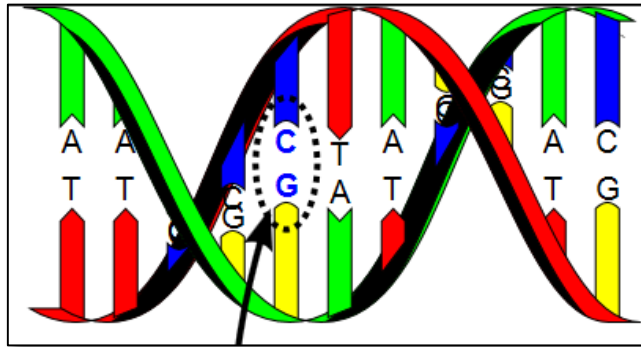


Next Generation Sequencing on the Ion Torrent PGM: New SNP Typing Applications



Future Trends in Forensic DNA Technology

August 6, 2014

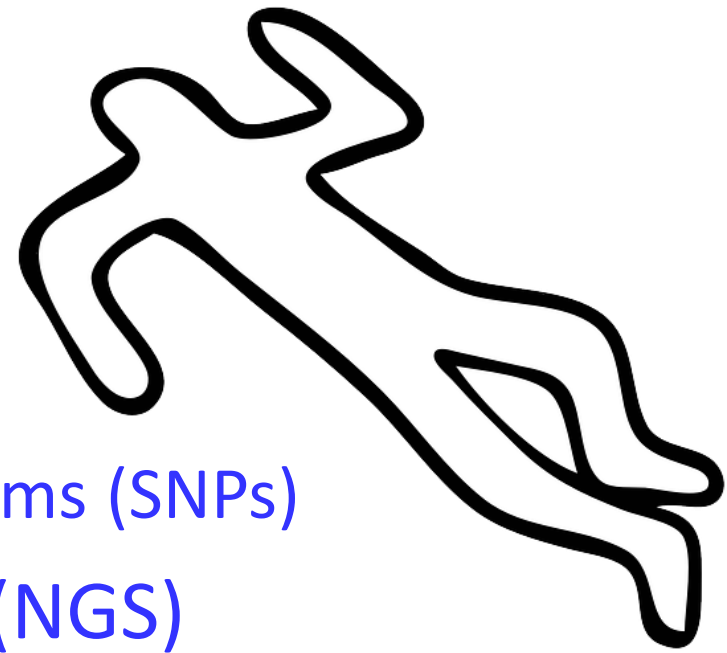
Kevin Kiesler, M.S.

Research Biologist, Applied Genetics Group
National Institute of Standards and Technology

Disclaimer

- Forensic DNA research conducted at NIST is supported by an interagency agreement between the National Institute of Justice and the NIST Law Enforcement Standards Office.
- 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. 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.

Outline

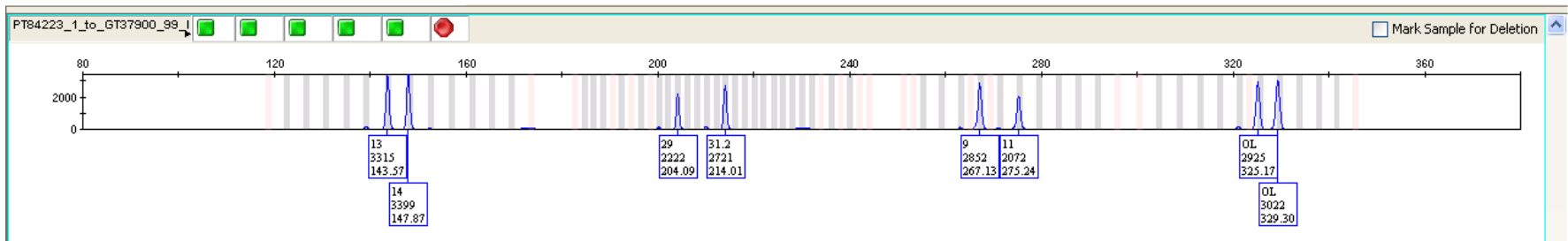


- Forensic DNA markers
 - Single Nucleotide Polymorphisms (SNPs)
- Next generation sequencing (NGS)
 - Personal Genome Machine (PGM) workflow
- Experimental data
 - HID-Ion Ampliseq Identity Panel
 - HID-Ion Ampliseq Ancestry Panel

Forensic DNA Markers

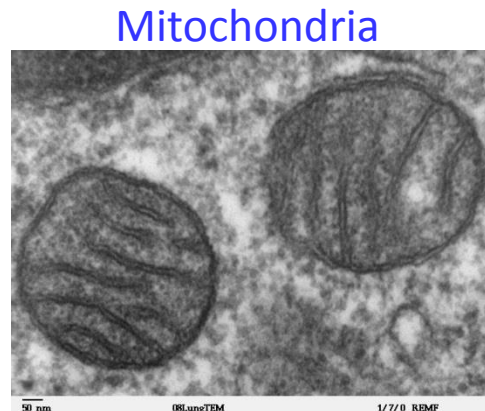
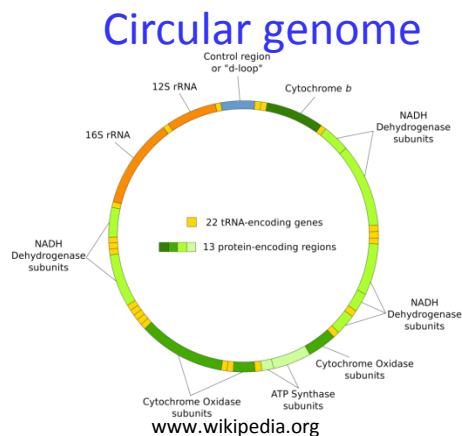
- Short Tandem Repeats (STRs)
 - PCR fragment-length polymorphisms
- Mitochondrial DNA (mtDNA)
 - Sanger sequencing
- Single Nucleotide Polymorphisms (SNPs)

Capillary electrophoresis electropherogram trace

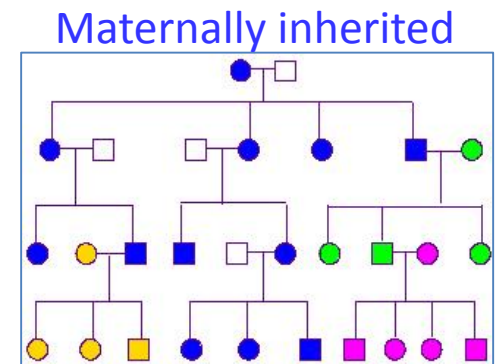


Forensic DNA Markers

- Short Tandem Repeats (STRs)
 - PCR fragment-length polymorphisms
- Mitochondrial DNA (mtDNA)
 - Sanger sequencing
- Single Nucleotide Polymorphisms (SNPs)



<http://remf.dartmouth.edu/images/mammalianLungTEM/source/8.html>



<http://www.orchidcellmark.ca>

Forensic DNA Markers

- Short Tandem Repeats (STRs)
 - PCR fragment-length polymorphisms
- Mitochondrial DNA (mtDNA)
 - Sanger sequencing
- Single Nucleotide Polymorphisms (SNPs)


Most methods are low throughput and/or require a lot of DNA.

NGS can analyze **many SNPs** for **many samples** in one sequencing run.

Categories of SNPs


- Individual Identification (IISNP)
- Ancestry Informative (AISNP)
- Lineage Informative (LISNP)
- Phenotype Informative (PISNP)

Available online at www.sciencedirect.com

 ScienceDirect

 ELSEVIER

Forensic Science International: Genetics Supplement Series 1 (2008) 471–472

 FSI
GENETICS
SUPPLEMENT SERIES

www.elsevier.com/locate/FSIGSS

Research article

Report on ISFG SNP Panel Discussion

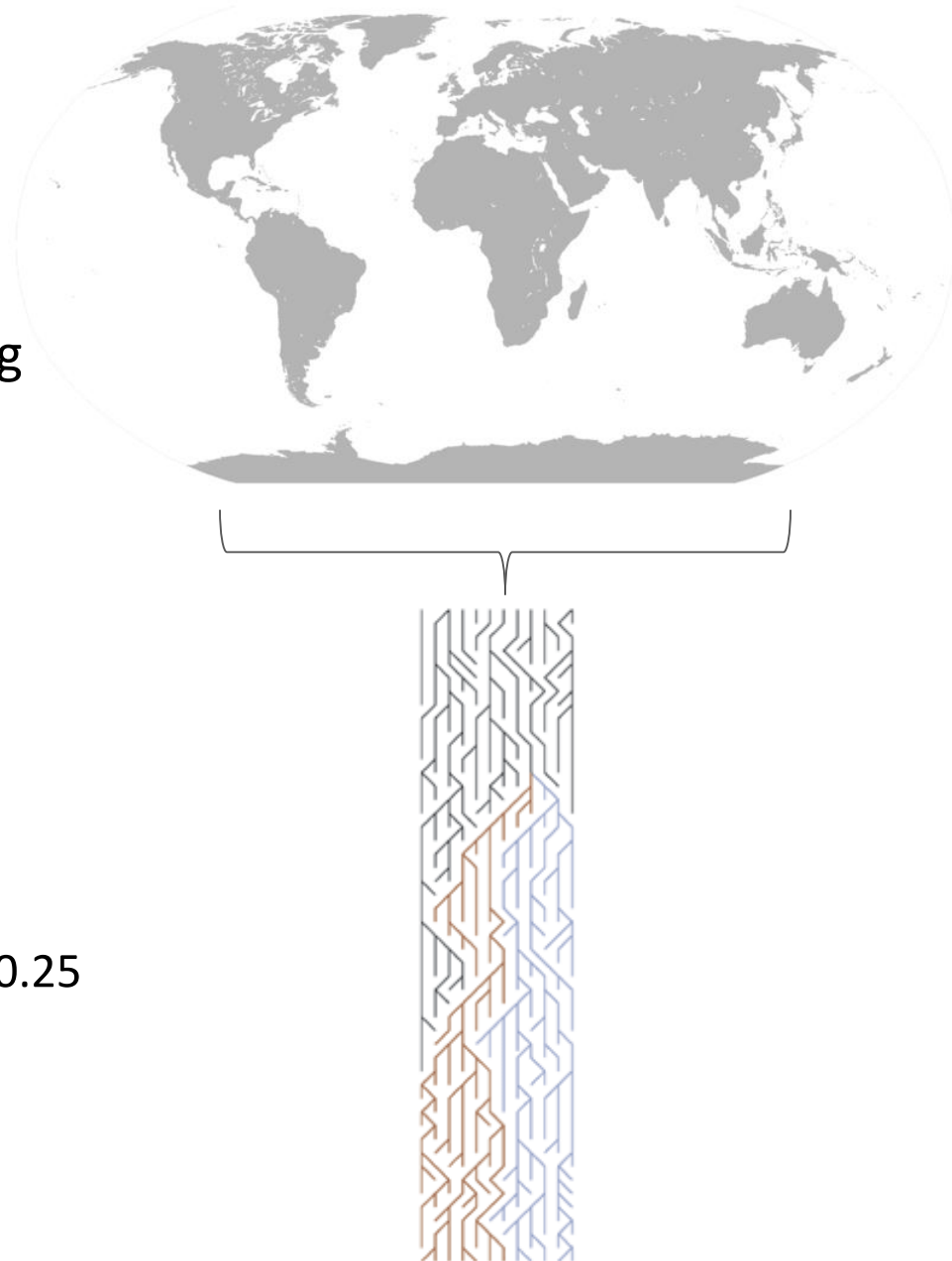
J.M. Butler^{a,*}, B. Budowle^b, P. Gill^c, K.K. Kidd^d, C. Phillips^e,
P.M. Schneider^f, P.M. Vallone^a, N. Morling^g

^a National Institute of Standards and Technology, Gaithersburg, MD, USA
^b FBI Laboratory, Quantico, VA, USA
^c Forensic Science Service, Birmingham, UK
^d Yale University, New Haven, CT, USA
^e University of Santiago de Compostela, Spain
^f University of Cologne, Germany
^g University of Copenhagen, Denmark

Received 25 September 2007; accepted 11 October 2007

SNP Information

- Individual Identification
 - Low Fixation Index (F_{ST}) among worldwide populations
 - Alleles are evenly distributed
 - Balancing has occurred in all populations
 - High heterozygosity
 - e.g. AA = 0.25, AG = 0.5, GG = 0.25



HID-Ion Ampliseq Identity Panel (*version 2.3*)

- 90 autosomal SNPs
 - SNPforID 52 (Phillips et al. 2007)
 - Kidd Lab 45 (Pakstis *et al.* 2010)
- 30 Y-chromosome SNPs
 - Hammer Lab (Karafet *et al.* 2008)
- RMP $\approx 4 \times 10^{-36}$



Hum Genet (2010) 127:315–324
DOI 10.1007/s00439-009-0771-1

ORIGINAL INVESTIGATION

SNPs for a universal individual identification panel

Andrew J. Pakstis · William C. Speed · Rixun Fang ·
Fiona C. L. Hyland · Manohar R. Furtado ·
Judith R. Kidd · Kenneth K. Kidd

Available online at www.sciencedirect.com


 ScienceDirect 

Forensic Science International: Genetics 1 (2007) 180–185
www.elsevier.com/locate/bsfig

Evaluation of the Genplex SNP typing system and
a 49plex forensic marker panel

C. Phillips^{a,*}, R. Fang^b, D. Ballard^c, M. Fondevila^a, C. Harrison^c, F. Hyland^b,
E. Musgrave-Brown^c, C. Proff^d, E. Ramos-Luis^a, B. Sobrino^a, A. Carracedo^a,
M.R. Furtado^b, D. Syndercombe Court^c, P.M. Schneider^d

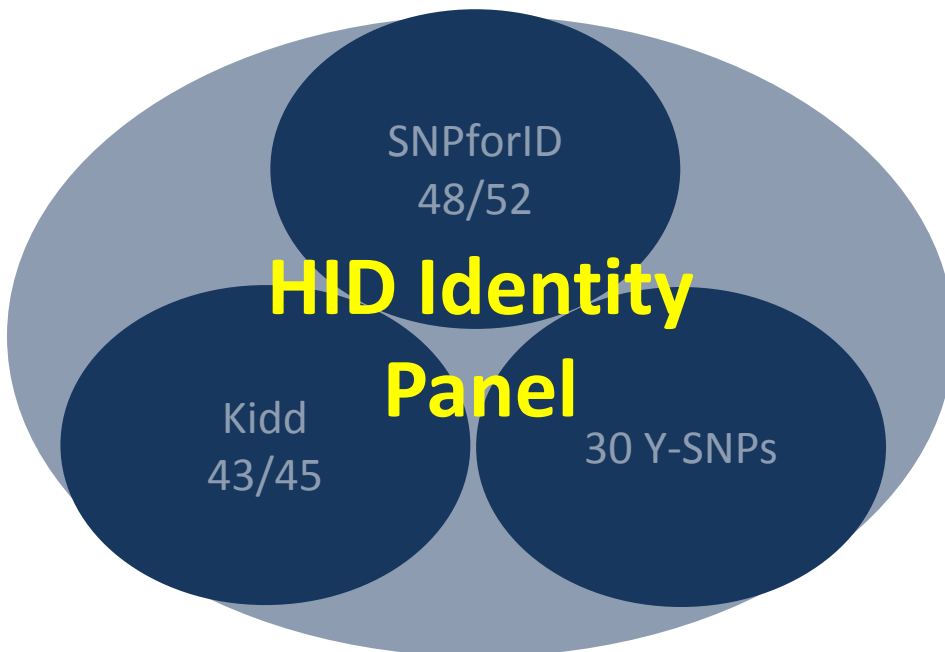
The SNPforID Consortium

 GENOME RESEARCH

**New binary polymorphisms reshape and increase resolution of the
human Y chromosomal haplogroup tree**

Tatiana M. Karafet, Fernando L. Mendez, Monica B. Meilerman, et al.

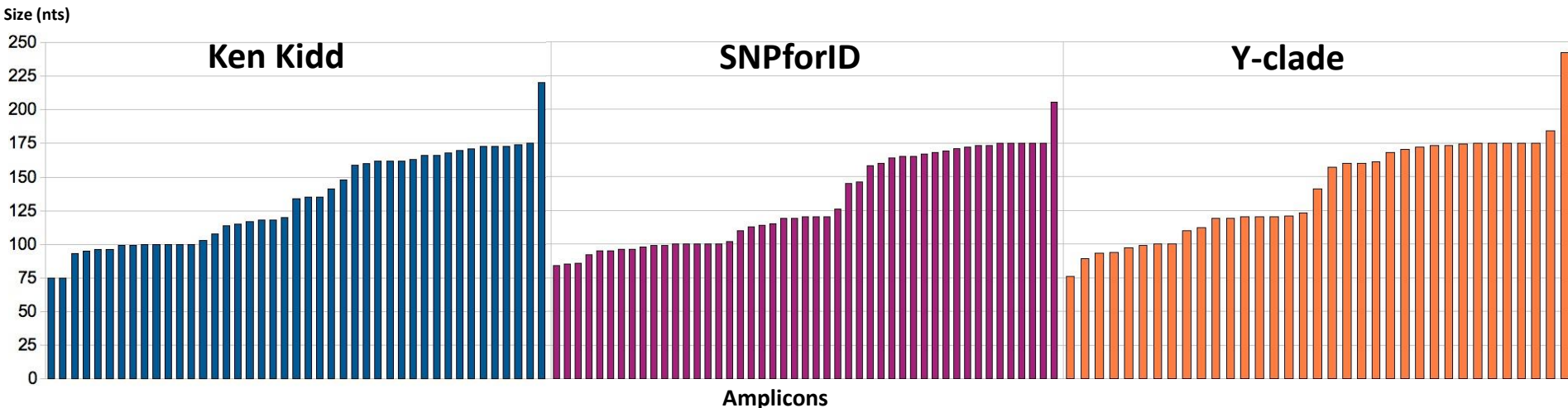
Genome Res. 2008 18: 830–838 originally published online April 2, 2008
Access the most recent version at [doi:10.1101/gr.7172008](https://doi.org/10.1101/gr.7172008)



HID-Ion AmpliSeq™ Identity Panel

version 4.0

- **124 markers** *(for degraded samples, samples with loss of STR data)*
 - 90 autosomal markers (1 shared)
 - 43 Ken Kidd¹
 - 48 SNPforID²
 - average autosomal marker size of 132 nts
 - 34 upper Y-clade markers³
 - average Y-clade marker size 141 nts



