Lessons learned from the characterization of a large set of population samples: identifying and addressing discordance

Carolyn R. Steffen, Katherine B. Gettings, John M. Butler, Michael D. Coble, Peter M. Vallone

Leader, Applied Genetics Group

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

MATERIAL MEASUREMENT LABORATORY

Outline

- Background on the 2013 “1036” publication
- Recent 1036 sequencing project
- Corrections to the “1036” allele calls
  - High level
  - Detailed examples
- Impact on Random Match Probabilities (RMPs)
- Dissemination
- Lessons learned and moving forward

NIST U.S. population data

Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. (2013) U.S. population data for 29 autosomal STR loci. *Forensic Sci. Int. Genet.* 7: e82-e83

Forensic Science International: Genetics 7 (2013) e82–e83

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

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

Letter to the Editor

U.S. population data for 29 autosomal STR loci

Dear Editor,

run and population statistics were confirmed using the Power-Marker v3.25 statistics program [10].

There were 14 instances where statistically significant devia-

aka NIST “1036”

1036 unrelated U.S. population samples

African American (342), Caucasian (361), Hispanic (236), Asian (97)

Genotypes and allele frequencies for 29 autosomal STR loci

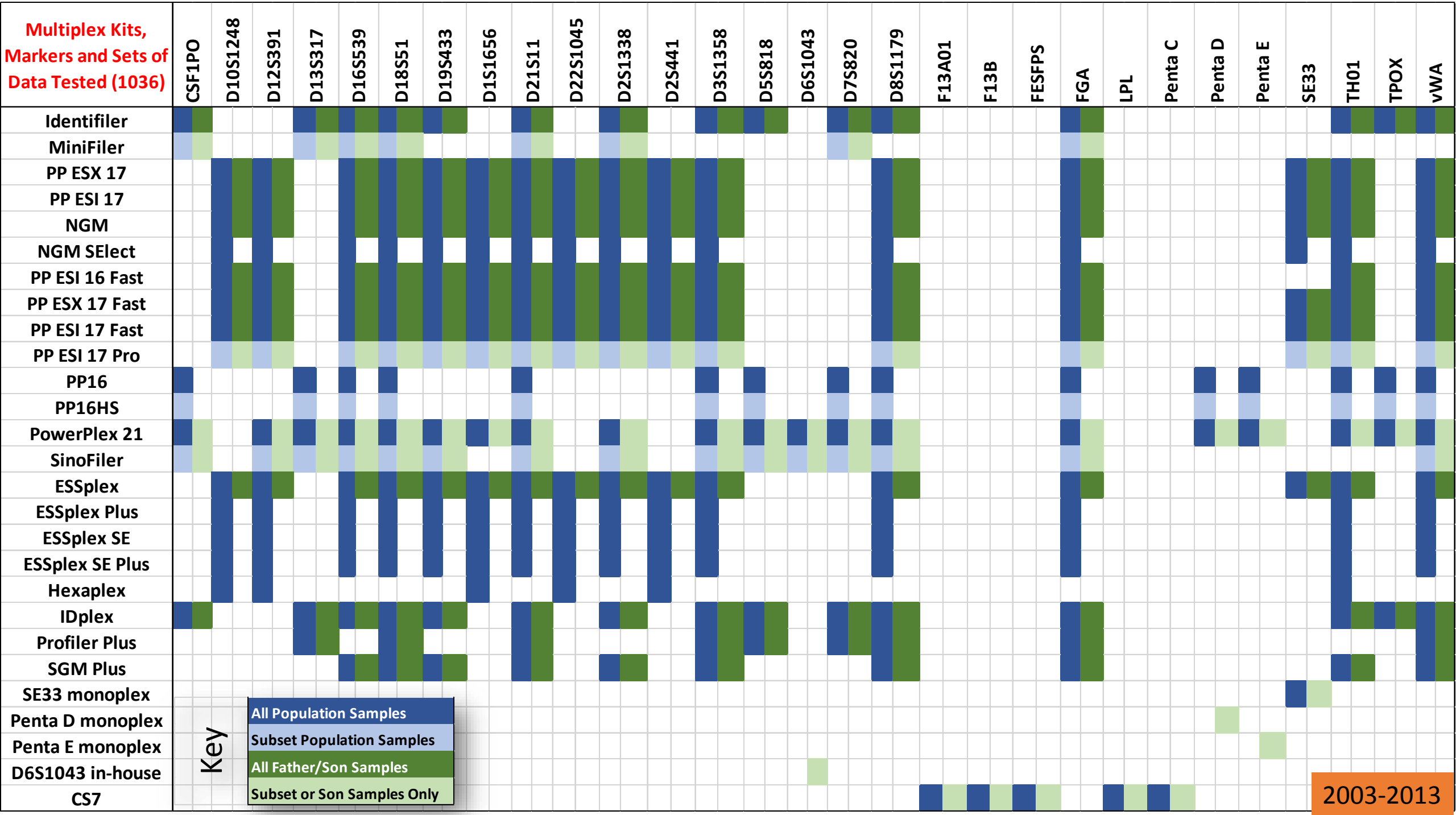
Source of samples

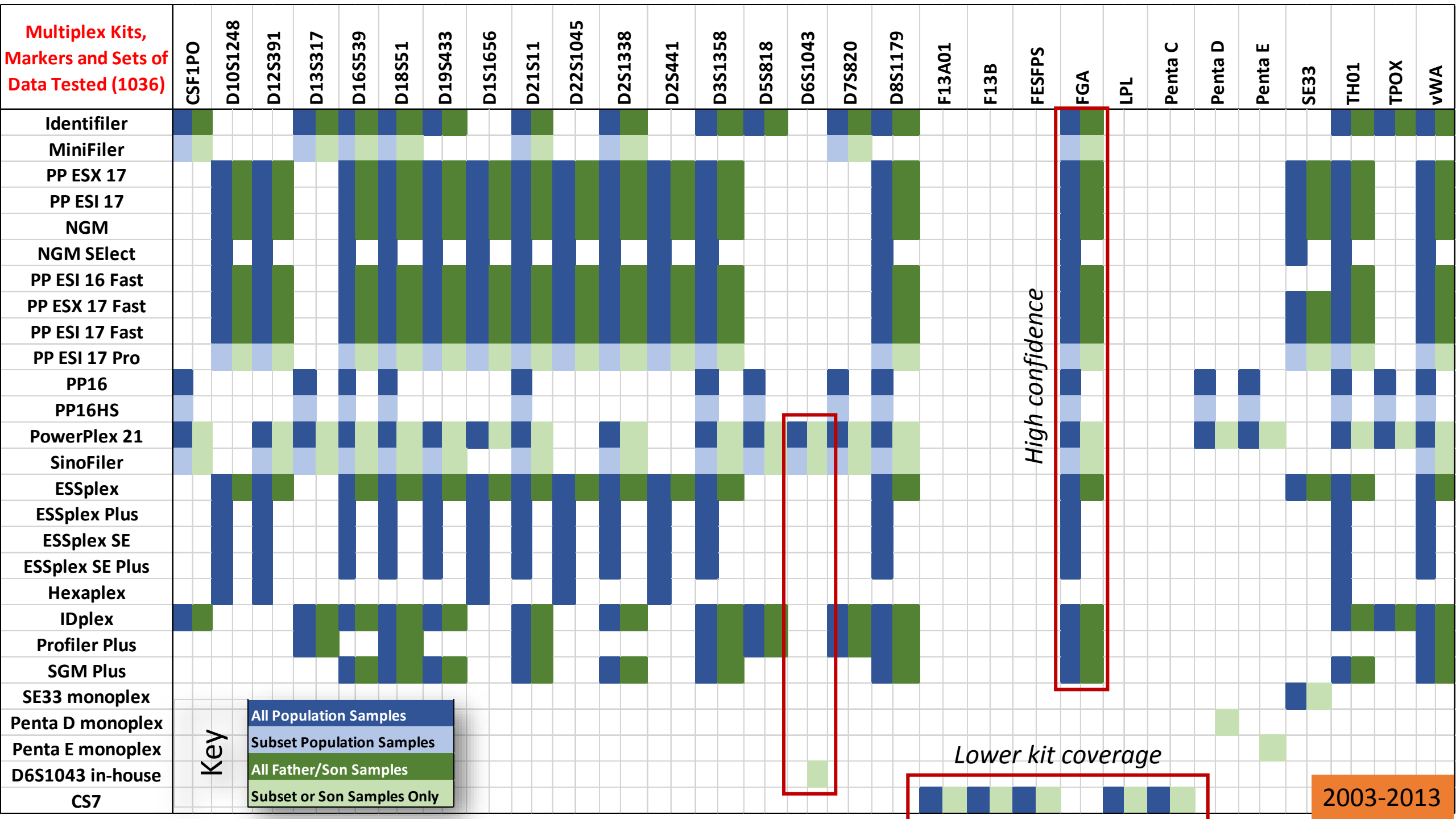
Anonymous blood samples

“Father-Son pairs” buccal swabs

Typing of the 1036 set

- 1036 samples
- 29 autosomal STR loci
- Experiments performed *over ~10 years (2003-2013)*
- 30,044 genotypes
- Data is derived from multiple STR kits (capillary electrophoresis)
- *There is not a single 29plex kit*
  - *Data was compiled from 27 kits or assays*


Multiplex Kits, Markers and Sets of Data Tested (1036)
CSF1PO
D10S1248
D12S391
D13S317
D16S539
D18S51
D19S433
D1S1656
D21S11
D22S1045
D2S1338
D2S441
D3S1358
D5S818
D6S1043
D7S820
D8S1179
F13A01
F13B
FESFPS
FGA
LPL
Penta C
Penta D
Penta E
SE33
TH01
TPOX
vWA
Identifiler
MiniFiler
PP ESX 17
PP ESI 17
NGM
NGM SElect
PP ESI 16 Fast
PP ESX 17 Fast
PP ESI 17 Fast
PP ESI 17 Pro
PP16
PP16HS
PowerPlex 21
SinoFiler
ESSplex
ESSplex Plus
ESSplex SE
ESSplex SE Plus
Hexaplex
IDplex
Profiler Plus
SGM Plus
SE33 monoplex
Penta D monoplex
Penta E monoplex
D6S1043 in-house
CS7
Key
All Population Samples
Subset Population Samples
All Father/Son Samples
Subset or Son Samples Only
2003-2013


Multiplex Kits, Markers and Sets of Data Tested (1036)
CSF1PO
D10S1248
D12S391
D13S317
D16S539
D18S51
D19S433
D1S1656
D21S11
D22S1045
D2S1338
