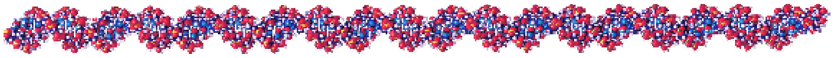
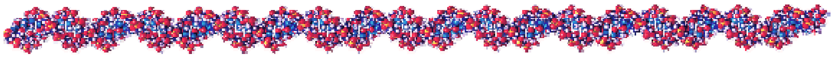


The Use of SNPs and Non-CODIS Genetic Markers: Beyond Identification



Peter M. Vallone and Kristen O’Connor

Applied Genetics Group

National Institute of Standards and Technology
Gaithersburg, Maryland

Technology Focus Day
In-Q-Tel Arlington, VA
February 9, 2011

Outline

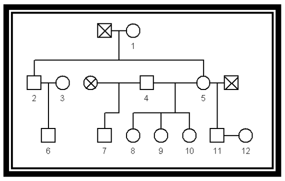
- Types of genetic variation
  - STRs & SNPs

.....(GAGA)(GAGA)(GAGA).....
.....(G/A).....

- Estimation of biogeographical ancestry

- Prediction of phenotypic traits

- Applications to kinship testing



Types of Genetic Variation

CODIS Loci are STRs

Length Variation

- **short tandem repeats (STRs)**

  CTAGTCGT**(GATA)(GATA)(GATA)**GCGATCGT

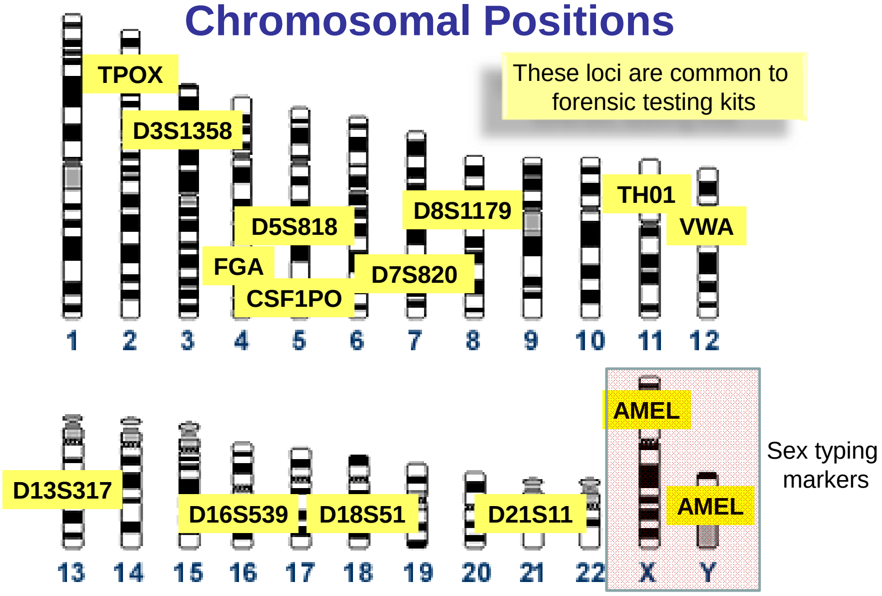
Sequence Variation

- insertions/deletions
- **single nucleotide polymorphisms (SNPs)**

  GCTAGTCGATGCTC**(G/A)**GCGTATGCTGTAGC

---

13 CODIS Core STR Loci with Chromosomal Positions



These loci are common to forensic testing kits

Sex typing markers

Use of CODIS Core STR loci

- STRs
  - – 9.6 million STR profiles in the national database
  - – Excellent for one-to-one matching
    - 1 in 100 trillion (13 CODIS)
  - – Paternity testing (M, F, Child)
- Well defined/uniform protocols for typing
  - – Sensitive and robust forensic use
  - – Target 0.5 to 1 ng of DNA (80 to 150 copies)

STR loci are not typically used for

- Predicting phenotypic traits
  - – Eye color, hair color, skin color (pigmentation)
  - – Height
  - – Facial structure

Externally visible characteristics

- Biogeographical Ancestry
  - – Estimate of continental genetic ancestry

Types of Genetic Variation

Length Variation

- **short tandem repeats (STRs)**

  CTAGTCGT**(GATA)(GATA)(GATA)**GCGATCGT

Sequence Variation

- insertions/deletions
- **single nucleotide polymorphisms (SNPs)**

  GCTAGTCGATGCTC**(G/A)**GCGTATGCTGTAGC

SNP Classifications

- Individual Identification SNPs (IISNPs): SNPs that collectively give very low probabilities of two individuals having the same multi-locus genotype
- Ancestry Informative SNPs (AISNPs): SNPs that collectively give a high probability of an individual's ancestry being from one part of the world or being derived from two or more areas of the world
- Phenotype Informative SNPs (PISNPs): SNPs that provide a high probability that the individual has particular phenotypes, such as a particular skin color, hair color, eye color, etc.
- Lineage Informative SNPs (LISNPs): Sets of tightly linked SNPs that function as multi-allelic markers that can serve to identify relatives with higher probabilities than simple bi-allelic SNPs

Budowle, B. and van, Daal. A. (2008) Forensically relevant SNP classes, *Biotechniques 44*, 603-8, 610.
Butler, J.M., Budowle, B., Gill, P., Kidd, K.K., Phillips, C., Schneider, P.M., Vallone, P.M., Morling, N. (2008) Report on ISFG SNP panel discussion. *Forensic Science International: Genetics Supplement Series (Progress in Forensic Genetics 12*) 1: 471-472.

