

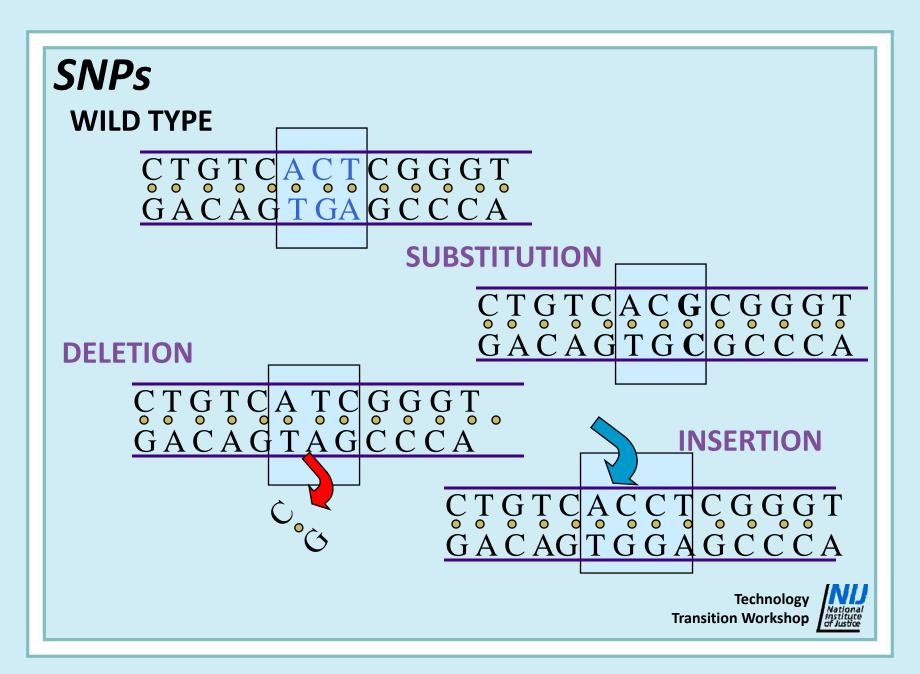
Technology Transition Workshop | Bruce Budowle, Ph.D.

Single Nucleotide Polymorphisms (SNPs)

Definition

SNPs are single nucleotide base substitutions (or an insertion or a deletion) in the genome and account for 85% of the genetic variability in humans.





SNPs

Person 1

GCA AGA GAT AAT TGT

Ala Arg Asp Asn Cys

Person 2

Synonymous

GCG AGA GAT AAT TGT

Ala Arg Asp Asn Cys

Person 3

Non-Synonymous

GCA GGA GAT AAT TGT

Ala Gly Asp Asn Cys

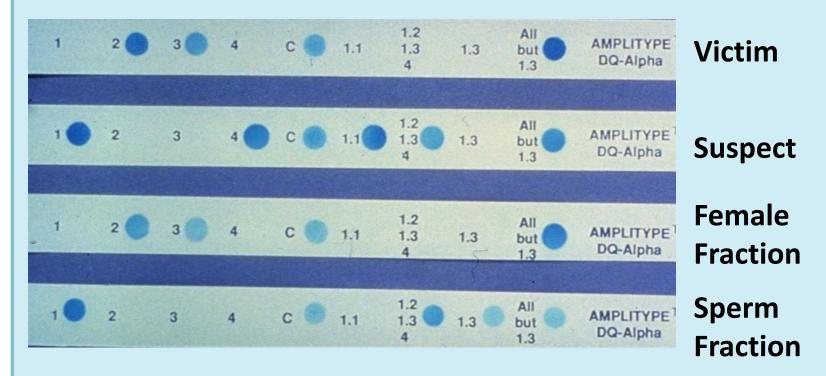


DQ Alpha Typing

- Located on chromosome 6
- 242 bp amplicon
- Seven alleles could be detected / inferred:
 - **1.1, 1.2, 1.3**
 - 2
 - 3
 - **4.1, 4.2/4.3**



DQ Alpha Typing (First Forensic SNP Assay)



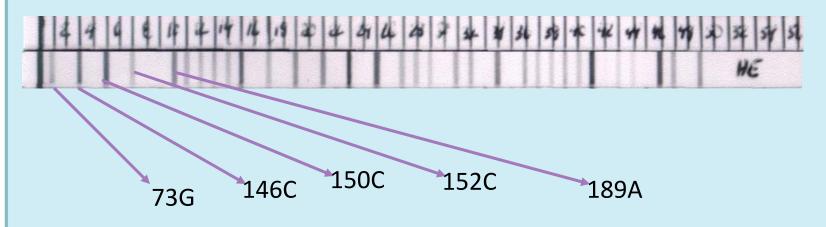
Sequence specific oligonucleotides bound to nylon membrane (SSO test strips)



SNP Assay - Hybridization Based

Rapid mtDNA SNP typing

Sequence specific oligonucleotides bound to nylon membrane (SSO test strips)





SNPs as Forensic Markers – Advantages

- Abundant
- Small amplicon size
 - As small as 45 to 55 bp the length of the two PCR primers
 - Very useful for severely degraded samples
- Low mutation rate
 - About 10⁻⁸ versus 10⁻³ for STRs



SNPs as Forensic Markers – Advantages

- Bi-allelic nature
 - More amenable to automation
 - Allele typing interpretation is simpler (e.g., no stutter)



SNPs as Forensic Markers – Limitations

- Mixture interpretation
- Lower power of discrimination (PD)
 - Multiplexing
- Low mutation rate
 - Population substructure
- Privacy concerns
 - Linkage to other genetic information



Mixtures

- More loci will be needed
- Quantitation
- Mixture deconvolution



Types of SNPs for Forensic Applications

- Identity Testing SNPs individualization, high heterozygosity, low F_{st}
- Ancestry Informative SNPs high probability of an individual's geographical ancestry
- Lineage Informative SNPs sets of tightly linked SNPs that function as multiallelic markers to identify relatives (missing persons)
- Phenotype Informative SNPs high probability that the individual has particular phenotype, such as skin color, hair color, eye color, etc.
- Pharmacogenetic SNPs molecular autopsy, personalized medicine

How Many SNP Loci Would Equal the Power of the Combined CODIS 13 STR Loci?

1682

Electrophoresis 1999, 20, 1682-1696

Review

Ranajit Chakraborty¹
David N. Stivers¹
Birg Su¹
Yixi Zhong¹
Bruce Budowle²

The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems

- Genomic location
- Allele frequency distribution
- Genetic substructure
- Random match probability
 - (1 in 10⁹ to 1 in 10¹⁵)



How Many SNP Loci Would Equal the Power of the Combined CODIS 13 STR Loci?

