Forensic DNA Standards for Next Generation Sequencing Platforms

7th Annual
Sequencing, Finishing, and Analysis in the Future Meeting
Santa Fe, New Mexico
June 8, 2012

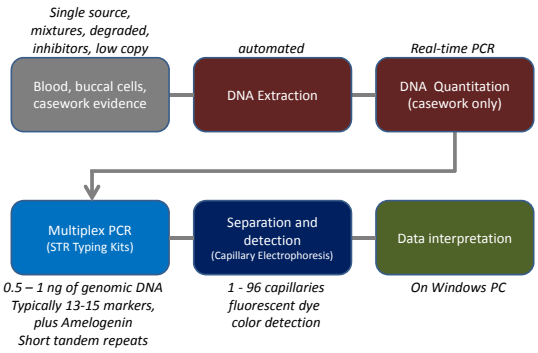
Dr. Peter M. Vallone, Carolyn R. Hill, Erica L.R. Butts, David L. Duewer, John M. Butler, and Margaret C. Kline

Biochemical Science Division
National Institute of Standards and Technology

Next Generation Sequencing Workshop held January 31, 2012

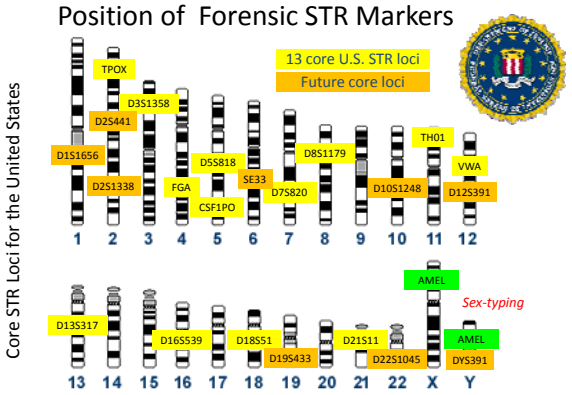
- **Interagency Workshop on the use of Next-Generation DNA Sequencing for Human Identification and Characterization**
- Discussion of forensic applications of NGS (NIST, DoD, FBI, DHS) – materials can be found at:
  - http://www.nist.gov/mml/biochemical/genetics/ngs_hid_workshop.cfm
- Beginning to assess platforms for characterizing forensic markers (STRs, mitochondrial genome, SNPs)
- Evaluate accuracy, reproducibility, **identify initial requirements for a NGS forensic reference material**

Forensic DNA Typing Process



http://www.cstl.nist.gov/biotech/strbase/pub_pres/Vallone_SFAF_June_8.pdf

Position of Forensic STR Markers



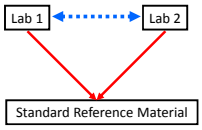
NIST Standard Reference Materials

http://www.cstl.nist.gov/biotech/strbase/srm_tab.htm

Traceable standards to ensure accurate measurements in our nation's crime laboratories

Helps meet QAS Std. 9.5.5 and ISO 17025

Just released 2391c with expanded CODIS core loci and Y-STRs



SRM 2391c – CODIS core STRs
SRM 2392 & 2392-I – mtDNA
SRM 2395 – Y-STRs
SRM 2372 – Human DNA quantitation

Calibration with SRMs enables confidence in comparisons of results between laboratories

Description of Components in SRM 2391c

Component	Description
A	50 µL of anonymous **female** genomic DNA
B	50 µL of anonymous **male** genomic DNA
C	50 µL of anonymous **male** genomic DNA
D	50 µL of **mixed-source** (Components A and C)
E	Two 6 mm punches of CRL-1486 cells spotted on **903 paper**
F	Two 6 mm punches of HTB-157 cells spotted on **FTA paper**

Liquid components ≈ 2 ng/µL
Paper 75,00 cells per spot

2391c contains confirmed genotypes for:

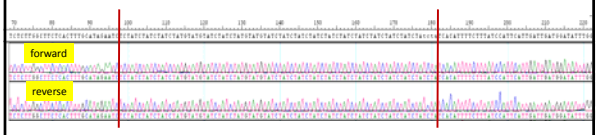
51 autosomal STR loci (includes FBI core loci, European core loci, others)

17 Y-STR loci

Allele calls (repeats) are confirmed by Sanger sequencing performed at NIST

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Vallone_SFAF_June_8.pdf

SRM 2391c, Component C, DYS635



- Repeat motif is confirmed by Sanger sequencing
- Loci amplified (heterozygous samples separated on a gel) and sequenced

Component C (21) =
$[TCTA]_4[TGTA]_2[TCTA]_2[TGTA]_2[TCTA]_{11}$

Next Generation Sequencing

Forensic Applications

- Going in depth into STR loci and beyond
  - STRs are useful for legacy (databases)
  - SNPs within STRs identify 'sub-alleles'
  - Millions of bases of sequence variants (SNPs)
- Opens up new human identity applications: biogeographical ancestry, externally visible traits, complex kinship, degraded samples, mixtures, low template, and other applications

Applications are currently being addressed by the forensic genetics community (*Kayser and deKnijff 2011)*

Next Generation Sequencing

- Challenges
  - Repeating sequences (STRs) and read lengths
  - Sample amount requirements (> 1 ng)
  - **Cost** and **time** per unit of information
  - Data analysis (storage, assembly, interpretation)
  - Policy, privacy, disease related markers
  - Validation
  - Standards/reference materials
    - Nomenclature
    - Accuracy of sequence information
    - Errors, platform and bioinformatics-based bias

Requirements for a NGS forensic SRM?

