Use of NGS for forensic applications

Highly-parallel/high-throughput next-generation sequencing technologies provide the ability to directly sequence forensically relevant targets

Issues: sample input amounts, back compatibility, new workflows, cost, throughput, etc

- Mitochondrial whole genome analysis
  - Potential for improved sensitivity, mixture detection, multiplex sequencing of full mitochondrial genomes
  - Detection of minor SNP variants – heteroplasmy
- Going in depth into STR loci
  - STRs are useful for legacy (databases)
  - SNPs within STRs identify 'sub-alleles'
- Forensically relevant SNPs: newer human identity applications: biogeographical ancestry, externally visible traits, complex kinship, degraded samples, low template

Initial Goals

- To characterize NIST forensic SRMs with NGS
  - Further characterizes the materials with a new technique
  - Supports adoption of NGS in forensic community
  - SRM 2391c: PCR Based DNA Profiling Standard
    - Not all STR loci have full sequence information
  - SRMs 2392 and 2392-I: Mitochondrial DNA Sequencing
    - Confirm Sanger data with a high coverage sequencing technology
  - Understand bias between NGS platforms: chemistry and bioinformatics
- Is there a need for a new material?
  - Forensic validation

NIST Standard Reference Materials

http://www.nist.gov/srm/

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

Human Identity SRMs

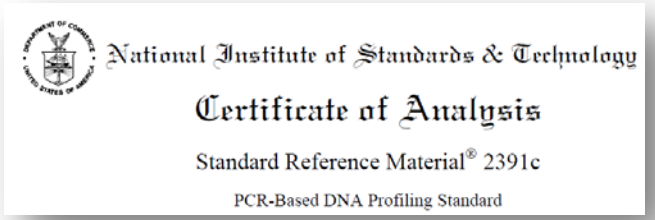
SRM 2391c – PCR-Based DNA Profiling

SRM 2392 & 2392-I – mitochondrial DNA

SRM 2395 – Y-STR DNA Profiling

SRM 2372 – Human DNA quantitation

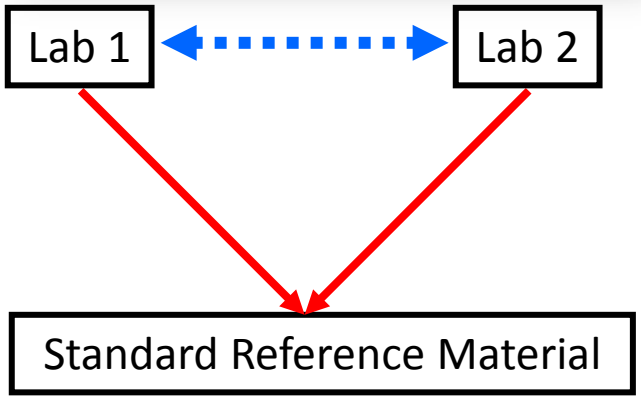
Genomic DNAs characterized for the expanded CODIS core loci and Y-STRs


National Institute of Standards & Technology

Certificate of Analysis

Standard Reference Material® 2391c

PCR-Based DNA Profiling Standard



Calibration with SRMs enables confidence in comparisons of results between laboratories and technologies