D2S441
D3S1358
D5S818
D6S1043
D7S820
D8S1179
F13A01
F13B
FESFPS
FGA
LPL
Penta C
Penta D
Penta E
SE33
TH01
TPOX
vWA
Identifiler
MiniFiler
PP ESX 17
PP ESI 17
NGM
NGM SElect
PP ESI 16 Fast
PP ESX 17 Fast
PP ESI 17 Fast
PP ESI 17 Pro
PP16
PP16HS
PowerPlex 21
SinoFiler
ESSplex
ESSplex Plus
ESSplex SE
ESSplex SE Plus
Hexaplex
IDplex
Profiler Plus
SGM Plus
SE33 monoplex
Penta D monoplex
Penta E monoplex
D6S1043 in-house
CS7
High confidence
Lower kit coverage
Key
All Population Samples
Subset Population Samples
All Father/Son Samples
Subset or Son Samples Only
2003-2013

Current 1036 Sequencing Project

- Goal: to sequence commonly used STR alleles and provide their allele frequencies
- Started January 2016

- As a result of the 1036 ***sequencing*** project
  - A second methodology was attempted
  - 24/29 STR loci were sequenced
  - Further concordance checking was performed
  - The remaining five loci were independently re-reviewed by two researchers (CE data)

Loci Sequenced	Reported in 1036
CSF1PO	CSF1PO
D10S1248	D10S1248
D12S391	D12S391
D13S317	D13S317
D16S539	D16S539
D17S1301	
D18S51	D18S51
D19S433	D19S433
D1S1656	D1S1656
D20S482	
D21S11	D21S11
D22S1045	D22S1045
D2S1338	D2S1338
D2S441	D2S441
D3S1358	D3S1358
D4S2408	
D5S818	D5S818
D6S1043	D6S1043
D7S820	D7S820
D8S1179	D8S1179
D9S1122	
	F13A01
	F13B
	FESFPS
FGA	FGA
	LPL
	Penta_C
PentaD	Penta_D
PentaE	Penta_E
(SE33)	SE33
TH01	TH01
TPOX	TPOX
vWA	vWA

1036 Sequencing Project

- Discordant allele calls were detected for 13 STR loci (4 U.S. core)
  - **Non-U.S. core:** D6S1043, F13A01, F13B, FESFPS, LPL, Penta D, Penta E, Penta C, SE33
  - **U.S. core:** D5S818, D13S317, D7S820, TPOX
- The discordancies were detected in 12 unique samples

Reasons were categorized as:

1 PCR primer design differences

2 Change in the reporting of tri-alleles

3 Laboratory error

4 Data analysis error

High level overview

- 12 samples affected
- 13 loci affected (4 U.S. core)

Samples: 12/1036 = 1.16%
Genotypes: 37/30,044 = 0.123%
Alleles: 39/60,088 = 0.065%*
*not including tri-alleles

*The magnitude of the effect on **allele frequencies***
Min/max change in allele frequency

N=97

	AfAm	Cauc	Hisp	Asian
min	0.00001	0.00138	0.00001	0.00515
max	0.00146	0.00277	0.00705	0.01030
average	0.00087	0.00166	0.00233	0.00562
std	0.00062	0.00057	0.00198	0.00155

High level overview - 12 samples, 13 loci (4 U.S. core)

# of loci

	Loci Affected			
	AfAm	**Cauc**	**Hisp**	**Asian**
	3	7	8	4
1	D5S818			D5S818
2	D6S1043	D6S1043	D6S1043	
3				D7S820
4				D13S317
5		F13A01	F13A01	
6		F13B	F13B	
7		FESFPS	FESFPS	
8		LPL	LPL	
9		Penta_C	Penta_C	
10			Penta_D	
11			Penta_E	
12		SE33		
13	TPOX			TPOX

