SNPCHAT: The forensic marker that could be your new BFF

Topics to Cover

What is a SNP?
What info do SNPs provide?
How can identity SNPs be used?
How are ancestry and phenotype SNP panels developed?
How do current prediction methods perform?

Single Nucleotide Polymorphism

Allele 1: TAGGATCGTGCCGATGACTG
Allele 2: TAGGATCGTACCGATGACTG

A/A  A/G  G/G
Homozygous  Heterozygous  Homozygous
**Single Nucleotide Polymorphism**

Population A

Population B

---

**Single Nucleotide Polymorphism**

Population A, B, C...

Population Z

---

**SNP Information**

- **IISNP**: Individual Identification SNP
- **AISNP**: Ancestry Informative SNP
- **PISNP**: Phenotype Informative SNP
- **LISNP**: Lineage Informative SNP
SNP Information
LISNP-Lineage

- Balancing has occurred in all populations
- Allele frequencies are similar within and between populations
- High heterozygosity

IISNP Benefits for Degraded DNA
ForenSeq STR and IISNP Statistics

Length-based allele frequencies
- Green Bar = sample result
- Gray Bar = scaled to range in population
- Gray Line = source attribution threshold

Can IISNP and auSTR results be combined?

CODIS 13
- CODIS 7
- Commonly Used

27 Autosomal STRs in ForenSeq

94 IISNPs in ForenSeq

44 from Kidd
50 from SNPforID
43 SNPs are spread across the 22 human chromosomes and show very little or no genetic linkage with each other
LD in casework

Collaboration to evaluate in NIST 1036 with Andreas Tillmar, National Board of Forensic Medicine

Ideally confirm with multiple sample sets from same population, multiple methods

Designing panel/assay: Evaluate LD, eliminate loci as needed based on informativeness

Implementing established panel/assay:

Best – Determine haplotype frequency for pair or block
  • For polymorphic loci the sample size would be unfeasible
  • Keep the more informative, similar to assay design

Alternative – Exclude one of the two markers during validation
  • RMP vs Kinship

Problematic – Exclude one of the two markers case-by-case
  • Still working

Ancestry Information

Ancestry Information

– Population specific fixation has occurred
– Low heterozygosity

Examples
– Malaria resistance in Sub-Saharan Africa
– Lighter skin pigmentation in Europe

Ancestry SNP Assay Evaluation Criteria

Does the SNP panel target the populations of interest?
  – Varies by country/region

How do the SNPs perform?
  – Interlocus balance
  – Heterozygote balance
  – Concordance

Does the interpretation model provide reliable predictions?
  – Dependent on appropriate model training data
Ancestry/Ethnicity of US Population (2010 Census)

- White
- Hispanic or Latino
- Black or African American
- East Asian
- Native Hawaiian and Other Pacific Islander
- Some Other Race
- Two or More Races

SNP Information

AISNP - Ancestry

- S5 Ancestry SNP Panel contains
  - Seldin 128 (with some exceptions)
  - Kidd 55
- ForenSeq contains Kidd 55
SNP Information
PISNP-Phenotype
- 24 Hirisplex SNPs
  - ForenSeq
  - S5 Phenotype Panel

SNP Information
192 SNPs - Identity, Ancestry & Phenotype
Contains (with a few exceptions):
- Kidd 55 for Ancestry
- Kidd 45 + SNPforID52 for Identity
- Hirisplex 24

ForenSeq SNPs
172 SNPs - Identity, Ancestry & Phenotype
Contains (with a few exceptions):
- Kidd 55 for Ancestry
- Kidd 45 + SNPforID52 for Identity
- Hirisplex 24

Left side y-axis = average SNP coverage
- ranges from 23X to 3167X (>150 fold)
Right side y-axis = average het balance
- ranges from 0.42 to 0.94

Precision ID Ancestry Panel
165 SNPs - Ancestry
Contains Kidd 55 and Seldin 128 (with exceptions)

Left side y-axis = average SNP coverage
- ranges from 264X to 2000X (7.6 fold)
Right side y-axis = average het balance
- ranges from 0.43 to 0.98
**Precision ID Ancestry Panel**

- ForenSeq software uses Hirisplex Model for Phenotype Prediction
- And 1000 Genome data for Continental level ancestry prediction

**ForenSeq AISNP + PISNP Panel**

- 1000 Genomes: A Deep Catalog of Human Diversity

**2391d Component A**

- Precision ID 151 AISNPs
**Precision ID 151 AISNPs**

**2391d Component B**

**ForenSeq – Kidd 55**

<table>
<thead>
<tr>
<th>Population</th>
<th>Abbreviation</th>
<th>Count in Training Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luhya in Webuye, Kenya</td>
<td>LWK</td>
<td>183</td>
</tr>
<tr>
<td>Americans of African Ancestry in SW USA</td>
<td>ASW</td>
<td>2350</td>
</tr>
</tbody>
</table>

Individuals Grouped by Self Described Eye Color

- **Brown**
  - 100%
  - 99%
  - 98%
  - 99%
  - 100%
  - 99%
  - 98%
  - 99%
  - 100%
  - 99%
- **Blue**
  - 100%
  - 99%
  - 98%
  - 99%
  - 100%

**SNP Assays - Conclusions**

- SNP genotyping performance is different from CE-STR assays
- Identity SNP panels are useful for degraded samples
- Identity SNP data can be combined with STR data
- Linkage Disequilibrium needs to be evaluated
- Labs must carefully consider how to convey ancestry and phenotype prediction information