Characterization of the existing SRMs

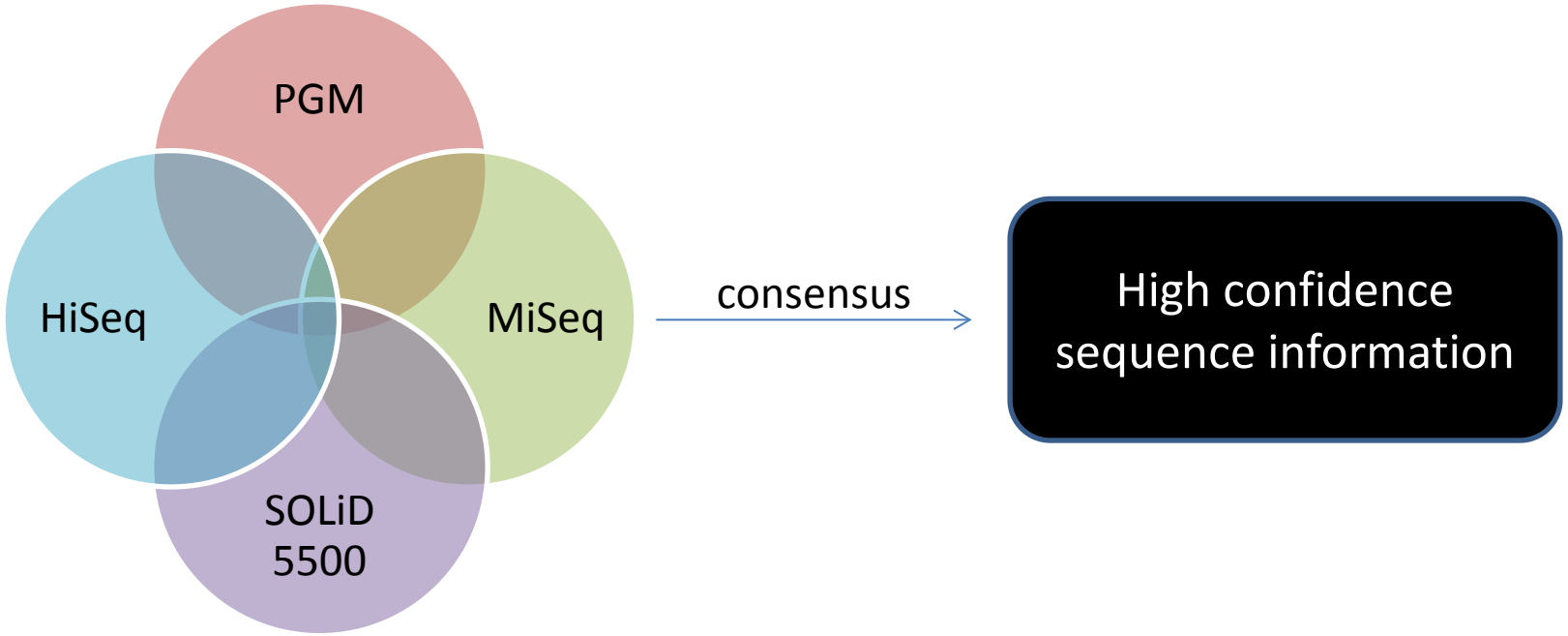
Current Status

- 2391c **PCR Based DNA profiling standard**
  - 68 STR markers (51 autosomal + 17 Y chromosome)
  - STR repeat lengths (alleles) were certified using multiple (unique) PCR primer sets
  - Sanger sequencing was only performed for loci without multiple PCR primer sets (only 10% of markers)

- 2392 & 2392-I **Mitochondrial DNA sequencing standard**
  - Entire mtGenome (≈16,569 bp) was certified by Sanger sequencing
  - Multiple F/R strand coverage across the mtGenome

Multiple NGS Platforms

- Use of multiple platforms to obtain a consensus sequence for the SRMs
  - Identify and reduce the false positives and negatives
  - Identify and account for bias in a specific chemistry and/or informatics pipeline

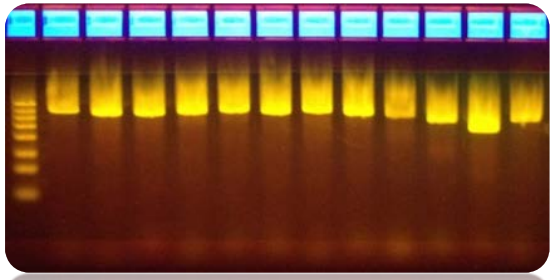


Sequencing Studies

Performed on four NGS platforms for SRM 2392 and 2392-I

- **Ion Torrent PGM**
  - Experiments performed at NIST
  - Edge Biosystems (outsourced)
- **Illumina HiSeq 2000**
  - Beckman-Coulter Genomics (outsourced)
- **Illumina MiSeq**
  - Edge Biosystems (outsourced)
- **SOLiD 5500**
  - Experiments performed at NIST

12 Amplicon PCR
0.8 kb to 1.9 kb



Illumina MiSeq and HiSeq platforms will be online at NIST by the end of this year

Data Processing Pipeline:

Signal Processing, Alignment, and Variant Calling

	Ion Torrent PGM	Illumina MiSeq	Illumina HiSeq	SOLiD 5500
Signal Processing Output: FASTQ	Torrent Server	MiSeq Reporter	HiSeq Control	LifeScope
Read Mapping Output: BAM	Torrent Server	Novoalign	BWA	LifeScope
Variant Calling Output: VCF	Torrent Server	GATK	GATK	GATK

Abbreviations:
FASTQ – Unaligned reads in text format with quality scores
BAM – Binary Alignment Map (Aligned reads)
VCF – Variant Call File
BWA – Burrows Wheeler Aligner
GATK – Genome Analysis Tool Kit

Sequence Coverage Summary

Experiment	Average Read Depth (AQ20*)	Experiment Design
EdgeBio PGM	280 x	Seven mtGenomes + spike-in controls**
NIST PGM Run 1	6,500 x	Three mtGenomes
NIST PGM Run 2	9,000 x	Three mtGenomes
Illumina MiSeq	49,000 x	Seven mtGenomes
Illumina HiSeq	41,000 x	Seven mtGenomes + spike-in controls**
NIST SOLiD	29,000 x	Seven mtGenomes + spike-in controls**

* AQ20 = reads with alignment quality score of 20 or above
= less than 1 error per 100 bases

**Spike-in control was NIST SRM 2374: DNA Sequence Library for External RNA Controls

False Positives and False Negatives

Using platform specific informatics pipeline

		PGM 1	PGM 2	PGM 3	HiSeq	MiSeq	5500
9947A	FP	1	5	3	21	9	11
	FN	3	4	3	3	3	3
CHR	FP	2	6	10	21	9	10
	FN	3	5	4	3	3	4
HL-60	FP	1	8	8	20	9	8
	FN	1	2	1	1	1	1
Avg Coverage		280	6,500	9,000	49,000	41,000	29,000

Calls made to the rCRS

On average 0.04 % error rate

False Positives and False Negatives

Using platform specific informatics pipeline

		PGM 1	PGM 2	PGM 3	HiSeq	MiSeq	5500
9947A	FP	1	5	3	21	9	11
	FN	3	4	3	3	3	3
CHR	FP	2	6	10	21	9	10
	FN	3	5	4	3	3	4
HL-60	FP	1	8	8	20	9	8
	FN	1	2	1	1	1	1
Avg Coverage		280	6,500	9,000	49,000	41,000	29,000

False negatives were concentrated in C stretch regions of the genome
The FN sites 13,759 and 5,228 were due to low coverage