	# of Alleles Affected in Freq Tables				
	AfAm	**Cauc**	**Hisp**	**Asian**	**Locus**
1	2			2	D5S818
2	4	2	5		D6S1043
3				3	D7S820
4				2	D13S317
5		2	2		F13A01
6		3	2		F13B
7		2	2		FESFPS
8		2	2		LPL
9		2	3		Penta_C
10			**15**		Penta_D
11			4		Penta_E
12		2			SE33
13	**8**			4	TPOX
	bold - triallele, all alleles affected				

High level overview -12 samples, Loci 13 (4 U.S. core)

Unique Samples	Population	Sample Name	Locus	1036 (2013)	1036 (2017)	
1	African American	C28B	D5S818	12,12	**7**,12	1
2	African American	OT05588	TPOX	9,11	removed	2
3	Hispanic	C88H	Penta D	11,14	removed	2
4	African American	C37B	D6S1043	20,20	**18**,20	3
5	African American	C63B	D6S1043	15,15	**13**,15	3
6	Asian	C66A	TPOX	8,11	**9,10**	4
	Asian	C66A	D5S818	12,13	**11**,12	4
	Asian	C66A	D7S820	11,11	**8,10**	4
	Asian	C66A	D13S317	8,11	11,**12**	4

1 PCR primer design differences | 2 Change in reporting of tri-alleles | 3 Laboratory error | 4 Data analysis error

"C82H" = Child 82 Hispanic

There exists a "AF82H" = Alleged Father 82 Hispanic (but not in the 1036 set of samples)

After a second extraction of a buccal swab the sample names were switched on the final collection tubes

For 8 loci the genotypes were switched between alleged father and son

3 Laboratory error

Unique Samples	Population	Sample Name	Locus	1036 (2013)	1036 (2017)	
7	Hispanic	C82H	D6S1043	11,14	11,**19**	3
	Hispanic	C82H	Penta C	5,11	5,**12**	3
	Hispanic	C82H	Penta D	11,12	**10**,11	3
	Hispanic	C82H	Penta E	7,12	7,**15**	3
	Hispanic	C82H	F13A01	6,7	**7**,7	3
	Hispanic	C82H	F13B	8,8	8,8	3
	Hispanic	C82H	FESFPS	11,13	11,**11**	3
	Hispanic	C82H	LPL	10,11	10,11	3
8	Hispanic	C84H	D6S1043	12,20.3	12,**15**	3
	Hispanic	C84H	Penta C	13,13	**12**,13	3
	Hispanic	C84H	Penta D	10,12	10,**10**	3
	Hispanic	C84H	Penta E	12,17	**15**,17	3
	Hispanic	C84H	F13A01	3.2,6	3.2,**5**	3
	Hispanic	C84H	F13B	9,9	**6**,9	3
	Hispanic	C84H	FESFPS	11,13	**12**,13	3
	Hispanic	C84H	LPL	10,10	10,**12**	3
9	Hispanic	C86H	D6S1043	13,14	13,**21.3**	3
	Hispanic	C86H	Penta C	13,13	**12**,13	3
	Hispanic	C86H	Penta D	10,12	10,**10**	3
	Hispanic	C86H	Penta E	7,9	9,**11**	3
	Hispanic	C86H	F13A01	5,7	**7**,7	3
	Hispanic	C86H	F13B	10,10	**9**,10	3
	Hispanic	C86H	FESFPS	12,13	12,**12**	3
	Hispanic	C86H	LPL	10,11	10,11	3

High level overview -12 samples, Loci 13 (4 U.S. core)

Unique Samples	Population	Sample Name	Locus	1036 (2013)	1036 (2017)	
10	Caucasian	OT07767	D6S1043	11,12	11,**13**	4
11	Caucasian	MT97180	SE33	18,20.2	**18.3**,20.2	4
12	Caucasian	C67C	F13A01	6,6	**5**,6	4
	Caucasian	C67C	F13B	10,10	**6,8**	4
	Caucasian	C67C	FESFPS	11,12	**10**,11	4
	Caucasian	C67C	LPL	10,10	**11,11**	4
	Caucasian	C67C	Penta C	11,11	**9**,11	4

1 PCR primer design differences | 2 Change in reporting of tri-alleles | 3 Laboratory error | 4 Data analysis error

Specific Examples

D5S818-C28B-AfAm

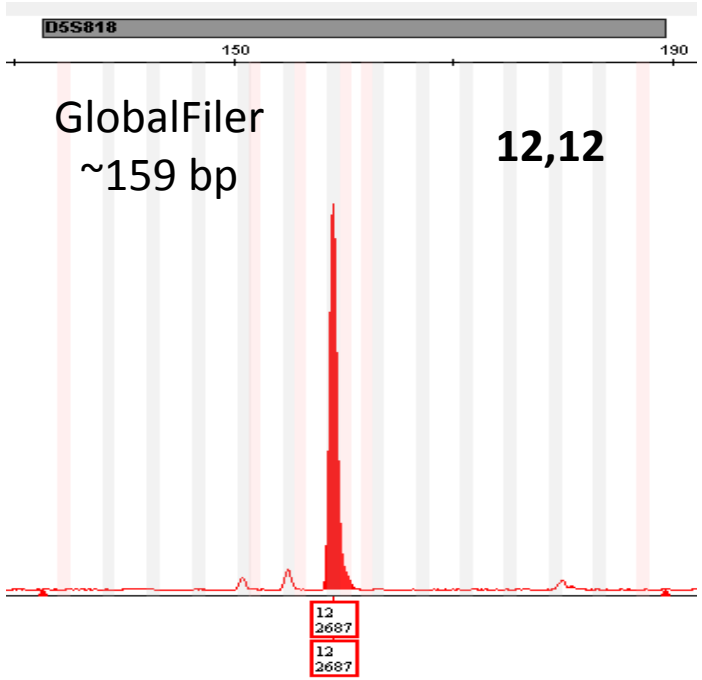
1036	1036 revised
12,12	7,12

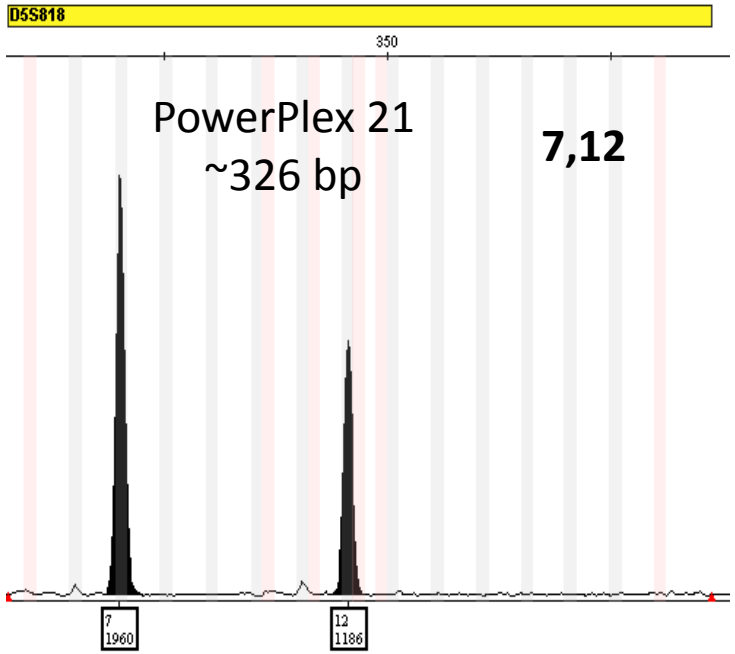
	AfAm		
D5S818	**1036**	**1036 Revised**	**delta**
7	0.0015	0.0029	0.0015
12	0.3699	0.3684	-0.0015