Information gathering stage

- Materials:
  - Genomic DNAs?
  - Cell line DNAs?
  - Use current forensic SRMs?
  - PCR amplicons of forensically relevant markers?
- How many components?
  - Family samples – paternity trio?
  - Mixtures?
- How much material is needed?
- Certify additional markers beyond core STR loci

- Is a full genome standard needed for forensic applications?
- Is it enough to fully characterize 'core' loci and have good/high confidence in other non-core loci?

Ongoing discussion with Multiplex Biomolecular Science group 'Genome in a bottle' consortium – Justin Zook and Marc Salit

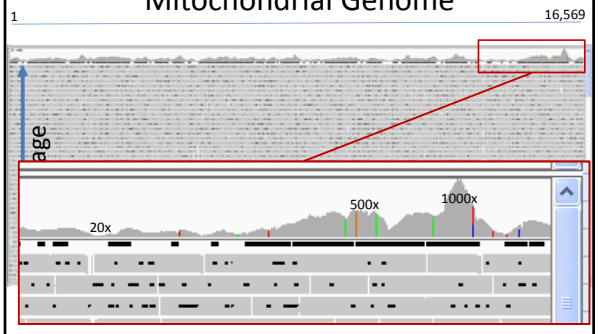
Characterization of a NGS forensic SRM?

- Involve multiple platforms/technologies?
  - Short and long read technologies
  - Genomic versus targeted (PCR product) sequencing
  - Interlaboratory studies (pilot prior to SRM production)
- For the core forensic markers (STRs, mito)
  - Continue to include Sanger confirmation
  - Develop specific primer sets for core loci?
- Call SNPs based on **consensus** from multiple platforms and assembly algorithms

Current Focus - Characterization of SRMs 2392 and 2392-I

- Used for mitochondrial DNA sequencing
- The mtGenome was amplified using an overlapping primer pair strategy
- The amplicons were purified, quantitated and pooled in an equimolar mixture (100 ng)
- Platforms
  - **Ion Torrent (Edge Biosystems)**
  - 5500xl (in house, to be completed this month)
  - Illumina (Beckman, results soon)
  - Hope to have multiple systems in house (2013)

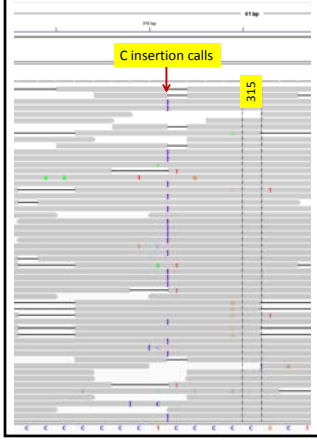
Coverage of the Mitochondrial Genome



Human HL-60

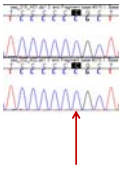
- Sole component of NIST SRM 2392-I
- NGS accurately called all sequence variants compared to the rCRS (33 variants)

- 315.1 C insertion (homopolymer stretch)
- Heteroplasmy at 12,071 (correctly called)
- Heteroplasmy at 16,069? (artifact of PCR)



C Insertion at 315.1

Sanger call
315.1 C insertion



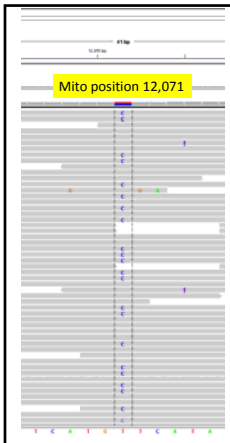
NGS call
'No call'

BUT

A C insertion is detected (about 60 %) after position 310 (but not called by the software)

Issues with homopolymers

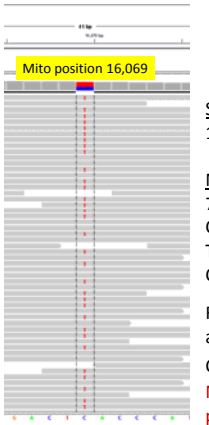
http://www.cstl.nist.gov/biotech/strbase/pub_pres/Vallone_SFAF_June_8.pdf



Heteroplasmy at 12,071

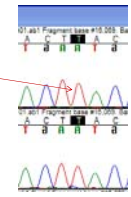
Sanger call
12,071 T/C heteroplasmy

NGS call
321 total reads
C: 168 calls
T: 153 calls
T:C ratio ≈ 0.9



Heteroplasmy at 16,069?

Sanger Call
16,069 T



NGS Call
726 reads
C: 320
T: 440
C:T ratio ≈0.7

Reverse primer for PCR (R16084) contains a G at position 16,069
CGGTTGTTGATGGGT**G**AGTC
Not heteroplasmy, but an artifact of the PCR primers used to enrich the materials

Thank you for your attention!

Questions?
peter.vallone@nist.gov
301-975-4872

Acknowledgements
Kevin Kiesler and Mike Coble

Outside funding agencies:
FBI - Evaluation of Forensic DNA Typing as a Biometric Tool
NIJ – Interagency Agreement with the Office of Law Enforcement Standards

http://www.cstl.nist.gov/biotech/strbase/pub_pres/Vallone_SFAF_June_8.pdf