STR Typing Workflows on the Ion Torrent S5 XL Sequencing Platform

Kevin Kiesler, M.S.
Research Biologist, Applied Genetics Group, NIST
Green Mountain DNA Conference
August 1, 2018

Outline

- Using NGS
- Sequencing STR loci
- S5-Ion Chef workflows
- Using Converge software for CE-NGS data comparisons

Sequencing STRs: Plusses and Minuses

+ Plus
- Smaller amplicons
  - Not constrained by CE size separation
  - Better for degraded DNA
- Improved mixture deconvolution *potential*
  - Discriminate shared alleles & stutter masking by sequence differences

- Minus
- Higher cost vs CE
- More complex procedure
- Novel analysis methods, nomenclature
- Longer time to result
- New methods require training, validation, implementation

Why Use S5 XL & Ion Chef?

- Modular panels -vs- megaplex
  - Lower complexity → improved balance & sensitivity
    - Large multiplexes give more information, but tricky to optimize
- Thermo Fisher Panels
  - STR
    - Precision ID GlobalFiler NGS (v2 panel for S5 instrument)
  - SNP
    - Precision ID Ancestry Panel
    - Precision ID Identity Panel (includes Y clade SNPs)
  - Mitochondrial DNA
    - Precision ID Whole Genome Panel
    - Precision ID Control Region Panel
- Ion Chef
  - Full automation of library construction and templating procedures
  - Developed and supported by ThermoFisher

NIST's Experiences Sequencing STRs at Scale

- 2014 Promega PowerSeq (24-plex)
  - 183 samples from NIST population set
    - African American, Asian, Caucasian, Hispanic
  - Collaboration with Battelle Institute
  - Wet lab time: ~ two weeks
  - Goal: initial characterization of sequence variation

2016 paper

- 2016 ForenSeq Signature Kit
  - 1036 samples from NIST population set
  - Wet lab time: ~ nine months
  - Goal: U.S. population allele frequencies

2018 paper

NIST's Experiences Sequencing STRs at Scale

- 2017 Promega PowerSeq 46GY – collaboration with Promega
  - 665 samples from NIST population set
  - Time to complete sequencing: ~ two months
  - Goal: assist with assay development, informatics
- 2017 Precision ID Globalfiler NGS (v2) – collaboration with ThermoFisher
  - 88 Japanese from 5 prefectures
  - Time to complete sequencing: ~ one month
  - Goal: population allele frequencies - Japanese
- 2018 Precision ID Globalfiler NGS (v2) – collaboration with ThermoFisher
  - 235 samples (n = 123 Asian, 98 Hispanic, 14 "Challenging")
  - Time to complete sequencing: ~ two months
  - Goal: population allele frequencies – U.S. Asian & Hispanic, also assay development

Library Construction With Ion Chef

8 samples/run

Run ≈ 7.5 hours

Fully automated

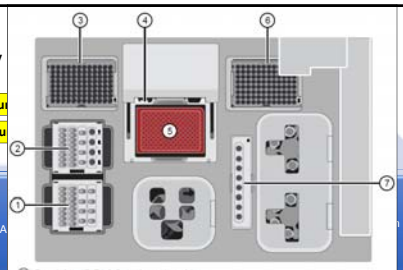
1 ng DNA Input → PCR Amplification → Partially Digest Amplicons → Ligate Adaptor/Barcode → Purification → Pool Libraries (8)

Manual Step → Library Quantitation qPCR

Library

8 samples/ru

Run ≈ 7.5 hou



① Precision ID DL8 Solutions cartridge
② Precision ID DL8 Reagents cartridge
③ Ion AmpliSeq™ Tip Cartridge L8
④ Framed PCR Foil Seal
⑤ Precision ID IonCode™ Barcode Adapters 96 Well PCR Plate
⑥ Empty Tip Cartridge L8

1 ng DNA Input → … → Pool Libraries (8)

Manual Step → Library Quantitation qPCR

Library Constr

8 samples/run

Run ≈ 7.5 hours

1 ng DNA Input → PCR Amplification

GlobalFiler Loci in CE AND NGS kits	In NGS NOT CE	In CE NOT NGS
D1S1656	D1S1677	SE33
TPOX	D2S1776	
D2S441	D3S4529	
D2S1338	D4S2408	
D3S1358	D5S2800	
FGA	D6S1043	
D5S818	D6S474	
CSF1PO	D12ATA63	
D7S820	D14S1434	
D8S1179	Penta_E	
D10S1248	Penta_D	
TH01	SRY	
vWA		
D12S391		
D13S317		
D16S539		
D18S51		
D19S433		
D21S11		
D22S1045		
AMELX		
AMELY		
DYS391		
Yindel		