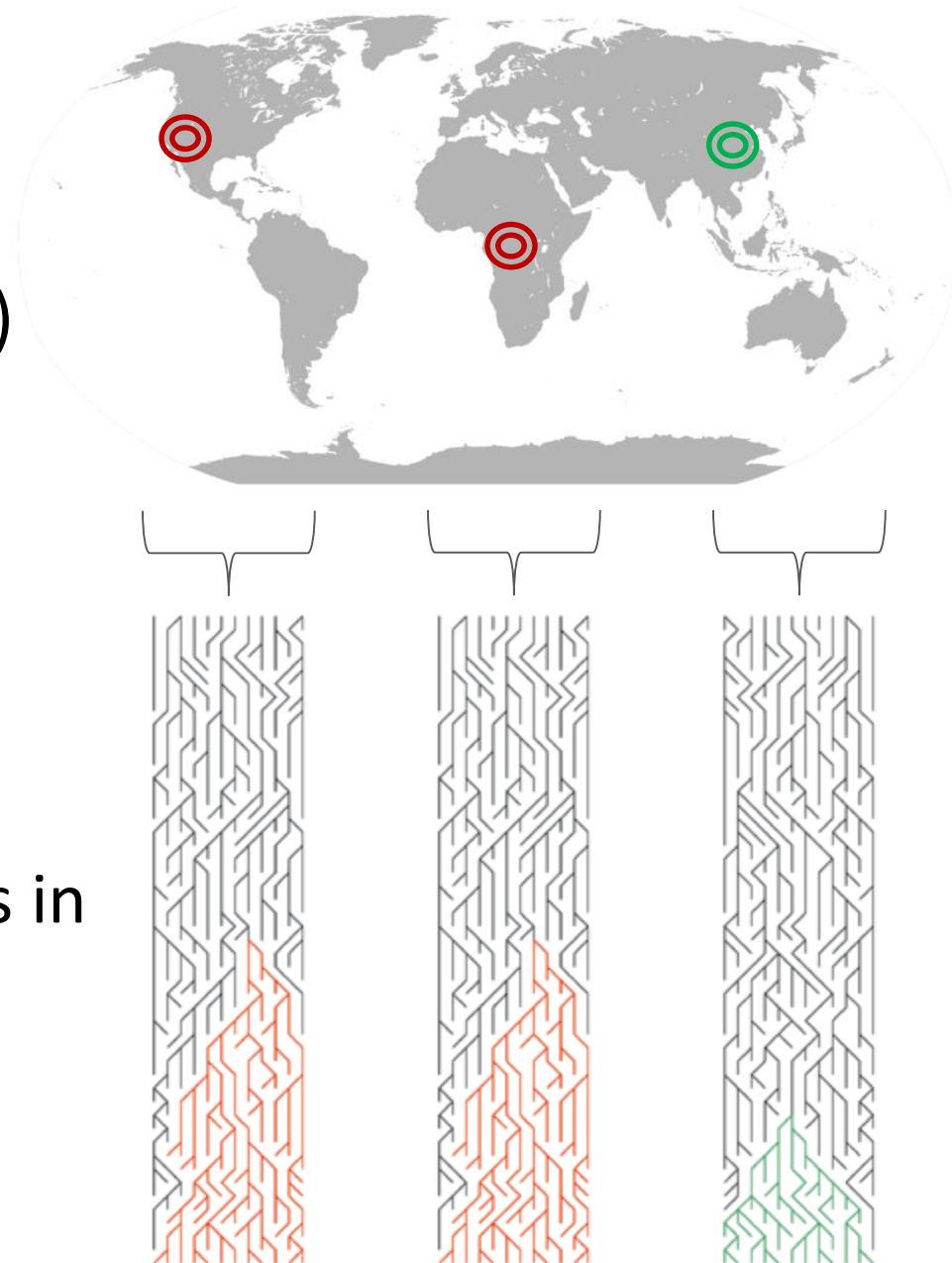
¹ Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C., Furtado, M. R., Kidd, J. R., & Kidd, K. K. (2010). SNPs for a universal individual identification panel. *Human Genetics*, 127(3), 315–324.

² Phillips, C., Fang, R., Ballard, D., Fondevila, M., Harrison, C., Hyland, F., et al. (2007). Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. *Forensic Science International: Genetics*, 1(2), 180–185.

³ Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., & Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research*, 18(5), 830–838.

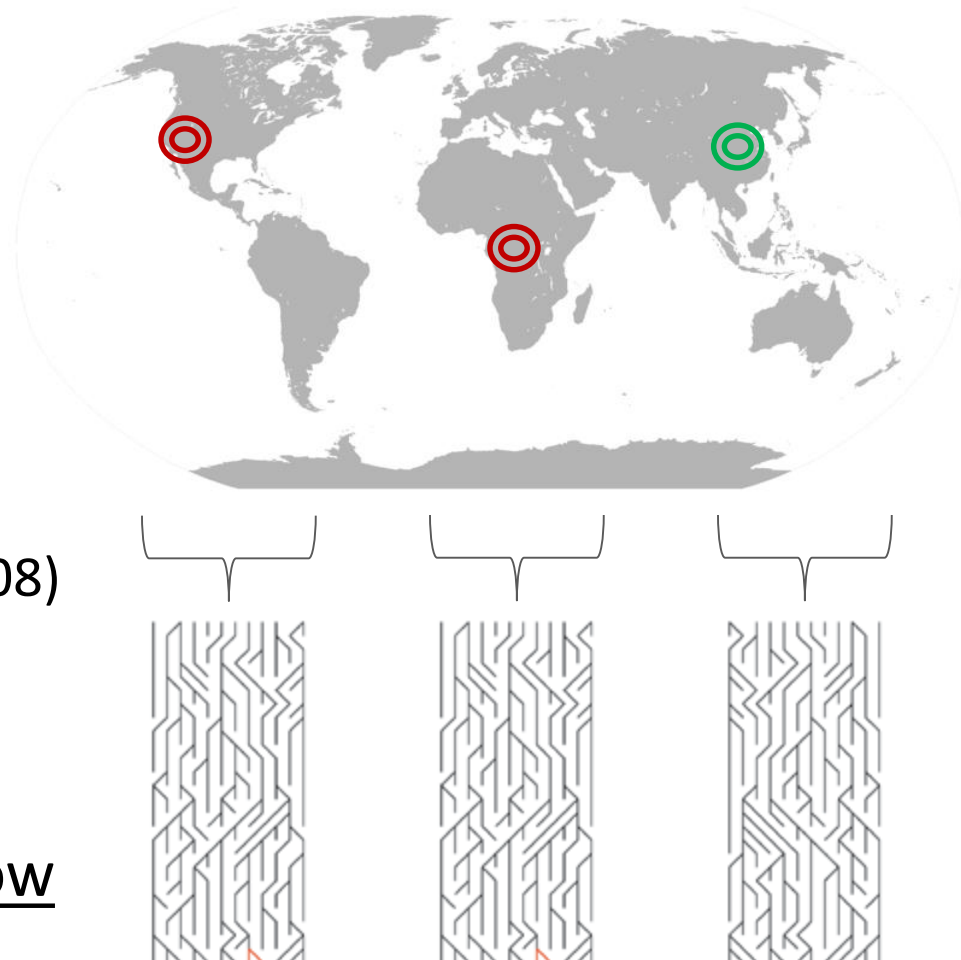
SNP Information

- Ancestry Markers
 - High Fixation Index (F_{ST})
 - Population specific fixation has occurred
 - Low heterozygosity
- Example
 - Malaria resistance SNPs in Sub-Saharan Africa



SNP Information

- HID Ancestry Panel
 - Beta version 3.0
 - 170 loci
 - Derived from
 - Seldin lab (Kosoy *et. al* 2008)
 - 128 SNP markers
 - Kidd lab (Kidd *et. al* 2014)
 - 55 SNP markers
 - V4.0 publicly available now



RESEARCH ARTICLE

Human Mutation

OFFICIAL JOURNAL
HGV
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Ancestry Informative Marker Sets for Determining Continental Origin and Admixture Proportions in Common Populations in America

Roman Kosoy,¹ Rami Nassir,¹ Chao Tian,¹ Phoebe A. White,² Lesley M. Butler,³ Gabriel Silva,⁴ Rick Kittles,⁵ Marta E. Alarcon-Riquelme,⁶ Peter K. Gregersen,⁷ John W. Belmont,⁸ Francisco M. De La Vega,² and Michael F. Seldin^{1*}

Forensic Science International: Genetics 10 (2014) 23–32

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

ELSEVIER

FSI GENETICS

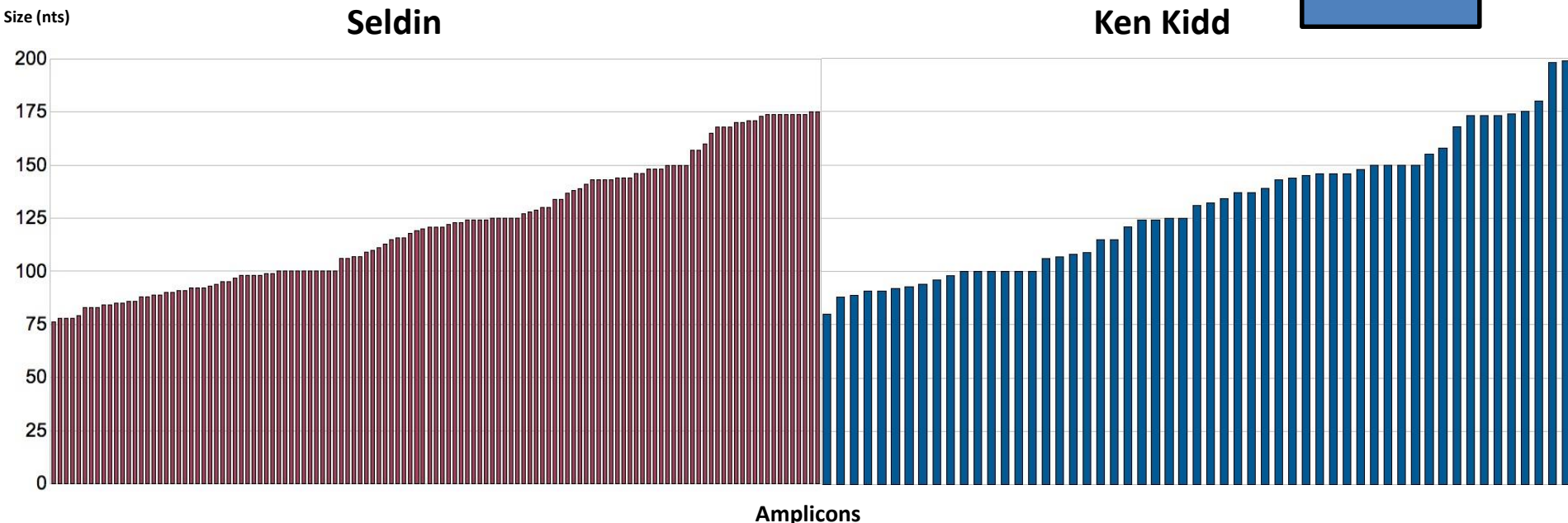
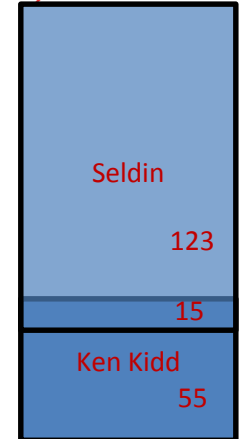
Progress toward an efficient panel of SNPs for ancestry inference[☆]

Kenneth K. Kidd^{a,*}, William C. Speed^a, Andrew J. Pakstis^a, Manohar R. Furtado^{b,1}, Rixun Fang^b, Abeer Madbouly^c, Martin Maiers^c, Mridu Middha^{c,2}, Françoise R. Friedlaender^d, Judith R. Kidd^a

CrossMark

HID-Ion AmpliSeq™ Ancestry Panel

- 165 autosomal markers (*for casework samples, investigative leads*) 165 SNPs
 - 55 Ken Kidd (55 AIM set ¹)
 - average marker size of 130 nts
 - 123 Seldin ²
 - average marker size of 122 nts

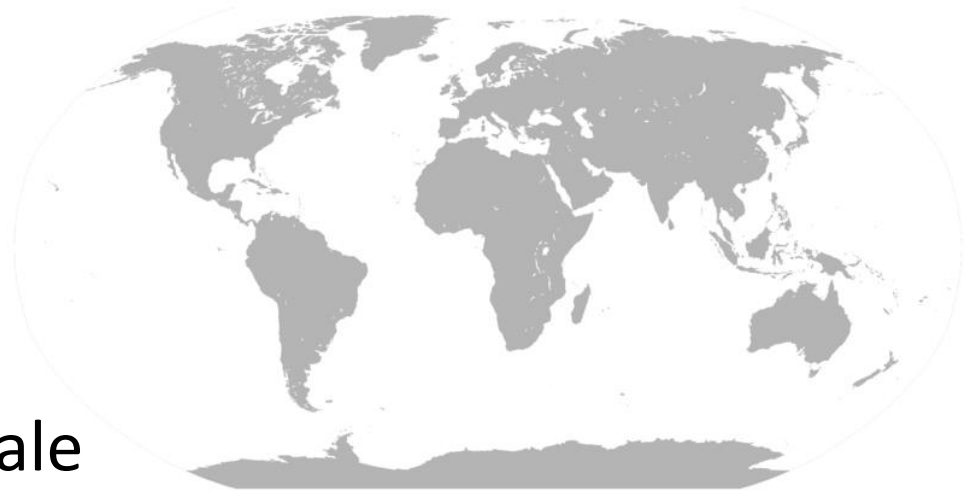


¹ Kidd, K. K. (2012). Better SNPs for Better Forensics: Ancestry, Phenotype, and Family Identification. Poster

² Nassir, R., Kosoy, R., Tian, C., White, P. A., Butler, L. M., Silva, G., et al. (2009). An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. *BMC Genetics*, 10(1), 39.

SNP Information

- RMP/LR calculations
 - FROG KB
 - From Dr. Kidd's Lab at Yale



FROG-kb Forensic Resource On Genetics knowledge base

DANIELE PODINI'S LIST OF 32 AISNPS

Daniele Podini's list of 32 AISNPs - Probability of Genotype in each Population

Indicates the values are within an order of magnitude of the highest likelihood.

Population (Region, Sample Size 2N)	Probability of Genotype in each Population	Likelihood Ratio
Komi-Zyrian (Asia,94)	● 2.3E-8	
Khanty (Asia,100)	● 1.4E-8	1.7
Chuvash (Europe,84)	● 1.0E-8	2.2
Hazara_HGDP-CEPH (Asia,50)	● 5.7E-9	4.0
Finns (Europe,72)	● 3.7E-9	6.1
Russians_Archangelsk (Europe,68)	1.4E-9	16
Burusho_HGDP-CEPH (Asia,50)	9.4E-10	24
Brahui_HGDP-CEPH (Asia,50)	8.9E-10	26
Adygei (Europe,108)	7.9E-10	29
Russians (Europe,96)	7.9E-10	29
Sardinian_HGDP-CEPH (Europe,56)	5.2E-10	44

FROG-kb is supported by National Institute of Justice grant 2010-DN-BX-K226

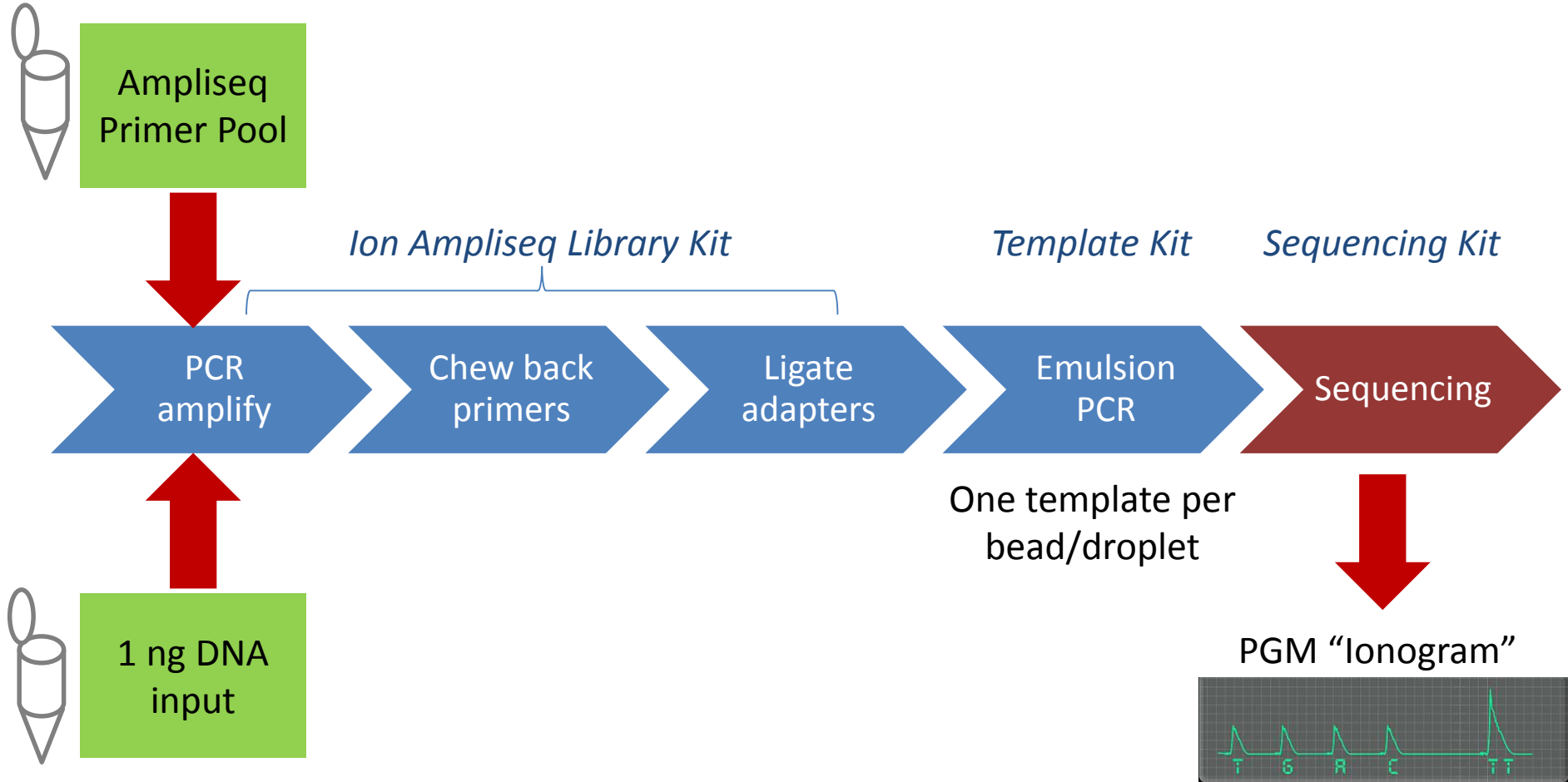
EastAsia,80)	6.8E-19	3.4E+10
(EastAsia,132)	3.5E-19	6.6E+10
San Francisco Chinese (EastAsia,124)	1.9E-19	1.2E+11
Rondonia (SouthAmerica,100)	2.3E-20	9.9E+11