1036

Follow up typing

	Genotype	Assay/Kit
1	12,12	Identifiler
2	12,12	Profiler Plus
3	8,12	ForenSeq (NGS)
4	7,12	PowerPlex 21
5	7,12	PowerPlex Fusion
6	7,12	Qiagen IDPlex
7	12,12	Globalfiler





Reason: 1 PCR primer design differences

Assuming a 4 base deletion, to be confirmed

TPOX-OT05588-AfAm

1036	1036 revised
9,11	removed

1036

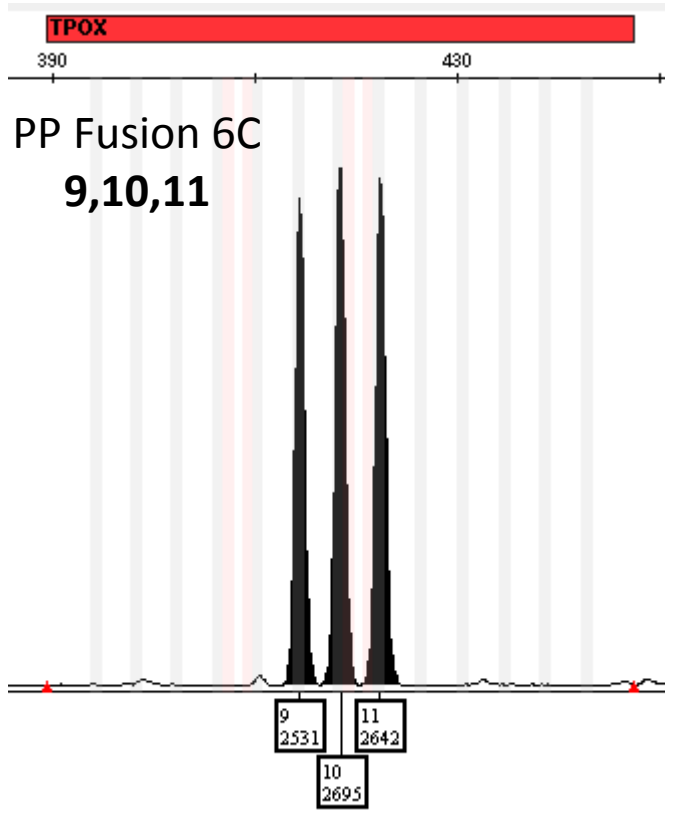
Follow up typing

	Genotype	Assay/Kit
1	9,10,11	Identifiler
2	9,10,11	PowerPlex 16
3	9,10,11	PowerPlex 16HS
4	9,10,11	PowerPlex 21
5	9,10,11	PowerPlex Fusion
6	9,10,11	PowerPlex Fusion 6C
7	9,10,11	Qiagen IDPlex
8	9,10,11	Qiagen 24plexQS
9	9,10,11	ForenSeq (NGS)

Reason: 2 Change in reporting of tri-alleles

N = 342 decreased to 341 for TPOX

	AfAm		
TPOX	1036	1036 Revised	delta
6	0.0892	0.0894	0.0003
7	0.0175	0.0176	0.0001
8	0.3670	0.3680	0.0011
9	0.1959	0.1950	-0.0009
10	0.0863	0.0865	0.0003
11	0.2164	0.2155	-0.0008
12	0.0263	0.0264	0.0001
13	0.0015	0.0015	0.0000

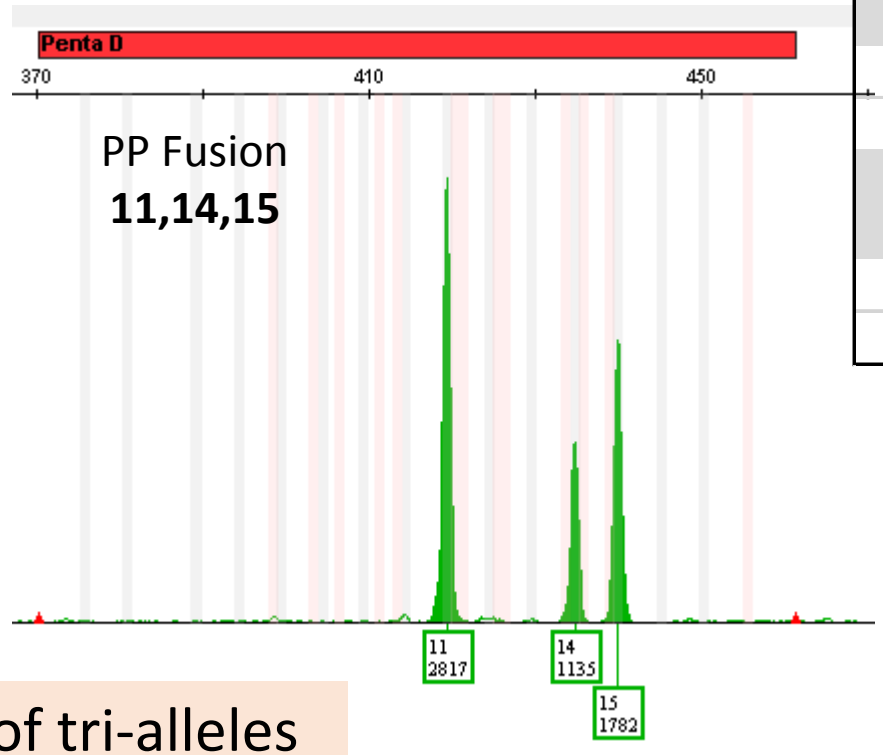


PentaD-C88H-Hisp

1036	1036 revised
11,14	removed

Follow up typing 1036

	Genotype	Assay/Kit
1	11,14,15	Penta D Monoplex
2	11,14,15	PowerPlex Fusion
3	11,14,15	ForenSeq (NGS)



N = 236 decreased to 235 for PentaD

	Hisp		
PentaD	1036	1036 Revised	delta
2.2	0.01695	0.01702	0.00007
3.2	0.00212	0.00213	0.00001
5	0.00636	0.00638	0.00002
6	0.00212	0.00213	0.00001
7	0.00212	0.00213	0.00001
8	0.01907	0.01915	0.00008
9	0.24153	0.24255	0.00102
10	0.15678	0.16383	0.00705
11	0.15678	0.15532	-0.00146
12	0.16314	0.15745	-0.00569
13	0.14407	0.14468	0.00061
14	0.07203	0.07021	-0.00182
15	0.01059	0.01064	0.00005
16	0.00424	0.00426	0.00002
17	0.00212	0.00213	0.00001