1682

Electrophoresis 1999, 20, 1682-1696

Review

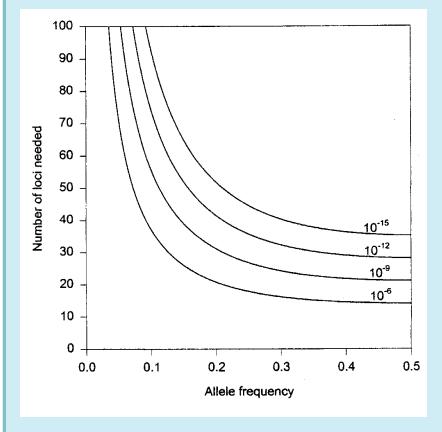
Ranajit Chakraborty¹
David N. Stivers¹
Birg Su¹
Yixi Zhong¹
Bruce Budowle²

The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems

- Biallelic
- Based solely as a function of allele frequency
- Average match probability is symmetrical around allele frequency of 0.5



How many SNPs?

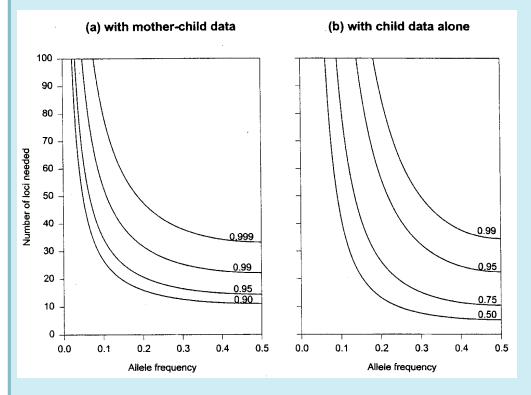


- 25 to 42 (p = 0.3, 0.7)
- 62 (p = 0.1, 0.9)
- 1 in 10⁹

Technology Transition Workshop



How Many SNPs are Needed for Paternity Testing



- 99.9% data on mother and child
- 33 81 loci (p = 0.5 p = 0.1)
- Data on child alone – 80%

Technology Transition Workshop



Identity SNPs - SNP Characteristics

- High heterozygosity
 - Maximizes information from each SNP
- Low F_{st} minimizes chance differences between populations
- Fewer SNPs and fewer population databases needed



Abundance

- Number of loci are limited by heterozygosity criterion
- F_{st}, linkage, chemistry
- Sanchez, et al. (2006) 52 SNPs
- Pakstis, et al. (2007) 40 SNPS***
- FBI / Orchid 120 SNPs

*** Pakstis, et al. seem to be the best based on population studies





Available online at www.sciencedirect.com



Forensic Science International 164 (2006) 20-32



Developing a SNP panel for forensic identification of individuals

Kenneth K. Kidd ^{a,*}, Andrew J. Pakstis ^a, William C. Speed ^a, Elena L. Grigorenko ^b, Sylvester L.B. Kajuna ^c, Nganyirwa J. Karoma ^c, Selemani Kungulilo ^d, Jong-Jin Kim ^e, Ru-Band Lu ^f, Adekunle Odunsi ^g, Friday Okonofua ^h, Josef Parnas ⁱ, Leslie O. Schulz ^j, Olga V. Zhukova ^k, Judith R. Kidd ^a

Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA
 Child Study Center and Department of Psychology, Yale University, New Haven, CT, USA
 The Hubert Kairuki Memorial University, Dar-es-Salaam, Tanzania
 Muhimbili University College of Health Sciences, Dar-es-Salaam, Tanzania
 DNA Analysis Division, National Institute of Scientific Investigation, Seoul, Korea
 Department of Psychiatry, College of Medicine and Hospital, National Cheng-Kung University, Tainan, Taiwan, ROC
 Department of Gynecological Oncology, Roswell Park Cancer Institute, Buffalo, NY, USA
 Department of Obstetrics and Gynecology, Faculty of Medicine, University of Benin, Benin City, Nigeria
 The Danish National Research Foundation, Center for Subjectivity Research, University of Copenhagen, Købmagergade 46, DK-1150 Copenhagen N, Denmark
 College of Health Sciences, University of Texas at El Paso, El Paso, TX, USA
 N.I. Vavilov Institute of General Genetics RAS, Moscow, Russia

Received 2 September 2005; received in revised form 3 November 2005; accepted 8 November 2005 Available online 19 December 2005