9947A (FN) = 309.1, 309.2, 315.1, 13,759
CHR (FN) = 309.1, 315.1, 16193.1, 16183, 16189
HL-60 (FN) = 315.1, 5,228

Analysis of False Positives for 9947A

- Surrounding HV1/HV2 homopolymer
- HV3 CA repeat
- Surrounding 3,107 placeholder deletion

PGM 1	PGM 2	PGM 3	HiSeq	MiSeq	5500
					298
					299
			300		300
			301		301
			302		
			302	302	302
				308	
			309		309
			310		
			310	310	310
					347
			360		
				513	
				515	
			1992		
			3103		
			3104		
			3105		
			3105		
	3106			3106	
					4722
			4796		
	5744				
			6419		
					6482
	7860				
			8163		
			9753		
		11512			
				12417	
				12418	
		13045			
	13058				
			14188		
14199	14199	14199	14199	14199	
			15259		
					15284
			15877		
1	5	3	21	9	11

Total false positives

PGM 1	PGM 2	PGM 3	HiSeq	MiSeq	5500
					298
					299
			300		300
			301		301
			302		
			302	302	302
				308	
			309		309
			310		
			310	310	310
					347
			360		
				513	
				515	
			1992		
			3103		
			3104		
			3105		
			3105		
	3106			3106	

Analysis of False Positives for 9947A

- Majority are low level (< 5%) in red
- 14,199 PCR primer artifact (consensus) in blue
- The remainder of sites are not reproducible across platforms
- Low confidence for false positive calls (no strong consensus)

PGM 1	PGM 2	PGM 3	HiSeq	MiSeq	5500
					298
					299
			300		300
			301		301
			302		
			302	302	302
				308	
			309		309
			310		
			310	310	310
					347
			360		
				513	
				515	
			1992		
			3103		
			3104		
			3105		
			3105		
	3106			3106	
					4722
			4796		
	5744				
			6419		
					6482
	7860				
			8163		
			9753		
		11512			
				12417	
				12418	
		13045			
	13058				
			14188		
14199	14199	14199	14199	14199	
			15259		
					15284
			15877		
1	5	3	21	9	11

Total false positives

					4722
			4796		
	5744				
			6419		
					6482
	7860				
			8163		
			9753		
		11512			
				12417	
				12418	
		13045			
	13058				
			14188		
14199	14199	14199	14199	14199	
			15259		
					15284
			15877		
1	5	3	21	9	11

Analysis of False Positives for CHR and HL-60

CHR					
PGM 1	PGM 2	PGM 3	HiSeq	MiSeq	5500
					298
			299		
					299
					300
			301		301
			302	302	302
			309		309
310			310	310	310
			310	310	
			360		
				515	
				639	
			1992		
			3103		
			3104		
		3105			
			3105		
			3105		
	3106	3106		3106	
					4547
					4722
			4796		
	5744	5744			
	6220	6220			
			6419		
			8163		
	8230	8230			
		9546	9753		
		11512			
					11826
		12417		12417	
				12418	
	12704				
		13045			
			14188		
14199	14199	14199	14199	14199	
			15259		
			15877		
			16182		
2	6	10	21	9	10

HL-60					
PGM 1	PGM 2	PGM 3	HiSeq	MiSeq	5500
				152	
			302		
			309		
					309
			310	310	310
			310	310	
			360		
				515	
			1992		
	2445	2445			
			3103		
			3104		
			3105		
			3105		
	3106			3106	
					3476
					4547
					4722
			4796		
	5149	5149			
	5297				
			6419		
					7508
			8163		
	8230	8230			
		8695			
		9541			
			9753		
		11512			
					11787
					11826
		12417			
				12417	
				12418	
	13058				
			14188		
14199	14199	14199	14199	14199	
			15259		
			15877		
	16361				
			16564		
			16565		
				16568	
1	8	8	20	9	8

Similar issues with homopolymers, low abundance variants, PCR primer artifacts

Again, the remainder of the false positive sites are not reproducible across platforms

Variant Calls – Concordance by Consensus

SRM 2392 Component B (9947A)

Nucleotide Position	rCRS Reference Sequence	SRM 2392 Component B Sanger Call	EdgeBio PGM	NIST PGM run 1	NIST PGM run 2	EdgeBio Illumina MiSeq	Beckman Genomics Illumina HiSeq	NIST SOLiD
93	A	G	G	G	G	G	G	G
195	T	C	C	C	C	C	C	C
214	A	G	G	G	G	G	G	G
263	A	G	G	G	G	G	G	G
309.1	:	C						
309.2	:	C						
315.1	:	C						
750	A	G	G	G	G	G	G	G
1393	G	G	G/A	G/A	G/A	G/A	G/A	G/A
1438	A	G	G	G	G	G	G	G
4135	T	C	C	C	C	C	C	C
4769	A	G	G	G	G	G	G	G
7645	T	C	C	C	C	C	C	C
7861	T	C	T/C	T/C	T/C	T/C	T/C	T/C
8448	T	C	C	C	C	C	C	C
8860	A	G	G	G	G	G	G	G
9315	T	C	C	C	C	C	C	C
13572	T	C	C	C	C	C	C	C
13759	G	A	A		A	A	A	A
15326	A	G	G	G	G	G	G	G
16311	T	C	C	C	C	C	C	C
16519	T	C	C	C	C	C	C	C

Variant Calls – Concordance by Consensus

SRM 2392 Component B (9947A)

