired-end read.

Dual indexing for high sample multiplexing is effective with higher-output sequencers

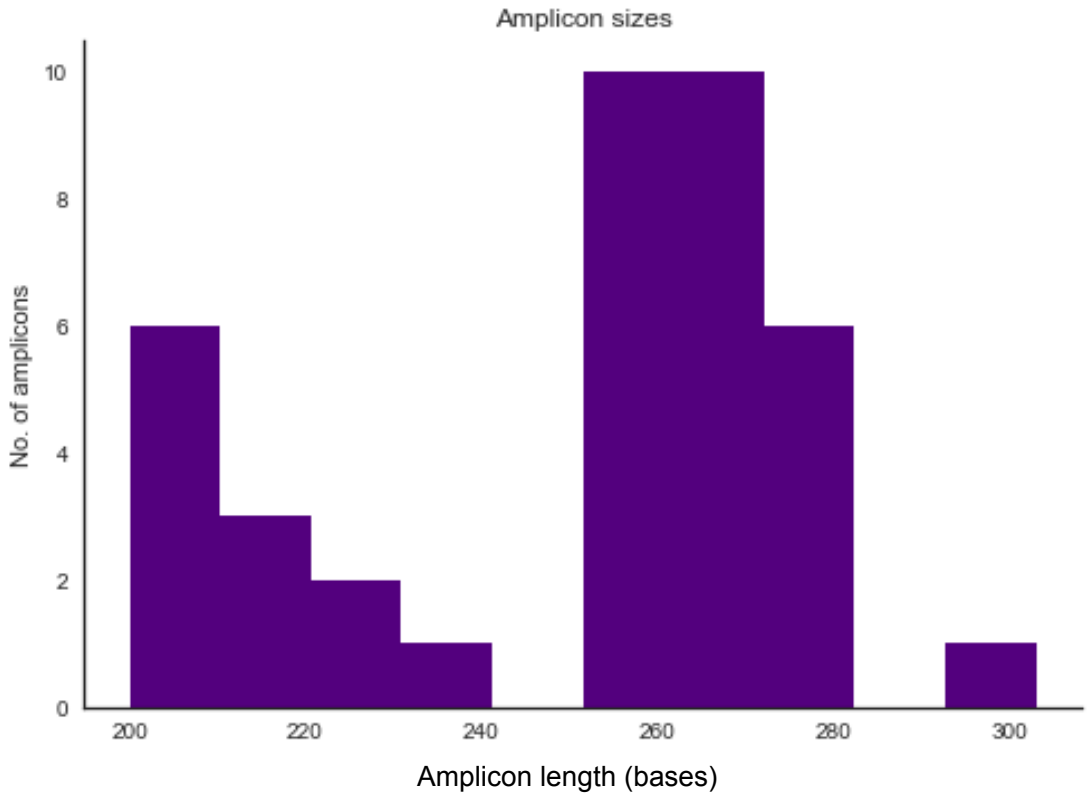
Number of Samples	Average Read Depth per Amplicon per Sample (MiSeq)	Average Read Depth per Amplicon per Sample (NextSeq)
384	500	14,200
768	200	7,100
1,152	170	4,700
1,536	100	3,500

Estimates are conservative. Factors used to determine the estimates:

- Use of the MiSeq Reagent Kit v2 and 15M total reads passing filter per lane
- Use of the NextSeq™ High Output Kit and 400M total reads passing filter per lane
- No negative control samples are included in these calculations.
- Assumes all 73 amplicons of CFTR panel amplified in all samples
- Average read depth per amplicon includes both ends of a paired-end read.

Estimated CFTR amplicon sizes

Amplicons for SNVs and small indels



Data from .bed file