Kidd, K. et al., Forensic Sci Int (2006) 164(1) 20-32

Technology Transition Workshop

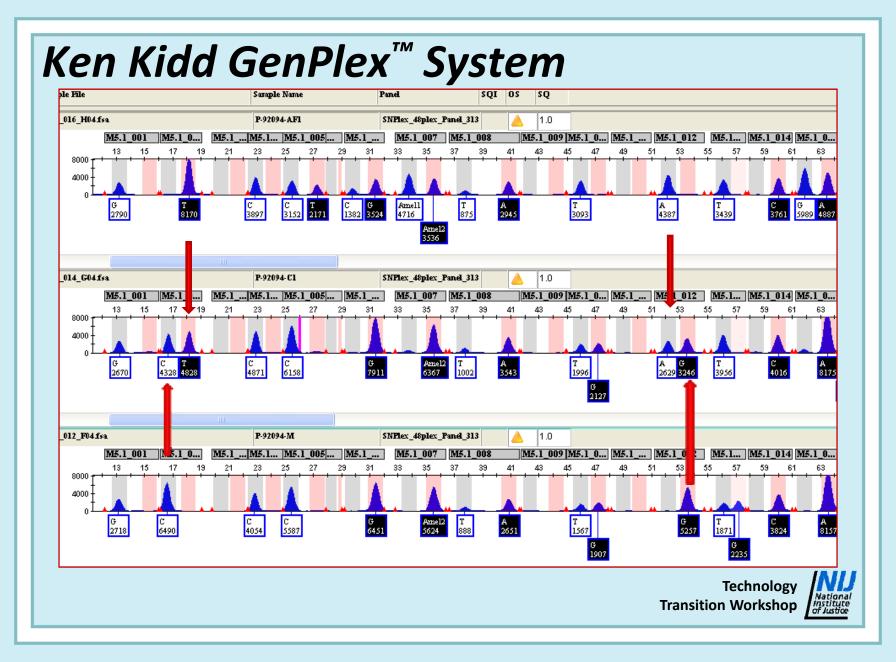


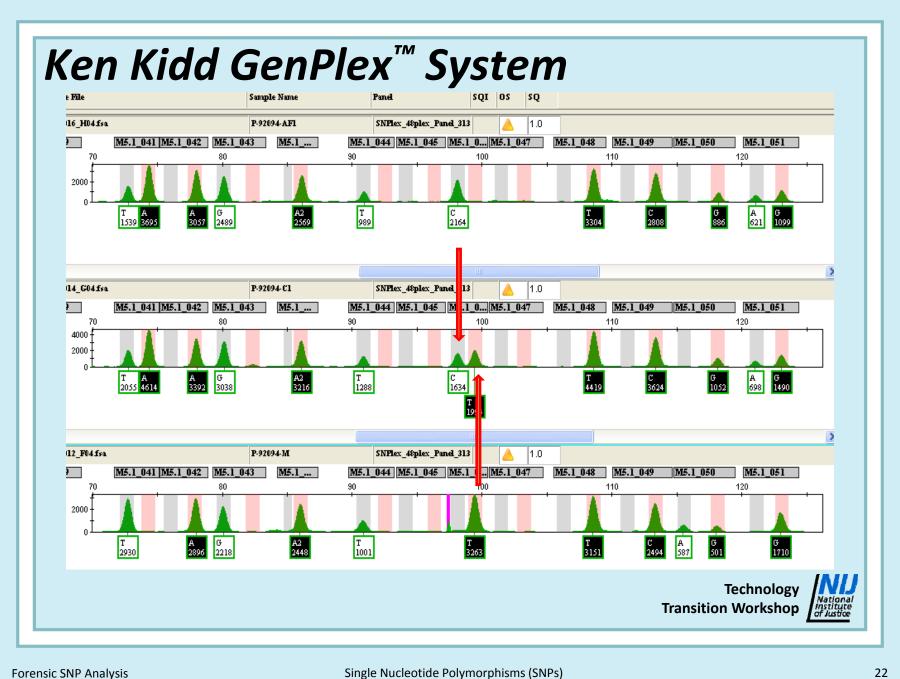
Identity SNPs

- SNP Screen
- 90,483 AB SNPs
- Allele frequencies in European American, African American, Chinese / Japanese
- 14,638 → avg. heterozygosity ≥ 0.45 per three populations
- F_{st} of 2,723 SNPs < 0.01 per three populations
- Chose SNPs > 1 Mb apart
- Screened 195 SNPs in seven populations











Available online at www.sciencedirect.com







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

Forensic Genetics Department, Genomic Medicine Group, University of Santiago de Compostela, Galicia, Spain
 b Applied Markets Group, Applied Biosystems, Foster City, CA, USA
 c Department of Haematology, ICMS, Queen Mary's School of Medicine & Dentistry, London E1 2AT, UK
 d Institute of Legal Medicine, University of Cologne, Germany

Received 29 January 2007; accepted 3 February 2007

Phillips, C. et al., Forensic Science International Genetics (2007) **1(2)** 180-185





The SNPforID Consortium

- A five lab consortium of groups from Innsbruck,
 Copenhagen, Mainz, Santiago & Barts, London
- Examining SNP analysis for forensic identification using high through-put techniques
- Collaborative framework formed from several EDNAP academic groups
- Funded for three years under EU Framework 5
 "competitive and sustainable growth"
- Open source intending rapid dissemination of data to forensic community

SNPforID and Identifiler[®]: European Population

	Average Probability of Identity	1 – (Average Probability of Paternity Exclusion)
SNPforID 52 SNPs	3.0 x 10 ⁻²¹	4.46 * 10 ⁻⁵
Identifiler®	1.19 * 10-18	5.39 * 10 ⁻⁷

SNPs are more informative for identity but less informative for paternity exclusion but lower mutation rate



52 SNPforID SNPs – Across Populations

Population	Average Probability of Identity	1 – (Average Probability of Paternity Exclusion)
AB_African_American	4.8 x 10 ⁻¹⁹	0.0001259
HapMap_Nigerian_Yoruba	3.1 x 10 ⁻¹⁶	0.0004643
SNPforID_Somalian	1.2 x 10 ⁻¹⁹	0.0000954
Sequenom_CEPH_African	1.0 x 10 ⁻¹⁵	0.0005886
AB_European	1.5 x 10 ⁻²¹	0.0000386
HapMap_Utah_Europeans	3.0 x 10 ⁻²¹	0.0000446
SNPforID_Dane_German	3.4 x 10 ⁻²¹	0.0000458
Sequenom_CEPH_European	6.9 x 10 ⁻²¹	0.0000525
AB_Asian	8.96 x 10 ⁻¹⁹	0.0001399
HapMap_Han_Chinese	5.2 x 10 ⁻¹⁸	0.0001993
SNPforID_Asian_Combined	3.3 x 10 ⁻¹⁸	0.0001812
Sequenom_CEPH_Chinese	8.5 x 10 ⁻¹⁹	0.0001385

Technology Transition Workshop



Privacy Concern Criterion for Identity SNPs

- No medical or sensitive personal information
- One can appreciate public apprehension over having medical information conveyed by the SNP alleles in a forensic database or case analysis
- Ethical concerns over identifying high likelihood of an individual developing a cancer, Alzheimer's disease, Huntington's disease, etc. should preclude using SNPs that would convey such information
- However, from a scientific perspective that does not generalize to precluding all SNPs from even those genes, much less any gene

 Technology Transition Workshop Transition Workshop

Privacy Concern Criterion for Identity SNPs

- The Mendelian disorders are rare
- SNPs with high heterozygosity will not convey significant information about the mutations for a Mendelian disorder even if there is complete linkage disequilibrium
- Multigenic disorders are less likely to be SNP informative



Linkage

- Legitimate privacy concerns; but what is reality?
- Consider Disease Gene (DG) is rare (e.g., 10,000 people in the United States who are afflicted with the condition)
- Assume that 10% of these individuals are convicted offenders whose SNP profiles are in the offender database
- Assume every one of these 10,000 people have common ancestor (disease allele arose once)
- Assume SNP C-allele is in complete linkage with mutant DG allele or resides within exon

Single Nucleotide Polymorphisms (SNPs)