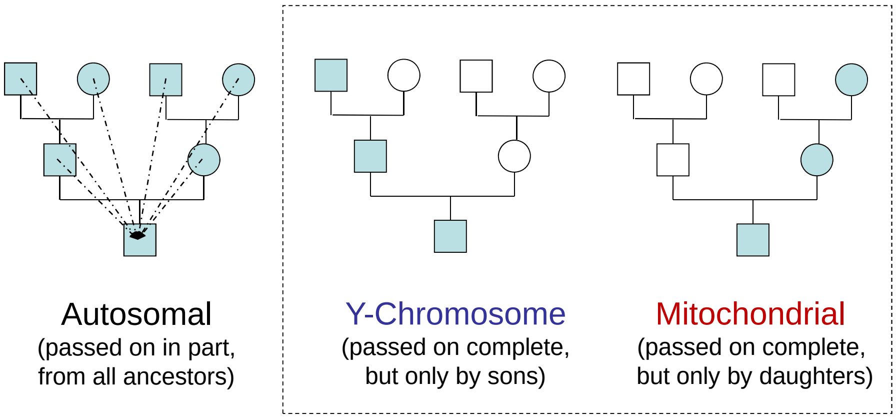
SNP Typing

- Forensic scale SNP assays
  - ~10 - 50 SNP markers
  - Utility with low amounts of sample (< 1ng)
- High throughput sequencing and DNA/SNP microarrays
  - Thousands to millions of SNPs
  - Greater input amounts of DNA are required
  - Higher level computational methods are required (data storage and analysis)

Multiple approaches and technologies exist for SNP typing

Different Inheritance Patterns

Lineage Markers



Autosomal
(passed on in part,
from all ancestors)

Y-Chromosome
(passed on complete,
but only by sons)

Mitochondrial
(passed on complete,
but only by daughters)

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.1, ©Elsevier Science/Academic Press

Ancestry Informative SNPs

Lao O, van Duijn K, Kersbergen P, de Knijff P, Kayser M. (2006) Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. Am J Hum Genet 78: 680-90.

10 SNPs

Phillips, C., Salas, A., Sanchez, J.J., Fondevila , M., Gomez-Tato, A., Alvarez-Dios, J., Calaza, M., Casares de Cal, M., Ballard, M., Lareu, M.V., Carracedo, A. (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. FSI: Genetics 1: 273-280.

34 SNPs

Halder, I., Shriver, M., Thomas, M., Fernandez, J.R., Frudakis, T. (2008) A Panel of Ancestry Informative Markers for Estimating Individual Biogeographical Ancestry and Admixture From Four Continents: Utility and Applications. Hum Mut 29: 648-658.