Nucleotide Position	rCRS Reference Sequence	SRM 2392 Component B Sanger Call	EdgeBio PGM	NIST PGM run 1	NIST PGM run 2	EdgeBio Illumina MiSeq	Beckman Genomics Illumina HiSeq	NIST SOLiD
93	A	G	G	G	G	G	G	G
195	T	C	C	C	C	C	C	C
214	A	G	G	G	G	G	G	G
263	A	G	G	G	G	G	G	G
309.1	:	C						
309.2	:	C						
315.1	:	C						
750	A	G	G	G	G	G	G	G
1393								G/A
1438								G
4135								C
4769								G
7645								C
7861								T/C
8448	T	C	C	C	C	C	C	C
8860	A	G	G	G	G	G	G	G
9315	T	C	C	C	C	C	C	C
13572	T	C	C	C	C	C	C	C
13759	G	A	A		A	A	A	A
15326	A	G	G	G	G	G	G	G
16311	T	C	C	C	C	C	C	C
16519	T	C	C	C	C	C	C	C

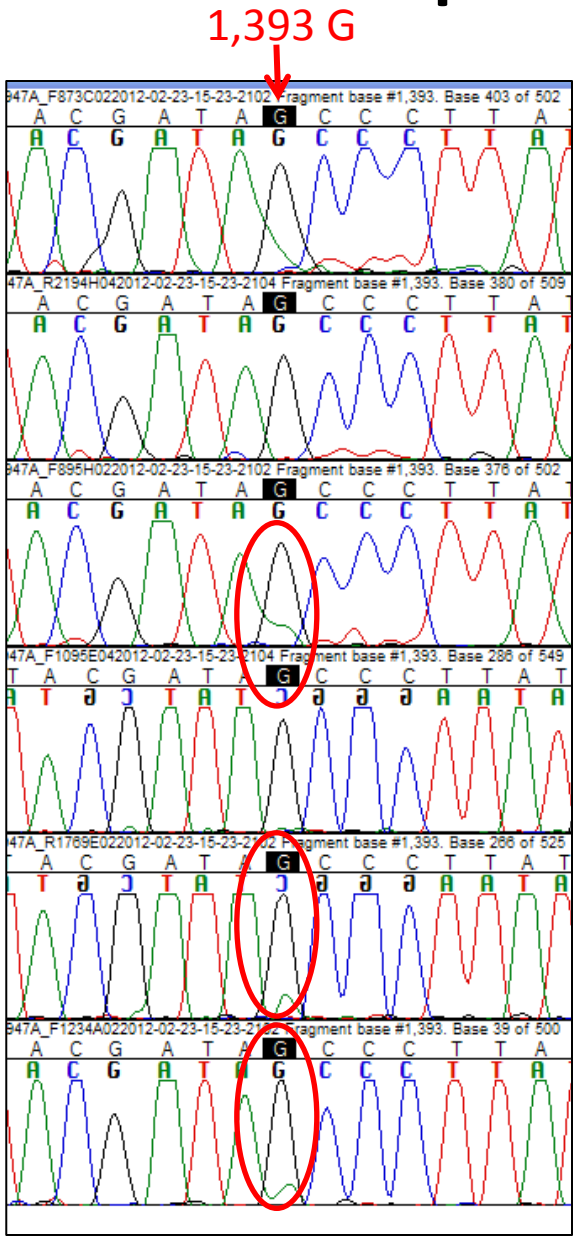
All polymorphisms from Sanger sequencing confirmed
Exception: C-stretch insertions/deletions
Future variant caller algorithms may improve in/del performance

Heteroplasmy at Positions 1,393 and 7,861

SRM 2392 Component B (9947A)

Nucleotide Position	rCRS Reference Sequence	SRM 2392 Component B Sanger Call	EdgeBio PGM	NIST PGM run 1	NIST PGM run 2	EdgeBio Illumina MiSeq	Beckman Genomics Illumina HiSeq	NIST SOLiD
93	A	G	G	G	G	G	G	G
195	T	C	C	C	C	C	C	C
214	A	G	G	G	G	G	G	G
263	A	G	G	G	G	G	G	G
309.1	:	C						
309.2	:	C						
315.1	:	C						
750	A	G	G	G	G	G	G	G
1393	G	G	G/A	G/A	G/A	G/A	G/A	G/A
1438	A	G	G	G	G	G	G	G
4135	T	C	C	C	C	C	C	C
4769	A	G	G	G	G	G	G	G
7645	T	C	C	C	C	C	C	C
7861	T	C	T/C	T/C	T/C	T/C	T/C	T/C
8448	T	C	C	C	C	C	C	C
8860	A	G	G	G	G	G	G	G
9315	T	C	C	C	C	C	C	C
13572	T	C	C	C	C	C	C	C
13759	G	A	A		A	A	A	A
15326	A	G	G	G	G	G	G	G
16311	T	C	C	C	C	C	C	C
16519	T	C	C	C	C	C	C	C

Heteroplasmy at 1,393?