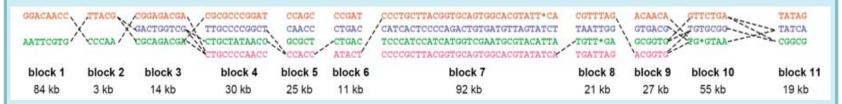
Technology Transition Workshop

Linkage

- Bias selection of SNPs for identity testing
- Assume SNP C-allele (f = 0.5) is used to predict the presence of DG allele for everyone / anyone in the database (N = 5,000,000)
- Assuming HWE, 75% of 5 million, or 3,750,000 people carry at least one copy SNP C (25% CC, 50% CT)
- Of these 3,750,000 positive predictions, only correct in 1000 cases
- For any particular positive prediction, the probability of the DG mutation is only 1000/3,750,000 = 0.00027
- The SNP locus has essentially no predictive value in the general population

Single Nucleotide Polymorphisms (SNPs)

Haplotype Block (Haploblock)

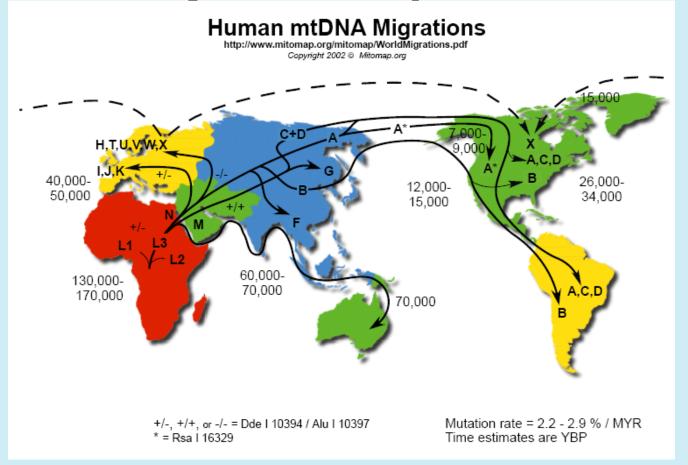


Haplotype structure across 500 kb on 5q31 (Daly, M.J., et al. 2001, Nat. Genet. 29: 229-232)

- Human genome is composed of block-like structures of low haplotype diversity (strong LD within block) separated by recombination hot spots
- Lineage marker like Y-chromosome and mtDNA
- Pseudo-STRs
- Kinship analysis



mtDNA SNPs for Ancestry



http://www.mitomap.org/WorldMigrations.pdf

Image courtesy of MITOMAP: A Human Mitochondrial Genome Database, 2009.

Technology Transition Workshop



Y Chromosome SNPs for Ancestry

Global Distribution of Y Haplogroups

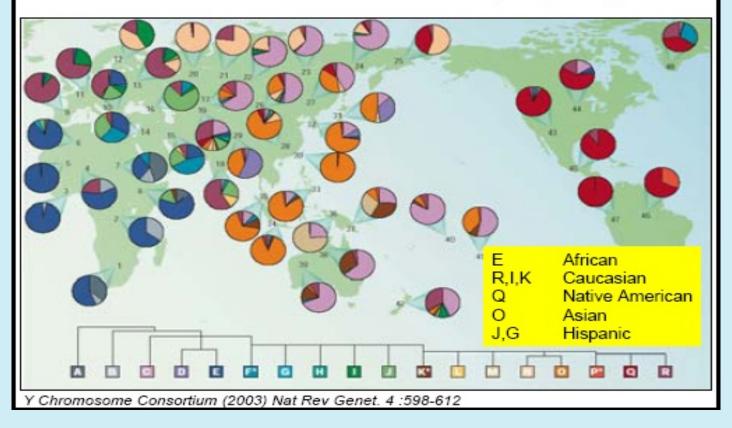


Figure 2 from: Jobling, M.A. and Tyler-Smith, C. *Nature Reviews Genetics* (2003) **4** 598-612 https://www.familytreedna.com/pdf/nrg1124 fs.pdf Tran

Technology
Transition Workshop



Ancestry

- Ancestry information
- Ancestry informative markers (AIMs)
 - Large differences in allele frequencies between / among world populations
 - May reveal geographic ancestral origin of a sample / person

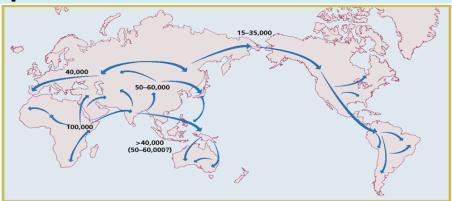


FIGURE 3 from: Cavalli-Sforza, L.L. and Feldman, M.W.

Nature Genetics Supplement (2003) **33** 266-275

http://hpgl.stanford.edu/publications/NGS 2003 v33 p266-275.pdf





DNAPrint – Mapping by Admixture Linkage Disequilibrium (MALD)

- MALD takes advantage of long-range haplotypes generated by gene flow among recently admixed groups
- Process used by DNAPrint Genomics was more consistent with MALD than ancestry informative markers (AIM)



Race

- Difficult to define
- Much discussion in literature
- Yet, there are some obvious differences that are associated with ancestral geography
- Forensic population data support the classifications



Basics

- Platform takes advantage of "genomic structure" or "population structure"
- A genome map of ancestry informative markers (AIMs) with delta > 0.4
- Population structure, sub-structure and microstructure demonstrated by measurement of AIMs
- Can accurately measure population structure within individuals as well cryptic structure between populations
- But some limitations



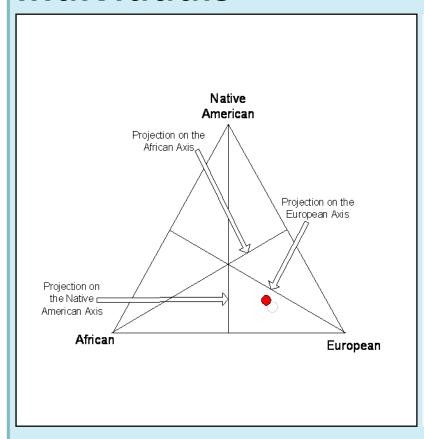
Basic Considerations

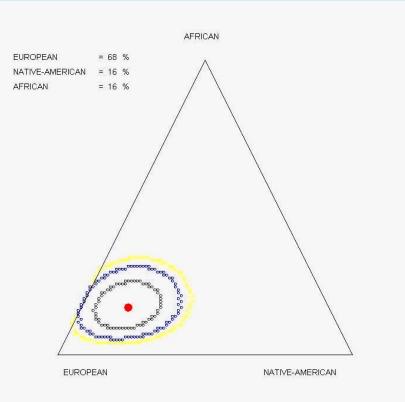
- Biogeographical Ancestry genetic structure is measurable and consistent with self-held notions of race
- Four main continental groups sub-Saharan, East Asian, IndoEuropean, Native American
- Crude geography
- Parental populations and self-reported population affiliation samples
- STRs may in some case be useful estimating major ancestral component, but were not selected for resolving population affiliation

