NIJ National Institute of Justice

Technology Transition Workshop| *Ranajit Chakraborty, Ph.D.*

Evaluation of Genome-wide SNP Haplotype Blocks for Human Identification Applications

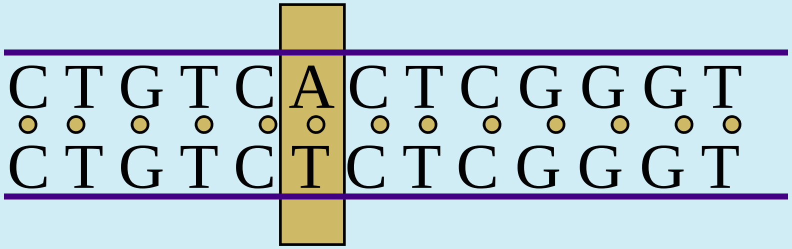
Overview

- Some brief remarks about SNPs
- Haploblock structure of SNPs in the human genome
- Criteria for selection of optimal SNP haploblocks for forensic applications
- Preliminary results of optimal parameter combinations from HapMap Data (Phase I and Phase II)
- Feasibility of SNP haploblock selection from human genome
- Strategies of interpretation of SNP haploblock-based forensic evidence
- Preliminary conclusions and future directions

Technology
Transition Workshop

NIJ National Institute of Justice

Single Nucleotide Polymorphism (SNP)



- Most SNPs are biallelic
- About three million SNPs in human genome (characterized)
- Provide more results from low quantity template DNA or degraded samples than STR typing
- Complete automation feasible
- Low mutation rates (10^{-8}/site/generation)
- Use of SNPs in forensics is not new (e.g., HLA-DQα)

Technology
Transition Workshop

NIJ National Institute of Justice

How Many SNPs Would be Needed for Forensic Applications?

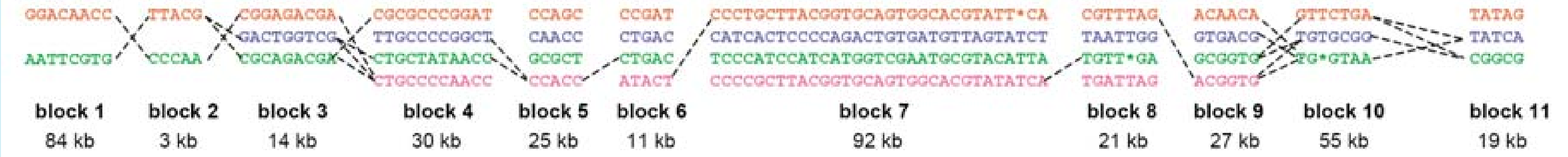
- **Answer depends upon allele frequencies at SNP sites, efficiency in different types of applications**
  - **For example, power of discrimination in identity testing; PE or PI in parentage analyses; LR in kinship assessment, etc.**
- **Chakraborty, et al. (1999, *Electrophoresis* 20: 1682-96) showed nomograms suggesting that the number of SNPs needed to equal the power of the current battery of STR loci would necessitate the use of several sets of syntenic SNPs**
  - **For example, SNPs residing on the same arm of several chromosomes**

Strategies for Improving Power of SNPs for Forensic Applications

- **Translate sets of SNPs into multiallelic markers**
- **Select a panel of SNP sets that satisfy conditions of the product rule**
  - **For example, statistically independent sets of SNPs**
- **Search for genome-wide availability of desired SNPs for feasibility of detection of such panels of SNPs**
- **Test the robustness of typing selected SNPs in forensic samples of compromised DNA quality**

Technology
Transition Workshop

Haplotype Block (Haploblock)



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

- **Linkage disequilibrium (LD): allelic association between two loci (for example, SNP sites)**
- **Closely linked SNPs with high LD → haplotype blocks**
- **Human genome is composed of block-like structures of low haplotype diversity (strong LD within block) separated by recombination hot spots**
- **Complete LD among *n* linked SNPs →(*n* + 1) haplotypes**