Reason: 2 Change in reporting of tri-alleles

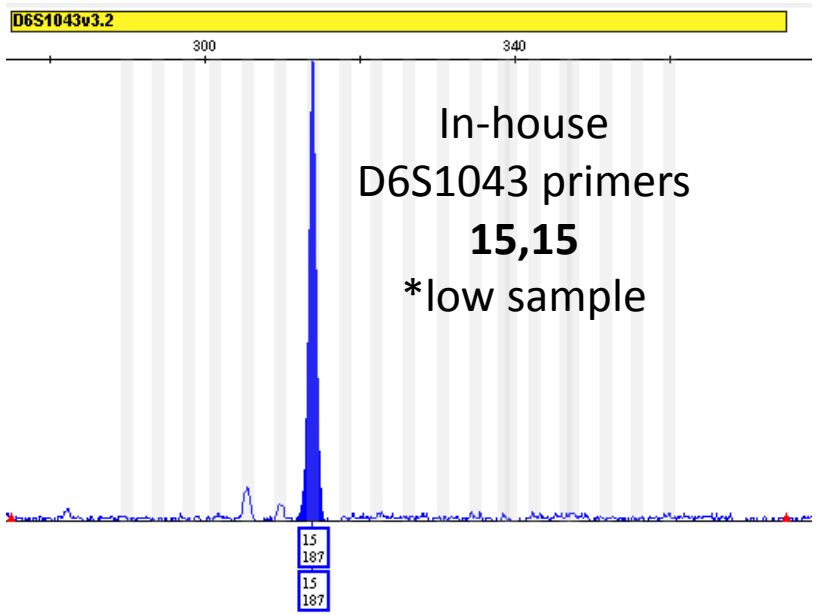
D6S1043-C63B-AfAm

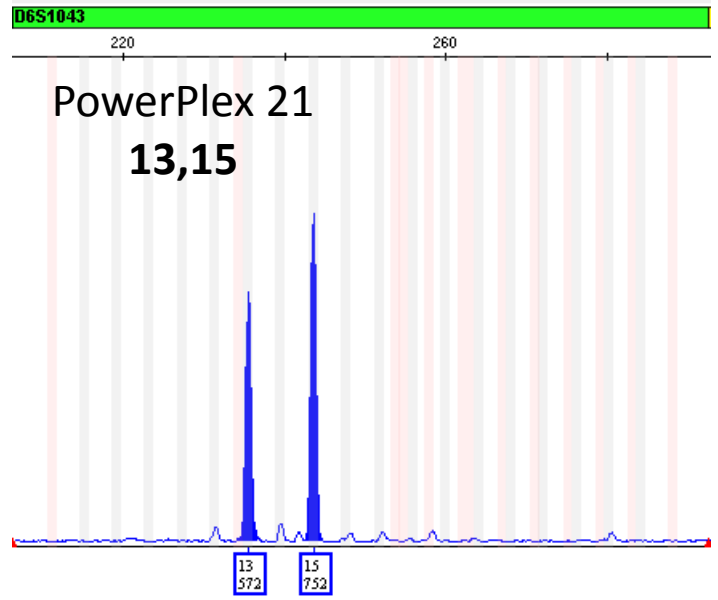
1036	1036 revised
15,15	13,15

	AfAm		
D6S1043	1036	1036 Revised	delta
13	0.0965	0.0980	0.0015
15	0.0541	0.0526	-0.0015
18	0.1067	0.1082	0.0015
20	0.0731	0.0716	-0.0015

Follow up typing 1036

	Genotype	Assay/Kit
1	15,15	In-house D6S1043
2	13,15	PowerPlex 21
3	13,15	In-house D6S1043
4	13,15	ForenSeq (NGS)





Sample re-extracted

Reason: 3 Laboratory error

D6S1043-OT07767-Cauc

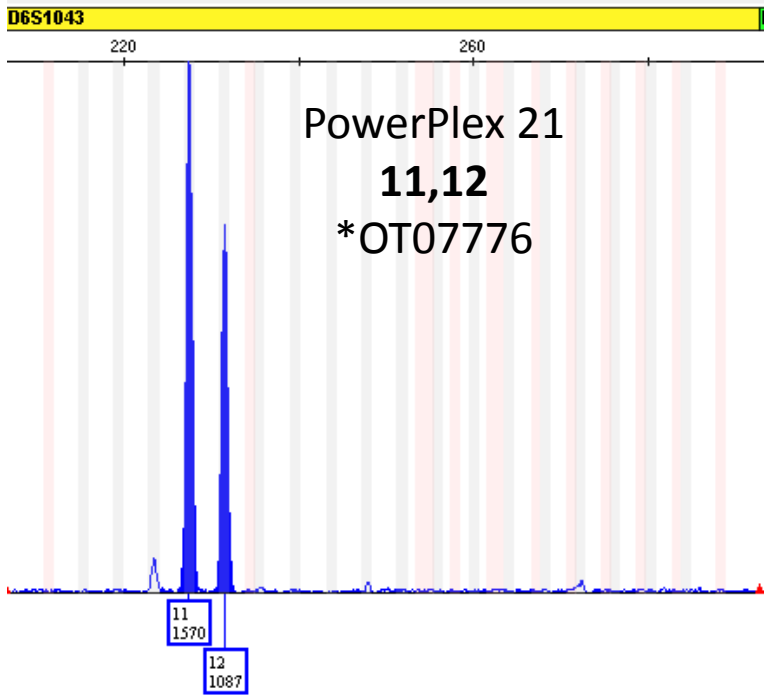
1036	1036 revised
11,12	11,13

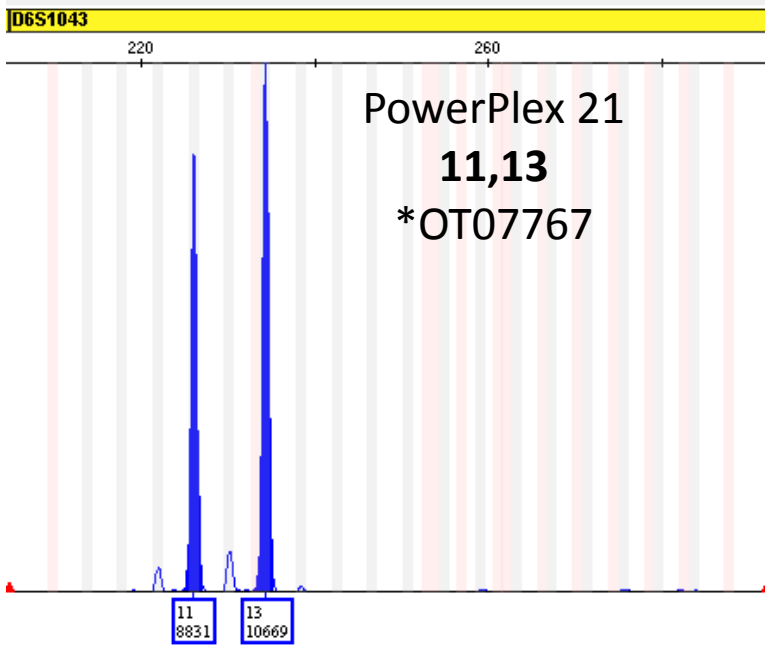
	Cauc		
D6S1043	**1036**	**1036 Revised**	**delta**
12	0.2368	0.2355	-0.0014
13	0.0859	0.0873	0.0014