- 6x coverage by Sanger
- 3/6 of reads indicate low-level heteroplasmy
  - Red circles
- Not reproducible in all reads
  - Not always detected by Sanger sequencing

Heteroplasmy detected by NGS at Site 1,393 and 7,861

- Agreement across platforms (high confidence)

  1,393 ≈ 18.0% (± 2.2%) minor component "A"

  7,861 ≈ 14.2% (± 2.9%) minor component "T"

Experiment	1,393 "G" (rCRS)	1,393 "A" (Var.)	Coverage	7,861 "T" (rCRS)	7,861 "C" (Var.)	Coverage
PGM at EdgeBio	77.3 %	22.7 %	97 x	14.1 %	85.9 %	71 x
PGM NIST Run 1	83.9 %	16.1 %	1385 x	20.0 %	80.0 %	191 x
PGM NIST Run 2	83.3 %	16.7 %	1571 x	15.0 %	85.0 %	571 x
NIST SOLiD	82.9 %	17.1 %	22,719 x	12.5 %	87.5 %	17,499 x
Illumina MiSeq	82.1 %	17.8 %	6,517 x	11.6 %	88.4 %	3,715 x
Illumina HiSeq	82.3 %	17.7 %	48,071 x	11.8 %	88.2 %	42,101 x
Average	**82.0 %**	**18.0 %**		**14.2 %**	**85.8 %**	

Summary of mtDNA Work

- The consensus data from the four NGS platforms for the mitochondrial SRMs agree with Sanger sequencing data
  - G/A heteroplasmy at 1,393 confirmed
  - T/C heteroplasmy at 7,861 confirmed
  - C insertions and deletions are issues (assemblers/variant callers)
  - The majority of false positives are of low abundance and not reproducible across platforms

- Continuing work
  - Experiments for setting a variant calling threshold
  - Evaluate a three amplicon approach for mitochondrial DNA enrichment
  - Sequence the mitoSRMs on the PacificBiosciences platform (Collaboration with Children's National Medical Center)
  - Benefit from a standardized (forensic) informatics pipeline (CLC bio software, NextGENe)
  - Evaluate improved variant callers from Life Technologies and Illumina

NIST SRM 2391c

PCR Based Profiling Standard

- Certified Reference Material for STR typing
  - Five components
    - A – Single-source female
    - B – Single-source male
    - C – Single-source male
    - D – Mixed-source (Components A and C)
    - E – Single-source female cells on 903 paper
    - F – Single-source male cells on FTA paper

- Components A, B, and C have been sequenced at NIST on the PGM

Sample Preparation

- PCR primers and conditions were from Kline *et al.* 2011
- All loci amplified in single-plex, purified, then pooled
- Libraries were barcoded to run all samples in one run

Forensic Science International: Genetics 5 (2011) 329–332

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Short communication

STR sequence analysis for characterizing normal, variant, and null alleles

Margaret C. Kline *, Carolyn R. Hill, Amy E. Decker [1], John M. Butler

National Institute of Standards and Technology, 100 Bureau Drive, M/S 8312, Gaithersburg, MD 20899, USA

ARTICLE INFO

Article history:
Received 5 April 2010
Received in revised form 23 July 2010
Accepted 8 September 2010

Keywords:
Short tandem repeat
STR typing
DNA sequencing
Allele dropout
Null allele
Variant allele

ABSTRACT

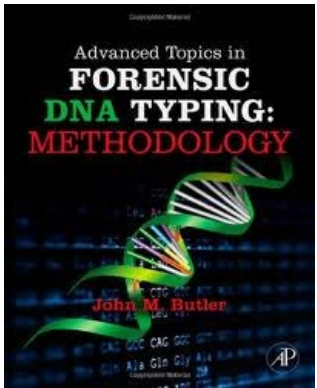
DNA sequence variation is known to exist in and around the repeat region of short tandem repeat (STR) loci used in human identity testing. While the vast majority of STR alleles measured in forensic DNA laboratories worldwide type as "normal" alleles compared with STR kit allelic ladders, a number of variant alleles have been reported. In addition, a sequence difference at a polymerase chain reaction (PCR) primer binding site in the DNA template can cause allele drop-out (i.e., a "null" or "silent" allele) with one set of primers and not with another. Our group at the National Institute of Standards and Technology (NIST) has been sequencing variant and null alleles supplied by forensic labs and cataloging this information on the NIST STRBase website for the past decade. The PCR primer sequences and strategy used for our STR allele sequencing work involving 23 autosomal STRs and 17 Y-chromosome STRs are described along with the results from 111 variant and 17 null alleles.

Published by Elsevier Ireland Ltd.

Analysis Methods

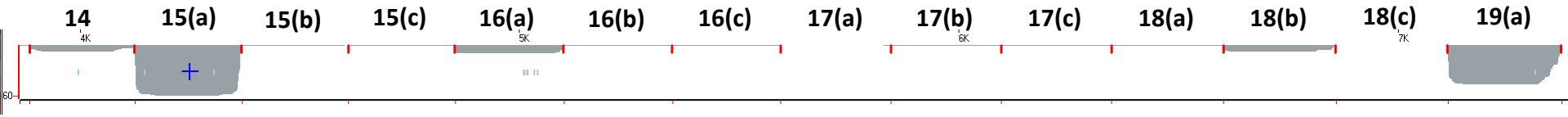
- We are using NextGENe for mapping reads to alleles
  - from Softgenetics
  - Forensic setting filters out reads < 80 % of reference
  - Reads must span all the way across repeat

- NextGENe needs a reference file for mapping reads
  - Create virtual allelic ladders for each locus using known sequence variants
  - Contains repeat structures for STR alleles
    - **…CAGGTG GATA GATA GATA GATA GATA TCATTG…**
    - **…CAGGTG GATA GATA GATA GATA GATA GATA TCATTG…**
    - **…CAGGTG GATA GATA GATA GATA GATA GATA GATA TCATTG…**
  - CODIS core loci plus Amelogenin (also D2, D19, Penta D & E)
    - 702 alleles (in Butler - Advanced Topics in Forensic DNA Typing: Methodology)
  - How much flanking sequence is needed?
    - This has an effect on the analysis
  - What about unknown alleles?
    - This is a limitation