HUMAN MUTATION Mutation in Brief (2010) Online

MUTATION IN BRIEF

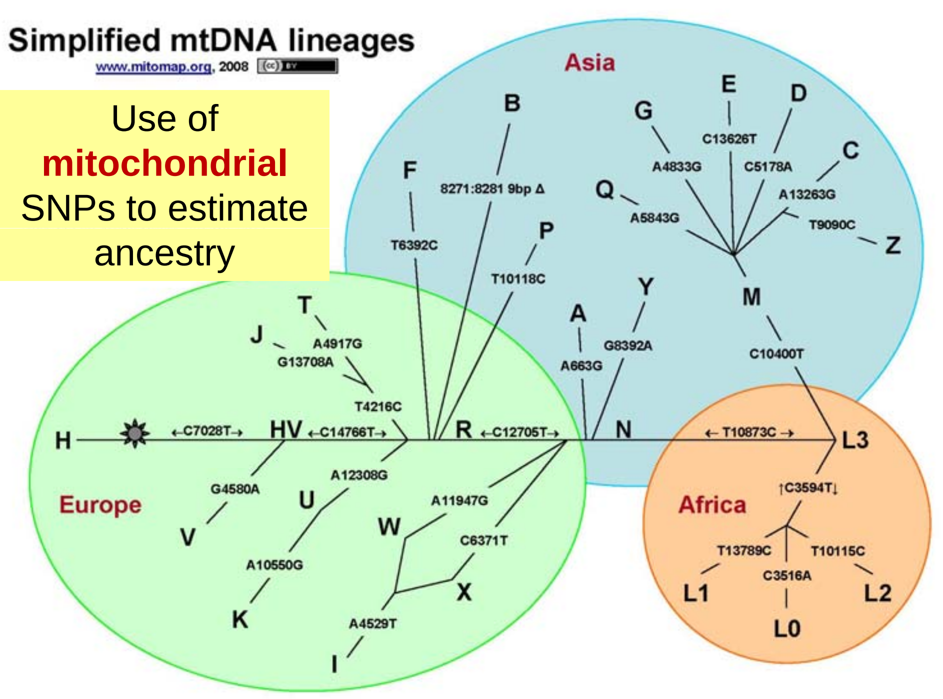
HUMAN MUTATION

24 SNPs

Evaluating Self-declared Ancestry of U.S. Americans With Autosomal, Y-chromosomal and Mitochondrial DNA

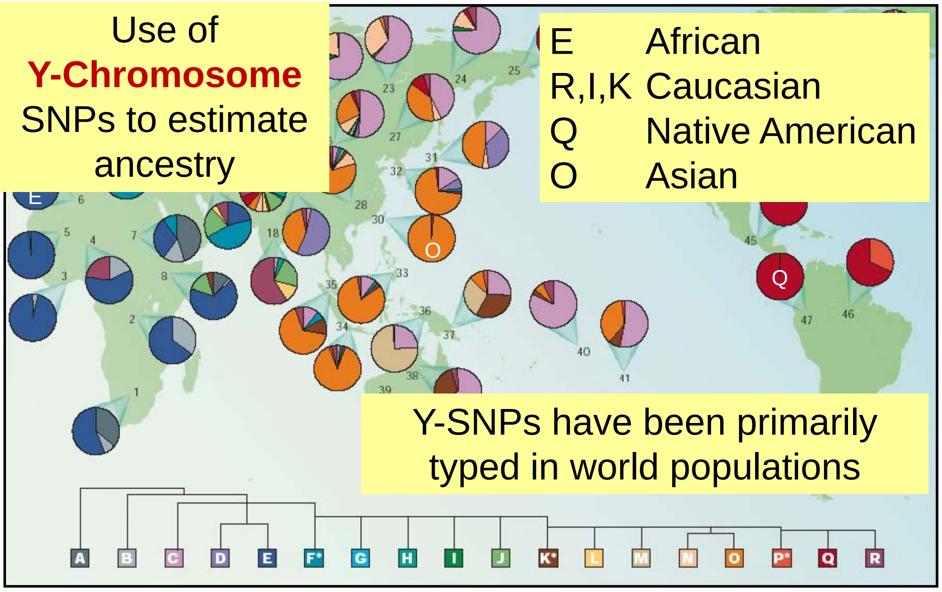
Oscar Lao[1,*], Peter M. Vallone[2,*], Michael D. Coble[3], Toni M. Diegoli[3], Mannis van Oven[1], Kristiaan J. van der Gaag[4], Jeroen Pijpe[4], Peter de Knijff[4], and Manfred Kayser[1*]

Ongoing and active area of research
Screening for sets of **autosomal** SNPs that will estimate ancestry
Global perspective



Use of **mitochondrial** SNPs to estimate ancestry

Global Distribution of Y Haplogroups



Use of
Y-Chromosome
SNPs to estimate
ancestry

E	African
R,I,K	Caucasian
Q	Native American
O	Asian

Y-SNPs have been primarily typed in world populations

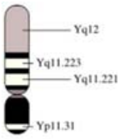
Y Chromosome Consortium (2003) Nat Rev Genet. 4 :598-612

Biogeographical Ancestry

Sample 1

Mitochondrial Hg
H
Europe

24 autosomal SNPs
94 % Caucasian
6 % (other)