<http://frog.med.yale.edu/FrogKB/>

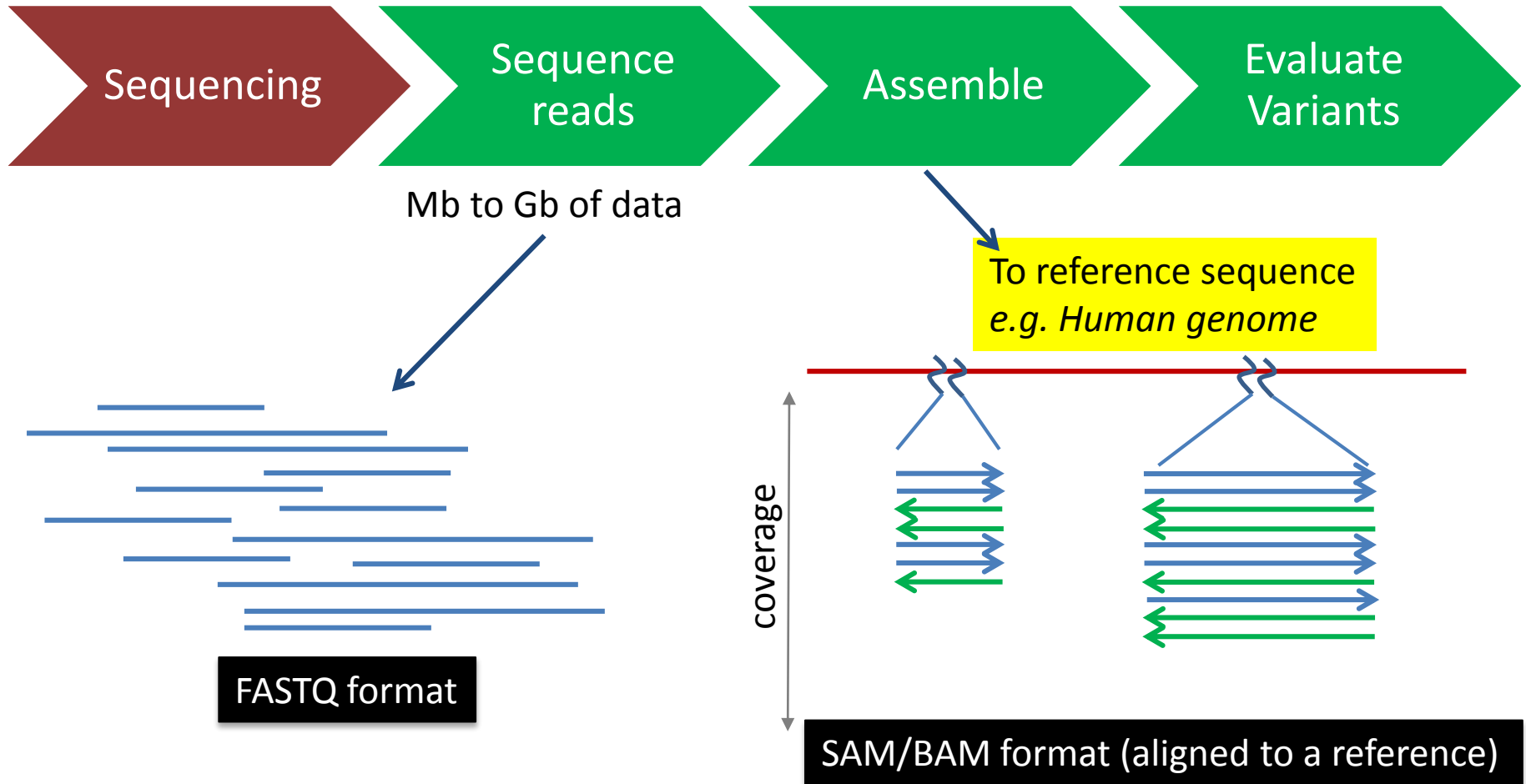
Pima (NorthAmerica,106)	3.9E-22	5.8E+13
American (Africa,182)	3.0E-22	7.6E+13
Yoruba_HGDP-CEPH (SouthAmerica,48)	2.3E-23	1.0E+15
Pima (NorthAmerica,104)	5.0E-25	4.6E+16
Yoruba (SouthAmerica,114)	6.3E-27	3.6E+18
Yoruba_HGDP-CEPH (Africa,14)	1.7E-31	1.4E+23
Yoruba (Africa,80)	6.0E-32	3.8E+23
Yoruba (Africa,44)	2.3E-32	1.0E+24
Yoruba (Africa,90)	4.5E-33	5.0E+24
Yoruba_HGDP-CEPH (Africa,48)	2.7E-33	8.4E+24
Yoruba (Africa,80)	1.1E-36	2.1E+28
Yoruba (Africa,78)	3.3E-37	7.0E+28
Yoruba_HGDP-CEPH (Africa,30)	3.0E-39	7.7E+30
Yoruba_HGDP-CEPH (Africa,50)	9.5E-41	2.4E+32
Yoruba (Africa,96)	4.1E-42	5.6E+33
Yoruba (Africa,78)	1.2E-44	1.9E+36
Yoruba (Africa,156)	3.7E-45	6.2E+36
Yoruba (Africa,140)	1.8E-45	1.3E+37

Sequencing Methods

Ampliseq Workflow



Analysis Workflow



Life Tech - Ion Torrent - PGM

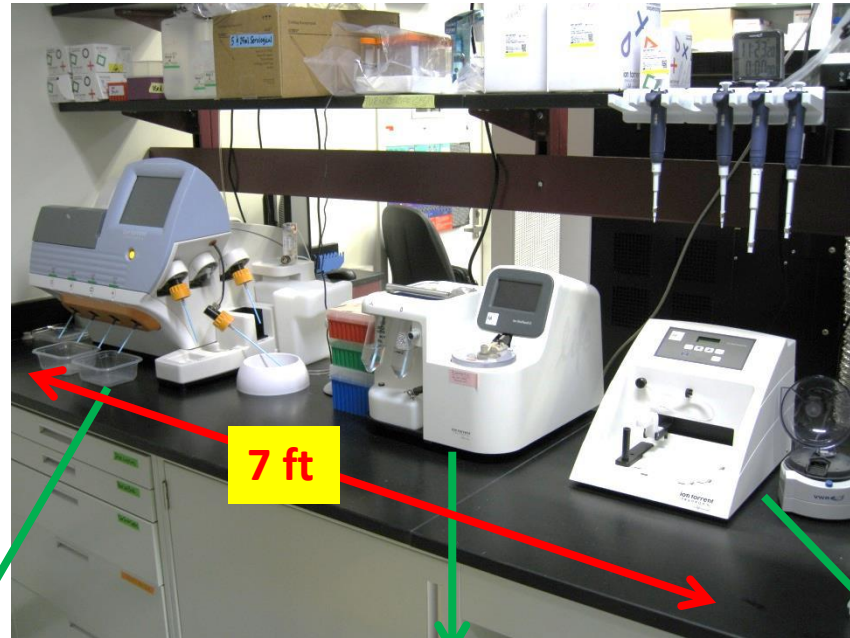
- Ion Torrent Personal Genome Machine (PGM)
 - Launched in 2010
- Ion Torrent sequencing:
 - Emulsion PCR for single copy reactors
 - Non-labeled nucleotide triphosphates
 - Flowed over a bead on a semiconductor surface
- Hydrogen Ion detection
 - pH change is detected
 - **No optics**



Ion Torrent PGM Workflow



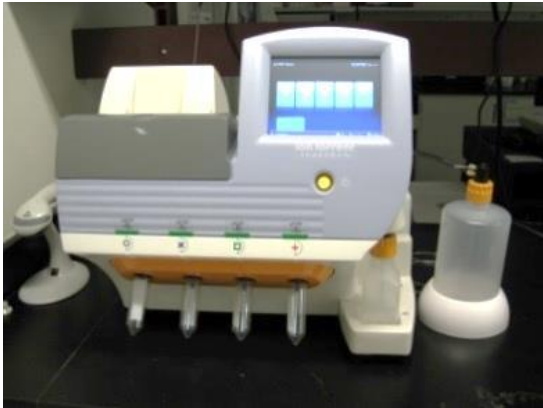
The PGM Instrument at NIST



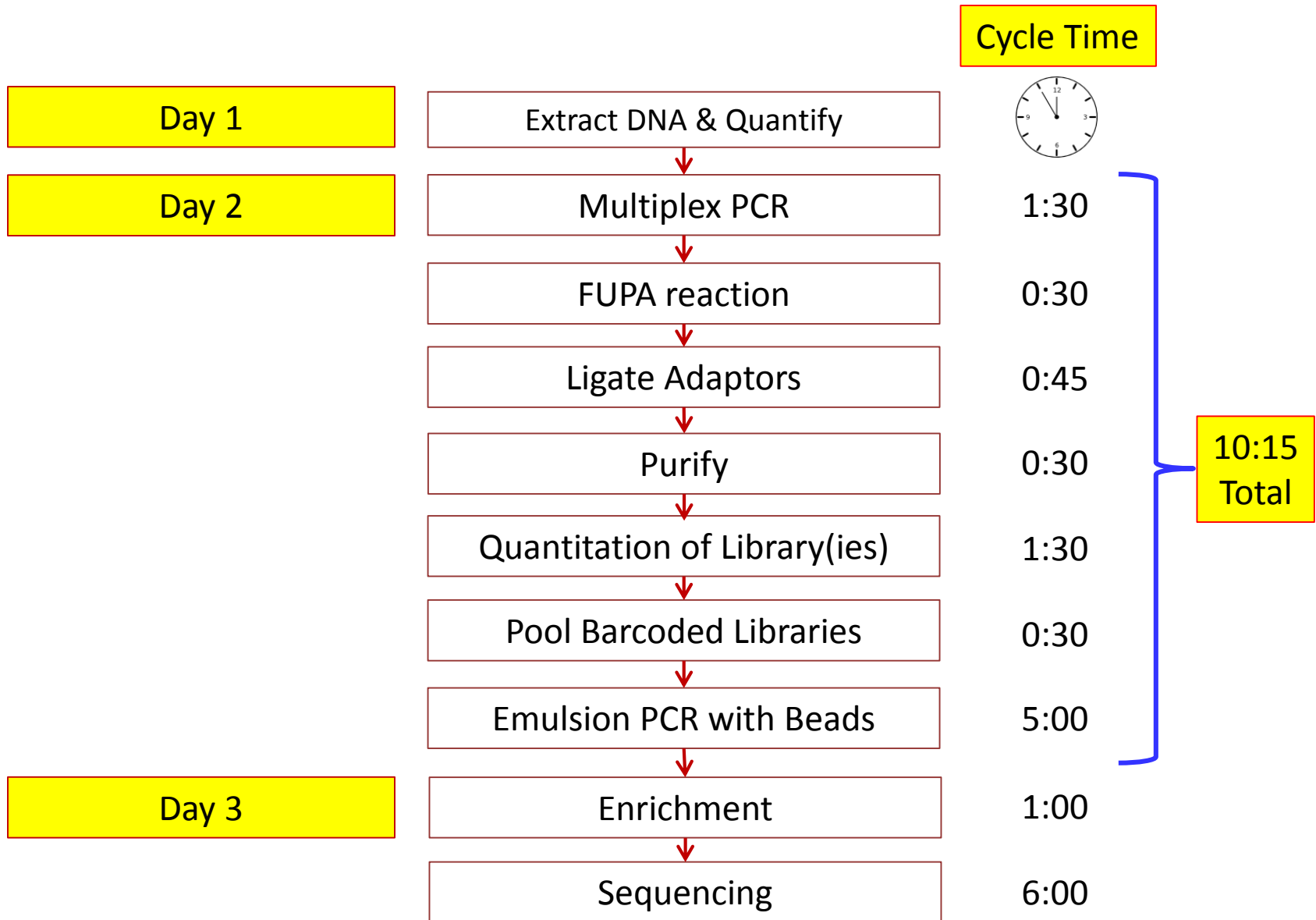
PGM
Sequencer

OneTouch 2
(Emulsion PCR)

OneTouch ES
(Enrichment)

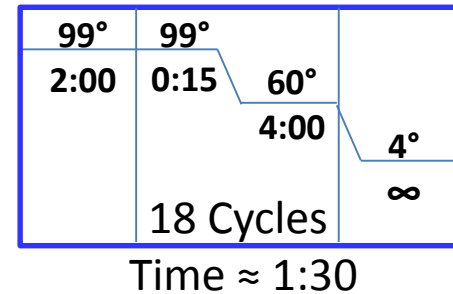


Workflow Overview



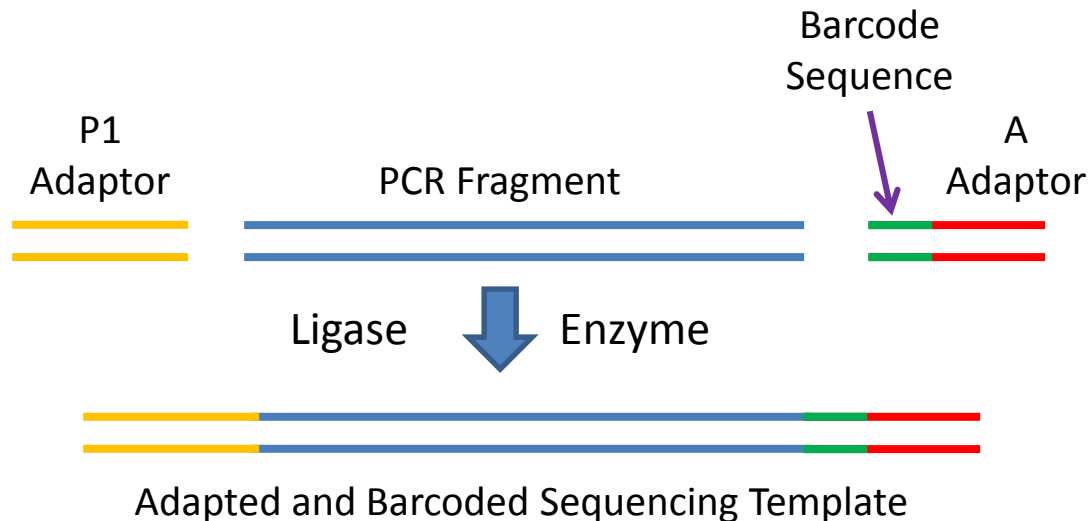
Front-End: Multiplex PCR

- HID-Ion Ampliseq **Identity Panel** (IISNP)
 - 120 markers in a single PCR reaction
 - Amplified regions **33 bp to 192 bp** long
- HID-Ion Ampliseq **Ancestry Panel** (AISNP)
 - 170 markers in a single PCR reaction
 - Amplified regions **34 bp to 136 bp** long
- **Small amplicons well suited to degraded or damaged DNA**



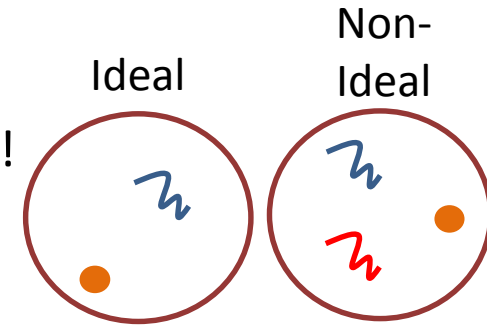
Digest Primer Regions & Ligate Adaptors

- Enzymatic digestion removes ≈ 25 bp from ends of amplicons
- Universal sequencing adaptors are ligated to DNA
 - Adaptors termed P1 and A
- **Barcoded** sequencing adaptors can be used in this step
 - Sequence multiple samples in one PGM run



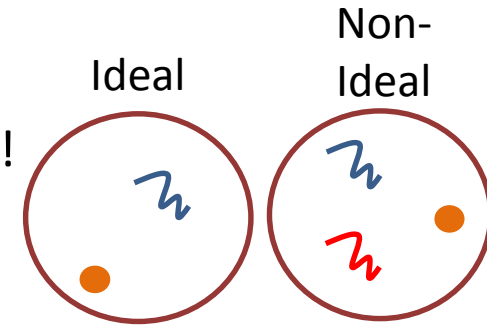
Prepare Ion Sphere Particles (ISPs)

- Libraries quantified by qPCR
 - Quantity of DNA going into emPCR is very important!
 - Goal: 10 % to 30 % template positive ISPs
 - Too much DNA → polyclonal ISPs (mixed read)



Prepare Ion Sphere Particles (ISPs)

- Libraries quantified by qPCR
 - Quantity of DNA going into emPCR is very important!
 - Goal: 10 % to 30 % template positive ISPs
 - Too much DNA → polyclonal ISPs (mixed read)



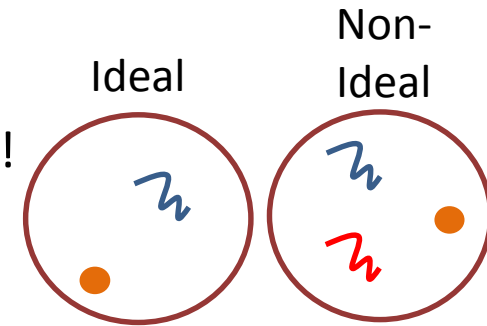
- Emulsion PCR
 - Nanoliter droplets of PCR reagents in oil
 - Attaches sequencing template to the ISP



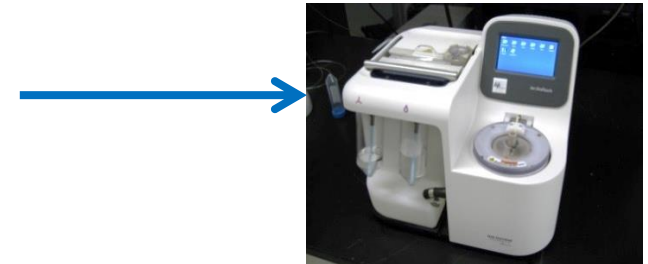
OneTouch 2

Prepare Ion Sphere Particles (ISPs)

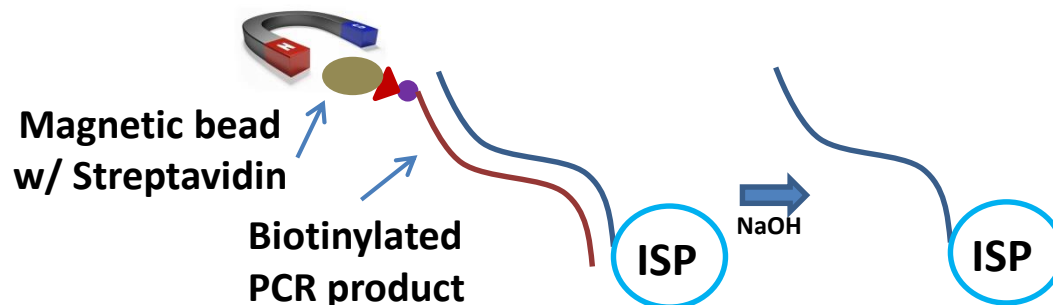
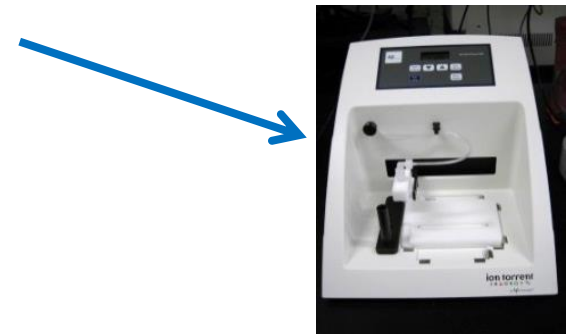
- Libraries quantified by qPCR
 - Quantity of DNA going into emPCR is very important!
 - Goal: 10 % to 30 % template positive ISPs
 - Too much DNA → polyclonal ISPs (mixed read)



- Emulsion PCR
 - Nanoliter droplets of PCR reagents in oil
 - Attaches sequencing template to the ISP