1036

Follow up typing

	Genotype	Assay/Kit
1	11,12	PowerPlex 21 *OT07776
2	11,13	PowerPlex 21
3	11,13	ForenSeq (NGS)





Reason: 4 Data analysis error

SE33-MT97180-Cauc

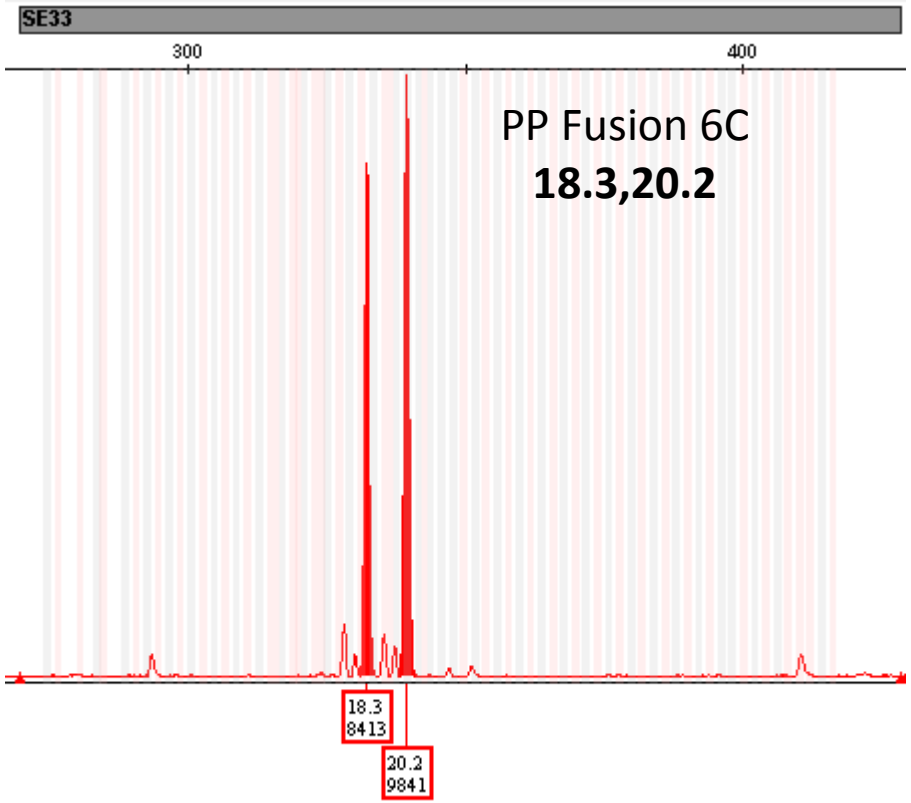
1036	1036 revised
18,20.2	18.3,20.2

1036

Follow up typing

	Genotype	Assay/Kit
1	18.3,20.2	PowerPlex ESX 17
2	18.3,20.2	PowerPlex ESI 17
3	18.3,20.2	SE33 Monoplex
4	18.3,20.2	NGM SElect
5	18.3,20.2	Qiagen ESSplex SE
6	18.3,20.2	Qiagen 24plexQS
7	18.3,20.2	PowerPlex Fusion 6C
8	18.3,20.2	ForenSeq (NGS)

	Cauc		
SE33	**1036**	**1036 Revised**	**delta**
18	0.0734	0.0720	-0.0014
18.3	0.0000	0.0014	0.0014



Reason: 4 Data analysis error

D5S818, D7S820, D13S317, TPOX-C66A-Asian

Markers	1036	1036 revised
D5S818	12,13	11,12
D7S820	11,11	8,10
D13S317	8,11	11,12
TPOX	8,11	9,10
CSF1PO	10,12	10,12

Follow up typing 1036

	Assay/Kit
1	Identifiler
2	PowerPlex Fusion
3	GlobalFiler
4	ForenSeq (NGS)

*Genotypes reported for C66A for these 5 loci were transposed from C90A: Concordance data from European kits (PP ESX 17 & PP ESI 17) did not contain data from these 5 loci (not present in European kits) to catch the transposition error

	Asian		
D5S818	**1036**	**1036 Revised**	**delta**
11	0.2680	0.2732	0.0052
13	0.1650	0.1598	-0.0052
D7S820	**1036**	**1036 Revised**	**delta**
8	0.1289	0.1340	0.0051
10	0.2577	0.2629	0.0052
11	0.3608	0.3505	-0.0103
D13S317	**1036**	**1036 Revised**	**delta**
8	0.2217	0.2165	-0.0052
12	0.2062	0.2113	0.0051
TPOX	**1036**	**1036 Revised**	**delta**
8	0.5516	0.5464	-0.0052
9	0.0773	0.0825	0.0052
10	0.0258	0.0309	0.0052
11	0.2990	0.2938	-0.0052

Reason: 4 Data analysis error

D6S1043, Penta C, Penta D, Penta E, F13A01, F13B, FESFPS, and LPL–C82H, C84H, C86H-Hisp

	Markers	1036			1036 revised		
		C82H	C84H	C86H	C82H	C84H	C86H
	D6S1043	11,14	12,20.3	13,14	11,19	12,15	13,21.3
	Penta C	5,11	13,13	13,13	5,12	12,13	12,13
	Penta D	11,12	10,12	10,12	10,11	10,10	10,10
	Penta E	7,12	12,17	7,9	7,15	15,17	9,11
	F13A01	6,7	3.2,6	5,7	7,7	3.2,5	7,7
	F13B	8,8	9,9	10,10	8,8	6,9	9,10
	FESFPS	11,13	11,13	12,13	11,11	12,13	12,12
	LPL	10,11	10,10	10,11	10,11	10,12	10,11

1036

Follow up typing

	Assay/Kit
1	Penta D Monoplex
2	Penta E Monoplex
3	In-house D6S1043
4	PowerPlex CS7
5	PowerPlex CS7
6	PowerPlex 21
7	ForenSeq (NGS)

The “Father” sample was switched with the “Son” sample after a second extraction of a buccal swab - the sample names were switched on the final collection tubes and the genotypes were switched for these 8 loci.