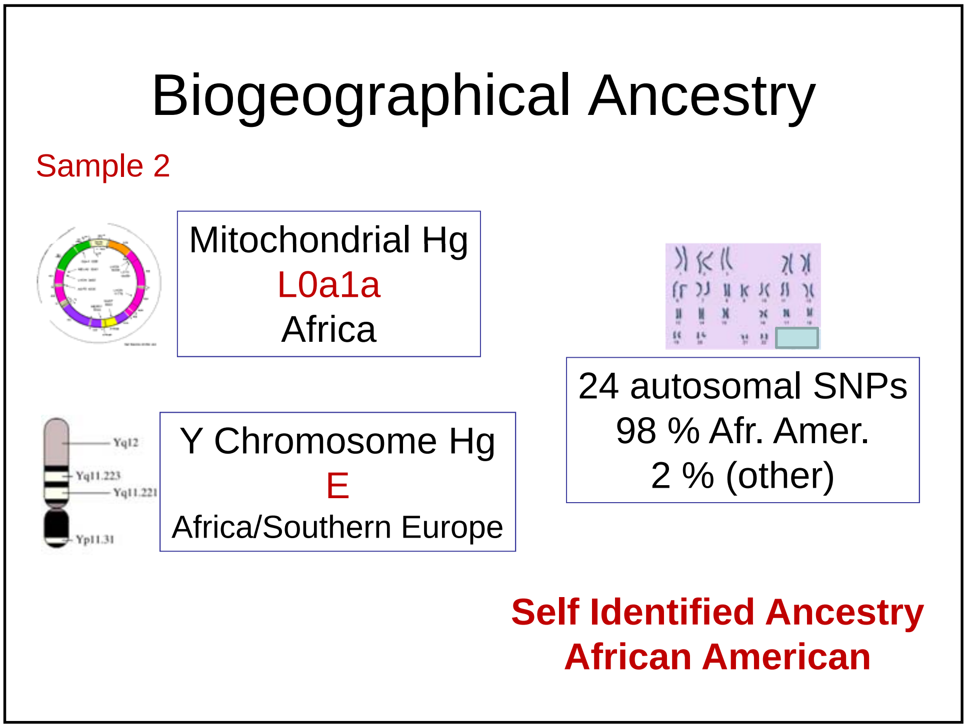
Y Chromosome Hg
R1b1b2
Western Europe

MUTATION IN BRIEF

HUMAN MUTATION

Evaluating Self-declared Ancestry of U.S. Americans With Autosomal, Y-chromosomal and Mitochondrial DNA

Self Identified Ancestry
Caucasian



Biogeographical Ancestry

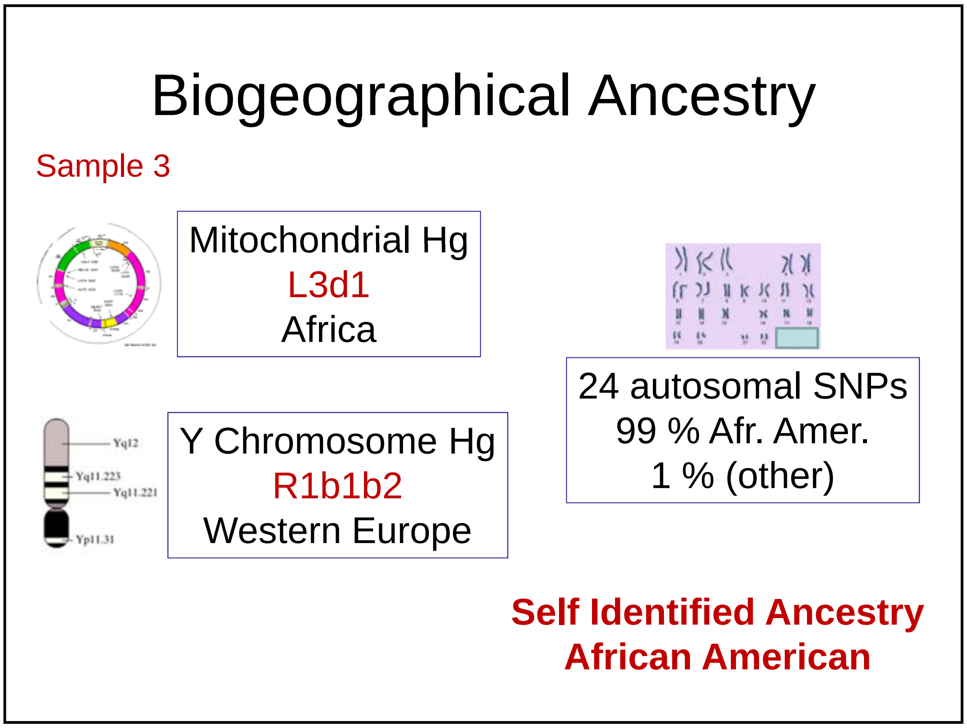
Sample 2

Mitochondrial Hg
L0a1a
Africa

Y Chromosome Hg
E
Africa/Southern Europe

24 autosomal SNPs
98 % Afr. Amer.
2 % (other)

