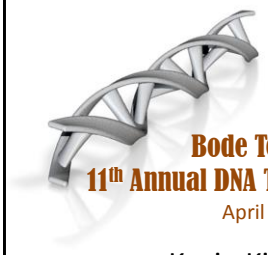
Next Generation Sequencing on the Ion Torrent PGM:



New SNP Typing Applications

Bode Technology
11th Annual DNA Technical Workshop
April 17, 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

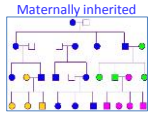
Forensic DNA Markers

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

Circular genome

Mitochondria

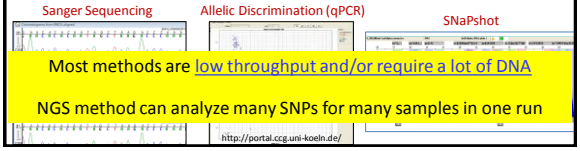
Maternally inherited



Forensic DNA Markers

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

Sanger Sequencing | Allelic Discrimination (qPCR) | SNaPshot



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

NGS method can analyze many SNPs for many samples in one run

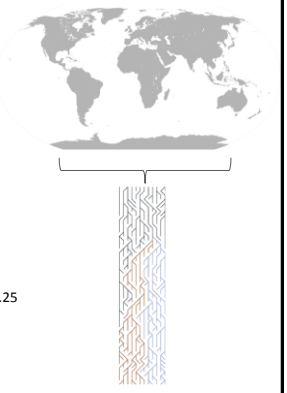
Categories of SNPs

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


Available online at www.sciencedirect.com
ScienceDirect
Research article
Report on ISFG SNP Panel Discussion

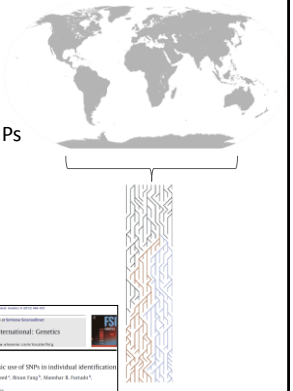
SNP Information

- 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



SNP Information

- Individual Identification
  Pakstis 2010, Kidd 2012
  - Panel of 45 unlinked SNPs
  - F_{ST} below ≈ 0.07
  - Avg het > 0.4
  - RMP 10^{-15} to 10^{-18}
    in 44 populations

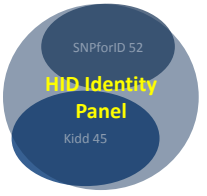


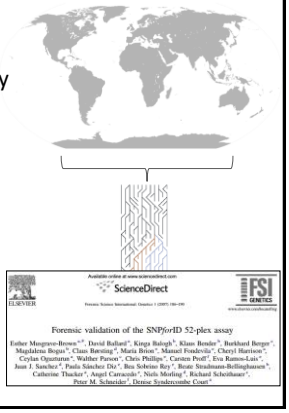

SNPs for a universal individual identification panel

Forensic Science International: Genetics
Expanding data and resources for forensic use of SNPs in individual identification

SNP Information

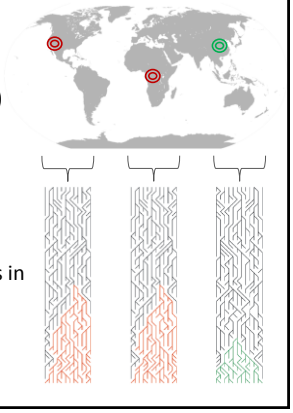
- HID-Ion Ampliseq Identity Panel *(version 2.3)*
  - 90 autosomal SNPs
  - 30 Y-chromosome SNPs
  - RMP 10^{-18} to 10^{-20}




Available online at www.sciencedirect.com
ScienceDirect
Forensic validation of the SNPforID 52-plex assay

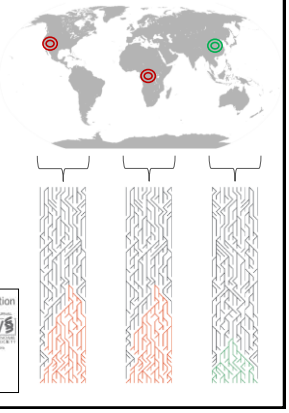
SNP Information

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



SNP Information

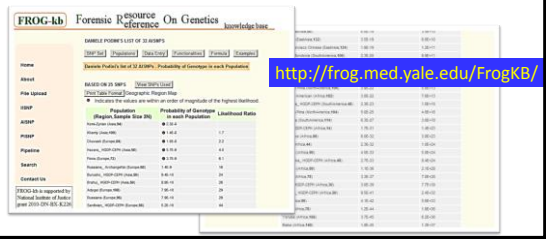
- HID Ancestry Panel
  - Beta version 3.0
  - Publicly available soon
  - 170 loci
  - Derived from
    - Kosoy *et. al* (2008)
      - 128 SNP markers
    - Kidd *et. al* (2014)
      - 55 SNP markers