Single Nucleotide Polymorphisms (SNPs)

Transition Workshop

Measure Population Structure within Individuals







Louisiana Serial Killer Case

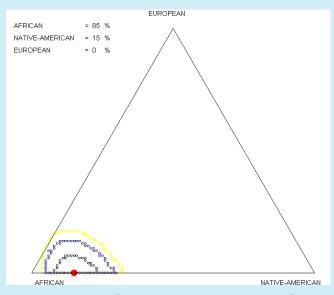
DNAPrint's DNA Witness Test provided break in the Louisiana multi-agency homicide task force serial killer case – world's first genomics-derived test for forensics redirected investigation with

dramatic results



Serial killers profile prior to DNAWitness testing, based on eye-witness reports and psychological profiling.

March 2002 to March 2003



BGA profile similar to that of serial killer obtained from DNA found at one of the crime scenes.

March 2003



May 2003





Physical Trait SNPs

- Current forensic DNA testing requires suspect with a "matching" profile
- DNA markers that describe appearance traits will allow genetic prediction of probable appearance for investigative lead
- Then type suspect for standard DNA markers
- Facial reconstructions



Phenotype Informative SNPs

- Skin color
 - SLC24A5, MATP, TYR, P, RABGGTA,
 MLPH, MYO5A, MC1R, ATP7B



- Hair color
 - EGFR, SLC24A5, MATP, TYR, RABBGTA,
 AP3B1, P, MLPH, MC1R, ATRN



- Eye color
 - SLC24A5, MATP, TYR, P, MYO7A, MC1R, ATP7B



Images courtesy of the National Eye Institute, National Institutes of Health



Privacy Concerns??

- Analysis on anonymous crime scene samples
- Sample not linked to specific person so no privacy concern
- Predictive / investigative of external traits only
- On suspect arrest obtain sample for STR typing
- Not racial profiling



Phenotype Informative SNPs Physical Appearance

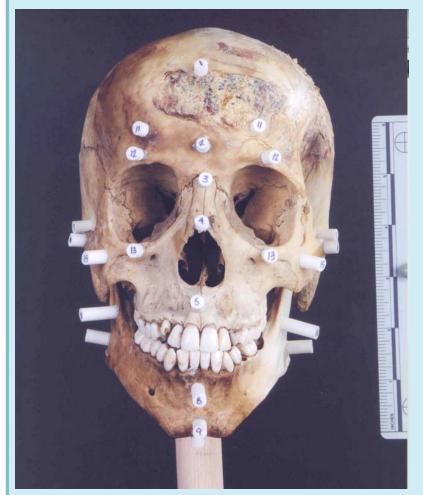
- High heritability traits
 - Pigmentation
 - Height
 - Facial morphology

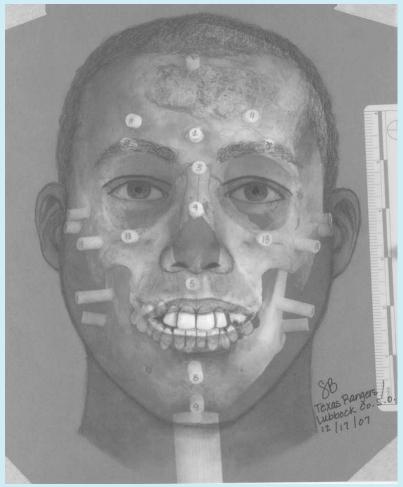


Badly Decomposed Human Skull Found Near Lubbock, Texas in December 2005



Facial Reconstruction Process





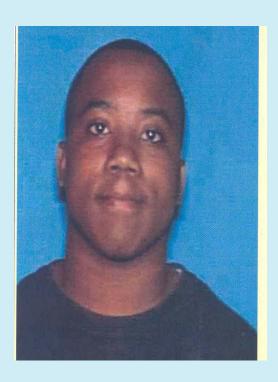
Images courtesy of the University of North Texas Center for Human Identification and the Lubbock County Texas Rangers



Identification of Human Skull Aided by Facial Reconstruction



Facial reproduction developed from the skull discovered December 2005



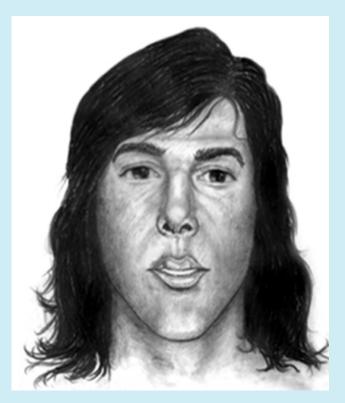
Bernard Wilson's Texas ID photo taken 12/2004