Technology
Transition Workshop

Advantages of Haploblock as Forensic Marker

- Can be typed in highly degraded samples
  - Where no results from STR analysis may be obtained
  - Improves the limited discrimination power of individual SNPs
- Haploblock can be considered as “pseudo STRs”
  - One haploblock → one “STR” locus
  - Different haplotypes → different “alleles”
- Each haplotype treated as a lineage marker like Y-chromosome and mtDNA
  - Exception – possible transmission from both parents following standard Mendelian principles

Technology
Transition Workshop

HapMap Project (www.hapmap.org)

- **Three major populations (90 Caucasian, 90 African, 45 Chinese and 45 Japanese)**
- **Phase II data: > 3,000,000, SNPs**
  - **LD information: D', r2**
  - **Phase information**
  - **Genotype information**

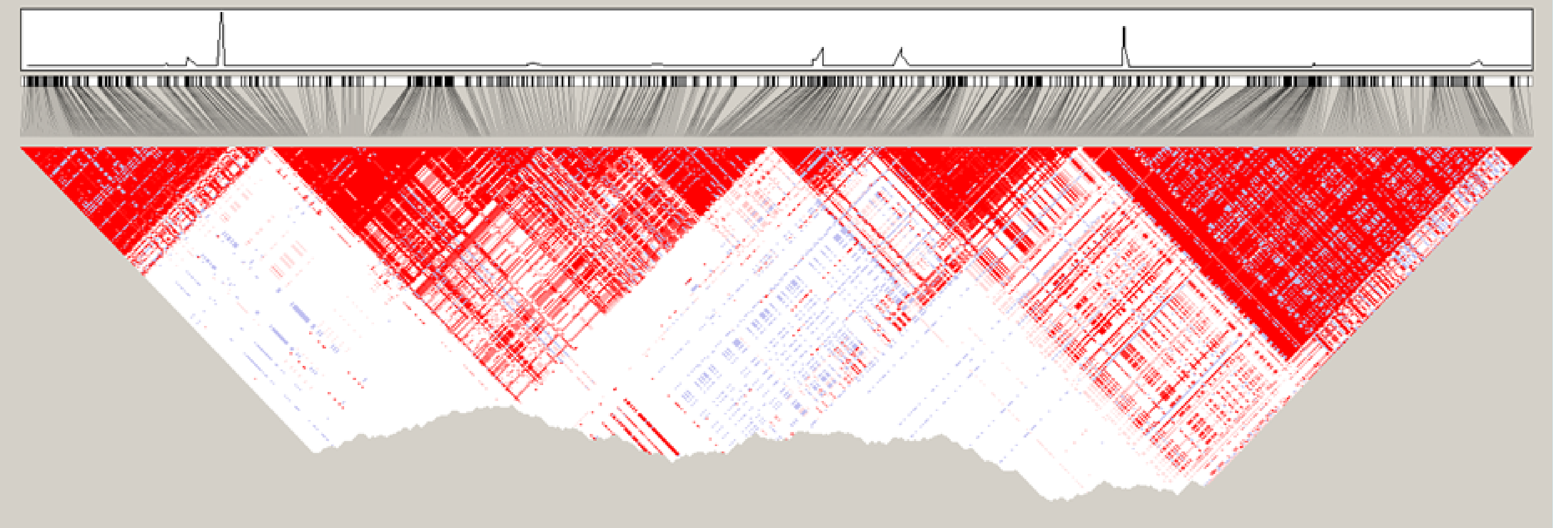


Image courtesy of http://hapmap.ncbi.nlm.nih.gov/

Technology Transition Workshop

NIJ National Institute of Justice

Haploblock Selection Criteria

- Exist in all major populations (Caucasian, East Asian, African)
- Higher discrimination power (for example, lower match probability) than that of the individual SNPs within the block
- Hardy-Weinberg Equilibrium for each block
- No significant LD between blocks
- Sufficient number of candidate haploblocks in the whole genome

Parameters Used in Selection

- **Maximum match probability reduction per haploblock (mmpr)**
- **Minimum LD between SNPs: r^2**
- **Population substructure: maximum F_{st}**
- **Minimum heterozygosity (MinHet)**
- **Minimum number of haplotypes in each population (MinHap)**
- **Minimum number of SNPs per haploblock (MinSNP)**