RESEARCH ARTICLE
Human Mutation
Ancestry Informative Marker Sets for Determining Continental Origin and Admixture Proportions in Common Populations in America

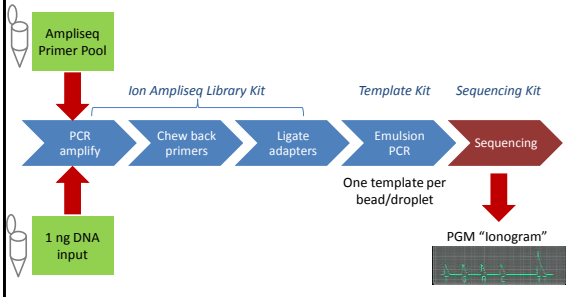
SNP Information

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

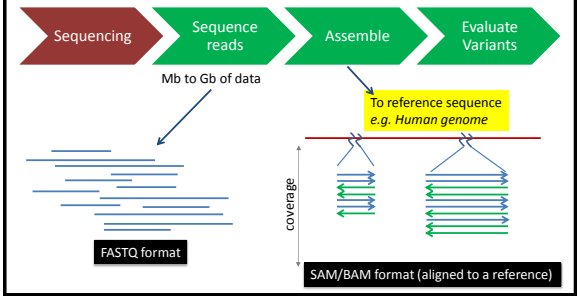


Sequencing Methods

Ampliseq Workflow



Analysis Workflow



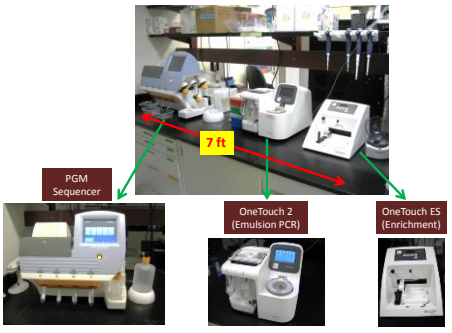
Life Tech - Ion Torrent - PGM

- 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**

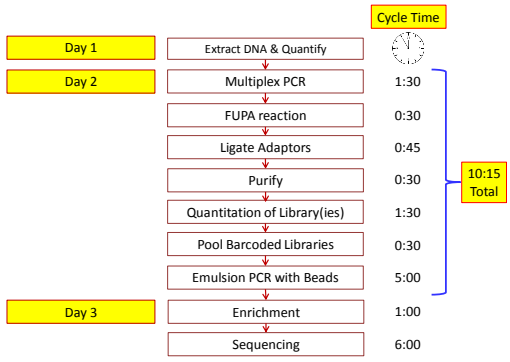
Ion Torrent PGM Workflow

http://www.youtube.com/watch?v=MxkYa9XCvBQ

The PGM Instrument at NIST

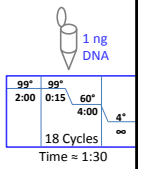


Workflow Overview



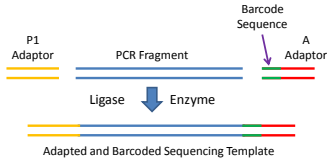
Front-End: Multiplex PCR

- 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



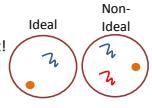
Digest Primer Regions & Ligate Adaptors

- 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



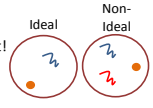
Prepare Ion Sphere Particles (ISPs)

- 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)

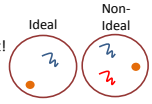


- Emulsion PCR
  - Nanoliter droplets of PCR reagents in oil
  - Attaches sequencing template to the ISP


OneTouch 2

Prepare Ion Sphere Particles (ISPs)

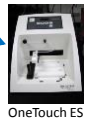
- Libraries quantified by qPCR
  - Quantity of DNA going into emPCR is very important!
  - Goal: 10 % to 30 % template positive ISPs
    - Too much DNA → polyclonal ISPs (mixed read)