Results from NextGENe

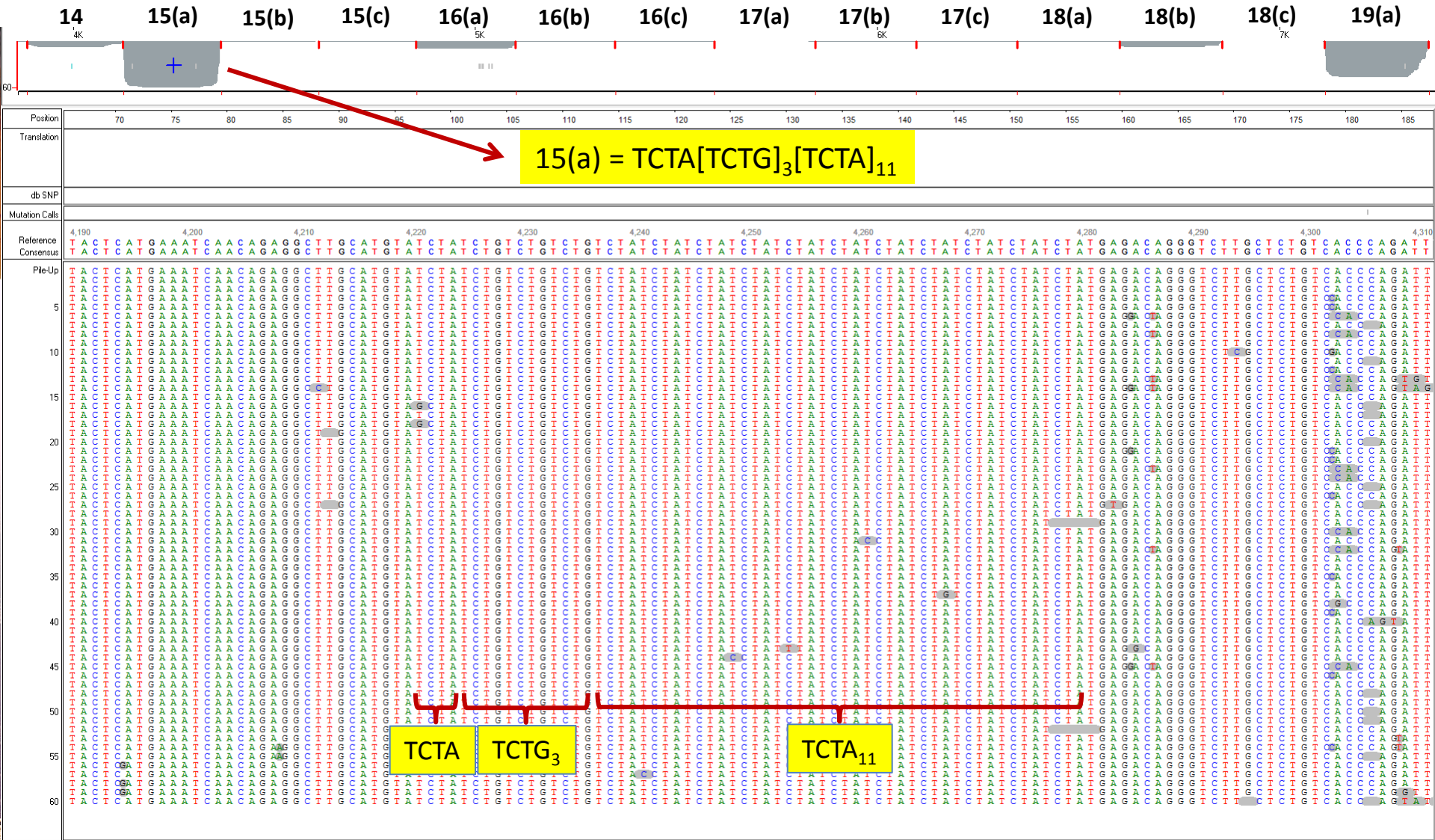
D3S1358 (15, 19) for SRM Component B



Results from NextGENe

D3S1358 (15, 19)

60x mapping coverage



Results for SRM 2391c Component A

- Alleles sequenced for all CODIS loci
  - All alleles confirmed and agree with Sanger sequence
  - D2 and D19 dropped out at the PCR stage