Images courtesy of

http://www.txdps.state.tx.us/director staff/public information/annrep2007.pdf



Forensic Facial Imaging

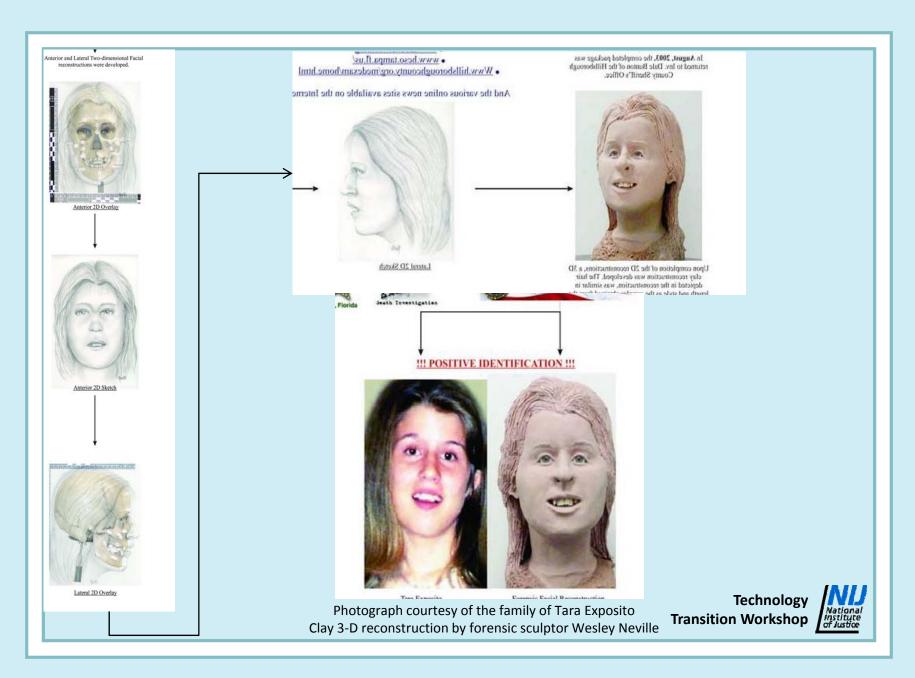


The victim was identified shortly after this drawing was released

Images courtesy of the University of North Texas Center for Human Identification





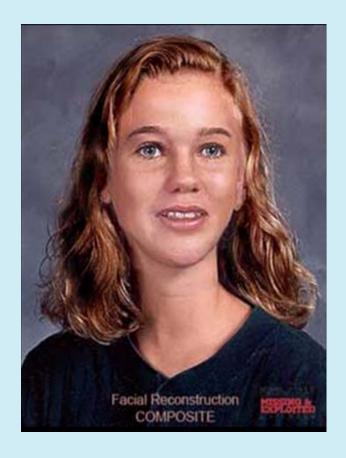


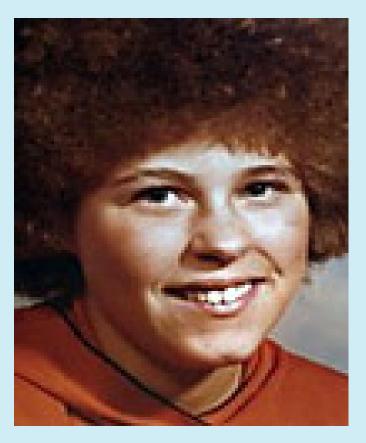
As Seen on TV

 On September 26, 1979, a young female body was found on a beach in Marin County, California. The body had been stabbed 43 times with an ice pick and doused with acetone and then set on fire. The female had also been shot in the head. For over 27 years the body remained unidentified.



As Seen on TV





Forensic sketch provided by the National Center for Missing & Exploited Children

Transition Workshop Photograph provided by the family of Tammy Vincent

Technology



Utilization of Phenotype Informative SNPs to Provide Additional Information for Facial Reconstructions

Fort Myers, Florida – eight unidentified remains



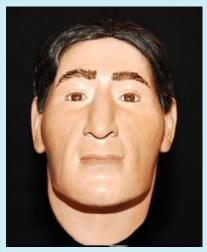
Images courtesy of the Fort Myers (Florida) Police Department and the University of North Texas Center for Human Identification

Technology Transition Workshop

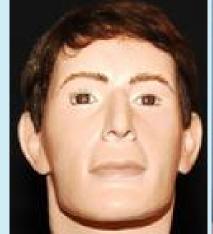


Utilization of Phenotype Informative SNPs to Provide Additional Information for Facial Reconstructions

Fort Myers, Florida — UNTCHI has identified two of the remains









John Blevins

Erik Kohler

Images courtesy of the Fort Myers (Florida) Police Department and the University of North Texas Center for Human Identification



Pharmacogenetic SNPs

- Molecular autopsy
 - Postmortem analysis to help resolve some cases initially believed to be suicides or classified as sudden unexplained deaths
 - Poisoning, incapacitation, inebriation, or certain diseases, such as epilepsy, depression, cardiac diseases or diabetes, where pharmacotherapy is an essential treatment, are factors in the cause of death
- Some people can metabolize a drug better or worse than others due to pharmocogenetic SNPs in or around specific encoded enzymes



Cytochrome p450s

- Cytochrome P450s (CYP) are a large group of monooxygenase enzymes responsible for the metabolism of numerous compounds
- The CYPs are a superfamily of enzymes, all of which contain a molecule of haem that is noncovalently bound to the polypeptide chain

For additional information regarding CYP, refer to: Goodman & Gilman's: The Pharmacological Basis of Therapeutics, 10th Ed, 2001, McGraw-Hill Professional, New York, NY.



Cytochrome p450 (CYP) 2D6

Human cytochrome P450 monooxygenase superfamily

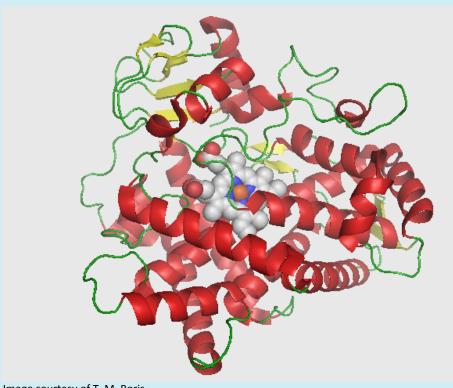


Image courtesy of T. M. Boris



Inhibitors of 2D6

http://medicine.iupui.edu/flockhart/table.htm

Strong

bupropion

fluoxetine

paroxetine

quinidine

cocaine

Moderate

duloxetine

terbinafine

Weak

amiodarone

cimetidine

sertraline

Unclassified

celecoxib

chlorpheniramine

chlorpromazine

cinacalcet

citalopram

clemastine

clomipramine

diphenhydramine

doxepin

doxorubicin

escitalopram

goldenseal

halofantrine

histamine H1 receptor

antagonists

hydroxyzine

levomepromazine

methadone

metoclopramide

mibefradil

midodrine

moclobemide

perphenazine

ranitidine

red-haloperidol

ritonavir

ticlopidine

tripelennamine



General Analytical Criteria for Forensic SNP Use

- Easily typed
- Multiplexing
- Highly informative for the stated purpose



Technologies / Methodologies for SNP Detection

- Hybridization/Chip
- Luminex bead/ flow cytometry
- SNaP Shot
- SNPstream UHT
- Pyrosequencing
- Mass Spectrometry
- OLA
- Etc.

