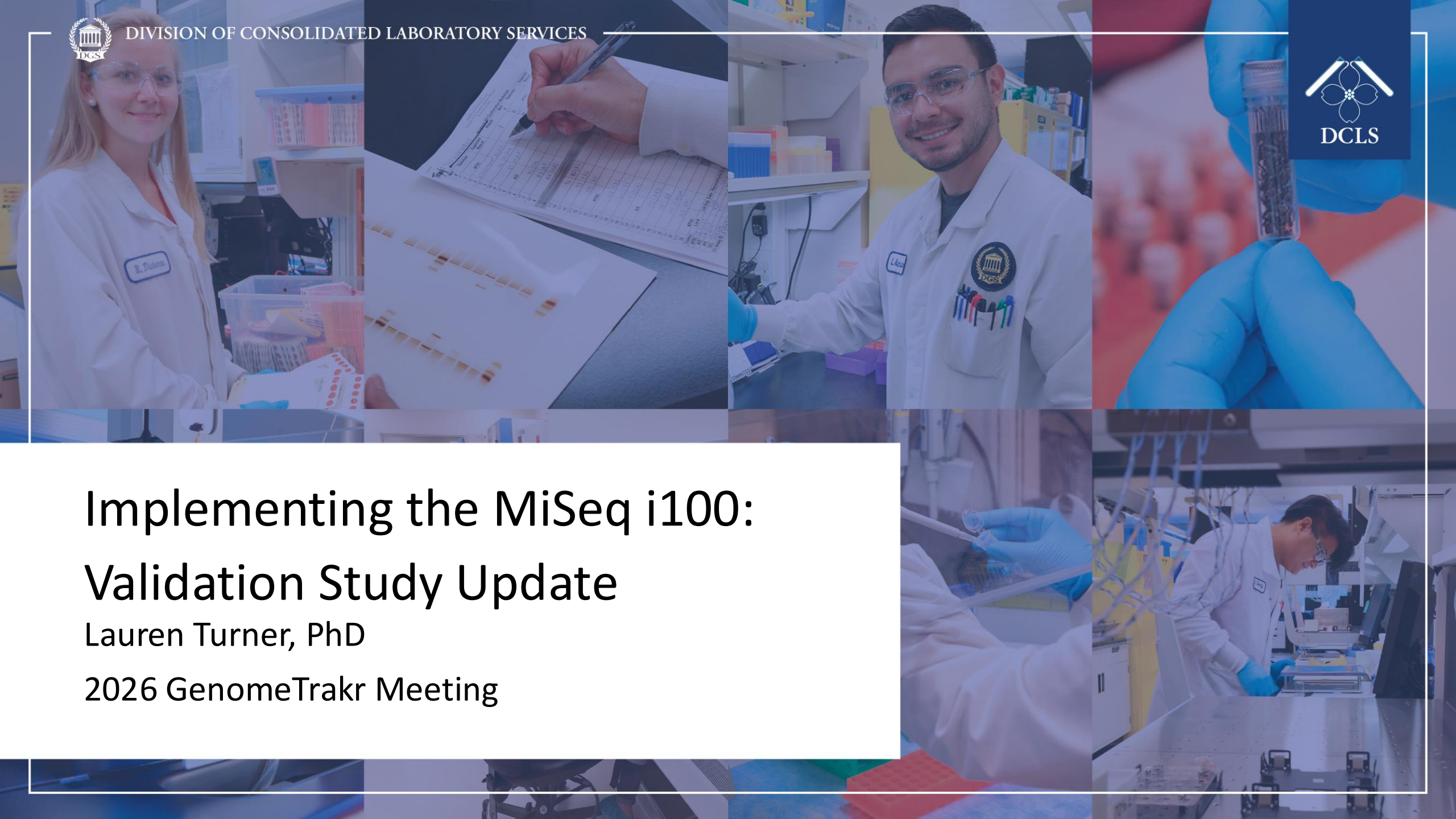


DIVISION OF CONSOLIDATED LABORATORY SERVICES

DCLS

Implementing the MiSeq i100: Validation Study Update

Lauren Turner, PhD

2026 GenomeTrakr Meeting

Background: Illumina Benchtop Sequencers

- MiSeq i100 short-read sequencing platform intended to replace the legacy MiSeq
  - Legacy system sunsetting at the end of 2029
- MiSeq i100 versus MiSeq i100 Plus
  - Different kit capacity
    - MiSeq i100 – Up to 15 Gb, 25 M Reads
    - MiSeq i100 Plus – Up to 30 Gb, 100M Reads