11 additional STRs + Y marker "SRY"

All Globalfiler Loci (Except SE33)

Image credit: Precision ID GlobalFiler™ NGS STR Panel v2 with the Ion S5™ System Application Guide

Library Construction With Ion Chef

8 samples/run

Run ≈ 7.5 hours

Automated library cost = $1800/run ≈ $225/sample

Manual library cost ≈ $120/sample + 8 hours labor

Fully automated

1 ng DNA Input → PCR Amplification → Partially Digest Amplicons → Ligate Adaptor/Barcode → Purification → Pool Libraries (8)

Next Step: Sequencing Template Preparation on Ion Chef

Time ≈ 12 hours (overnight) | **Up to 32 Samples Per chip (GlobalFiler)** | **Sequencing reagents ≈ $935/run Ion 530 chip (x2) ≈ $550 each (x2)**

Manual Step → Library Quantitation qPCR

Templating Process on Ion Chef

Time ≈ 12 hours (overnight)

Up to 32 samples Per chip (GlobalFiler)

Sequencing reagents ≈ $935/run Ion 530 chip (x2) ≈ $550 each (x2)

Transfer chip to S5

Fully automated

Input Libraries: 50 pM normalized SUPERPOOL → Emulsion PCR With Ion Sphere Particles (ISPs) → Emulsion Breaking → Enrich for Templated ISPs → Load Chips (2) → Sequence

Templating Process on Ion Chef

Time ≈ 12 hou (overnight)

Input Libraries: 50 pM normalized SUPERPOOL

Create Superpool with Normalized Libraries (@50 pM)

Library Build #1 (Barcodes 1-8) + Library Build #2 (Barcodes 9-16) + Library Build #3 (Barcodes 17-24) + Library Build #4 (Barcodes 25-32) → SUPERPOOL #1

Library Build #5 (Barcodes 1-8) + Library Build #6 (Barcodes 9-16) + Library Build #7 (Barcodes 17-24) + Library Build #8 (Barcodes 25-32) → SUPERPOOL #2

Templating Process

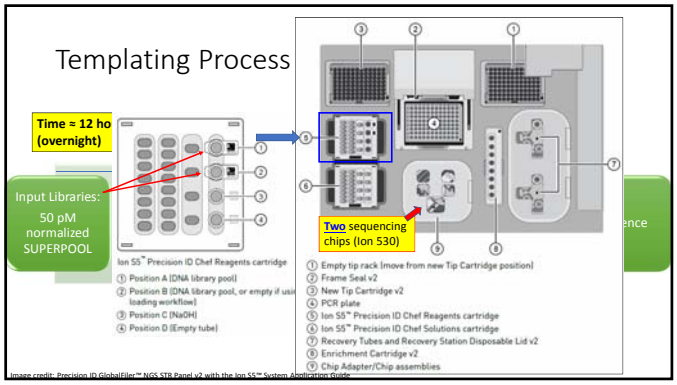
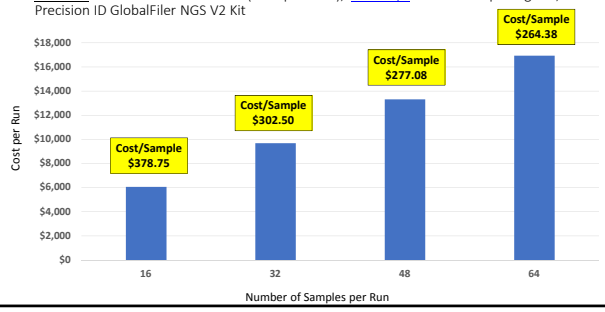


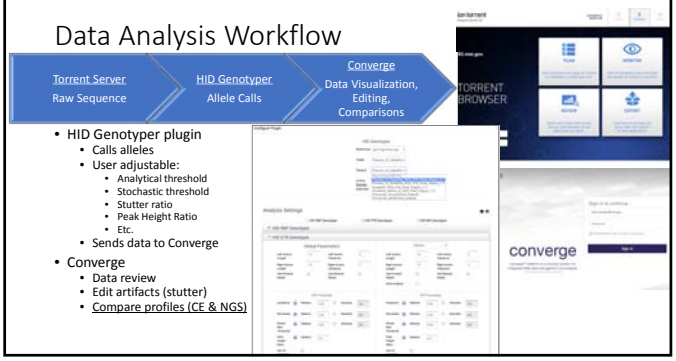
Image credit: Precision ID GlobalFiler™ NGS STR Panel v2 with the Ion S5™ System Application Guide