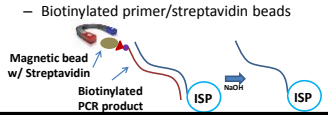


- Emulsion PCR
  - Nanoliter droplets of PCR reagents in oil
  - Attaches sequencing template to the ISP


OneTouch 2

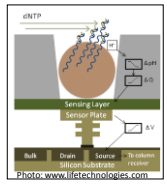
- Enrich for positive ISPs
  - Liquid handler removes non-templated ISPs
  - Biotinylated primer/streptavidin beads

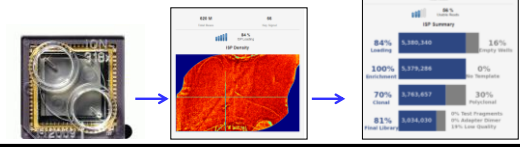

OneTouch ES



Sequencing & Data Analysis

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

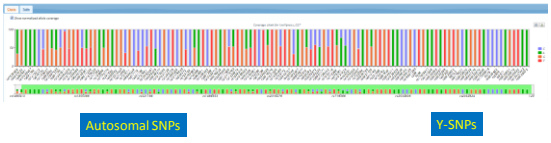




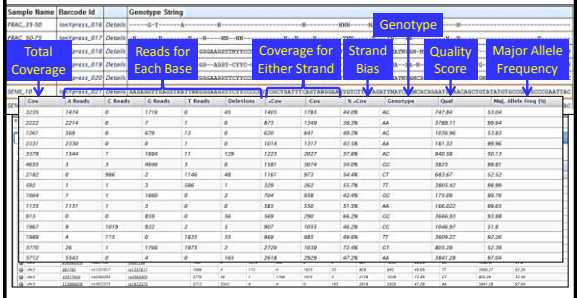
Data Analysis
HID SNP Genotyper Plugin

- Allele coverage histogram

- Normalized y-axis scale



Data Analysis
HID SNP Genotyper Plugin



Experimental Data

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

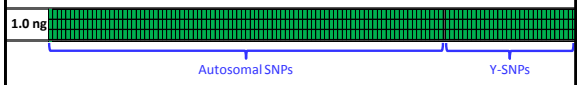
HID SNP Panel
Sensitivity Study

- 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

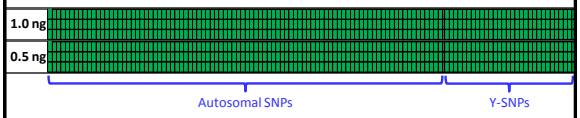
HID SNP Panel Sensitivity Study

- 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



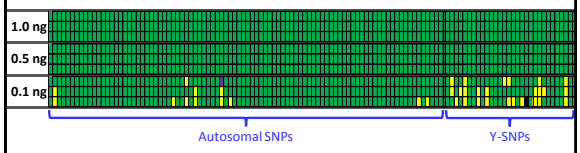
HID SNP Panel Sensitivity Study

- 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



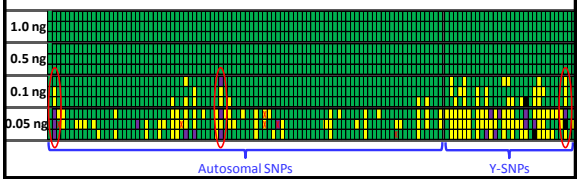
HID SNP Panel Sensitivity Study

- 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

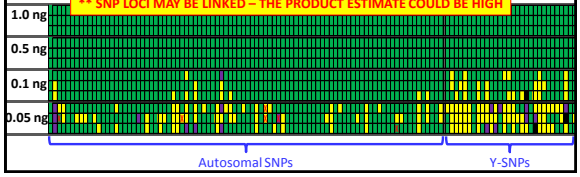
- Sequence coverage per locus
  - 200 or more reads = green
  - Below 200 reads = yellow
  - Below 50 reads = purple
  - Below quality cutoff = black
  - Incorrect genotype = red X