Flow Cell Type	5M	25M	50M[2]	100M[2]
1 x 100 bp	-	2.5 Gb	5 Gb	10 Gb
2 x 150 bp	1.5 Gb	7.5 Gb	15 Gb	30 Gb
2 x 300 bp	3 Gb	15 Gb	30 Gb	-
2 x 500 bp	-	25 Gb	-	-



[2]: 50M and 100M flow cells are available for the MiSeq i100 Plus System only

MiSeq i100 Details

- Same library preparation kits and indices may be used
- Instrument handling and operability improvements
  - Onboard denaturation
  - No washes required (cartridge-based fluidics)
  - Room temperature reagent cartridges
    - No onboard chiller
  - No formamide (reducing chemical waste)
  - Linux-based OS
    - Integrated secondary DRAGEN analysis
  - Options for local, or cloud-based data storage, or hybrid with an external storage

Run Time and Data Quality Improvements

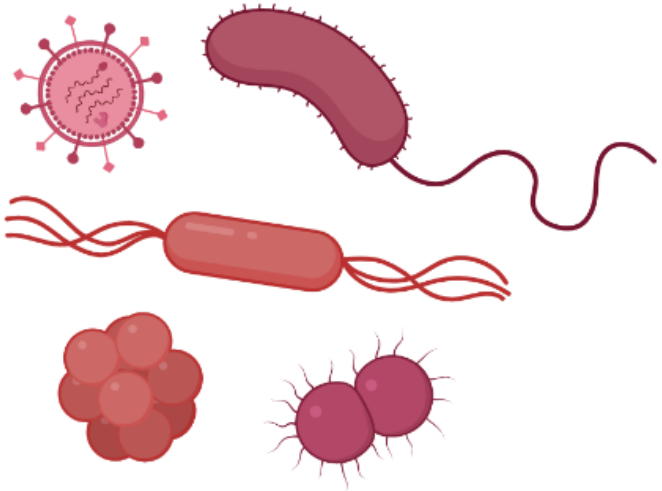
- Faster run times with improved read quality
  - Updates in XLEAP SBS technology, signal processing, and imaging

Flow Cell Type	5M	25M	50M[2]	100M[2]	Q30
1 x 100 bp	-	~ 4 hr	~ 4.5 hr	~ 5 hr	≥ 90% of bases
2 x 150 bp	~ 7 hr	~ 7 hr	~ 7.5 hr	~ 8 hr	≥ 90% of bases
2 x 300 bp	~ 15 hr	~ 15 hr	~ 15.5 hr	-	≥ 85% of bases
2 x 500 bp	-	~ 24 hr	-	-	≥ 85% of bases

[2]: 50M and 100M flow cells are available for the MiSeq i100 Plus System only

Meeting DCLS's Testing Needs

- Throughput
  - DCLS sequenced ~300 bacterial pathogen isolates monthly in 2026
    - 75 isolates per week
    - Optimal flow cell 25 M (28+ isolates per run)
- Flexibility
  - Mixed bacterial organism runs
  - Option for viral library spike-in
    - Norovirus, SARS-COV-2 (clinical and wastewater)



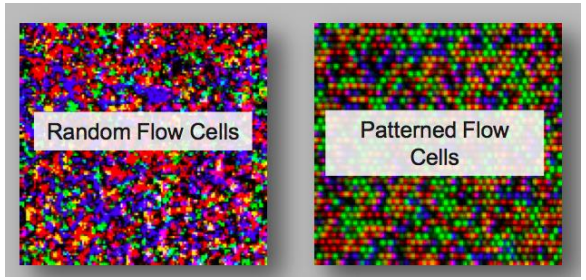
Validation Approach

- Undefined process for spiking pools to create combined bacterial and viral runs
- Proposed mixed organism testing approach
  - Libraries prepared separately by the pathogen-specific protocol
    - Bacterial libraries pooled, fragment analysis performed, adjusted to 4 nM
    - Viral libraries separately pooled and normalized to 4 nM and 4 uL spiked into the bacterial pool