Self Identified Ancestry
African American



Biogeographical Ancestry

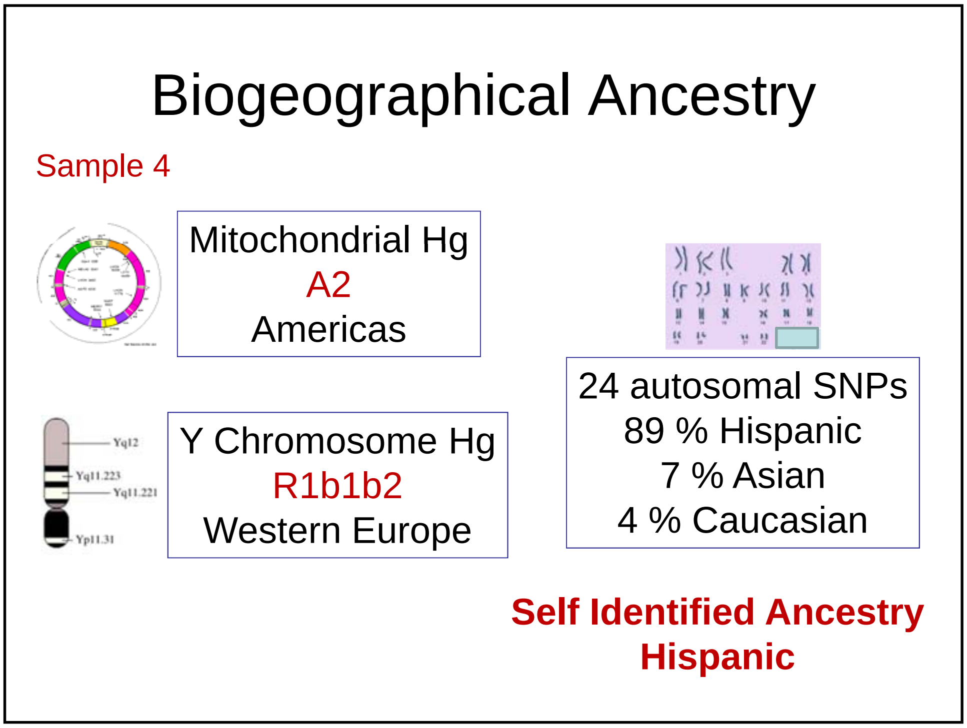
Sample 3

Mitochondrial Hg
L3d1
Africa

Y Chromosome Hg
R1b1b2
Western Europe

24 autosomal SNPs
99 % Afr. Amer.
1 % (other)

Self Identified Ancestry
African American



Biogeographical Ancestry

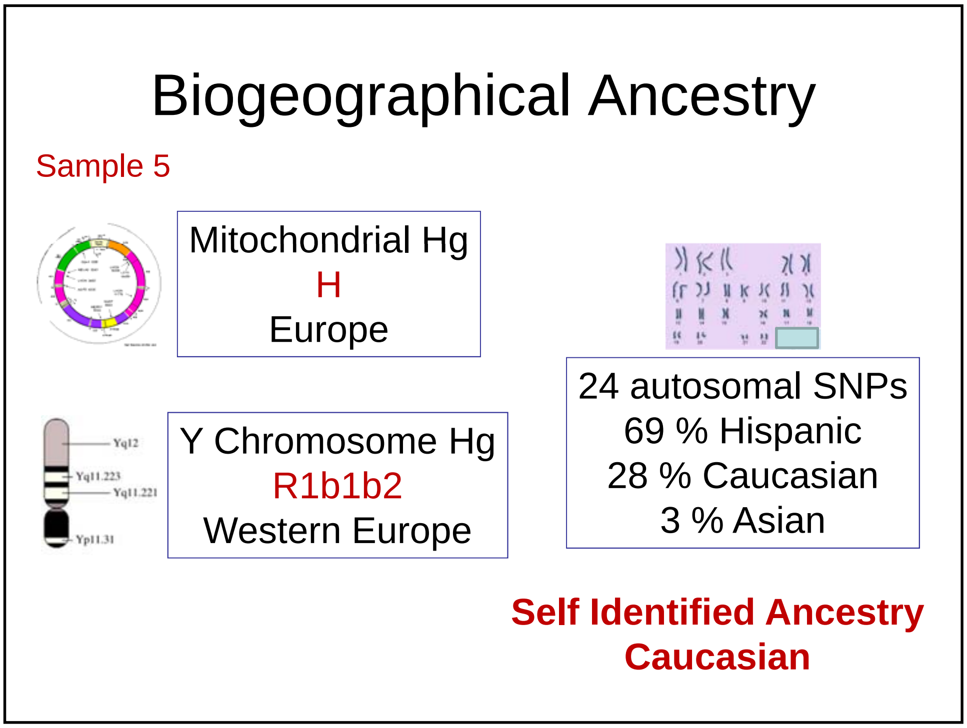
Sample 4

Mitochondrial Hg
A2
Americas

Y Chromosome Hg
R1b1b2
Western Europe

24 autosomal SNPs
89 % Hispanic
7 % Asian
4 % Caucasian

Self Identified Ancestry
Hispanic



Biogeographical Ancestry

Sample 5

Mitochondrial Hg
H
Europe

Y Chromosome Hg
R1b1b2
Western Europe

24 autosomal SNPs
69 % Hispanic
28 % Caucasian
3 % Asian

Self Identified Ancestry
Caucasian

Summary of Ancestry SNPs

Success estimating Self Identified Ancestry

n	U.S. sampling	Mitochondrial	Y Chromosome	24 Autosomal
259	**African American**	93%	69%	98%
262	**Caucasian**	97%	84%	81%
49	**Asians**	99%	100%	100%
140	**Hispanics**	NA	NA	74%

The appropriate populations of interest need to be studied

Issues with 'newer' admixed populations (e.g. US Hispanics)

Phenotype Informative SNPs

- Predict an observable trait – definitively
  - Eye, hair, and skin color
  - Height, stature
- Some key pigmentation genes have been characterized
  - Wide range of pigmentation in humans
  - Multiple genes involved, complex phenotype
- Gene discovery and characterization is ongoing
- Should not be predictive of disease

Tully G. Genotype versus phenotype: human pigmentation. Forensic Sci Int Genet. 2007 1: 105-10.