HID SNP Panel Sensitivity RMP With 200 Read Cutoff

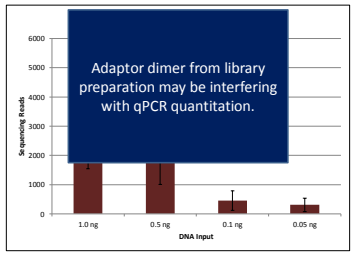
	% Below 200 Reads	RMP* Estimate SNPforID 52	RMP* Estimate Kidd 45	RMP Estimate Product **
1.0 ng	0 %	1 in 1.0 x 10^{15}	1 in 4.0 x 10^{14}	1 in 4.1 x 10^{29}
0.5 ng	0 %	1 in 1.0 x 10^{15}	1 in 4.0 x 10^{14}	1 in 4.1 x 10^{29}
0.1 ng	10 %	1 in 4.4 x 10^{14}	1 in 2.6 x 10^{13}	1 in 1.1 x 10^{28}
0.05 ng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

* RMP calculations from Frog KB (http://frog.med.yale.edu/FrogKB/)
** SNP LOCI MAY BE LINKED – THE PRODUCT ESTIMATE COULD BE HIGH



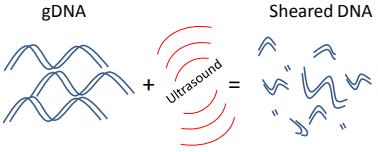
HID SNP Panel Sensitivity Study

- Sequencing coverage decreases with DNA input
  - Library concentrations are normalized



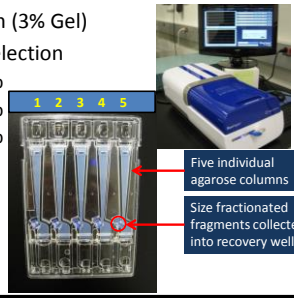
HID SNP Panel
Degraded DNA Study

- Total genomic DNA was sheared
  - Covaris S2 Focused Ultrasonicator



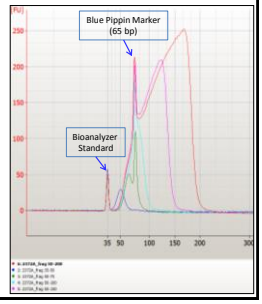
HID SNP Panel
Degraded DNA Study

- 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



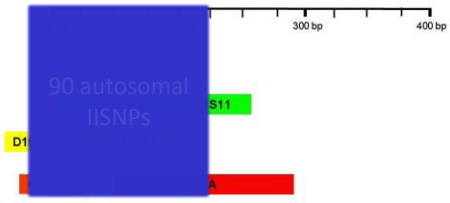
HID SNP Panel
Degraded DNA Study

- 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



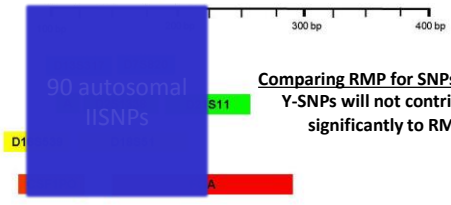
HID SNP Panel
Degraded DNA Study

HID SNP Panel



HID SNP Panel
Degraded DNA Study

HID SNP Panel

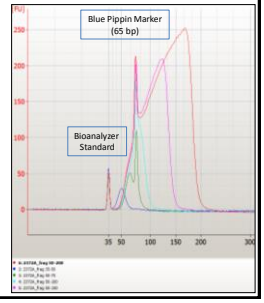


Comparing RMP for SNPs & STRs
Y-SNPs will not contribute significantly to RMP

HID SNP Panel
Degraded DNA Study

HID SNP Panel compared with MiniFiler

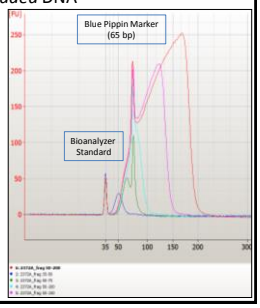
- 1 ng (PGM) or 0.5 ng (MF) *non-degraded* DNA
- 1 ng degraded DNA, 50-200
- 1 ng degraded DNA, 50-150
- 1 ng degraded DNA, 50-100
- 1 ng degraded DNA, 50-75
- 1 ng degraded DNA, 35-50