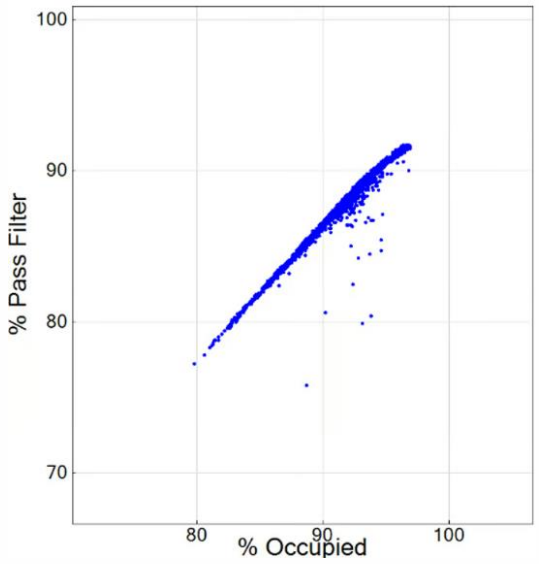
25M – 300 cycle (2 x 151)	25M – 600 cycle (2 x 301)
Bacterial isolates	Bacterial isolates
Norovirus	Norovirus
SARS-CoV-2 (clinical and wastewater)	-

Initial Testing: Dialing in Loading Concentration

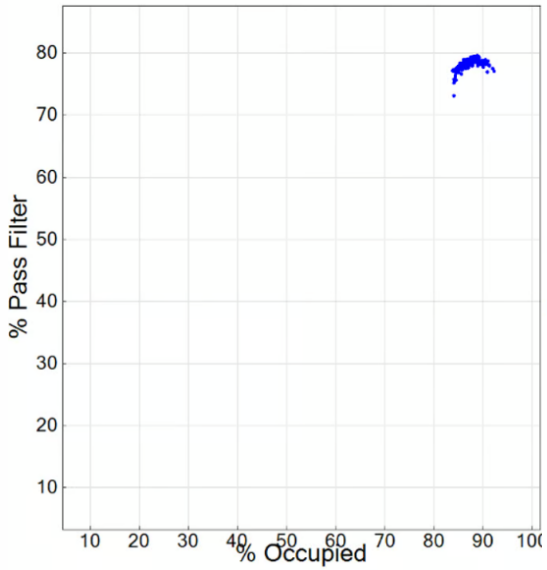
- Patterned flow cell design
  - Fixed number of nanowells where clustering can occur
  - Optimizing loading concentration - % passing filter vs. % occupied



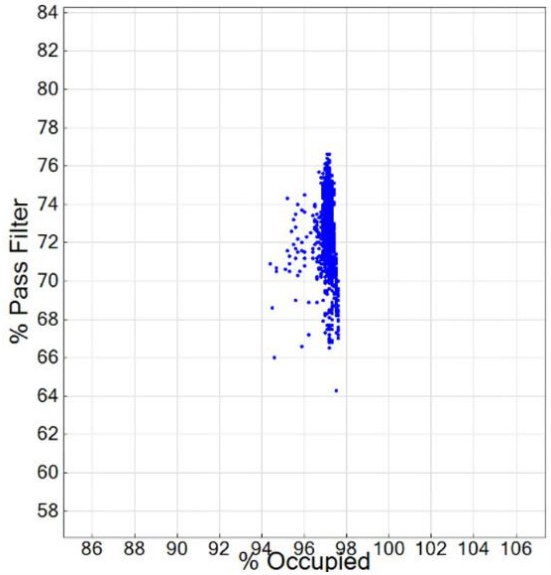
Underloaded



Optimally Loaded



Overloaded



Optimal Concentrations/Kit

Flow Cell – Cycles	Loading Concentration
25M – 300	100 pM
25M – 600	100 – 120 pM

Initial Testing: Dialing in Total Genome Load

- Testing with optimal genome load from CDC PulseNet interim guidance
- Reduced total genome load recommended for organisms requiring ≥ 40x coverage
  - *Vibrio*, STEC, healthcare associated infections (HAI) and antimicrobial resistant (AR)