Pigmentation Related Genes

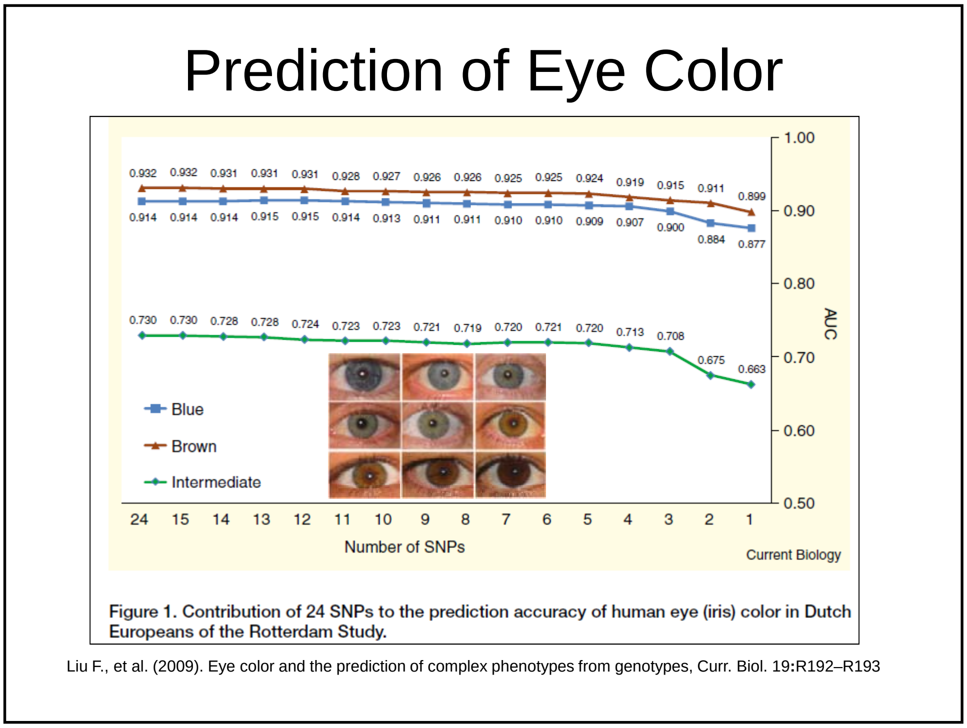
Eye	Hair	Skin
ASIP	ASIP	ASIP
		DCT
		DRD2
		EGFR
HERC2	HERC2	HERC2
IRF4	IRF4	IRF4
	KITLG	KITLG
	MATP	MATP
	MC1R	MC1R
OCA2	OCA2	
		MYO5A
SLC24A4	SLC24A4	SLC24A4
SLC45A2		
	TPCN2	SLC24A5
TYR	TYR	TYR

Overlap between eye, hair, and skin related genes

Potential for a panel of markers that could predict all three traits

Recent Work on PISNPs

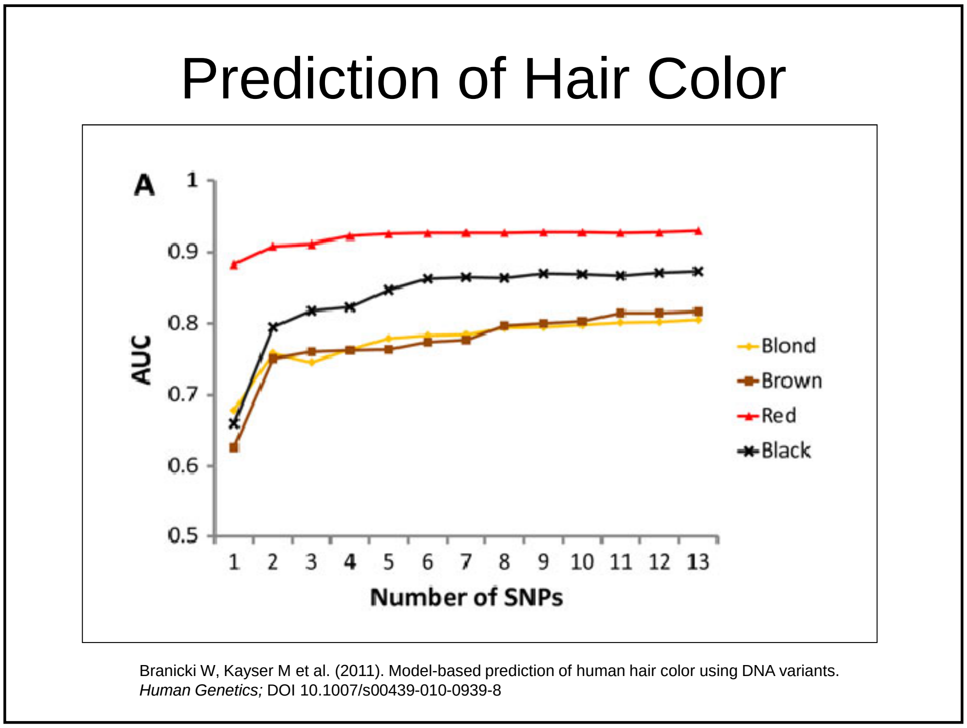
- Branicki W, Kayser M et al. (2011). Model-based prediction of human hair color using DNA variants. *Human Genetics;* DOI 10.1007/s00439-010-0939-8
- Walsh S., et al. (2010) IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci. Int. Genet. (Epub)
- Mengel-From J., et al. (2010) Human eye colour and HERC2, OCA2 and MATP. Forensic Sci. Int. Genet. 5: 323-8
- Kayser M., Schneider P.M. (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. Forensic Sci. Int. Genet. 3(3):154-61
- Liu F., et al. (2009). Eye color and the prediction of complex phenotypes from genotypes, Curr. Biol. 19**:**R192–R193