Reason: 3 Laboratory error

	Hisp		
D6S1043	**1036**	**1036 Revised**	**delta**
14	0.1356	0.1314	-0.0042
15	0.0297	0.0318	0.0021
19	0.0763	0.0784	0.0021
20.3	0.0127	0.0106	-0.0021
21.3	0.0403	0.0424	0.0021
PentaD	**1036**	**1036 Revised**	**delta**
2.2	0.01695	0.01702	0.00007
3.2	0.00212	0.00213	0.00001
5	0.00636	0.00638	0.00002
6	0.00212	0.00213	0.00001
7	0.00212	0.00213	0.00001
8	0.01907	0.01915	0.00008
9	0.24153	0.24255	0.00102
10	0.15678	0.16383	0.00705
11	0.15678	0.15532	-0.00146
12	0.16314	0.15745	-0.00569
13	0.14407	0.14468	0.00061
14	0.07203	0.07021	-0.00182
15	0.01059	0.01064	0.00005
16	0.00424	0.00426	0.00002
17	0.00212	0.00213	0.00001
PentaE	**1036**	**1036 Revised**	**delta**
7	0.1186	0.1165	-0.0021
11	0.0742	0.0763	0.0021
12	0.1737	0.1695	-0.0042
15	0.0911	0.0953	0.0042
PentaC	**1036**	**1036 Revised**	**delta**
11	0.3326	0.3305	-0.0021
12	0.2034	0.2098	0.0064
13	0.1081	0.1038	-0.0042
F13A01	**1036**	**1036 Revised**	**delta**
6	0.1716	0.1674	-0.0042
7	0.3030	0.3072	0.0042
F13B	**1036**	**1036 Revised**	**delta**
6	0.1186	0.1208	0.0021
10	0.4407	0.4386	-0.0021
FESFPS	**1036**	**1036 Revised**	**delta**
12	0.2140	0.2182	0.0042
13	0.0784	0.0742	-0.0042
LPL	**1036**	**1036 Revised**	**delta**
10	0.4852	0.4831	-0.0021
12	0.2119	0.2140	0.0021

Penta C, F13A01, F13B, FESFPS, and LPL–C67C-Cauc

Markers	1036	1036 revised
F13A01	6,6	5,6
F13B	10,10	6,8
FESFPS	11,12	10,11
LPL	10,10	11,11
Penta C	11,11	9,11

Follow up typing 1036

	Assay/Kit
1	PowerPlex CS7
2	PowerPlex CS7

*Genotypes reported for C67C for these 5 loci were transposed from C67A: only one kit (PowerPlex CS7) was available to type these markers so there was no concordance data to catch the transposition error

	Cauc		
F13A01	**1036**	**1036 Revised**	**delta**
5	0.1925	0.1939	0.0014
6	0.3504	0.3490	-0.0014
F13B	**1036**	**1036 Revised**	**delta**
6	0.0942	0.0956	0.0014
8	0.2452	0.2465	0.0014
10	0.3920	0.3892	-0.0028
FESFPS	**1036**	**1036 Revised**	**delta**
10	0.2812	0.2826	0.0014
12	0.2368	0.2355	-0.0014
LPL	**1036**	**1036 Revised**	**delta**
10	0.4252	0.4224	-0.0028
11	0.2618	0.2645	0.0028
PentaC	**1036**	**1036 Revised**	**delta**
9	0.1482	0.1496	0.0014
11	0.3961	0.3947	-0.0014

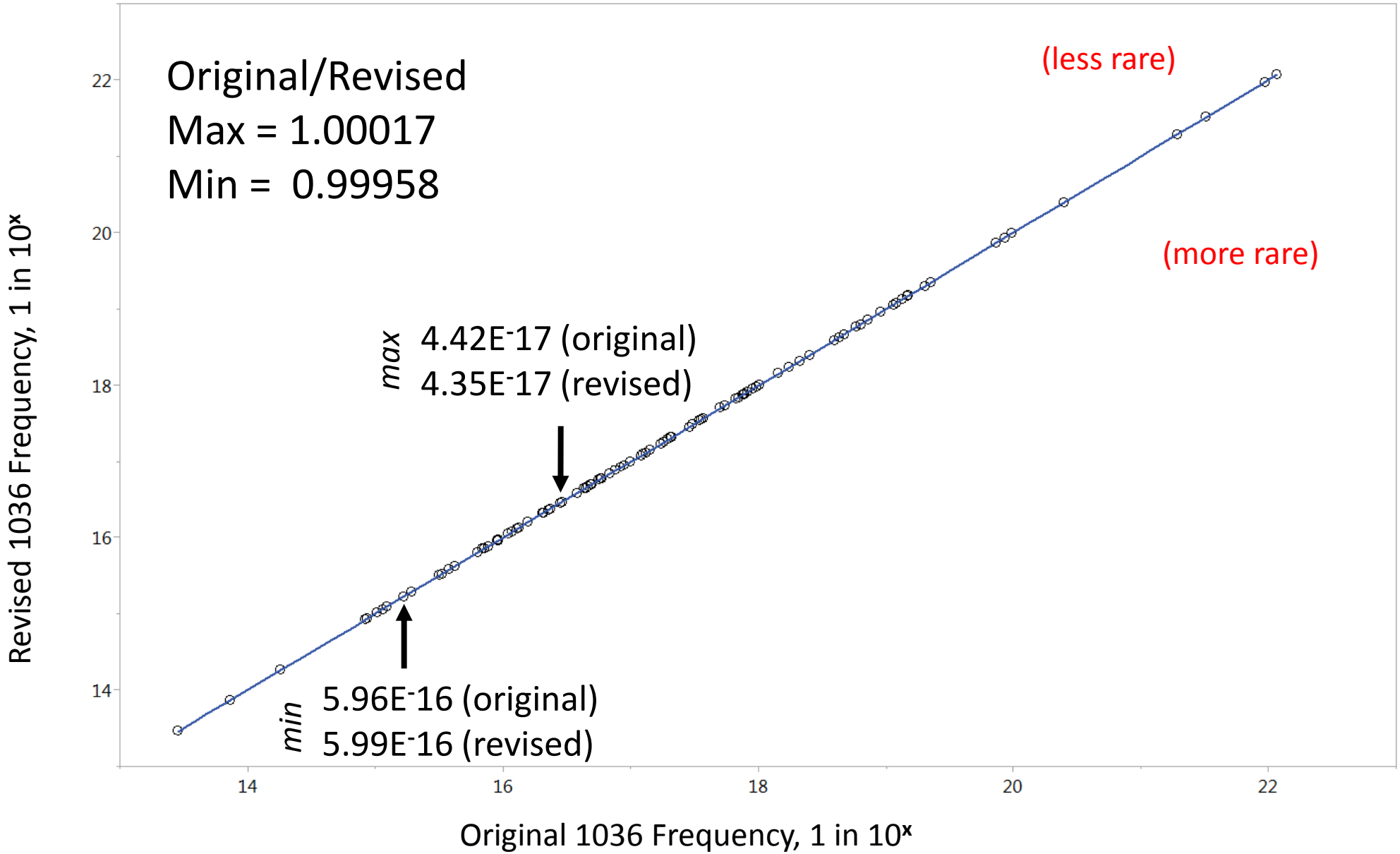
Reason: 4 Data analysis error

Effect on random match probabilities (RMP)