MiSeq i100 Sequencing Kit (Flow cell – Cycles)	Total Genome Load (Mb) – Without Isolates ≥ 40x	Total Genome Load (Mb) – With Isolates ≥ 40x	DCLS Tested
5M – 300	40	35	No
5M – 600	75	~55	No
25M – 300	140	135	Yes
25M – 600	~270	~250	Yes

Validation Design: Validation Panels

- 25M – 300 cycle, genome load 130 Mb

Organism type	Total Number of Isolates
Foodborne bacteria	21 isolates
ARLN bacteria	3 isolates
HAI bacteria	1 isolate
VPD bacteria	1 isolate
Respiratory bacteria	2 isolates
Norovirus	7 samples (including controls)
SARS-CoV-2 Clinical	11 clinical samples
SARS-CoV-2 WW	46 samples

- 25M – 600 cycle, genome load 250 Mb

Organism type	Total Number of Isolates
Foodborne bacteria	44 isolates
ARLN bacteria	9 isolates
HAI bacteria	2 isolates
Norovirus	6 samples

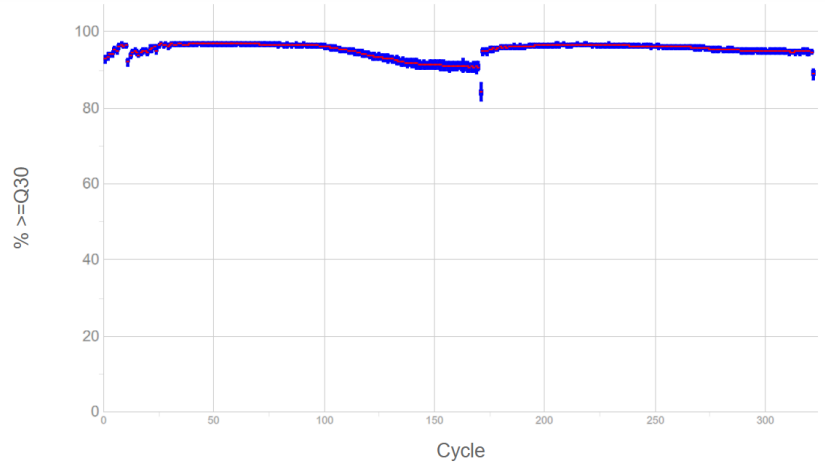
Validation Design: Performance Characteristics

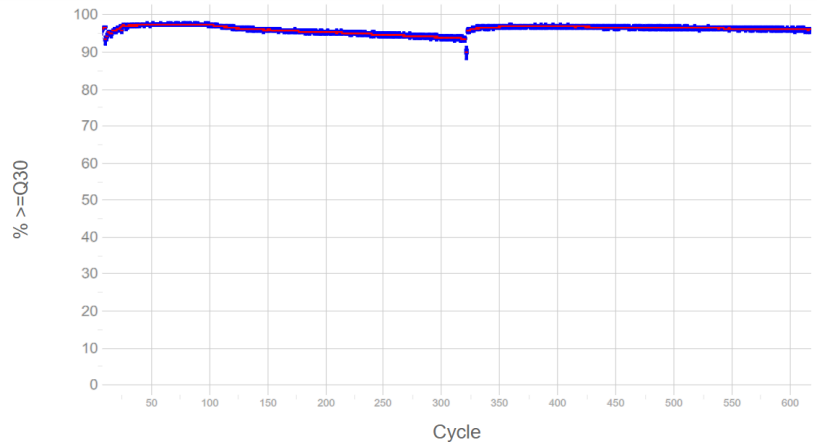
- Run-level metrics
  - Percent of reads Q score ≥ Q30,
  - Percent of reads R1/R2 ≥ Q30,
  - Percent of clusters passing filter
- Organism-level quality metrics
- Accuracy in taxonomic and genotypic predictions
- Precision assessment - replicate analysis of a subset of samples
  - Taxonomy, genome identity (allelic differences)
- Accuracy in cluster determination
  - Analysis of well-defined outbreak isolates