Prediction of Eye Color

Figure 1. Contribution of 24 SNPs to the prediction accuracy of human eye (iris) color in Dutch Europeans of the Rotterdam Study.

Liu F., et al. (2009). Eye color and the prediction of complex phenotypes from genotypes, Curr. Biol. 19**:**R192–R193

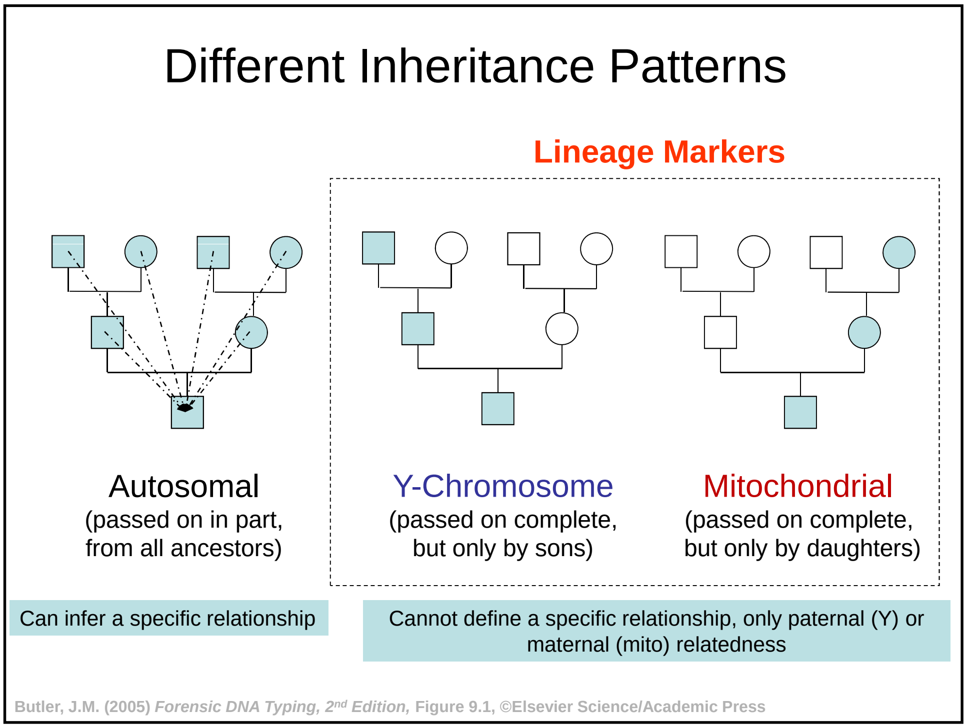


Prediction of Hair Color

Branicki W, Kayser M et al. (2011). Model-based prediction of human hair color using DNA variants. *Human Genetics;* DOI 10.1007/s00439-010-0939-8

Lineage Informative SNPs

- Much of the LISNP literature focused on Y-chromosome and mitochondrial DNA SNPs
- With genome-wide arrays, autosomal SNP typing for lineage analysis is possible
- Looking for blocks of DNA that have been transmitted unchanged from one generation to the next
- Useful in evolutionary studies and kinship analysis



Different Inheritance Patterns

Lineage Markers

Autosomal
(passed on in part, from all ancestors)

Y-Chromosome
(passed on complete, but only by sons)