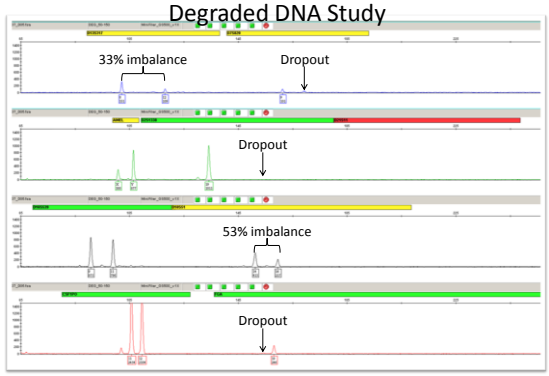
HID SNP Panel
Degraded DNA Study

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

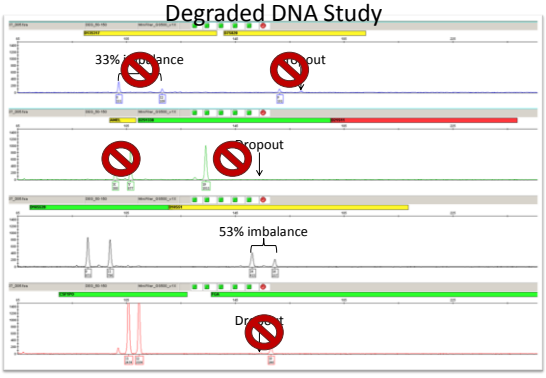
MiniFiler Thresholds
- 50 RFU analytical
- all loci heterozygous
- < 50% heterozygote balance



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

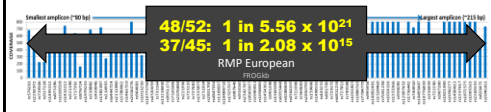
MiniFiler Thresholds
- 50 RFU analytical
- all loci heterozygous
- 50% heterozygote balance

PGM Thresholds
- 50X coverage "analytical"
- 100X coverage "stochastic"
- 60% heterozygote balance

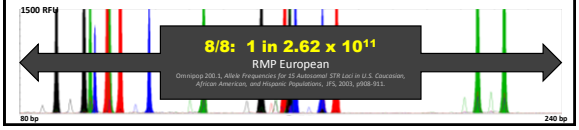
COMPARISON OF RMPs

HID SNP Panel
Degraded DNA Study

PGM IISNP – 1 ng *non-degraded* input DNA

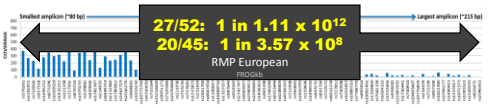


MiniFiler – 0.5 ng *non-degraded* input DNA

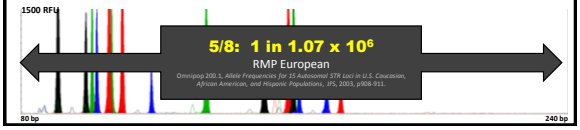


HID SNP Panel
Degraded DNA Study

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

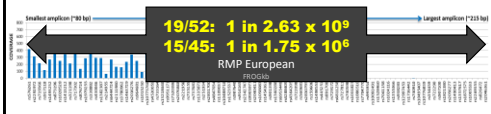


MiniFiler – 1 ng *degraded* input DNA, 50-200 bp size selected

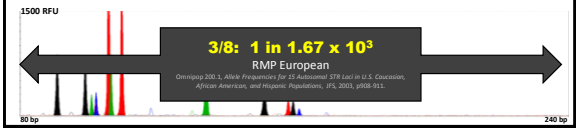


HID SNP Panel
Degraded DNA Study

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



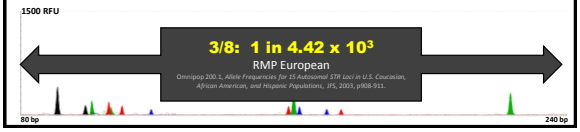
MiniFiler – 1 ng *degraded* input DNA, 50-150 bp size selected



HID SNP Panel
Degraded DNA Study

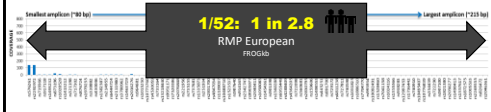
PGM IISNP – 1 ng *degraded* input DNA, 50-100 bp size selected

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



HID SNP Panel
Degraded DNA Study

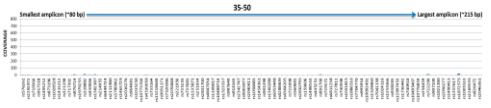
PGM IISNP – 1 ng *degraded* input DNA, 50-75 bp size selected



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

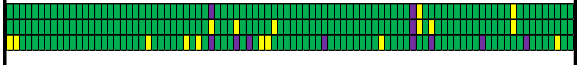
HID SNP Panel
Mixture Study

- Evaluate allelic ratios in a mixture sample
  - 3:1 mixture of two individuals
  - 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