How it's Going: Run-Level Metrics

Run Metrics	25M – 300 (2 x 151)	25M – 600 (2 x 301)
R1 % ≥ Q30 (St. Dev.)	94.94 (0.09)	95.9 (0.05)
R2 % ≥ Q30(St. Dev.)	96.08 (0.22)	96.83 (0.18)
Mean % ≥ Q30 (St. Dev.)	95.47 (0.16)	96.26 (0.05)
% Passing Filter (St. Dev.)	81.77 (0.1)	84.27 (1.57)





How it's Going: Sample-Level metrics

- Preliminary Observations: 25M – 300 cycle, 130 Mb

Organism type	Number Tested	Number Meeting Quality Metrics	Percent Meeting Quality Metrics
Foodborne bacteria	21 isolates	18 isolates	86%*
Non-foodborne bacteria (AR, HAI, VPD, Respiratory)	6 isolates	6 isolates	100%
Norovirus	5 samples	5 samples	100%
SARS-CoV-2 Clinical	11 clinical samples	10 clinical samples	90%
SARS-CoV-2 WW	46 samples	Data under review	Data under review

*All PN organism failures were STEC isolates (n = 3)

How it's Going: Sample-Level Metrics

- Preliminary Observations: 25M – 600 cycle, 250 Mb

Organism Type	Total Number of Isolates	Number of Isolates Meeting Quality Metrics	Percent Meeting Quality Metrics
Foodborne bacteria	44 isolates	41 isolates	93%*
Non-foodborne bacteria (AR, HAI)	11 isolates	11 isolates	100%
Norovirus	5 samples	2 samples	40%**

*All PN organism failures (n=3) were STEC isolates

**Likely a pooling issue, in repeat testing 5/5 passed quality metrics

Short Read Length Warning

- Increase in PulseNet 2.0 read-length warning for 600-cycle 2 x 301 paired-end reads
  - Likely due to patterned flow-cell binding affinity for shorter fragments
  - Resolved with libraries with longer average fragment sizes (800 – 1000 bp)
  - No impact on read-based or assembly-based quality metrics

Taxonomic and Genotypic Assessment

✓ = Accurate in taxonomic and genotypic predictions compared to MiSeq results

Characterization Type	*Salmonella spp.*	*Escherichia/ Shigella*	*Vibrio spp.*	*Campylobacter spp.*	*Listeria monocytogenes*
ANI Taxonomic Identification	✓	✓	✓	✓	✓
Serotype	✓	✓	✓	N/A	N/A
Lineage Type	N/A	N/A	N/A	N/A	✓
Pathotype	N/A	✓	N/A	N/A	N/A
Toxin Gene Identification	N/A	✓	✓	N/A	N/A
Virulence Factor	N/A	✓	N/A	N/A	N/A

How it's Going: Additional Observations

- Preparing new libraries for previously sequenced samples was the most efficient approach to avoid index overlap or insufficient volumes
- PN 2.0 has not scaled with the increased sample load possible for MiSeq i100, resulting in data analysis bottlenecks and delays
- .BCL convert must be run to generate fastq files – either through selection at the start of the run or using the DRAGEN application on BaseSpace
- Take time to orient to where data is accessible between the instrument and BaseSpace
  - Location of fastq files for secondary analysis

Next Steps

- Determine acceptability of read-lengths for foodborne pathogens and other bacteria
- Add accuracy assessments of sequence type (ST) and antimicrobial resistance genes
- Complete precision assessment
- Assess the accuracy of cluster detection from MiSeq i100 data for known outbreaks
- Investigate increased failures for SARS-CoV-2 libraries and expanding to include other viral targeted amplicon libraries
- Finalize the validation summary