2391c Component A				
Locus	**Certified Value**	**NGS Value**	**Repeat Structure - Allele 1**	**Repeat Structure - Allele 2**
D2S1338	18,23	TBD	To Be Determined	To Be Determined
D3S1358	15,16	15, 16	TCTA$[TCTG]_2[TCTA]_{12}$	TCTA$[TCTG]_3[TCTA]_{12}$
D5S818	11,12	11,12	$[AGAT]_{11}$	$[AGAT]_{12}$
D7S820	11,11	11,11	$[GATA]_{11}$	$[GATA]_{11}$
D8S1179	13,14	13,14	$[TCTA]_{13}$	$[TCTA]_2TCTG[TCTA]_{11}$
D13S317	8,8	8,8	$[TATC]_8$	$[TATC]_8$
D16S539	10,11	10,11	$[GATA]_{10}$	$[GATA]_{11}$
D18S51	12,15	12,15	$[AGAA]_{12}$	$[AGAA]_{15}$
D19S1443	13,14	TBD	To Be Determined	To Be Determined
D21S11	28,32.2	28,32.2	$[TCTA]_4[TCTG]_6$ {$[TCTA]_3TA[TCTA]_3TCA[TCTA]_2TCCATA$}$[TCTA]_{10}$	$[TCTA]_5[TCTG]_6$ {$[TCTA]_3TA[TCTA]_3TCA[TCTA]_2TCCATA$} $[TCTA]_{12}$ TA TCTA
CSF1PO	10,10	10,10	$[AGAT]_{10}$	$[AGAT]_{10}$
FGA	21,23	21,23	$[TTTC]_3TTTT$ TTCT$[CTTT]_{13}CTCC[TTCC]_2$	$[TTTC]_3TTTT$ TTCT$[CTTT]_{15}CTCC[TTCC]_2$
Penta D	9,13	9,13	$[AAAGA]_9$	$[AAAGA]_{13}$
Penta E	5,10	5,10	$[AAAGA]_5$	$[AAAGA]_{10}$
TH01	8,9.3	8,9.3	$[AATG]_8$	$[AATG]_6ATG[AATG]_3$
TPOX	8,8	8,8	$[AATG]_8$	$[AATG]_8$
VWA	18,19	18,19	TCTA$[TCTG]_4[TCTA]_{13}$	TCTA$[TCTG]_4[TCTA]_{14}$
AMEL	X, X	X, X	No Polymorphisms Observed	No Polymorphisms Observed

Results for SRM 2391c Component B

- Alleles sequenced for all CODIS loci
  - D18 allele 16 needs to be confirmed

★ Novel variant

2391c Component B				
Locus	**Certified Value**	**NGS Value**	**Repeat Structure - Allele 1**	**Repeat Structure - Allele 2**
D2S1338	17,17	TBD	To Be Determined	To Be Determined
D3S1358	15,19	15, 19	TCTA$[TCTG]_3[TCTA]_{11}$	TCTA$[TCTG]_3[TCTA]_{15}$
D5S818	12,13	12,13	$[AGAT]_{12}$	$[AGAT]_{13}$
D7S820	10,10	10,10	$[GATA]_{10}$	$[GATA]_{10}$
D8S1179	10,13	10,13	$[TCTA]_{10}$	$[TCTA]_{13}$
D13S317	9,12	9,12	$[TATC]_9$	$[TATC]_{12}$
D16S539	10,13	10,13	$[GATA]_{10}$	$[GATA]_{13}$
D18S51	13,16	13,TBD	$[AGAA]_{13}$	To Be Determined
D19S1443	16,16.2	TBD	To Be Determined	To Be Determined
D21S11	32,32.2	32, 32.2	★ **$[TCTA]_4[TCTG]_6$ {$[TCTA]_3TA[TCTA]_3TCA[TCTA]_2TCCATA$}$[TCTA]_{14}$**	$[TCTA]_5[TCTG]_6$ {$[TCTA]_3TA[TCTA]_3TCA[TCTA]_2TCCATA$}$[TCTA]_{12}$ TA TCTA
CSF1PO	10,11	10,11	$[AGAT]_{10}$	$[AGAT]_{11}$
FGA	20,23	20,23	$[TTTC]_3TTTT$ TTCT$[CTTT]_{12}CTCC[TTCC]_2$	$[TTTC]_3TTTT$ TTCT$[CTTT]_{15}CTCC[TTCC]_2$
Penta D	8,12	8,12	$[AAAGA]_8$	$[AAAGA]_{12}$
Penta E	7,15	7,15	$[AAAGA]_7$	$[AAAGA]_{15}$
TH01	6,9.3	6,9.3	$[AATG]_6$	$[AATG]_6ATG[AATG]_3$
TPOX	8,11	8,11	$[AATG]_8$	$[AATG]_{11}$
VWA	17,18	17,18	TCTA$[TCTG]_4[TCTA]_{12}$	TCTA$[TCTG]_4[TCTA]_{13}$
AMEL	X, Y	X, Y	No Polymorphisms Observed	No Polymorphisms Observed

Results for SRM 2391c Component C

- Alleles sequenced for all CODIS loci
  - D18 allele 16 needs to be confirmed
  - Novel alleles found at D8, D13