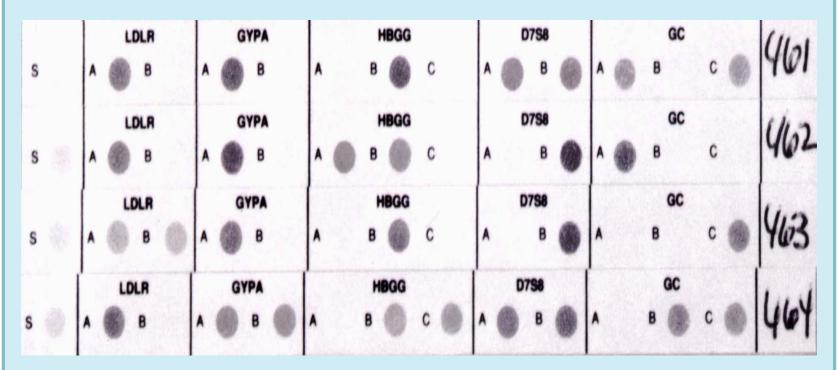


SSO Typing Methods

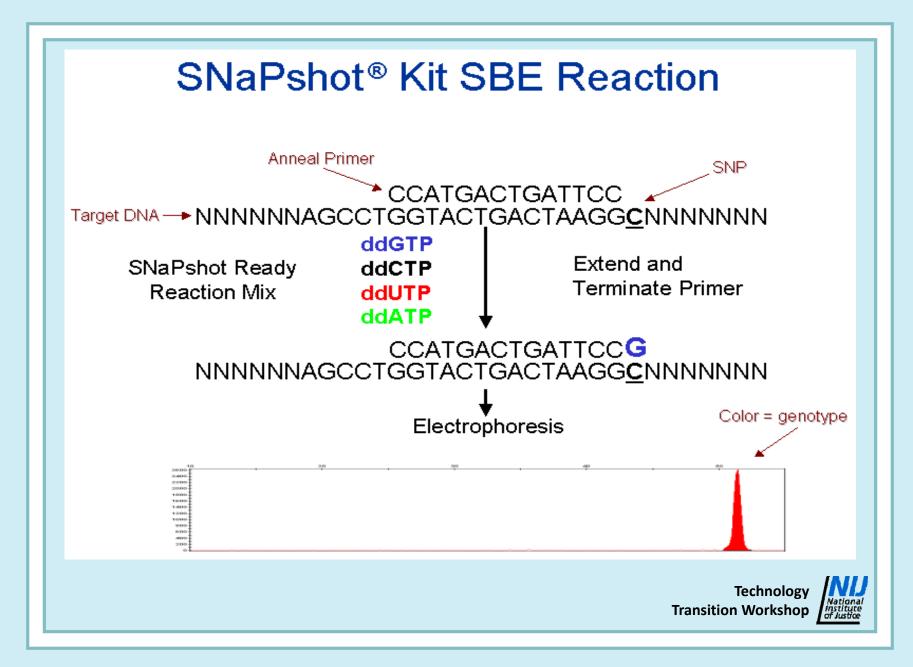
- Typing based on sequence differences (dots)
 - DQ alpha typing
 - Polymarker
 - mtDNA

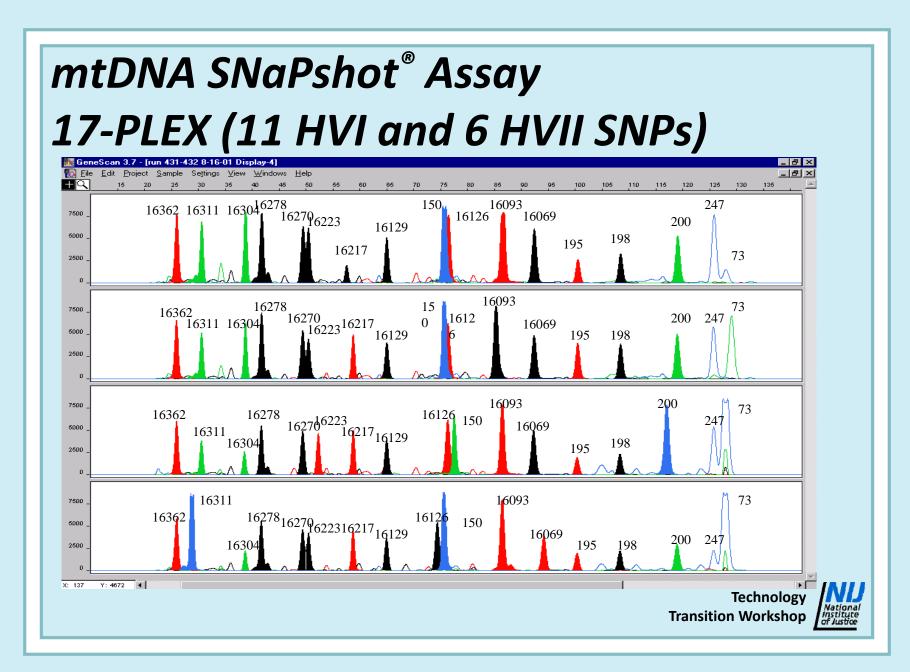


SNP Assay Hybridization Based HLA-DQA1 and Polymarker









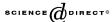
Electrospray Ionization MS



*<u>Triangulation ID</u> for <u>Genetic Evaluation of Risks</u> By IBIS



Available online at www.sciencedirect.com



Analytical Biochemistry 344 (2005) 53-69

ANALYTICAL BIOCHEMISTRY

www.elsevier.com/locate/yabio

Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: A novel tool for the identification and differentiation of humans

Thomas A. Hall ^a, Bruce Budowle ^b, Yun Jiang ^a, Lawrence Blyn ^a, Mark Eshoo ^a, Kristin A. Sannes-Lowery ^a, Rangarajan Sampath ^a, Jared J. Drader ^a, James C. Hannis ^a Patina Harrell ^a, Vivek Samant ^a, Neill White ^a, David J. Ecker ^a, Steven A. Hofstadler ^a,*

Hall, T. et al., Analytical Biochemistry (2005) 344 53-69



Electrospray Ionization Mass Spectrometry

- Formation of highly charged liquid droplets from which ions are desolvated / desorbed
- Generates multiple charge states of large analytes
 - Results in "folded-over" spectra, which can be recorded over narrower m/z range
- Very soft ionization technique
 - Applicable to labile molecules and noncovalent complexes
- High sensitivity
 - Applicable to analyte concentrations < 1 nM



Double stranded PCR product in solution

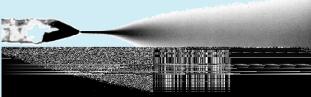


Image courtesy of Steven Hofstadler, Ph.D.