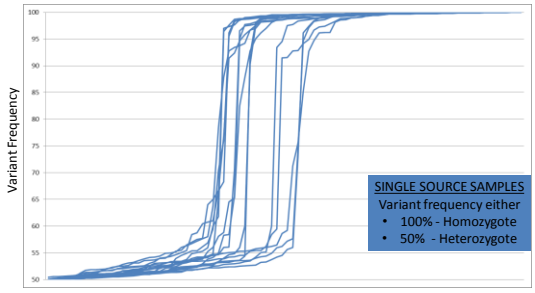
- Deviation from expected variant frequency
  - Less than 5 % = green
  - 5 % to 10 % = yellow
  - Above 10 % = Purple

HID SNP Panel
Mixture Study - Results

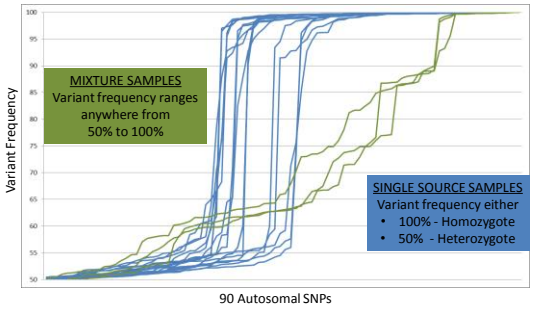
- Three replicates of SRM 2391c Component D
- Only autosomal loci considered

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

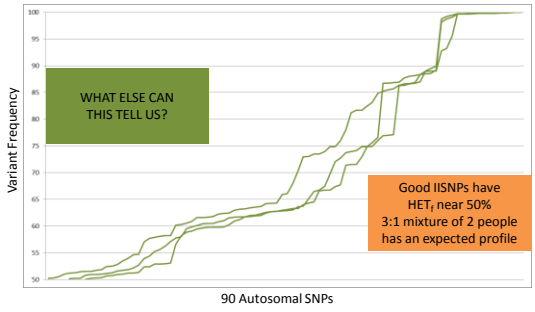
HID SNP Panel
Mixture Detection



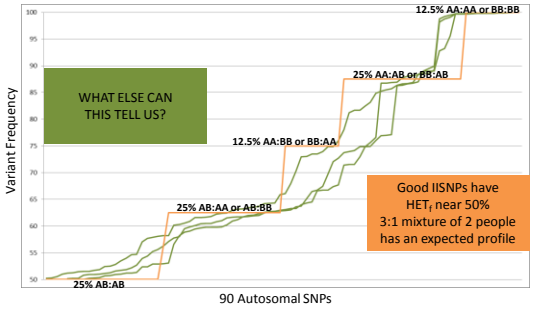
HID SNP Panel
Mixture Detection



HID SNP Panel
Mixture Detection



HID SNP Panel
Mixture Detection



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 extant 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 Beta)
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}

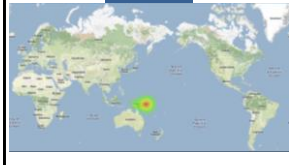
Kidd 55 SNPs

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}

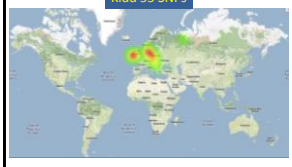
Kidd 55 SNPs

Seldin 128 SNPs

Ancestry Prediction
SRM 2391c Component F

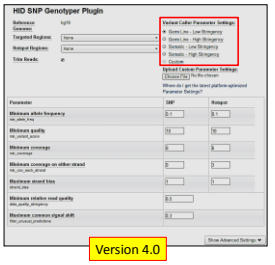
SRM 2391c Component	Gender	Ethnicity	Kidd 55 Prediction	Seldin 128 Prediction
F	Male	Caucasian	European 2.35×10^{31}	European 1.16×10^{55}

Kidd 55 SNPs



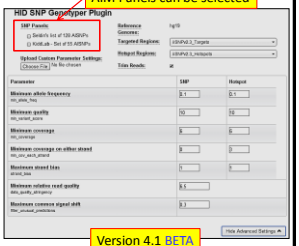
Seldin 128 SNPs

HID SNP Genotyper Plugin V4.1
Some Settings Are Locked Down



Version 4.0

AIM Panels can be selected



Version 4.1 BETA

Acknowledgements

Dr. Peter Vallone
Group Leader

THANK YOU

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

Dr. Katherine Gettings
Research Biologist

NIST Applied Genetics

Thank you for your attention!

Contact Info:
Kevin.Kiesler@nist.gov
301-975-4306