Technology Transition Workshop

NIJ National Institute of Justice

Best Parameter Set

- mmpr = 0.85
- r^2 = 0.7
  - No haploblock found with $r^2 \geq 0.8$
- F_{st} = 0.06
- MinHet = 0.2
- MinHap = 3
- MinSNP = 3

The best thresholds of parameters other than r^2 found on Chr1

Technology
Transition Workshop

Haploblocks with Best Parameter Set

Chromo -some	Num. blocks with PS	Num. blocks with PS & HWE	Num. blocks with PS & HWE & LD filters (*n*)	Avg. Cum. MP of blocks (*b*)	Cum. Min. MP of SNPs (*s*)	MP reduction per block (*mpr*)	Num. Of SNPs
1	9	9	0				
2	23	14	1	0.3287	0.4050	0.8117	6
3	12	10	2	0.1144	0.1617	0.8412	9
4	21	15	1	0.2926	0.3765	0.7773	6
5	16	12	3	0.02633	0.05480	0.7833	25
6	15	10	0				
7	16	9	2	0.1035	0.1465	0.8403	30
8	18	12	2	0.1025	0.1518	0.8215	7
9	8	6	0				
10	15	8	1	0.3527	0.4169	0.8460	4
11	14	12	3	0.03872	0.06700	0.8209	13
12	12	5	1	0.3036	0.3890	0.7806	5
13	17	14	3	0.0344	0.06409	0.8123	14
14	10	6	3	0.02339	0.04789	0.7876	11
15	9	4	0				
16	7	4	1	0.3310	0.4053	0.8167	3
17	5	4	0				
18	8	7	1	0.3123	0.3689	0.8465	5
19	5	4	0				
20	6	1	0				
21	6	3	0				
22	1	1	0				
Total	253	170	24	1.059E-12	1.566E-10	0.8121	138

Technology
Transition Workshop

NIJ National Institute of Justice

Haploblock Example – Chr2

Haplotype Frequencies

Haplotype	CEU	JPT+CHB	YRI
010011	0	0	0.0417
011000	0.0083	0	0.0417
001000	0.3417	0.5167	0.4833
000111	0	0.0056	0
110010	0	0.0056	0
111000	0	0	0.0167
110111	0.5833	0.4611	0.4167
101000	0.0083	0	0
110110	0.05	0.0056	0
110100	0.0083	0.0056	0

Num. SNPs = 6
Num. haplotypes = 10
Avg. Het. = 0.5499
MP of block = 0.3287
Min. MP of SNPs = 0.4050
MP reduction = 0.8117
F_{st} = 0.024

Technology
Transition Workshop

NIJ National Institute of Justice

Different Haploblock Structure Among Populations

- $r^2 = 0.7$ and MinSNP = 3
  - 11,741 haploblocks in Caucasian
  - 12,456 haploblocks in Chinese
  - 12,237 haploblocks in Japanese
  - 7,318 haploblocks in African
- Population-specific haploblock selection criteria may be necessary to obtain best performing systems

Evidence Interpretation Based on Haploblocks

- Transfer evidence
- Mixture interpretation
- Kinship analysis

One genotype → …(A/T)(A/T)…

Two possible haplotype combinations → TT+AA
TA+AT

Transfer Evidence

- Compared a single source profile from crime scene evidence with profile of the suspect
- Exclusion or inclusion → compare the genotypes
- If inclusion, random match probability is:

$$\Pr(G) = \sum_{\substack{\text{Haplotype combination} \\ (H_i, H_j) \text{ composes } G}} p_i p_j$$

- Mixture versus single source sample

Technology
Transition Workshop

NIJ National Institute of Justice

Transfer Evidence – Example

Haplotype	Frequency
TT	0.4
TA	0.3
AA	0.2
AT	0.1

Genotype **. . . (A/T)(A/T) . . .**

$$\textbf{Match Probability} = \begin{cases} \textbf{TT/AA:} & 2 \times 0.4 \times 0.2 = 0.16 \\ \textbf{TA/AT:} & 2 \times 0.3 \times 0.1 = 0.06 \end{cases}$$