Species are detected as single strands

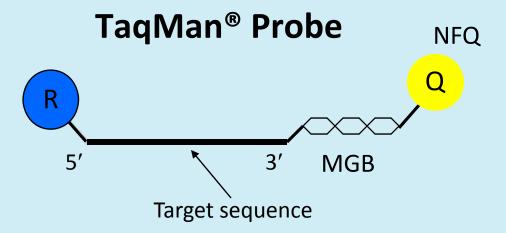


Mass Spectrometry

- No labeling
- Mass accuracy
- Multiplexing
- Mixture interpretation
- Automation



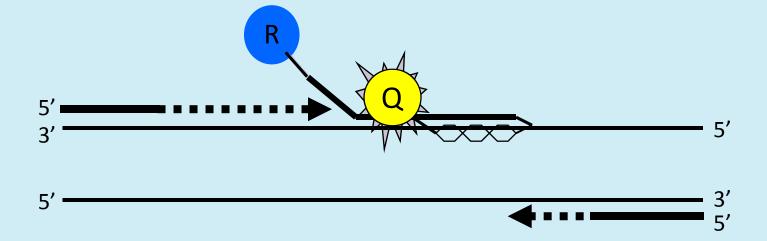
5' Nuclease Assay



- R = Reporter (FAM™ or VIC® Dyes)
- Q = Non-Fluorescent Quencher (NFQ)
 - Acts as energy transfer acceptor that does not emit a detectable fluorescent signal

5' Nuclease Assay **Forward** Hybridization Primer Reverse Primer **Excitation Technology Transition Workshop**

5' Nuclease Assay



Displacement



5' Nuclease Assay Cleavage **Technology Transition Workshop**

5' Nuclease Assay



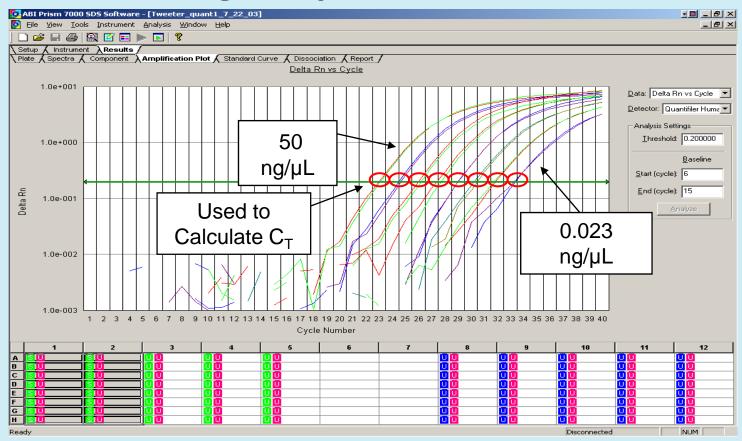




Polymerization completed



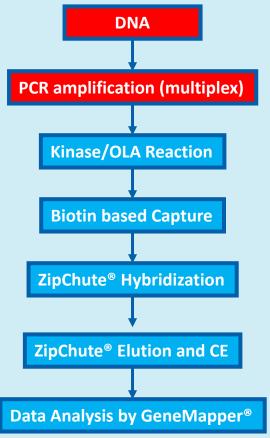
Amplification Plots for DNA Concentration Standards – Eight 3-fold Serial Dilutions

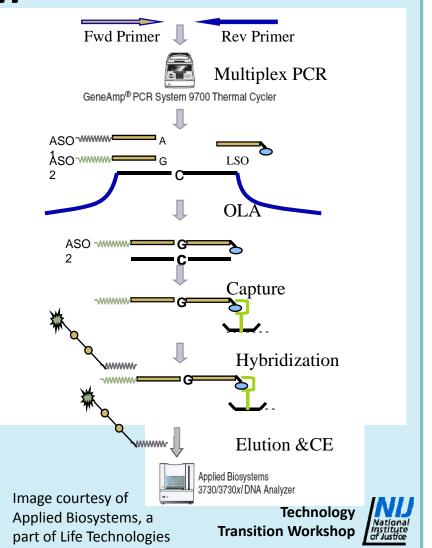




GenPlex™ HID System

The PCR-OLA Genotyping System





SNPs as Forensic Markers

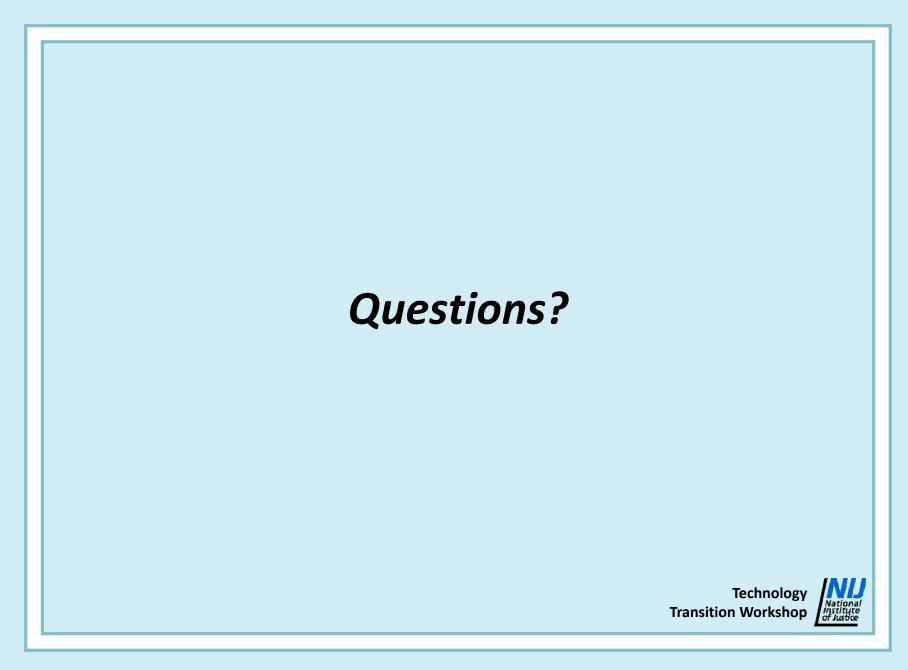
- Five classes of SNPs
- Abundant
- Low F_{st} and high F_{st}
- Identity SNPs will be primary focus for missing persons, other kinship applications, and CODIS
- Haploblock SNPs also for kinship analyses
- Phenotypic SNPs for investigative leads and facial reconstructions
- Pharmocogenetic SNPs for cause of death



SNPs as Forensic Markers

- Next Steps
- Select a consensus set(s)
- Design kits for platforms
- Establish collaborations
- Validation studies





Contact Information

Bruce Budowle, Ph.D.

bbudlowle@hsc.unt.edu

Images courtesy of Bruce Budowle, Ph.D. or his research assistants unless otherwise noted.