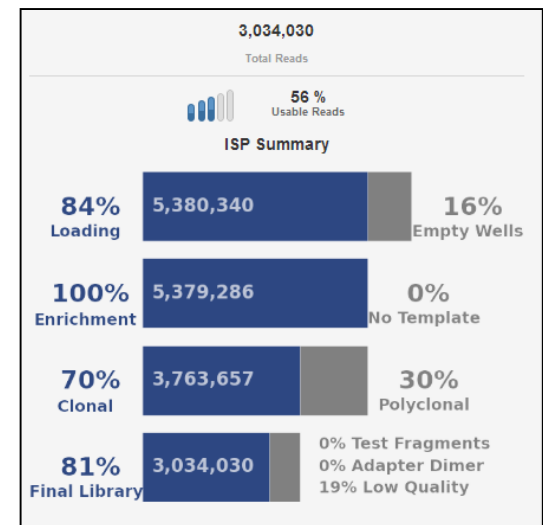
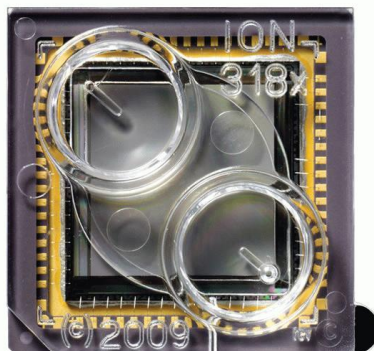
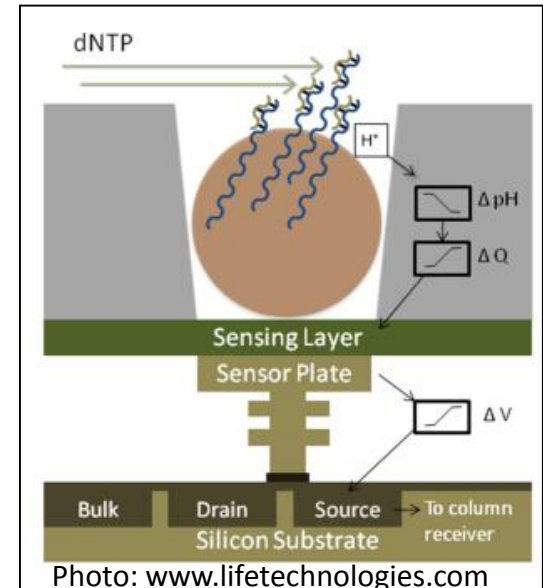


- Enrich for positive ISPs
 - Liquid handler removes non-templated ISPs
 - Biotinylated primer/streptavidin beads



Sequencing & Data Analysis

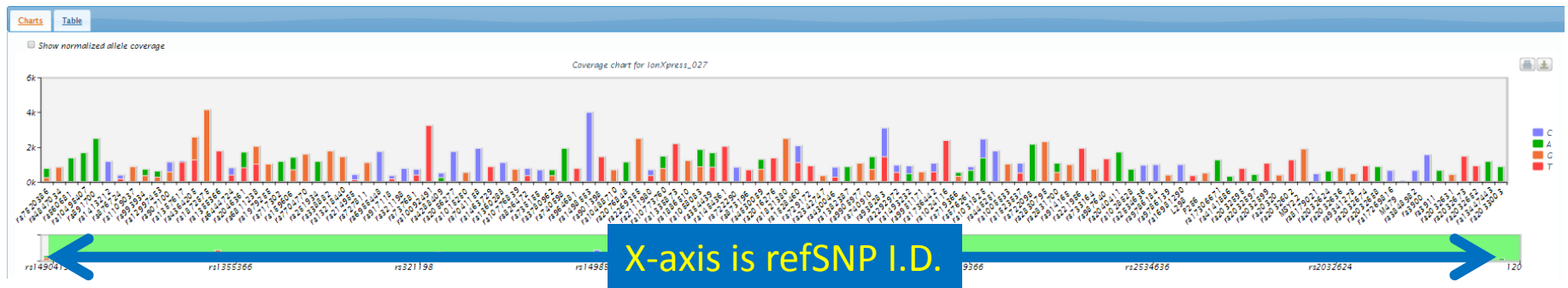
- Library ISPs loaded onto chip
- PGM runs flows & detects pH
- Torrent Server & Torrent Suite Software
 - Processes pH signal into base calls
 - Maps reads to reference genome
 - Displays run summary



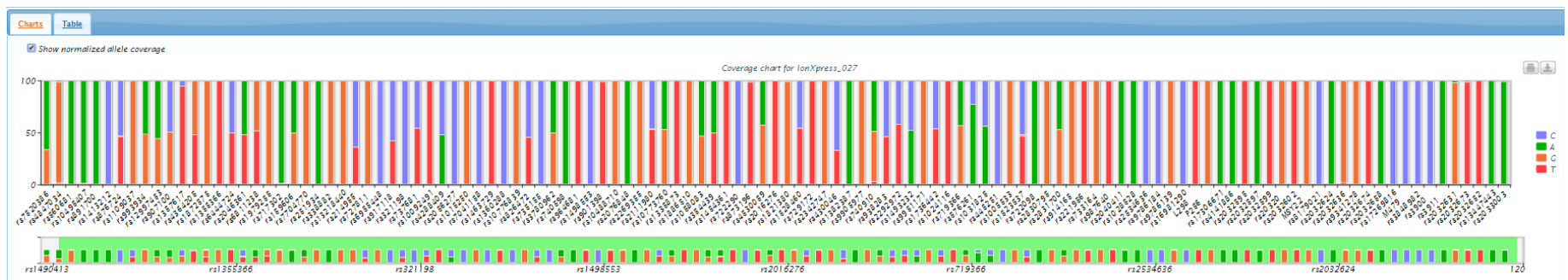
Data Analysis

HID SNP Genotyper Plugin

- Allele coverage histogram



- Normalized y-axis scale



Autosomal SNPs

Y-SNPs






Experimental Data

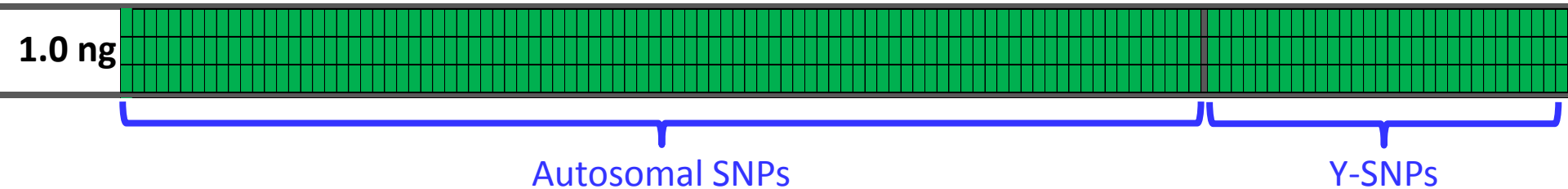
- HID Identity Panel
 - Sensitivity study
 - Degraded DNA study
 - Mixture study
- HID Ancestry Panel
 - Ancestry prediction

HID SNP Panel Sensitivity Study






- Dynamic range of DNA input to PCR
 - 1 ng is recommended
 - 10 ng (1 data point) – no problems were observed
 - 1 ng
 - 0.5 ng
 - 0.1 ng
 - 0.05 ng
- 3 Replicates
- Libraries were generated and pooled (n = 12)
 - Sequenced on PGM 318 chip (11 M wells)
 - 200 bp read chemistry

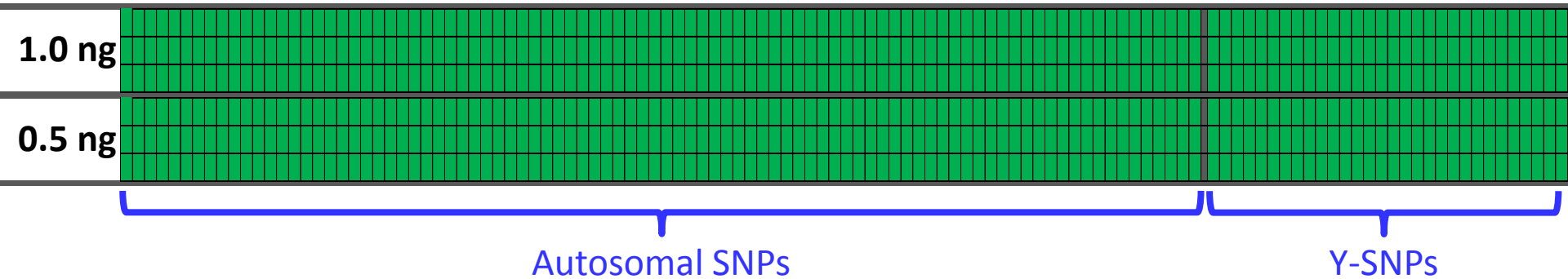
HID SNP Panel Sensitivity Study

- Sequence coverage per locus
 - 200 or more reads = green 
 - Below 200 reads = yellow 
 - Below 50 reads = purple 
 - Below quality cutoff = black 
 - Incorrect genotype = red X 








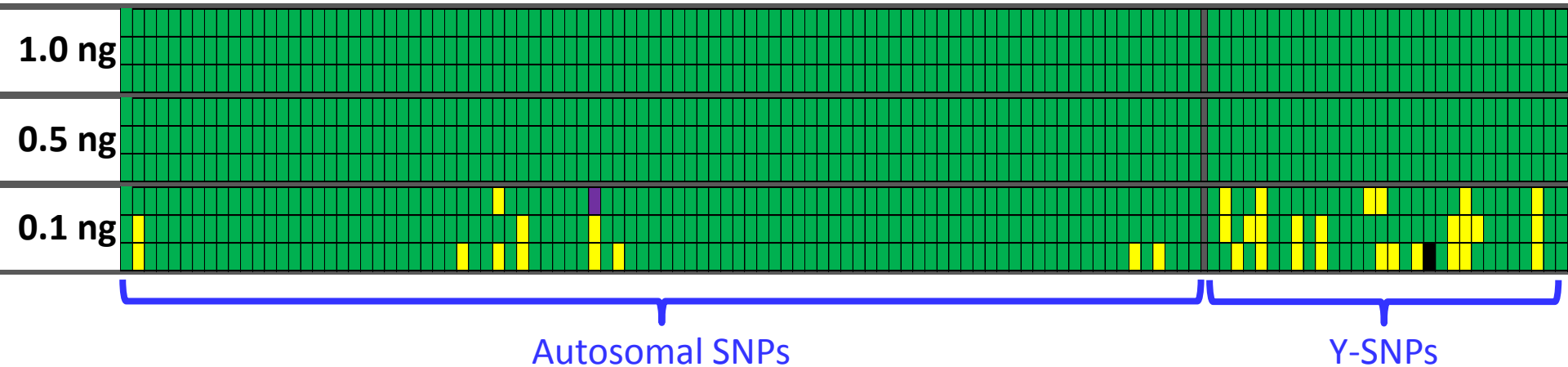
HID SNP Panel Sensitivity Study

- Sequence coverage per locus
 - 200 or more reads = green 
 - Below 200 reads = yellow 
 - Below 50 reads = purple 
 - Below quality cutoff = black 
 - Incorrect genotype = red X 








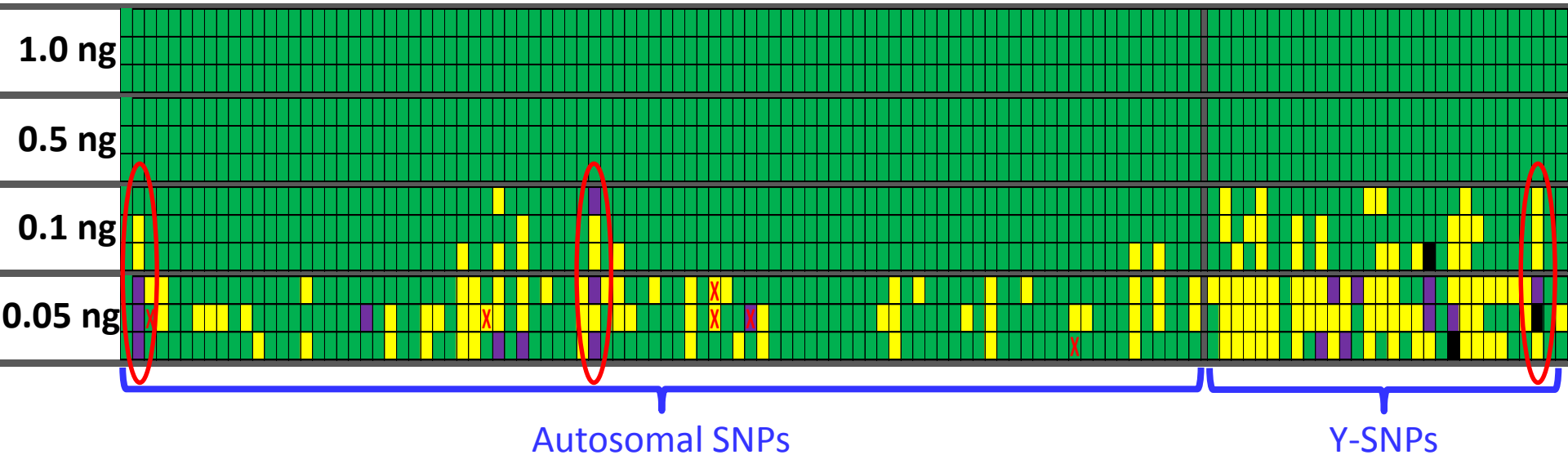
HID SNP Panel Sensitivity Study

- Sequence coverage per locus
 - 200 or more reads = green 
 - Below 200 reads = yellow 
 - Below 50 reads = purple 
 - Below quality cutoff = black 
 - Incorrect genotype = red X 



HID SNP Panel Sensitivity Study

- Sequence coverage per locus
 - 200 or more reads = green 
 - Below 200 reads = yellow 
 - Below 50 reads = purple 
 - Below quality cutoff = black 
 - Incorrect genotype = red X 



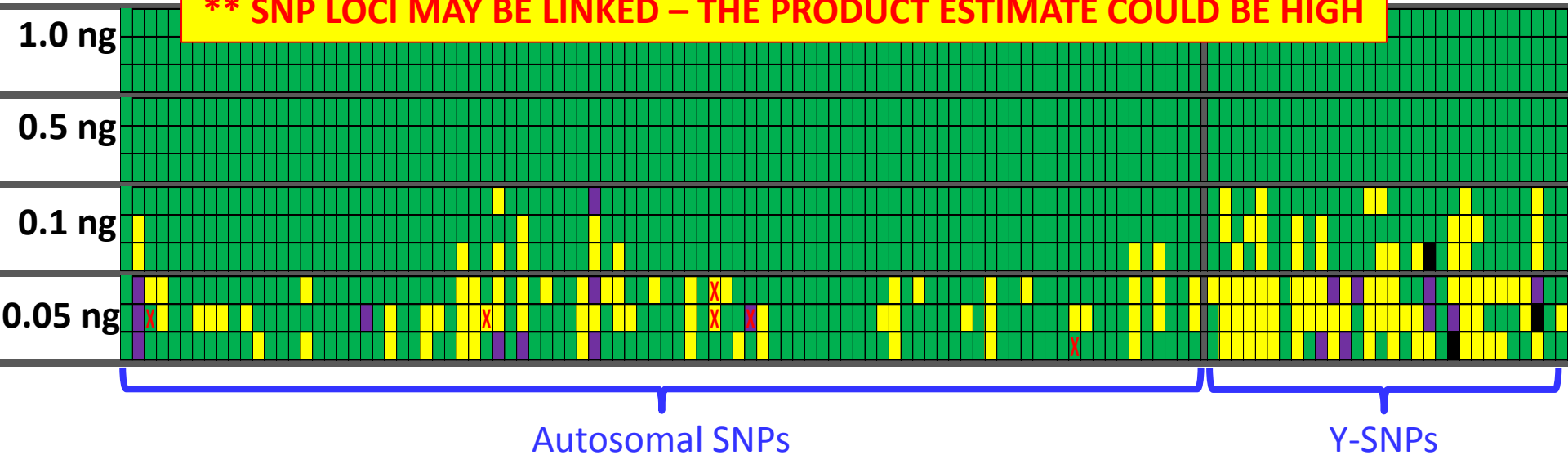
HID SNP Panel Sensitivity

RMP With 200 Read Cutoff

	% Below 200 Reads	RMP* Estimate SNPforID 52	RMP* Estimate Kidd 45	RMP Estimate Product **
1.0 ng	0 %	1 in 1.0×10^{15}	1 in 4.0×10^{14}	1 in 4.1×10^{29}
0.5 ng	0 %	1 in 1.0×10^{15}	1 in 4.0×10^{14}	1 in 4.1×10^{29}
0.1 ng	10 %	1 in 4.4×10^{14}	1 in 2.6×10^{13}	1 in 1.1×10^{28}
0.05 ng	35 %	1 in 7.4×10^{11}	1 in 1.7×10^{10}	1 in 1.2×10^{22}

* RMP calculations from Frog KB (<http://frog.med.yale.edu/FrogKB/>)

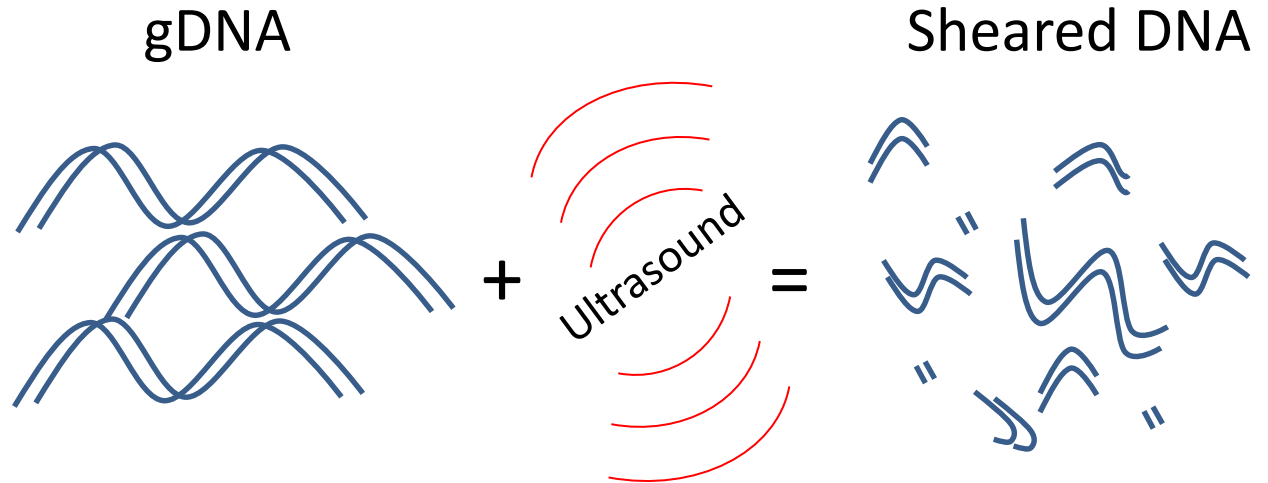
**** SNP LOCI MAY BE LINKED – THE PRODUCT ESTIMATE COULD BE HIGH**



HID SNP Panel

Degraded DNA Study

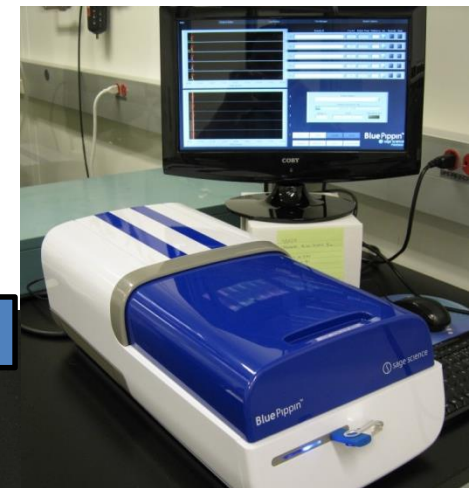
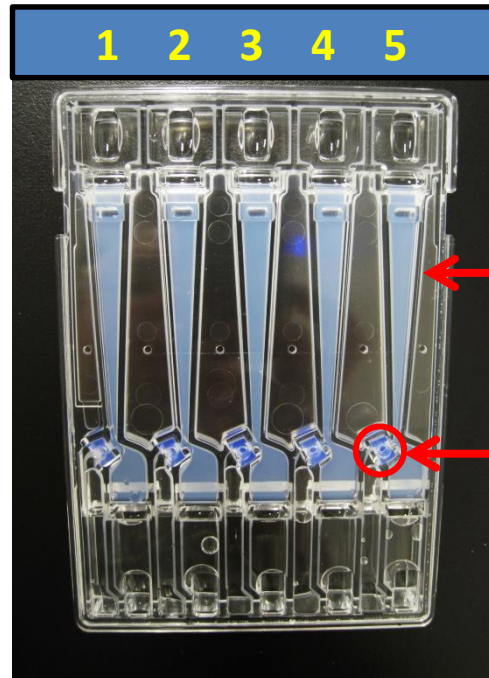
- Total genomic DNA was sheared
 - Covaris S2 Focused Ultrasonicator