= 0.22

Technology
Transition Workshop

NIJ National Institute of Justice

Mixture Detection

- **Multiple contributors → at least four haplotypes**
- **The probability of a genotype (G):**

$$\Pr(G) = \sum_{\substack{Haplotype\ combination \\ (H_k,\ \ldots\ ,\ H_l)\ composes\ G}} \prod_{i=k}^{l} p_i$$

- **The probability of a genotype (G) given number of contributors (N)**

$$\Pr(G \mid N = 1) = \sum_{\substack{Haplotype\ combination \\ (H_i,\ H_j)\ composes\ G}} p_i p_j$$

$$\Pr(G \mid N = 2) = \sum_{\substack{Haplotype\ combination \\ (H_i,\ H_j,\ H_k,\ H_l)\ composes\ G}} p_i p_j p_k p_l$$

$$\Pr(G \mid N = 3) = \sum_{\substack{Haplotype\ combination \\ (H_i,\ H_j,\ H_k,\ H_l,\ Hm,\ H_n)\ composes\ G}} p_i p_j p_k p_l p_m p_n$$

Technology
Transition Workshop

Exclusion Probability and Likelihood Ratio for Mixture Analysis

- **Probability of exclusion (PE)**

$$PE = 1 - \left(\sum_{H_i} p_i \right)^2 ,$$ **where Σ is over all H_i's that are contributors to *G***

- **Likelihood ratio (LR): S is suspect; V is victim; UN is an unknown contributor**

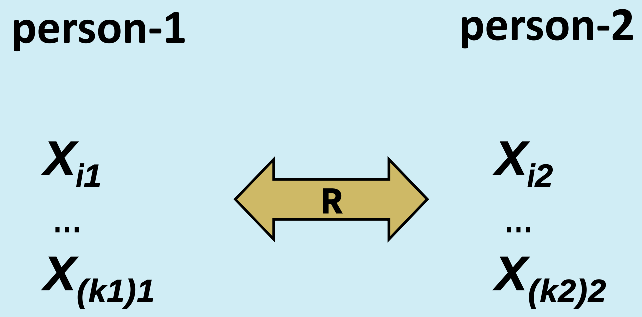
$$LR = \frac{\Pr(V + S)}{\Pr(V + UN)}$$

Technology
Transition Workshop

Pairwise Kinship Analysis

- **One genotype (G) has *k* haplotype combinations; $X_i = (H_{i1}, H_{i2})$ is i-th combination, with likelihood $P(X_i)$; w_i as the weight of X_i**

$$w_i = P(X_i) / \sum_{i=1}^{K} P(X_i)$$



Likelihood of these two persons given relationship (*R*):

$$L_{Block} = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} w_{i1} w_{j2} L(X_{i1}, X_{j2} \mid R)$$

Conclusions

- This is the first effort to assess the feasibility of genome-wide SNP haploblock structures for human identity testing encompassing all major forensic applications
- SNP haploblocks provide an alternative approach for forensic investigations, especially for highly degraded samples
- Haploblock selection depends on multiple criteria
- Consideration is needed for evidence interpretation based on haploblock results, because of multiple haplotype combinations that are possible for observed genotypes

Future Directions

- Portability/universality of efficient haploblocks to be tested with wider sets of genome data
- Alternatively, population-group specific panels of haploblocks have to be determined with validation data from anthropologically defined populations
- Robustness of genotyping in samples with compromised DNA quality (mimicking forensic samples) has to be tested

Technology
Transition Workshop

NIJ National Institute of Justice

Acknowledgements

- This work was jointly done with Dr. Jianye Ge, Dr. Huifeng Xi, Dr. Bruce Budowle, and Dr. John Plantz, who are co-authors of the manuscript to be submitted for publication
- Research for the work was partially funded by grants and contracts from the US National Institutes of Health and US National Institute of Justice

Technology
Transition Workshop

Questions?

Technology
Transition Workshop

NIJ
National
Institute
of Justice

Contact Information

Ranajit Chakraborty, Ph.D.

Professor, Dept. Environmental Health

University of Cincinnati College of Medicine

3223 Eden Avenue, Room K-108

Cincinnati, OH 45267-0056

Tel. (513) 558-4925; Fax (513) 558-4397

E-mail: ranajit.chakraborty@uc.edu