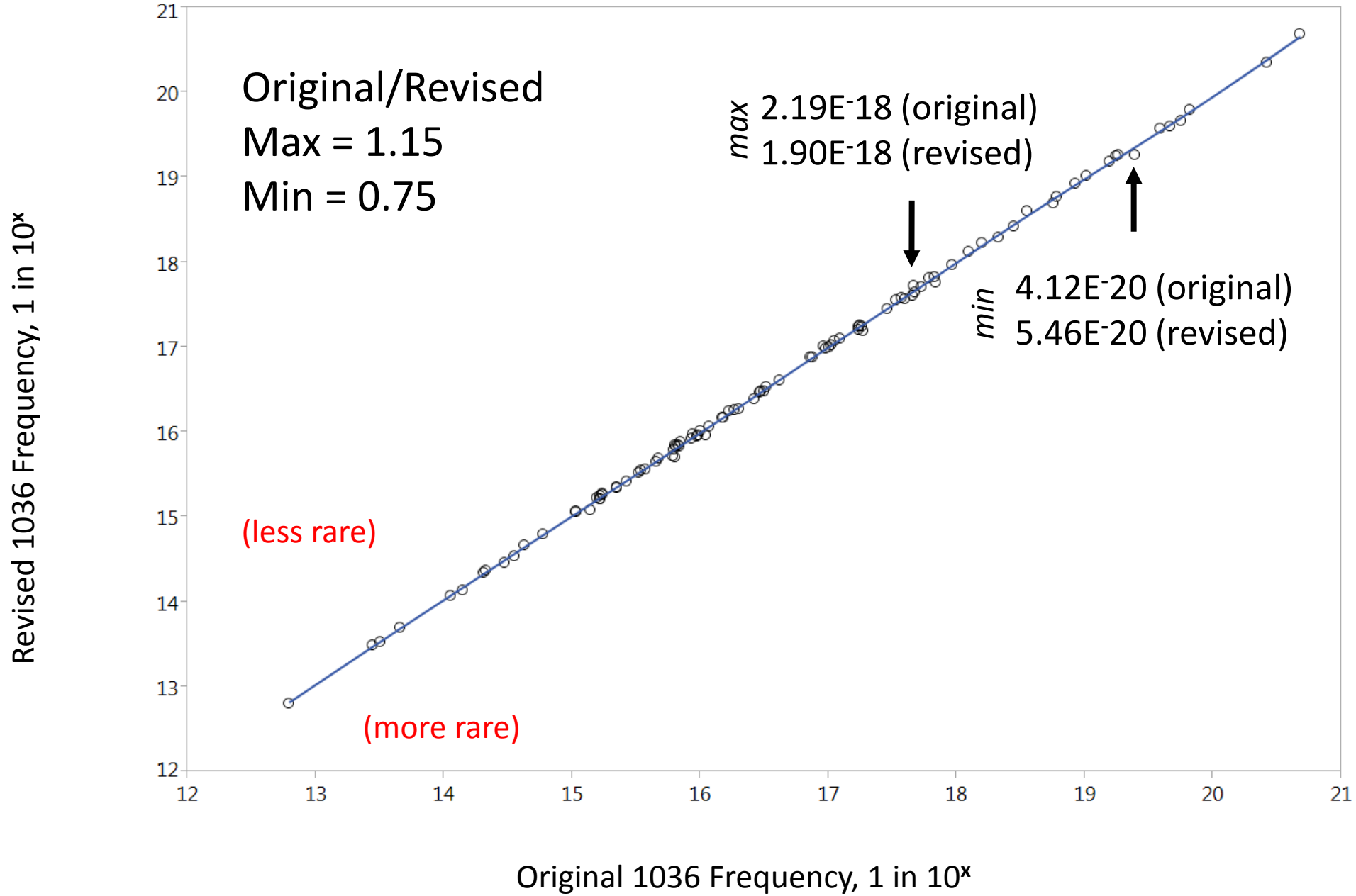
- RMPs for 1036 versus revised 1036 were calculated
  - For the ***13 U.S. core loci*** – note: none of the expanded loci were affected
  - Hisp and Cauc not affected at core loci

- 100 randomly generated profiles were used for calculations in AfAm and Asian populations
  - Random profiles generated using “DNA Profile Builder” – Dr. Chris Maguire at The University of Northumbria at Newcastle using original 1036 allele frequencies
  - Calculations performed by Mike Coble (LSAM – “Lisa Statistical Analysis Methods” software package http://www.ftechi.com/)

Effect on RMP (CODIS 13) – AfAm: D5S818 and TPOX revised



Effect on RMP (CODIS 13) – Asian: D5S818, D7S820, D13S317, TPOX revised



Disseminating the revisions

- The revised 1036 Genotypes and Allele Frequencies were provided to Dr. Douglas Hares at the FBI for Popstats [*pending analysis and public documentation*]

- The revised 1036 Genotypes and Allele Frequencies and this presentation will be available on STRBase
  - http://strbase.nist.gov/NISTpop.htm

- Drafting a “Letter to the Editor” to be submitted to Forensic Science International: Genetics describing the revisions

Disseminating the revisions

DNA Data [Autosomal Markers] [Y-Chromosome Markers] [Mitochondrial DNA]

NIST 1036 U.S. Population Dataset - 29 autosomal STR loci and 23 Y-STR loci

- Butler, J.M., Hill, C.R., Coble, M.D. (2012) Variability of new STR loci and kits in U.S. population groups. *Profiles in DNA*.
- Coble, M.D., Hill, C.R., Butler J.M. (2013) Haplotype data for 23 Y-chromosome markers in four U.S. population groups. *Forensic Sci. Int. Genet.* 7: e66-e68.
- Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. (2013) U.S. population data for 29 autosomal STR loci. *Forensic Sci. Int. Genet.* 7: e82-e83.

Please note new link to STRBase

http://strbase.nist.gov/NISTpop.htm

NIST 1036 Revised U.S. Population Dataset (July 2017) NEW

Revisions were made to 13 of the 29 autosomal STR loci reported in Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. (2013) U.S. population data for 29 autosomal STR loci. Forensic Sci. Int. Genet. 7: e82-e83.

- Excel file of 1036 revised Genotypes
- Excel file of 1036 revised Allele Frequencies
- Presentation describing the revisions in detail given at the NIST Forensic Science Error Management International Symposium July 25, 2017.
- Letter to the Editor in preparation, will be posted here upon publication.
- Original dataset remains available as a supplemental file to the publication, and may also be requested by contacting Peter Vallone.

Lessons Learned

- The role of concordance checking for allele calls
  - Testing multiple kits provides confidence (repeated measurements)
  - Not a replacement for validation and secondary review of data

- Benefits of testing new methods
  - Perform a QC service on existing data

- The collection of data over time
  - Instruments, versions of analysis software, researchers
  - Manual curation of large datasets can introduce errors

A centralized data repository would facilitate further QC

- The stringency of interpretation parameters: initial STR kit testing versus for implemented allele frequency datasets

Considerations

- The importance of acknowledging errors
- Carefully correcting and taking accountability
- Dissemination to the community
- Learning and improving from the experience

- *Balance*

- The desire to immediately correct errors
- In the rush we don’t want to make further mistakes or miss other errors
- What is the best route and timing for dissemination?
- Is there a threshold for reporting changes to frequency data?

Thank you for your attention

peter.vallone@nist.gov

Acknowledgements

Co Authors

Lisa Borsuk