HID SNP Panel

Degraded DNA Study

- Sheared DNA was fractionated by size range
 - Blue Pippin system (3% Gel)
 - Automated size selection
 - 1) 50 bp to 200 bp
 - 2) 50 bp to 150 bp
 - 3) 50 bp to 100 bp
 - 4) 50 bp to 75 bp
 - 5) 35 bp to 50 bp



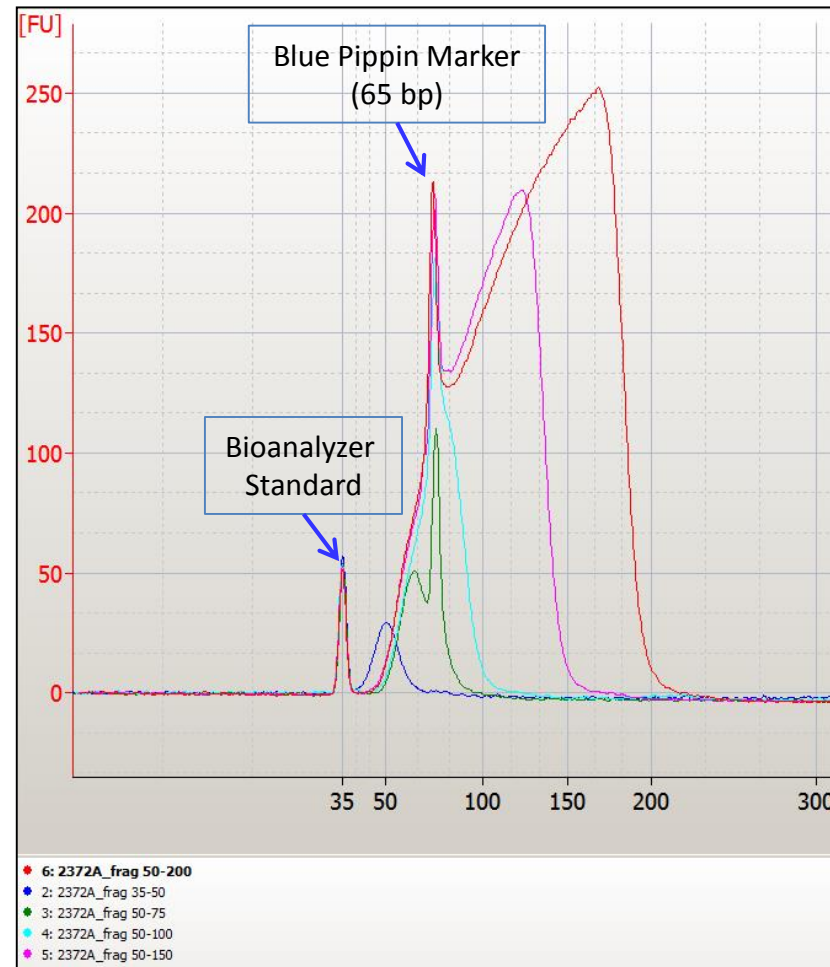
Five individual
agarose columns

Size fractionated
fragments collected
into recovery wells

HID SNP Panel

Degraded DNA Study

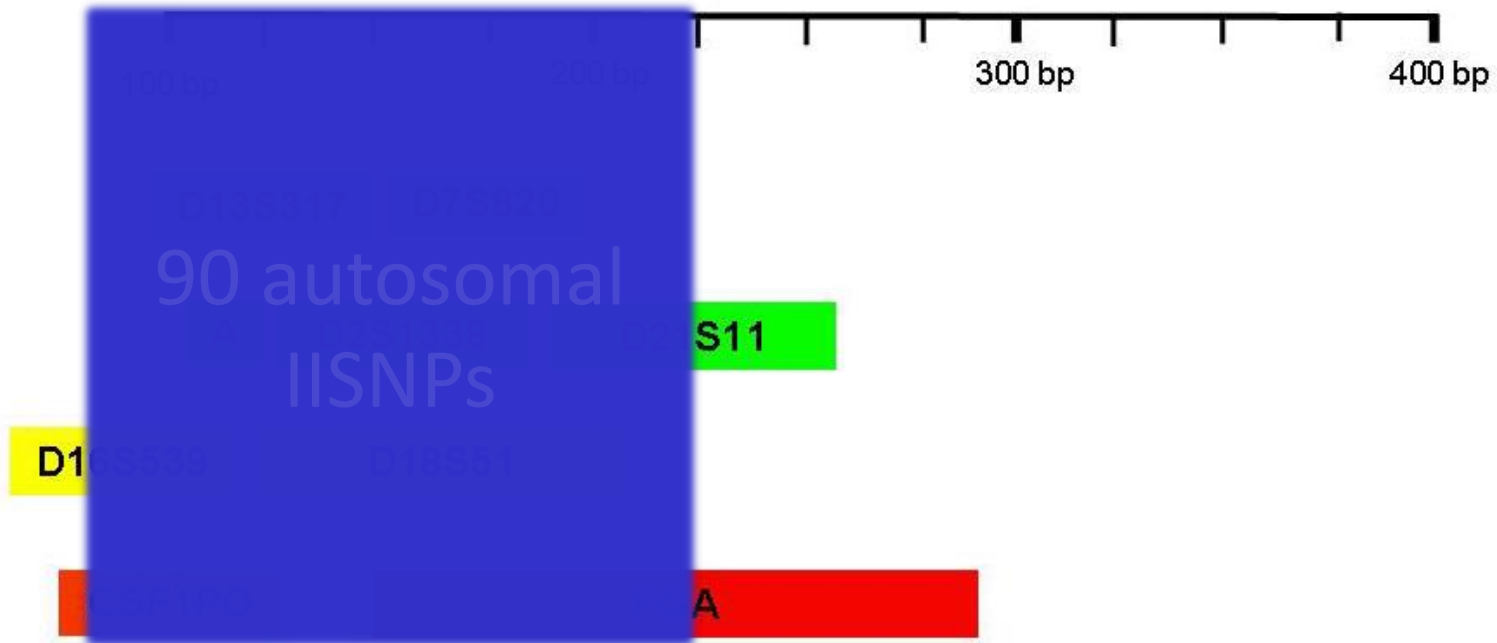
- Sheared DNA was fractionated by size range
 - Agilent Bioanalyzer Trace
 - Size selected sheared DNA
 - 50 bp to 200 bp
 - 50 bp to 150 bp
 - 50 bp to 100 bp
 - 50 bp to 75 bp
 - 35 bp to 50 bp
 - Input to HID Panel PCR
 - 1 ng DNA
 - Built libraries and sequenced



HID SNP Panel

Degraded DNA Study

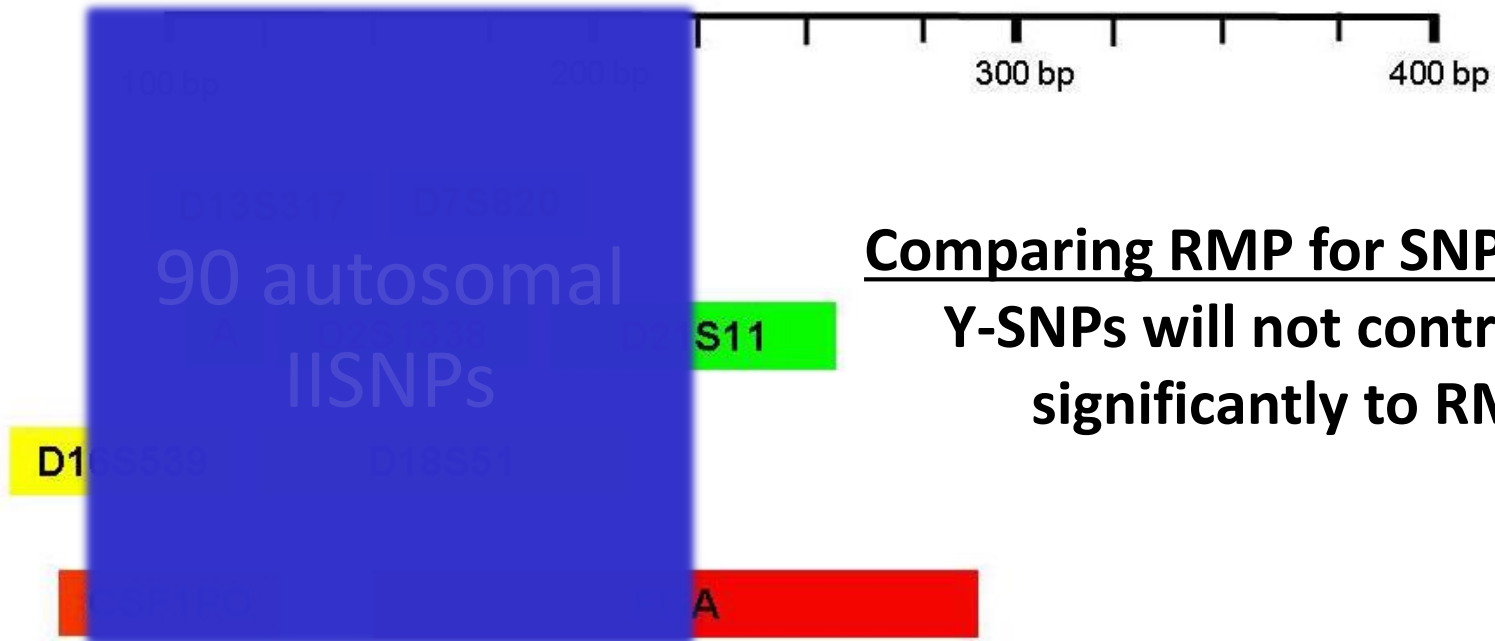
HID SNP Panel



HID SNP Panel

Degraded DNA Study

HID SNP Panel



Comparing RMP for SNPs & STRs

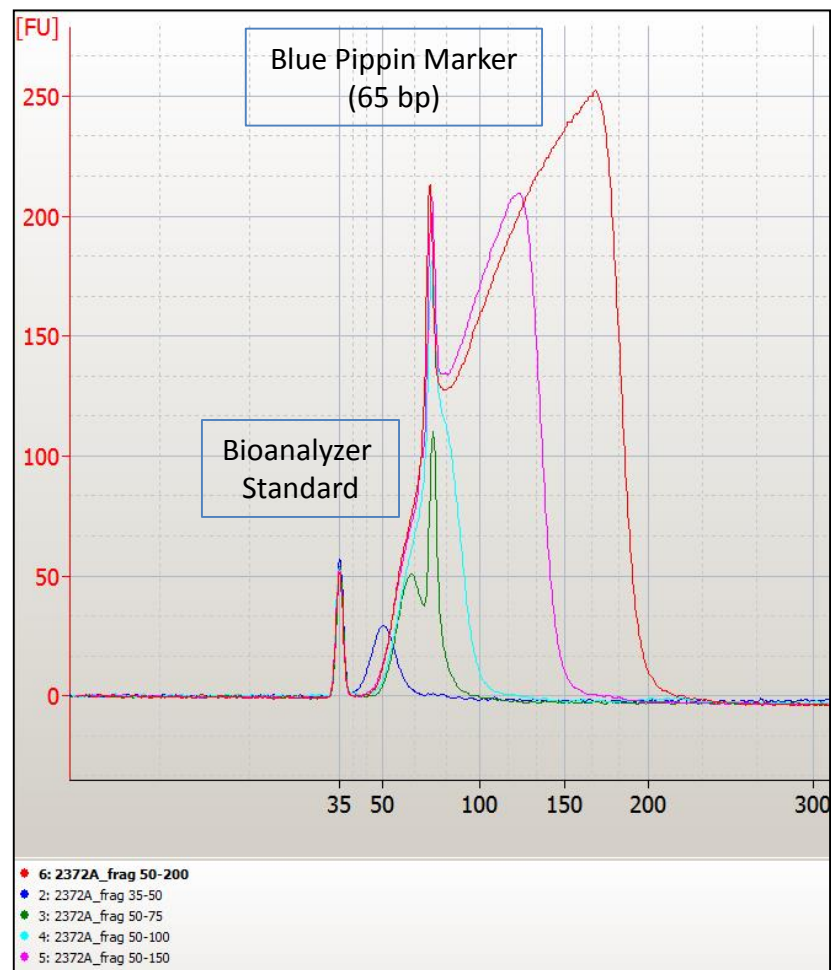
Y-SNPs will not contribute significantly to RMP

HID SNP Panel

Degraded DNA Study

HID SNP Panel compared with MiniFiler

- 1 ng (PGM) or 0.5 ng (MF) *non-degraded* DNA
- 1 ng degraded DNA, 50-200
- 1 ng degraded DNA, 50-150
- 1 ng degraded DNA, 50-100
- 1 ng degraded DNA, 50-75
- 1 ng degraded DNA, 35-50



HID SNP Panel

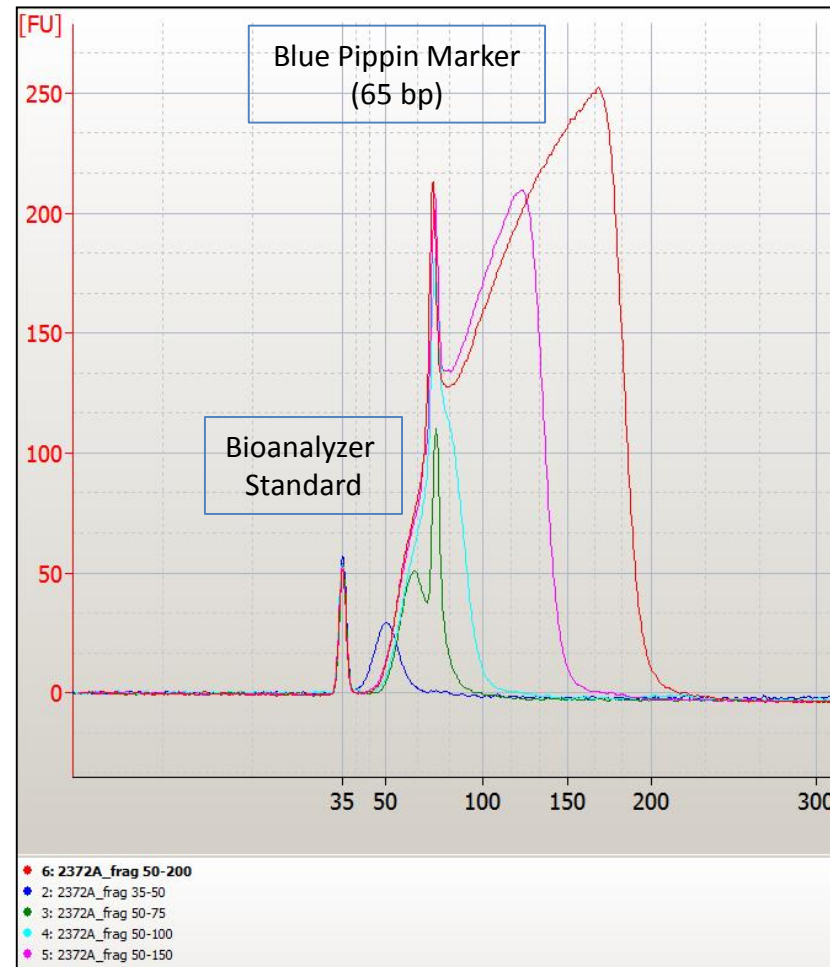
Degraded DNA Study

HID SNP Panel compared with MiniFiler

- 1 ng (PGM) or 0.5 ng (MF) *non-degraded* DNA
- 1 ng degraded DNA, 50-200
- 1 ng degraded DNA, 50-150
- 1 ng degraded DNA, 50-100
- 1 ng degraded DNA, 50-75
- 1 ng degraded DNA, 35-50

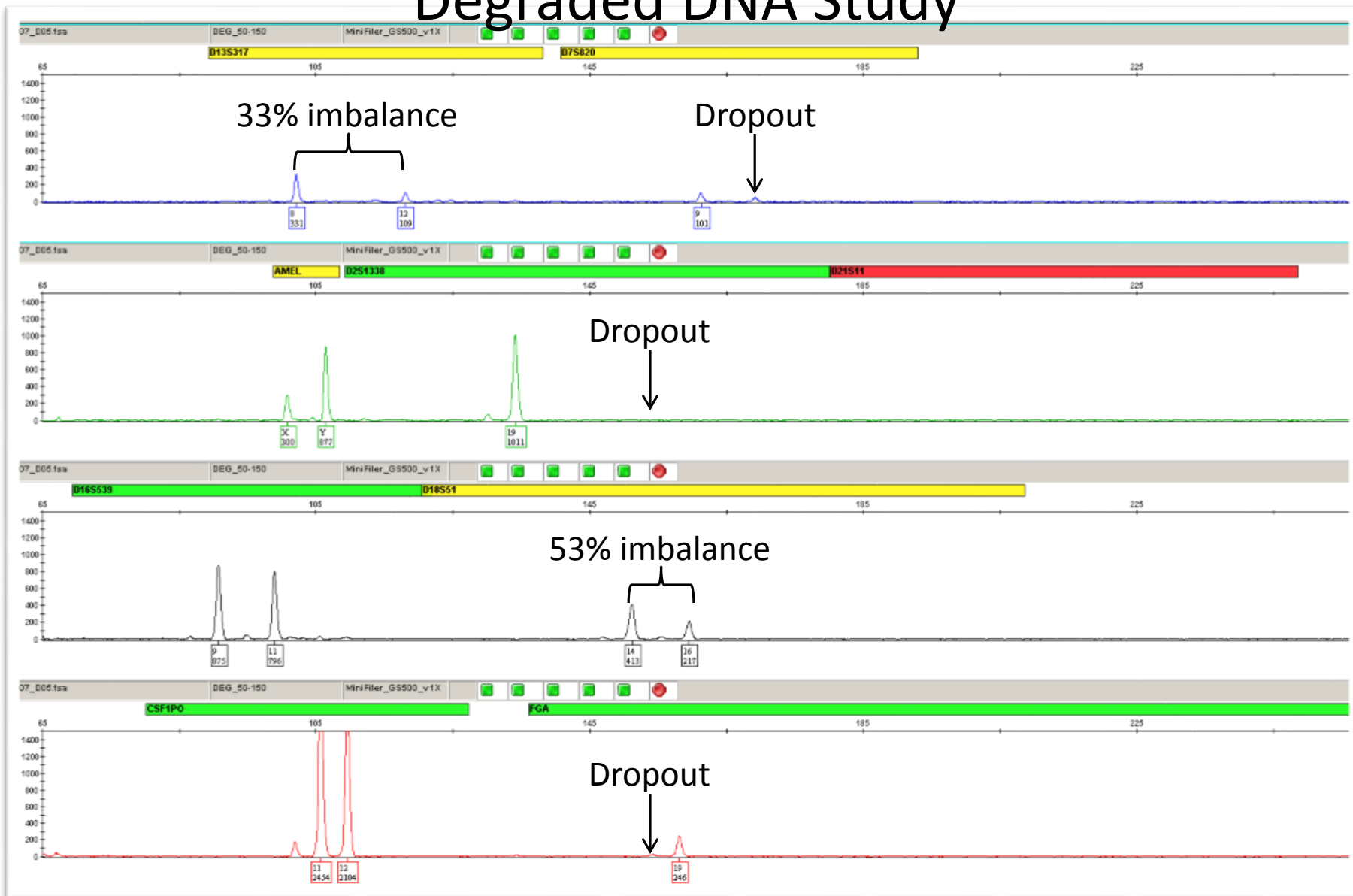
MiniFiler Thresholds

- 50 RFU analytical
- all loci heterozygous
- < 50% heterozygote balance



HID SNP Panel

Degraded DNA Study



HID SNP Panel

Degraded DNA Study



HID SNP Panel

Degraded DNA Study

HID SNP Panel compared with MiniFiler

- 1 ng (PGM) or 0.5 ng (MF) *non-degraded* DNA
- 1 ng degraded DNA, 50-200
- 1 ng degraded DNA, 50-150
- 1 ng degraded DNA, 50-100
- 1 ng degraded DNA, 50-75
- 1 ng degraded DNA, 35-50