Acknowledgements

DCLS Sequencing and Bioinformatics Group

Mary Toothman, MS

Mackenzie Murphy, MS

Lena Washington

Corey Russell

Brenham Howard

Brenay Friend

Josh Caythe

Darrius Sykes

Molly Creighton, MS

North Carolina State Laboratory of Public Health

Jordan Ferguson

PulseNet Working Group Coordinators

Amy Bryant, MPH

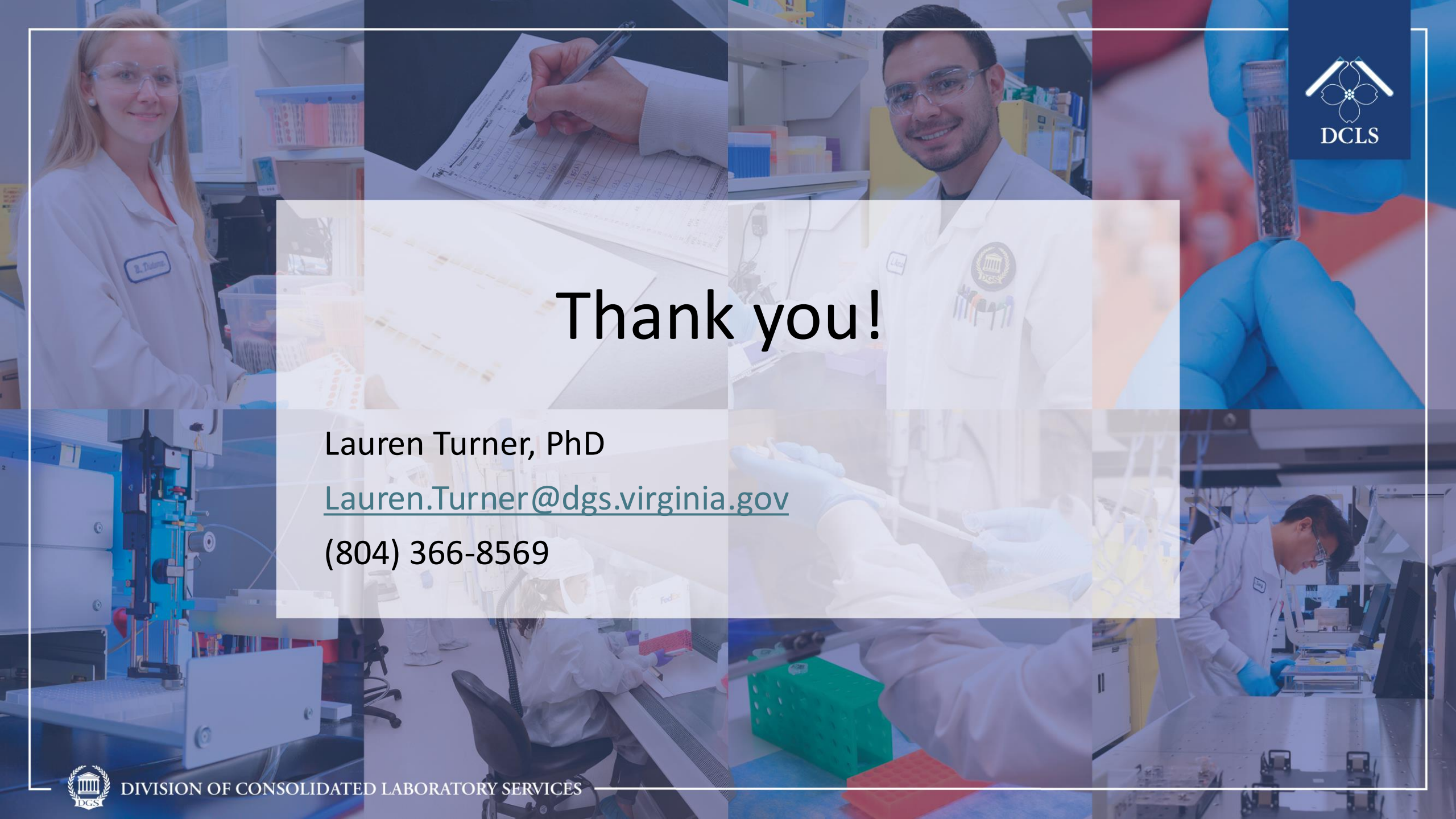
Courtney Wheeler, MPH

Patti LaFon

Illumina

Jason Smith

Kristal Young



Thank you!

Lauren Turner, PhD

Lauren.Turner@dgs.virginia.gov

(804) 366-8569

DCLS