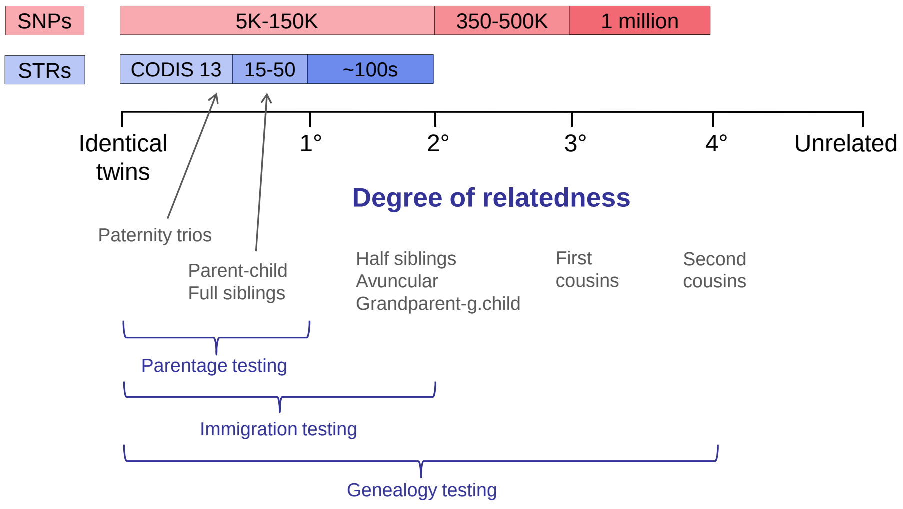
Mitochondrial
(passed on complete, but only by daughters)

Can infer a specific relationship

Cannot define a specific relationship, only paternal (Y) or maternal (mito) relatedness

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.1, ©Elsevier Science/Academic Press

Kinship Inference with Autosomal STRs or SNPs



SNPs: 5K-150K | 350-500K | 1 million

STRs: CODIS 13 | 15-50 | ~100s

Identical twins — 1° — 2° — 3° — 4° — Unrelated

Degree of relatedness

Paternity trios

Parent-child
Full siblings

Half siblings
Avuncular
Grandparent-g.child

First cousins

Second cousins

Parentage testing

Immigration testing

Genealogy testing

Kinship Analysis with Autosomal SNPs

- Supplement autosomal STRs for paternity and complex kinship analysis

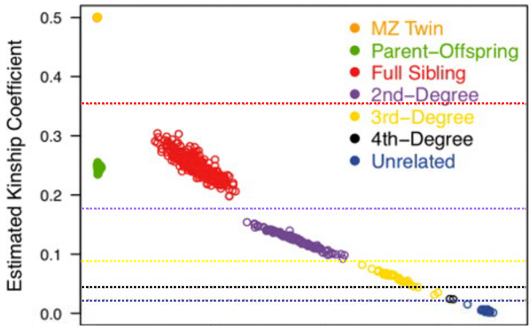
  C. Børsting, N. Morling. Forensic Sci. Int. Genet. 2010 Mar 29. [Epub ahead of print]
  C. Phillips et al. Forensic Sci. Int. Genet. 2008 Jun;2(3):198-204.

- Identity by state (AA, AB, or BB)
  - Identify longer blocks of DNA shared between related individuals than between random, unrelated individuals
  - No allele frequencies required

- Clustering based on identity by descent
  - Based on the likelihood of observing DNA shared from a common ancestor assuming a specific relationship question
  - Requires allele frequencies estimated from tested samples or population database
    - Population structure (admixture) can inflate inferred relatedness if not accounted for in the analysis

Relationship inference using 350K SNPs

602 individuals from 143 families with self-reported relationships



Degree of relatedness

More related ⟶ less related

Different relationships cluster together.
3° and 4° are not always distinguishable.
4° is not distinguishable from unrelated.

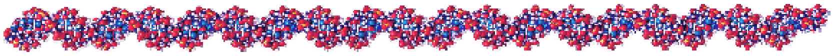
Adapted from Manichaikul et al. Bioinformatics 2010;26:2867-2873

Summary

- Non-CODIS markers can be used for estimating biogeographical ancestry, phenotype, and relatedness
- Level of certainty varies with application and markers
  - STRs/SNPs:
    - Identification (one-to-one matching): ~100% (10^{-16})
  - SNPs:
    - Biogeographical ancestry: ~70-90%
    - Phenotype: ~70-90%
    - Kinship: Variable (depends on relationship)
- Improvements in certainty may be achieved through more discriminating panels of markers
  - Ongoing studies of well-defined traits or with specific (relevant) population groups can help refine prediction

Questions?

Project Leader	NRC Postdoctoral Fellow
Peter Vallone	**Kristen O’Connor**
Rapid PCR & Biometrics	Kinship Analysis
peter.vallone@nist.gov 301-975-4872	kristen.oconnor@nist.gov 301-975-5205



Final version of this presentation available at:
http://www.cstl.nist.gov/strbase/NISTpub.htm

Forensic SNP Information:
http://www.cstl.nist.gov/strbase/SNP.htm