MiniFiler Thresholds

- 50 RFU analytical
- all loci heterozygous
- 50% heterozygote balance

PGM Thresholds

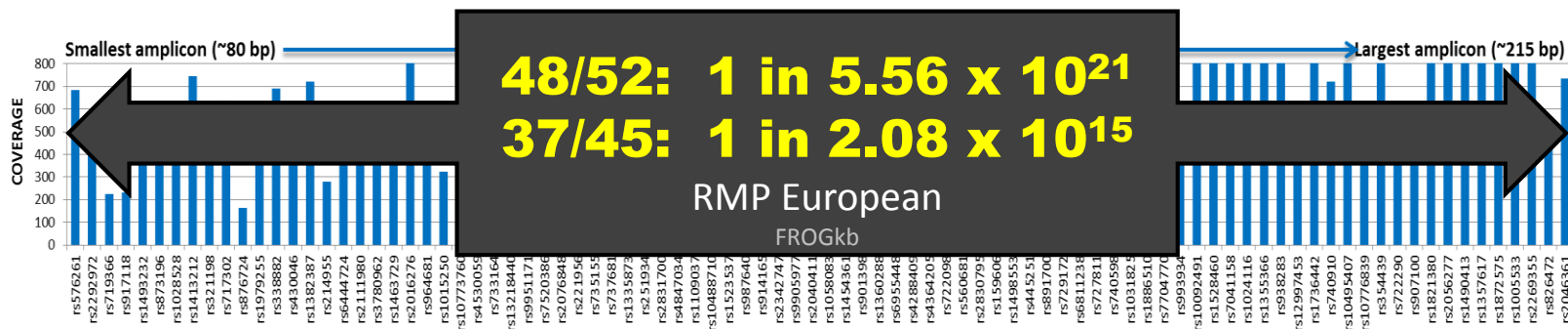
- 50X coverage “analytical”
- 100X coverage “stochastic”
- 60% heterozygote balance

COMPARISON OF RMPs

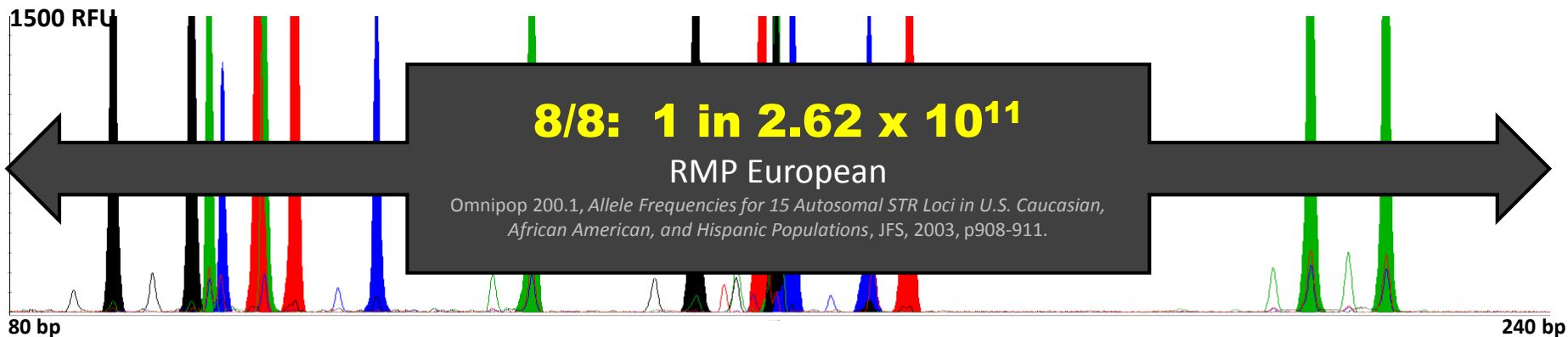
HID SNP Panel

Degraded DNA Study

PGM IISNP – 1 ng non-degraded input DNA



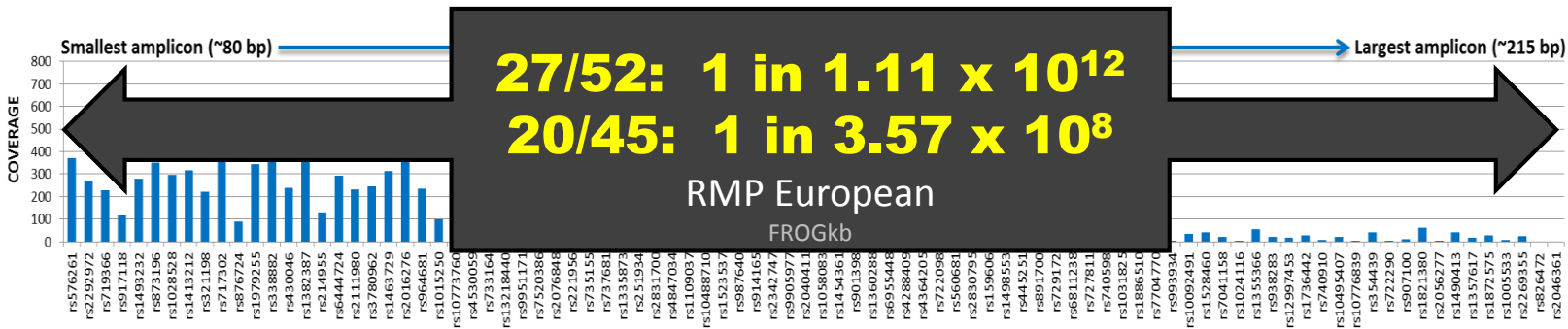
MiniFiler – 0.5 ng non-degraded input DNA



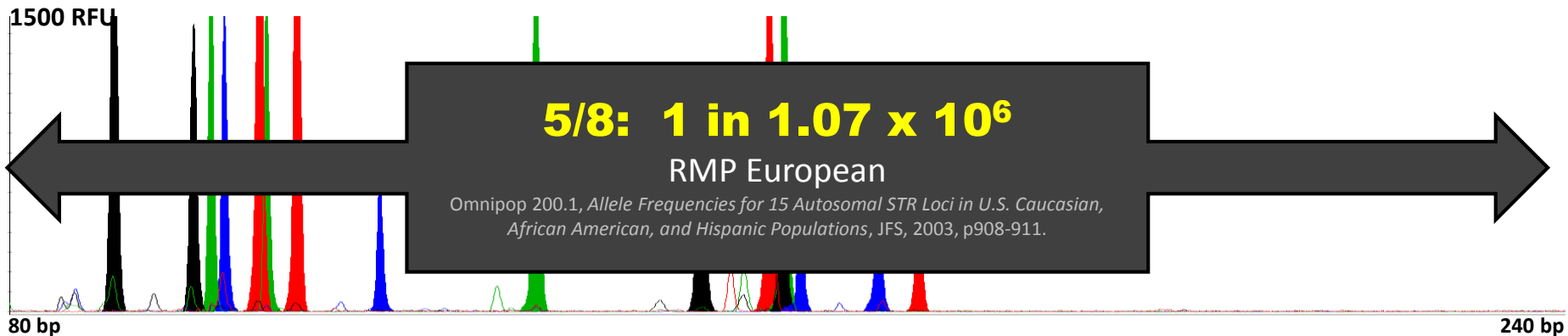
HID SNP Panel

Degraded DNA Study

PGM IISNP – 1 ng degraded input DNA, 50-200 bp size selected



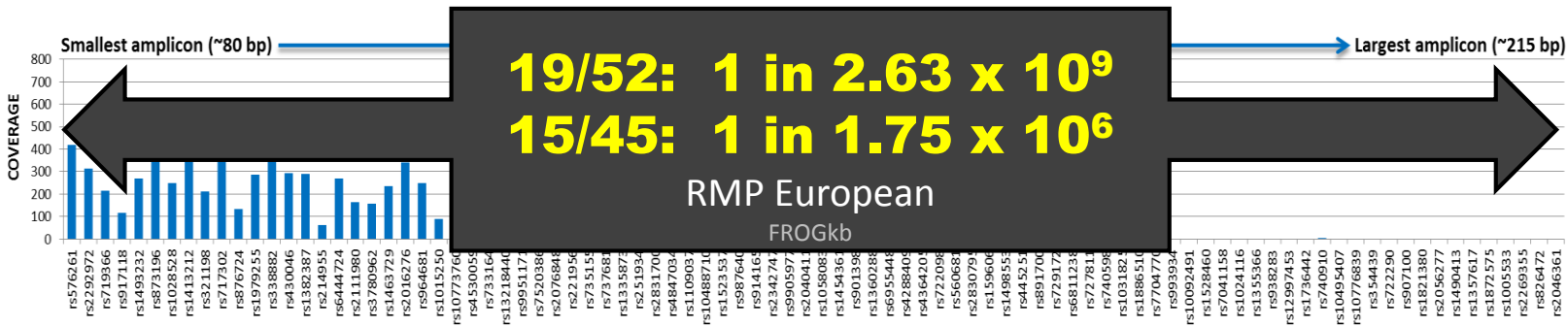
MiniFiler – 1 ng degraded input DNA, 50-200 bp size selected



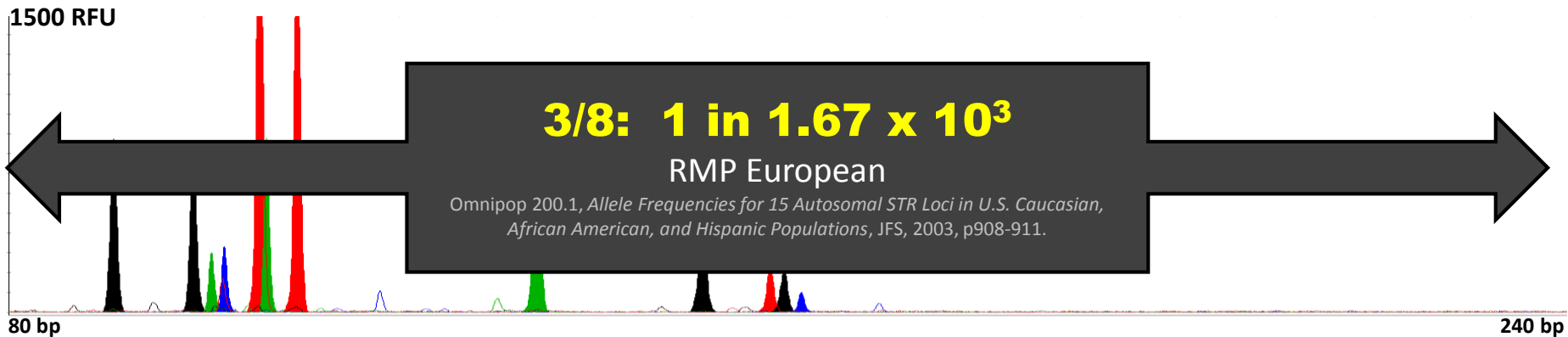
HID SNP Panel

Degraded DNA Study

PGM IISNP – 1 ng degraded input DNA, 50-150 bp size selected



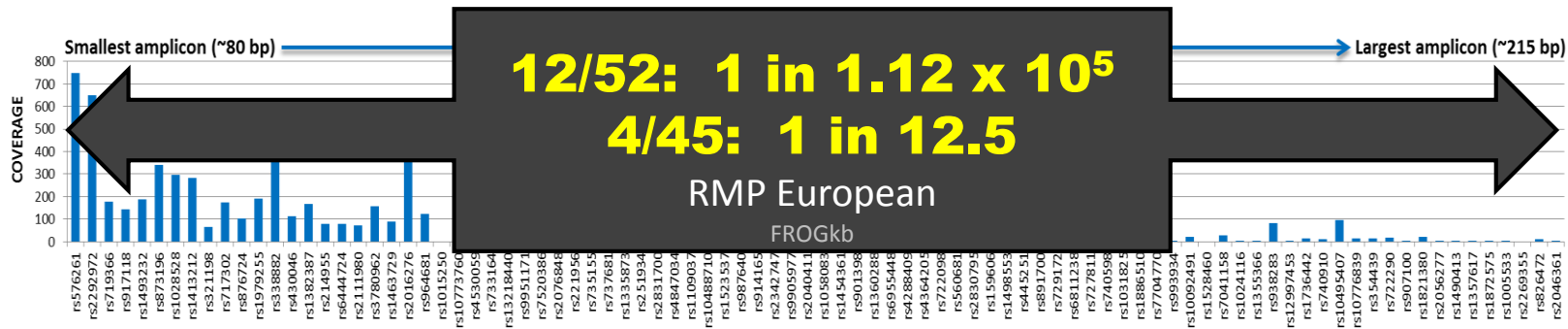
MiniFiler – 1 ng degraded input DNA, 50-150 bp size selected



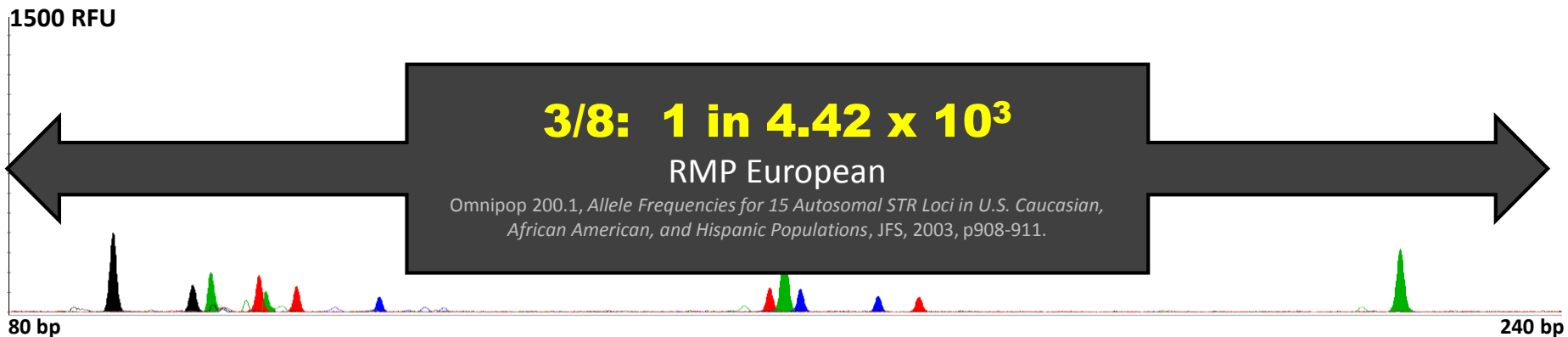
HID SNP Panel

Degraded DNA Study

PGM IISNP – 1 ng degraded input DNA, 50-100 bp size selected



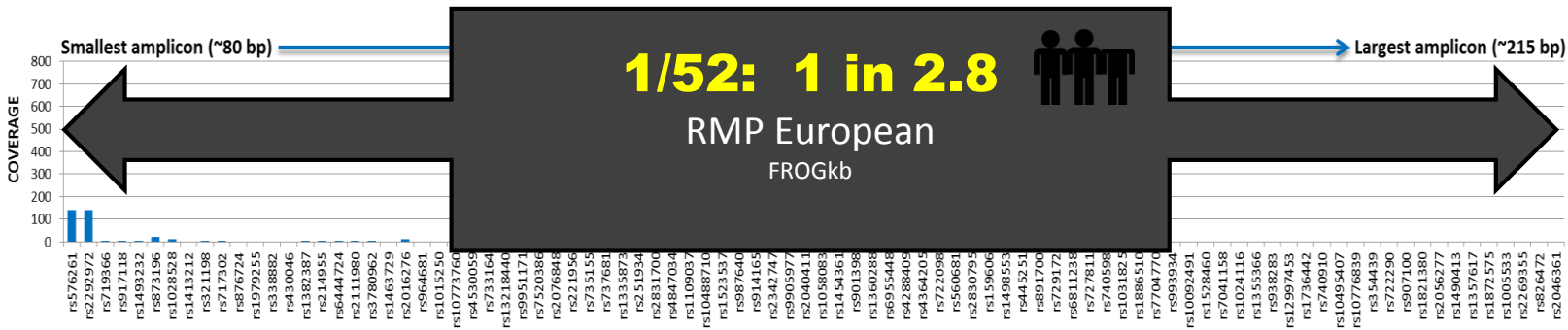
MiniFiler – 1 ng degraded input DNA, 50-100 bp size selected



HID SNP Panel

Degraded DNA Study

PGM IISNP – 1 ng degraded input DNA, 50-75 bp size selected



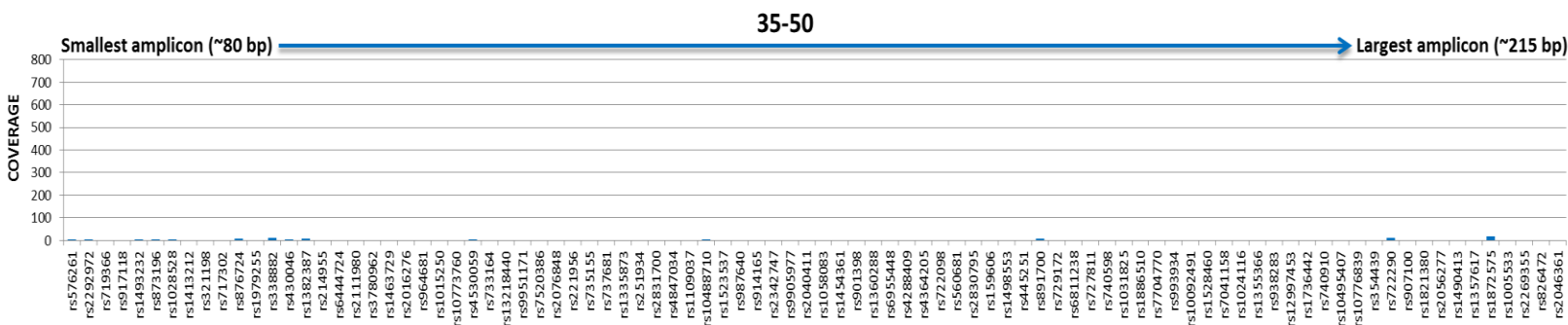
MiniFiler – 1 ng degraded input DNA, 50-75 bp size selected