★ Novel variant

2391c Component C				
Locus	**Certified Value**	**NGS Value**	**Repeat Structure - Allele 1**	**Repeat Structure - Allele 2**
D2S1338	19,19	TBD	To Be Determined	To Be Determined
D3S1358	16,18	16,18	TCTA[TCTG]$_3$[TCTA]$_{12}$	TCTA[TCTG]$_3$[TCTA]$_{14}$
D5S818	10,11	10,11	[AGAT]$_{10}$	[AGAT]$_{11}$
D7S820	10,12	10,12	[GATA]$_{10}$	[GATA]$_{12}$
D8S1179	10,17	10,17	[TCTA]$_{10}$	★ **[TCTA]$_2$[TCTG]$_1$[TCTA]$_{14}$**
D13S317	11,11	12,12	★ **[TATC]$_{12}$ Del ATCA 6 bp ds**	★ **[TATC]$_{12}$ Del ATCA 6 bp ds**
D16S539	10,10	10,10	[GATA]$_{10}$	[GATA]$_{10}$
D18S51	16,19	TBD,19	To Be Determined	[AGAA]$_{19}$
D19S1443	13.2,15.2	TBD	To Be Determined	To Be Determined
D21S11	29,30	29,30	[TCTA]$_4$[TCTG]$_6$ {[TCTA]$_3$TA[TCTA]$_3$TCA[TCTA]$_2$TCCATA}[TCTA]$_{11}$	[TCTA]$_6$[TCTG]$_5$ {[TCTA]$_3$TA[TCTA]$_3$TCA[TCTA]$_2$TCCATA}[TCTA]$_{11}$
CSF1PO	10,12	10,12	[AGAT]$_{10}$	[AGAT]$_{12}$
FGA	24,26	24,26	[TTTC]$_3$TTTT TTCT[CTTT]$_{16}$CTCC[TTCC]$_2$	[TTTC]$_3$TTTT TTCT[CTTT]$_{18}$CTCC[TTCC]$_2$
Penta D	10,11	10,11	[AAAGA]$_{10}$	[AAAGA]$_{11}$
Penta E	12,13	12,13	[AAAGA]$_{12}$	[AAAGA]$_{13}$
TH01	6,8	6,8	[AATG]$_6$	[AATG]$_8$
TPOX	11,11	11,11	[AATG]$_{11}$	[AATG]$_{11}$
VWA	16,18	16,18	TCTA[TCTG]$_4$[TCTA]$_{11}$	TCTA[TCTG]$_4$[TCTA]$_{13}$
AMEL	X, Y	X, Y	No Polymorphisms Observed	No Polymorphisms Observed

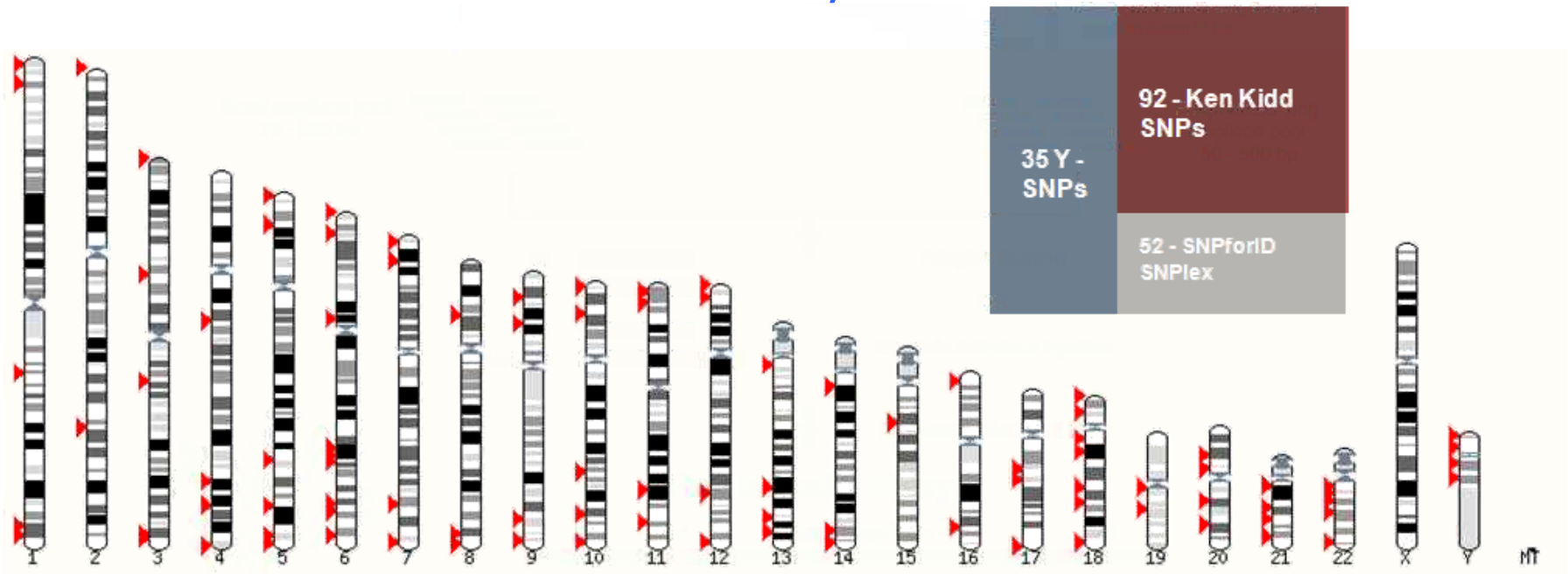
Summary of STR Work

- SRM 2391c Components A, B, and C sequenced
  - Used both 200 bp and 400 bp chemistry
  - Ion Torrent PGM – 314 (v2) and 316 (v2) chips
- Analysis parameters are being optimized
  - Some loci were more difficult to analyze
    - Changed flanking region sequence to 35 bp either side
    - Improved specificity to alleles, reduced # of reads mapping
  - Dropouts at D18 (16 allele) for SRM Components B and C
- Illumina MiSeq being installed at NIST next week

Life Technologies

Coming Soon for PGM

- HID SNP Panel v2.2
  - Autosomal loci chosen for high heterozygosity and low Fst
  - Genotype match probability 10^{-31} to 10^{-35}
  - 179 loci amplified in a single multiplex PCR
  - Short amplicons ≈ 150 bp
  - Commercial launch date not yet set



Life Technologies

Future Plans

- Ancestry informative and phenotypic SNP panel
- For generating investigative leads, subject exclusion
- 245 SNPs
  - 202 Ancestral SNPs
  - 45 Hair and eye color SNPs