Cost Per Sequencing Run vs # Samples

Assumes: DL8 automated libraries (multiples of 8), two chips on Chef templating run, Precision ID GlobalFiler NGS V2 Kit



Data Analysis Workflow



- HID Genotyper plugin
  - Calls alleles
  - User adjustable:
    - Analytical threshold
    - Stochastic threshold
    - Stutter ratio
    - Peak Height Ratio
    - Etc.
  - Sends data to Converge
- Converge
  - Data review
  - Edit artifacts (stutter)
  - Compare profiles (CE & NGS)

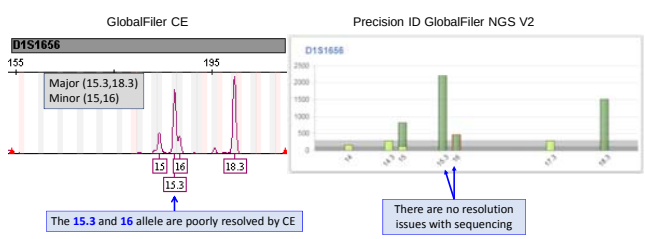
Profile Comparison: Concordance Testing

- Sample intentionally selected due to known deletion in D5S818:
  - Flanking 4 bp deletion results in dropout in GlobalFiler CE: (genotype 12, 12)
  - Correctly typed by GlobalFiler NGS: (genotype 8, 12)
    - Corroborated by Illumina ForenSeq (genotype 8, 12)

	GlobalFiler CE Genotype	GlobalFiler NGS Genotype
D5S818	12, 12	8, 12
D13S317	11, 13	11, 13
D16S539	9, 11	9, 11
D7S820	10, 12	10, 12

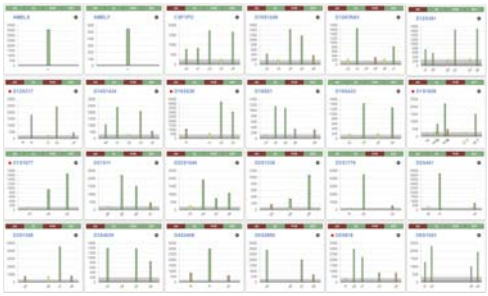
Sequencing Eliminates Single Nucleotide Resolution Issues

Locus: D1S1656

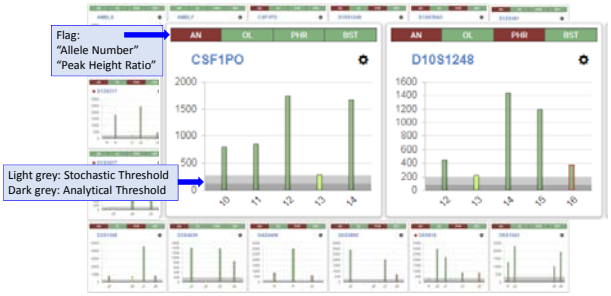


NIST SRM 2391d Component D (3:1 mixture), candidate material

Mixture Data (3 to 1 Ratio)

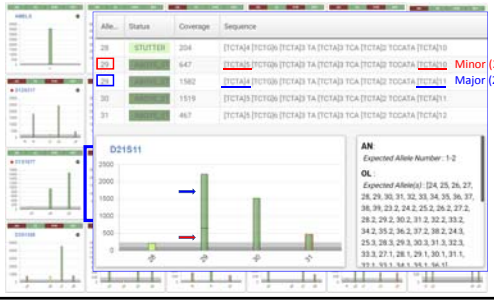


Mixture Data (3 to 1 Ratio)



Mixture Data - D21S11 Allele 29 "Isoallele"

Mixture Data - D21S11 Allele 29 "Isoallele"



Conclusions/summary

- STR Sequencing
  - Increased information – variation in STR, flanking SNPs
    - Useful with difficult casework samples
    - Mixtures, degraded DNA
- Ion Chef
  - Saves labor
    - Cost vs time trade-off
  - Increases reproducibility, reliability
- Converge
  - Visualize & edit data
  - Data comparison (CE/NGS)

Thank You!

Kevin.Kiesler@NIST.gov

√ Acknowledgements √

Thermo Fisher Scientific
Matt Gabriel
Rob Lagacé

NIST Applied Genetics Group
Dr. Peter Vallone, Group leader
Sarah Riman
Lisa Borsuk
Becky Steffen
Katherine Gettings

- Funding
  - NIST Special Programs Office: *Forensic DNA*
  - FBI Biometrics Center of Excellence: *Forensic DNA Typing as a Biometric tool.*

- **Disclaimer** - Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce or the Department of Justice. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

- All work presented has been reviewed and approved by the NIST Human Subjects Protections Office.