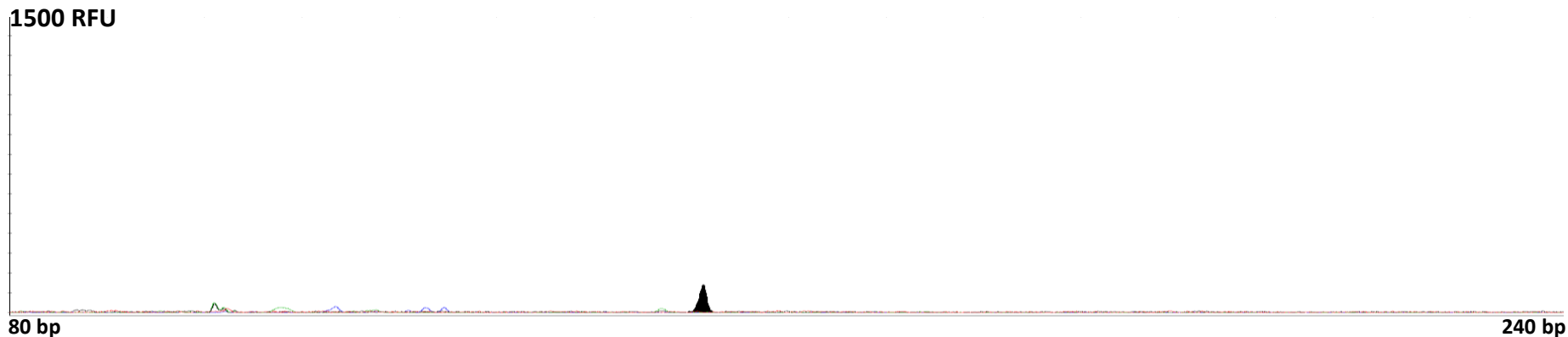
HID SNP Panel

Degraded DNA Study

PGM IISNP – 1 ng degraded input DNA, 35-50 bp size selected



MiniFiler – 1 ng degraded input DNA, 35-50 bp size selected



HID SNP Panel Mixture Study

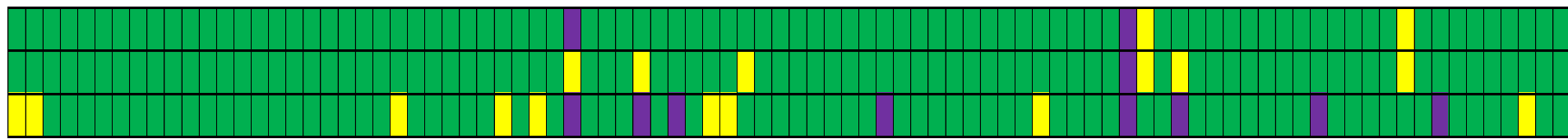
- Evaluate allelic ratios in a mixture sample
 - 3:1 mixture of two individuals (SRM 2391c)
 - Calculate expected ratio for bi-allelic SNPs
 - Examine deviation from expectation

Individual 1 (3x)	Individual 2 (1x)	% A	% B	Expected Variant Frequency (3:1 Mixture)
AA	AA	100	0	100 %
AA	AB	87.5	12.5	87.5 %
AA	BB	75	25	75 %
AB	AA	62.5	37.5	62.5 %
AB	AB	50	50	50 %
AB	BB	37.5	62.5	62.5 %
BB	AA	25	75	75 %
BB	AB	12.5	87.5	87.5 %
BB	BB	0	100	100 %

HID SNP Panel

Mixture Study - Results

- Three replicates of SRM 2391c Component D
- Only autosomal loci considered

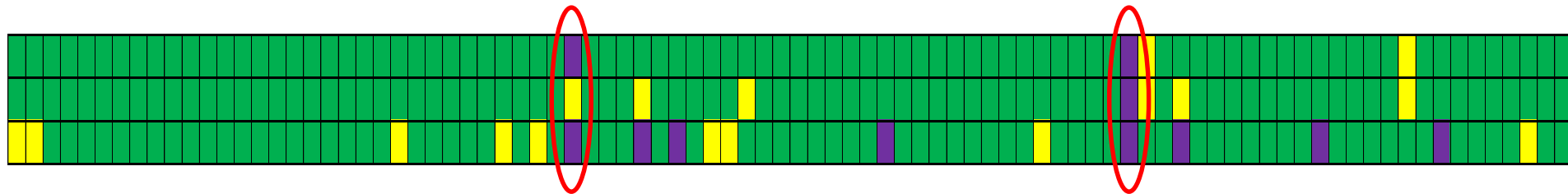


- Deviation from expected variant frequency
 - Less than 5 % = green ■
 - 5 % to 10 % = yellow ■
 - Above 10 % = Purple ■

HID SNP Panel

Mixture Study - Results

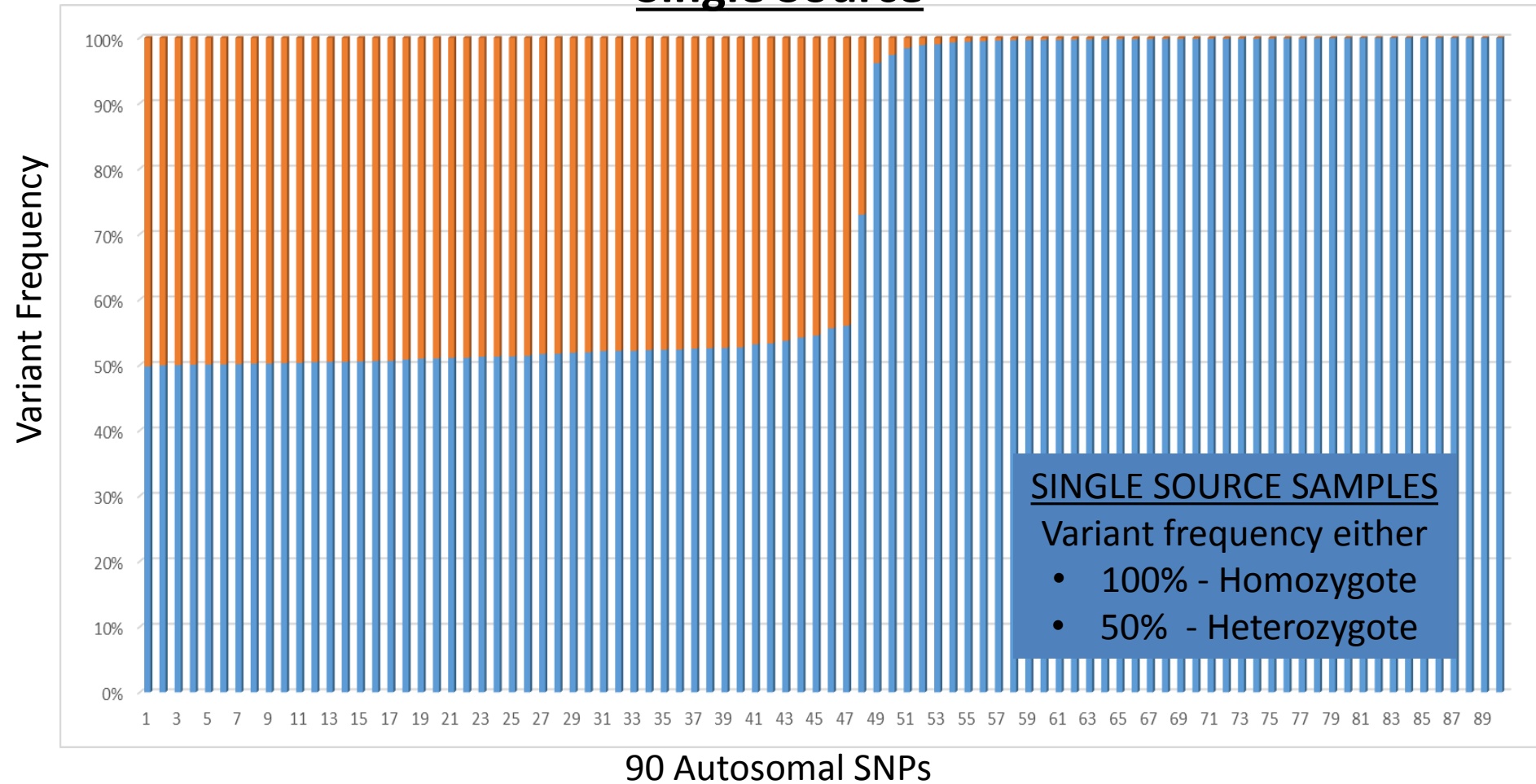
- Three replicates of SRM 2391c Component D
- Only autosomal loci considered



- Deviation from expected variant frequency
 - Less than 5 % = green ■
 - 5 % to 10 % = yellow ■
 - Above 10 % = Purple ■

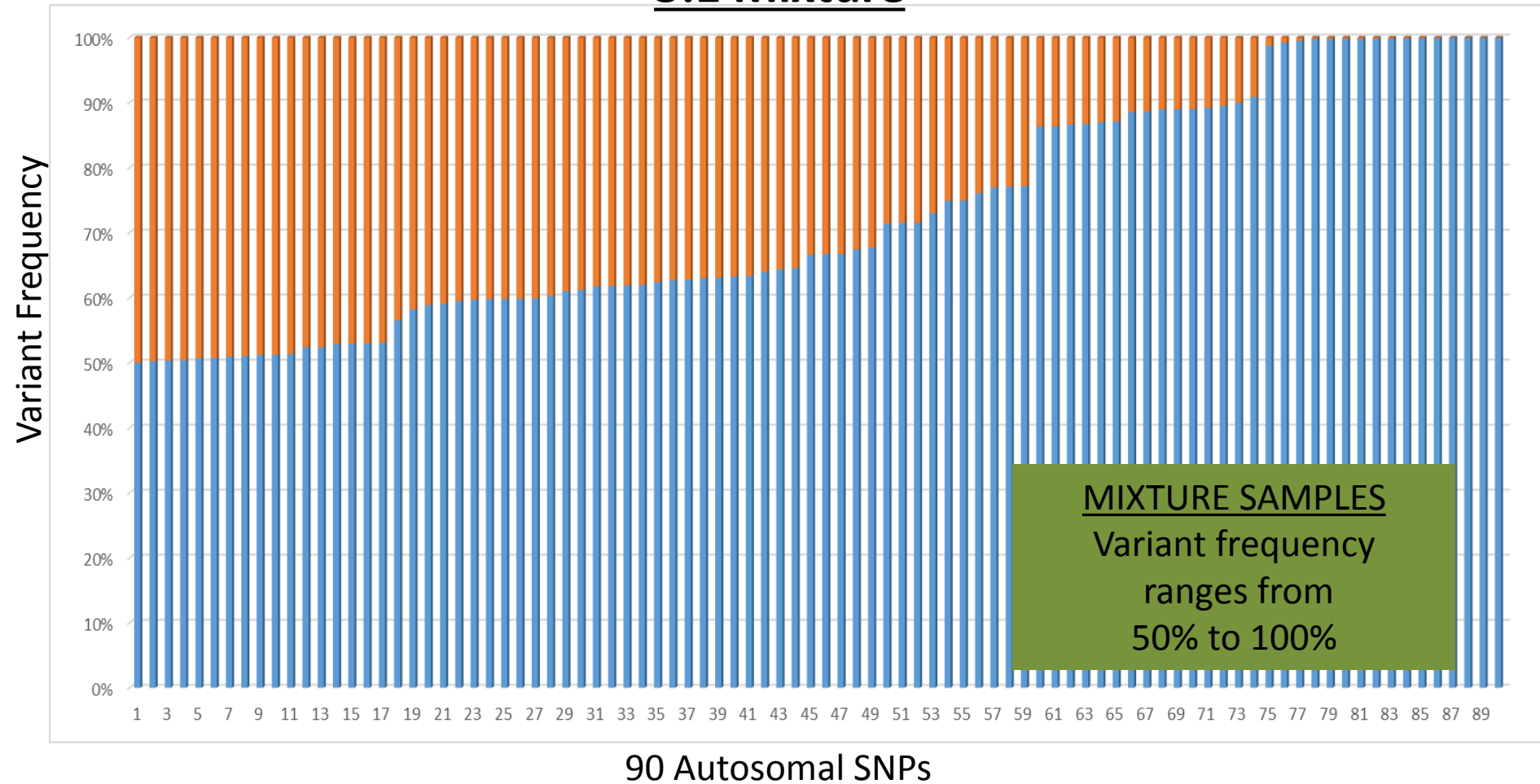
HID SNP Panel Mixture Detection

Single Source



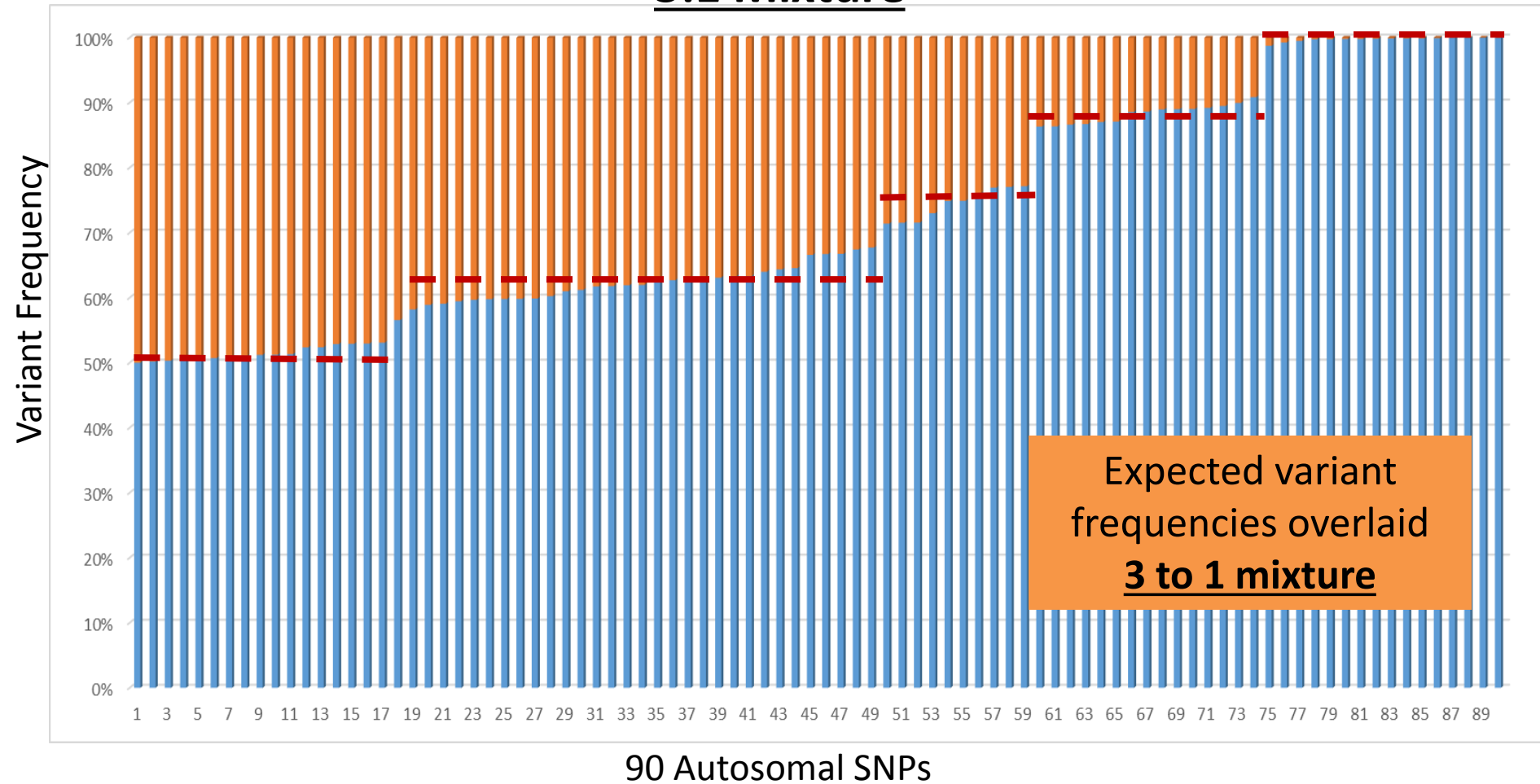
HID SNP Panel Mixture Detection

3:1 Mixture



HID SNP Panel Mixture Detection

3:1 Mixture



Experimental Data

- HID Identity Panel
 - Sensitivity study
 - Degraded DNA study
 - Mixture study
- HID Ancestry Panel
 - Ancestry prediction

AIM Panel

Ancestry Prediction – SRM 2391c

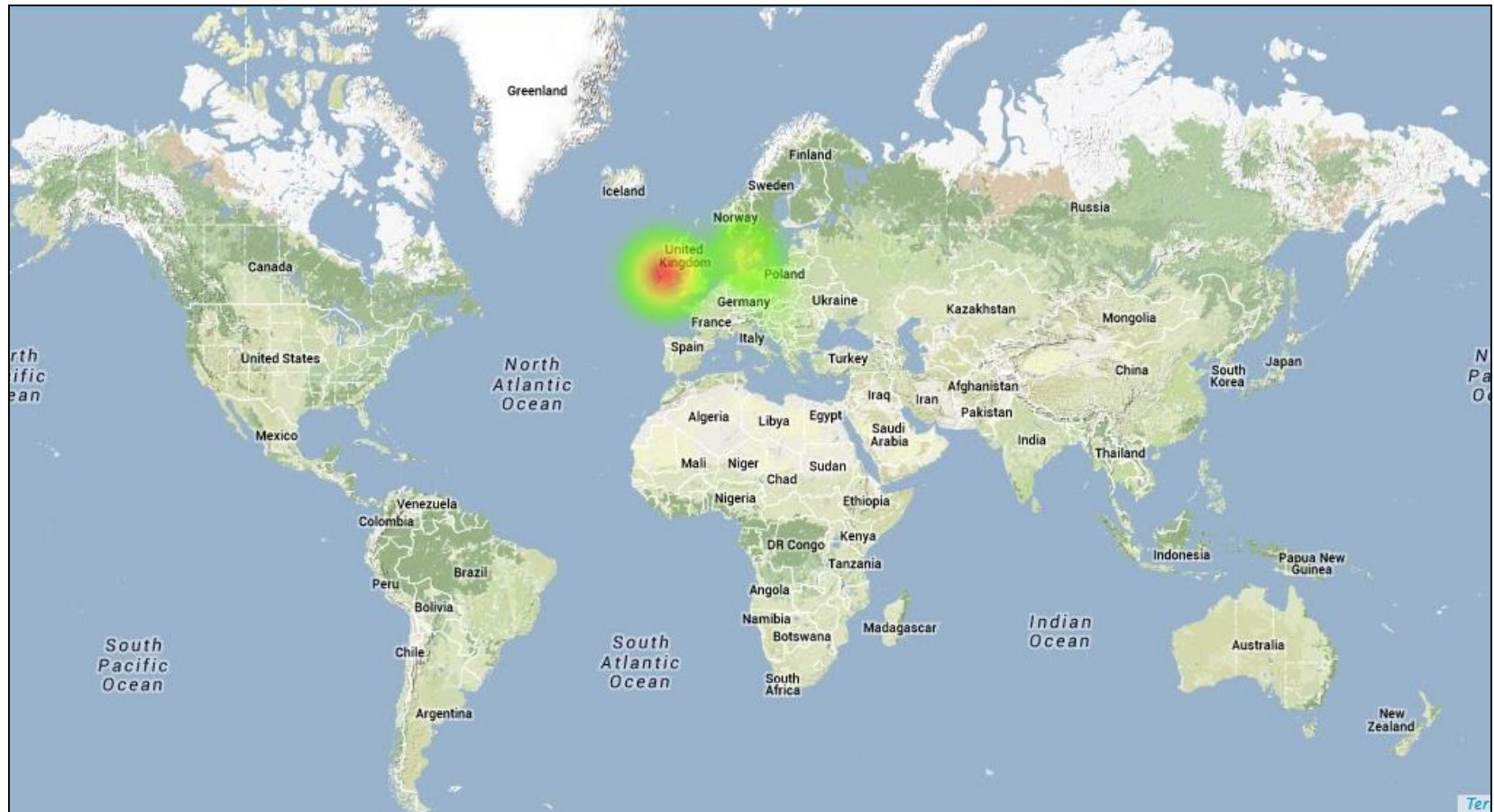
- Likelihood Ratio calculations
 - Four categories present in both Kidd and Seldin studies
 - Europeans, African Americans, Maya, and Han Chinese
 - Allows comparison of SNP sets' performance
 - Representative of major U.S. populations

SRM 2391c Component	Gender	Ethnicity (self declared)
A	Female	Not listed
B	Male	Mexican-American
C	Male	Melanesian
D	Female:Male	Mixed sample
E	Female	Not listed
F	Male	Caucasian

HID SNP Genotyper Plugin (v4.1)

New Feature – Ancestry Map

- Heatmap of highest probability of origin



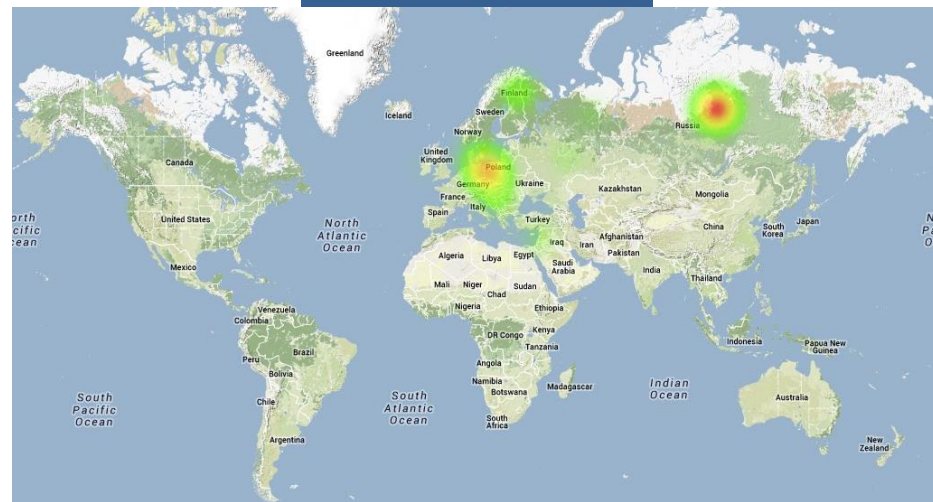
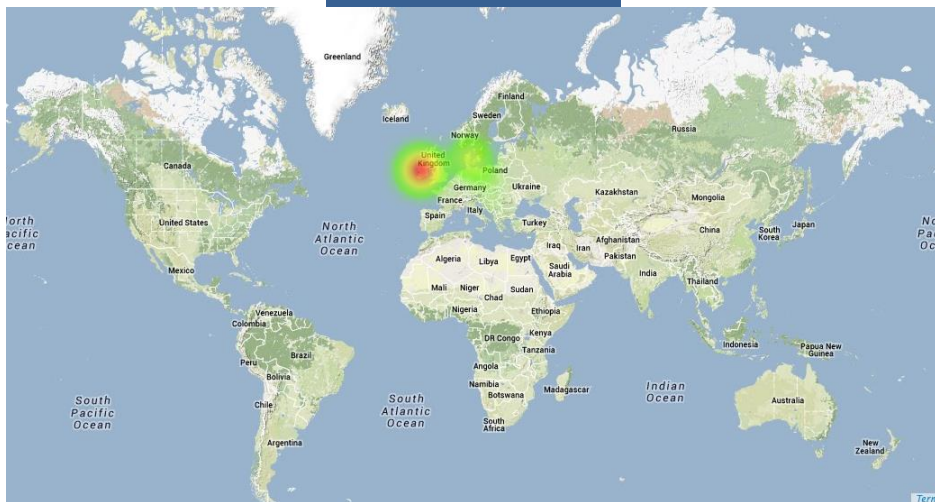
Ancestry Prediction

SRM 2391c Component A

SRM 2391c Component	Gender	Ethnicity	Kidd 55 Prediction	Seldin 128 Prediction
A	Female	Not listed	European 1.02×10^{33}	European 6.32×10^{66}

Kidd 55 SNPs

Seldin 128 SNPs

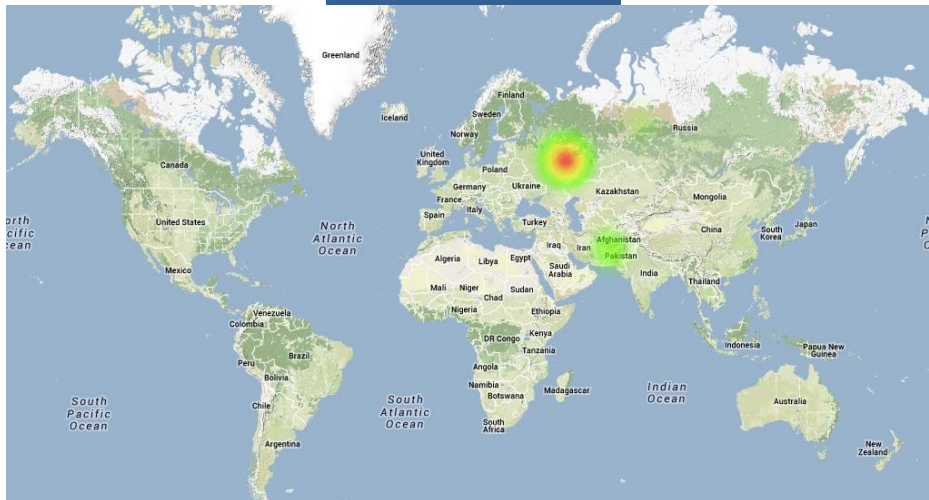


Ancestry Prediction

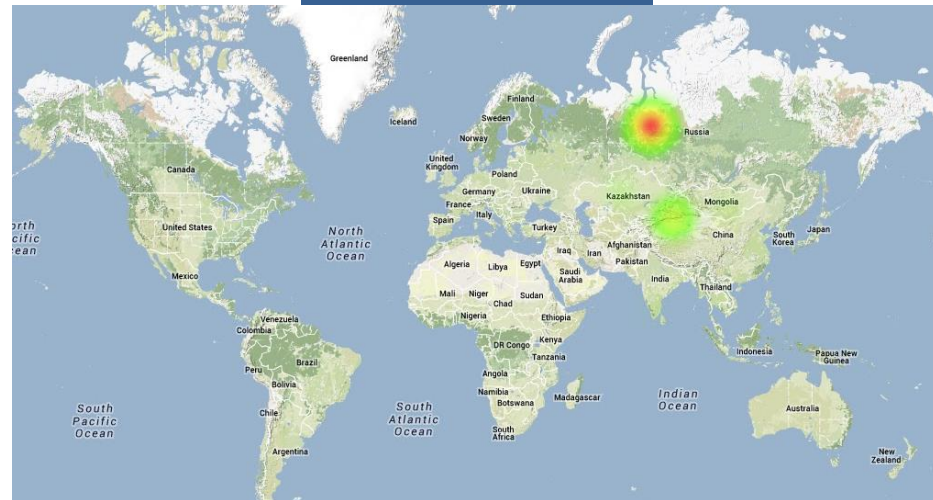
SRM 2391c Component B

SRM 2391c Component	Gender	Ethnicity	Kidd 55 Prediction	Seldin 128 Prediction
B	Male	Mexican-American	European 5.39×10^{12}	Han Chinese 1.48×10^{19}

Kidd 55 SNPs



Seldin 128 SNPs

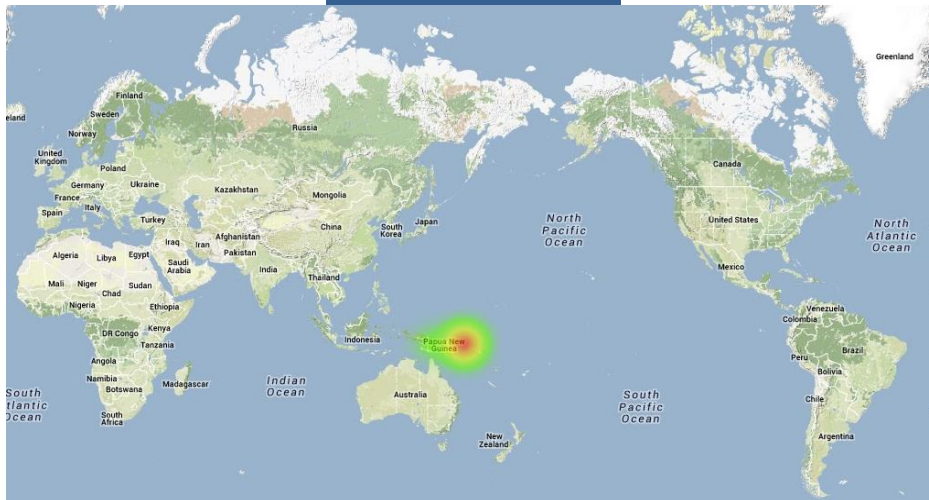


Ancestry Prediction

SRM 2391c Component C

SRM 2391c Component	Gender	Ethnicity	Kidd 55 Prediction	Seldin 128 Prediction
C	Male	Melanesian	Han Chinese 1.54×10^{14}	Han Chinese 6.67×10^{28}

Kidd 55 SNPs



Seldin 128 SNPs

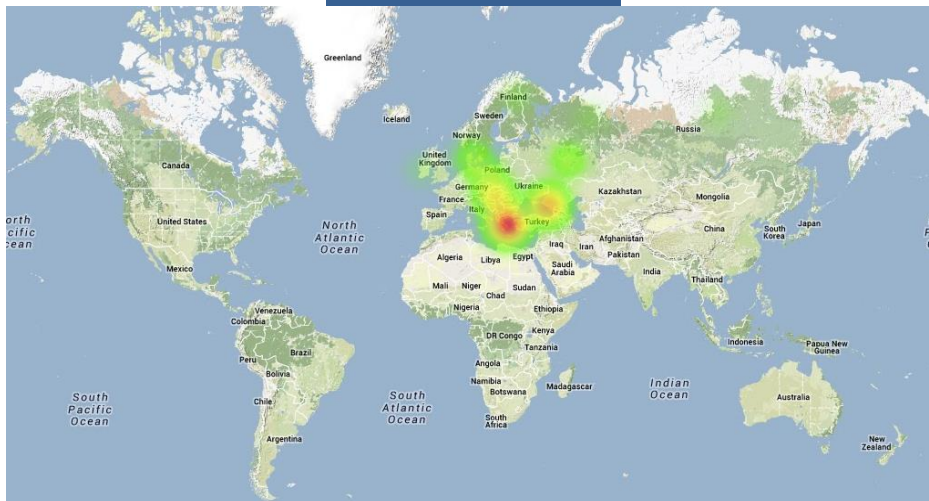


Ancestry Prediction

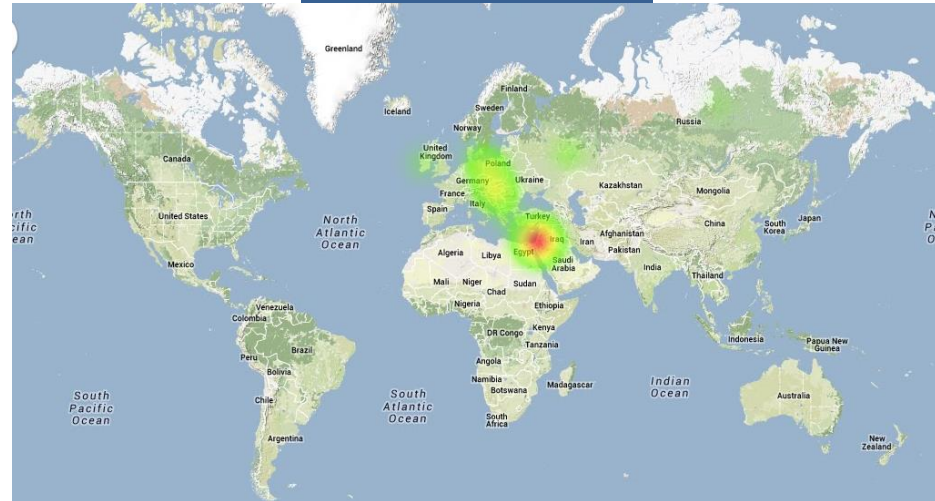
SRM 2391c Component E

SRM 2391c Component	Gender	Ethnicity	Kidd 55 Prediction	Seldin 128 Prediction
E	Female	Not listed	European 5.41×10^{21}	European 3.92×10^{50}

Kidd 55 SNPs



Seldin 128 SNPs



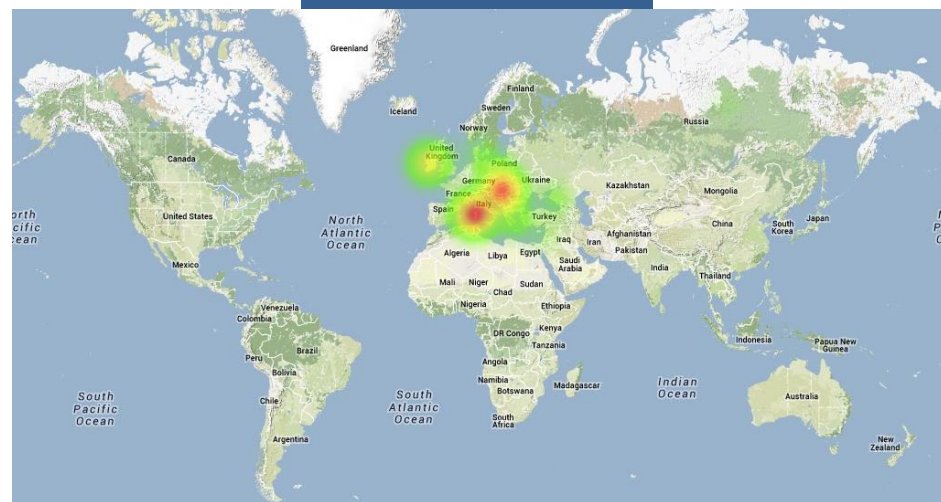
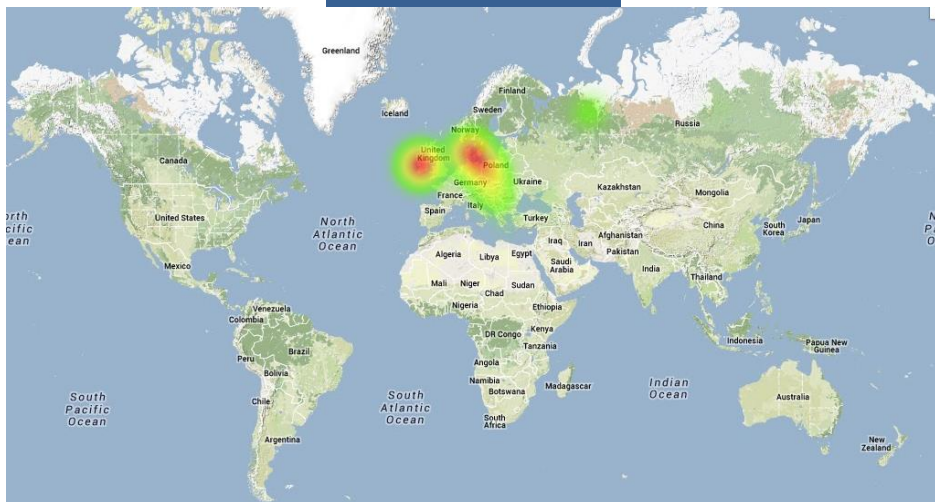
Ancestry Prediction

SRM 2391c Component F

SRM 2391c Component	Gender	Ethnicity	Kidd 55 Prediction	Seldin 128 Prediction
F	Male	Caucasian	European 2.35×10^{31}	European 1.16×10^{55}

Kidd 55 SNPs

Seldin 128 SNPs



HID SNP Genotyper Plugin V4.1

Some Settings Are Locked Down

AIM Panels can be selected

HID SNP Genotyper Plugin

Reference Genome: hg19

Targeted Regions:

Hotspot Regions:

Trim Reads:

Variation Caller Parameter Settings:

- Germ Line - Low Stringency
- Germ Line - High Stringency
- Somatic - Low Stringency
- Somatic - High Stringency
- Custom

Upload Custom Parameter Settings:
 No file chosen

Where do I get the latest platform-optimized Parameter Settings?

Parameter	SNP	Hotspot
Minimum allele frequency min_allele_freq	<input type="text" value="0.1"/>	<input type="text" value="0.1"/>
Minimum quality min_variant_score	<input type="text" value="10"/>	<input type="text" value="10"/>
Minimum coverage min_coverage	<input type="text" value="6"/>	<input type="text" value="6"/>
Minimum coverage on either strand min_cov_each_strand	<input type="text" value="0"/>	<input type="text" value="3"/>
Maximum strand bias strand_bias	<input type="text" value="1"/>	<input type="text" value="1"/>
Minimum relative read quality data_quality_stringency	<input type="text" value="6.5"/>	
Maximum common signal shift filter_unusual_predictions	<input type="text" value="0.3"/>	

▾

Version 4.0

HID SNP Genotyper Plugin

SNP Panels:

- Seldin's list of 128 AISNPs
- KiddLab - Set of 55 AISNPs

Reference Genome: hg19

Targeted Regions:

Hotspot Regions:

Trim Reads:

Upload Custom Parameter Settings:
 No file chosen

Parameter	SNP	Hotspot
Minimum allele frequency min_allele_freq	<input type="text" value="0.1"/>	<input type="text" value="0.1"/>
Minimum quality min_variant_score	<input type="text" value="10"/>	<input type="text" value="10"/>
Minimum coverage min_coverage	<input type="text" value="6"/>	<input type="text" value="6"/>
Minimum coverage on either strand min_cov_each_strand	<input type="text" value="0"/>	<input type="text" value="3"/>
Maximum strand bias strand_bias	<input type="text" value="1"/>	<input type="text" value="1"/>
Minimum relative read quality data_quality_stringency	<input type="text" value="6.5"/>	
Maximum common signal shift filter_unusual_predictions	<input type="text" value="0.3"/>	

▲

Version 4.1 BETA

Acknowledgements

THANK YOU!



Dr. Peter Vallone
Group Leader



Funding from the **FBI
Biometrics Center of
Excellence** 'Forensic DNA
Typing as a Biometric Tool'



Dr. Katherine Gettings
Research Biologist

Our gratitude to:
The R&D team at Life Technologies/Thermo Fisher

Thank you for your attention!

Contact Info:

Kevin.Kiesler@nist.gov

301-975-4306

