# Characterization of 114 insertion/deletion (INDEL) polymorphisms, and selec- tion for a global INDEL panel for human identification

**15 authors**, including:

Bobby Larue

University of North Texas HSC at Fort Worth

30 PUBLICATIONS 201 CITATIONS

SEE PROFILE

Jianye Ge

University of North Texas HSC at Fort Worth

56 PUBLICATIONS 515 CITATIONS

SEE PROFILE

Jonathan L King

University of North Texas HSC at Fort Worth

61 PUBLICATIONS 323 CITATIONS

SEE PROFILE

Ranajit Chakraborty

University of North Texas HSC at Fort Worth

203 PUBLICATIONS 6,459 CITATIONS

SEE PROFILE

# Accepted Manuscript

## Characterization of 114 insertion/deletion (INDEL) polymorphisms, and selection for a global INDEL panel for human identification

Bobby L. LaRue, Robert Lagacé, Chien-Wei Chang, Allison Holt, Lori Hennessy, Jianye Ge, Jonathan L. King, Ranajit Chakraborty, Bruce Budowle

PII: S1344-6223(13)00122-3
DOI: http://dx.doi.org/10.1016/j.legalmed.2013.10.006
Reference: LEGMED 1072

To appear in: *Legal Medicine*

Received Date: 8 February 2013
Revised Date: 19 August 2013
Accepted Date: 22 October 2013

# Characterization of 114 Insertion/Deletion (INDEL) Polymorphisms, and Selection for a Global INDEL Panel for Human Identification.

Bobby L. LaRue[a]* Bobby.larue@unthsc.edu

Robert Lagacé[b] Robert.lagace@lifetech.com

Chien-Wei Chang[b] Chien-wei.chang@lifetech.com

Allison Holt[b] Allison.holt@lifetech.com

Lori Hennessy[b] Lori.hennesy@lifetech.com

Jianye Ge[a] Jianye.ge@unthsc.edu

Jonathan L. King[a] Jonathan.king@unthsc.edu

Ranajit Chakraborty[a] Ranajit.chakraborty@unthsc.edu

Bruce Budowle[a,c] Bruce.budowle@unthsc.edu

[a] Institute of Applied Genetics, Department of Forensic and Investigative Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Forth Worth, Texas, 76107, USA

[b] Life Technologies, 850 Lincoln Centre Drive, Foster City, California, 94404, USA

[c] Center of Excellence in Genomic Medicine (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

**Keywords**

INDEL, Human genotyping, Identity testing, Degraded DNA, SNP, STR, Population Genetics

**Abstract**

Bi-Allelic Insertions and Deletions (INDELs) are a powerful set of genetic markers for Human Identification (HID). They have certain desirable features, such as low mutation rates, no stutter, and potentially small amplicon sizes that could prove effective in some circumstances. In this study, we analyzed the distribution of 114 INDELs in four North American populations (Caucasian, African American, Southwest Hispanic, and Asian) to estimate their distribution in major global populations. Of the 114 INDELs a primary panel of 38 candidate markers was selected that met the criteria of 1) a minimum allele frequency of greater than 0.20 across the populations studied; 2) general concordance with Hardy-Weinberg equilibrium (HWE) expectations; 3) relatively low $F_{ST}$ based on the major populations; 4) physical distance between markers greater than 40 Mbp; and 5) a lack of linkage disequilibria between syntenic markers. Additionally, another 11 supplemental markers were selected for an expanded panel of 49 markers which met the above criteria, with the exception that they are separated at least by 20 Mbp. The resulting panels had Random Match Probabilities that were at least $10^{-16}$ and $10^{-19}$, respectively, and combined $F_{ST}$ values of approximately 0.02. Given these findings, these INDELs should be useful for HID.

## 1. Introduction

Small bi-allelic insertion and deletion (INDEL) markers have generated interest for human identification (HID) as an adjunct or viable alternative to short tandem repeat (STR) or single nucleotide polymorphism (SNP)-based approaches [1-12]. Various HID panels utilizing INDELs have been developed and described [2-10, 12]. To augment these existing panels, it would be desirable to seek INDELs that apply well to HID on a more global basis which demonstrate high discrimination power and low inter-population diversity (e.g., low $F_{ST}$).

In the study herein, genotype and allele frequency distributions were generated for 114 candidate INDELs in four major population groups (Caucasian, African, Asian, and Southwest Hispanic) from North America. Criteria were set to select those INDELs that would be best suited for HID. Two subpanels of INDELs (a primary and a secondary set) were derived from the 114 markers that may be well-suited for use in a global INDEL panel for HID.

## 2. Materials and Methods

### 2.1 Marker Selection

INDEL candidates were selected from NCBI using NCBI's dbSNP [13] search web page (http://www.ncbi.nlm.nih.gov/SNP/). The following criteria were used to select INDELs from dbSNP 132:

*Organism: Homo sapiens*
*Chromosomes: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, x, y*
*Function Class: intron*
*SNP Class: in-del*
*Validation Status: by-cluster, by-frequency, by-2hit-allele*
*Heterozygosity: 0-10, 10-20, 20-30, 30-40, & 40-50*

The resulting XML file was parsed and filtered (in-house computer programs were written in PERL to process NCBI's data files) for INDELs of length three or more, validated by a method other than computed and were designated as unique. Population data for African, African American, American Indian, Asian, Chinese, European, Hispanic, Japanese, and Sub-Saharan African populations were gathered from files downloaded from NCBI's dbSNP ftp site and INDELs with a minor allele frequency of greater than or equal to 0.20 were selected.

The candidate INDELs were characterized by analyzing the INDEL and its surrounding genomic DNA using the program mreps [14]. INDELs that were shown to have a repetitive element, those where the INDEL sequence was seen to repeat 2.5 times or more, were excluded as being possible STRs and originating from a method other than an insertion or deletion[15-19]. INDELs of four and five nucleotides were given priority for integrating into a multiplex and the candidates with the highest minor allele frequencies were tested against an internal panel of population samples.

## 2. 2 Primer Design and Preliminary Multiplex Optimization

Four multiplex PCR assays for the detection of a total of 114 INDEL loci were developed using the primer design function of the Primer3 software [20], and the five-dye technology from Applied Biosystems. Simultaneous amplification of the INDEL markers was performed on the GeneAmp® PCR Systems 9700 in a reaction volume of 25 µl using 0.2 µM concentration of each primer and 1x AmpFlSTR® Identifiler® Direct Master Mix supplemented with 17.5 nmol $MgCl_2$, 18 nmol dNTP, and 8.1 U AmpliTaq Gold® enzyme.

## 2. 3 Samples

Buccal swabs from unrelated individuals (80 African Americans, 85 Asians, 98 Caucasians, and 86 Southwestern Hispanics) residing in the United States were kindly provided by Genetic Testing Laboratories (Las Cruces, NM). The samples were collected and anonymized in accordance with methods approved by the Institutional Review Board for the University of North Texas Health Science Center in Fort Worth, Texas. Additional anonymous, unrelated human samples were obtained from the University of California at San Francisco or purchased as whole blood from the Interstate Blood Bank, Inc. (Memphis, TN) or Boca Biolistics (Coconut Creek, FL) (166 African Americans, 202 Asians, 166 Caucasians, and 167 Southwestern Hispanics).

## 2.4 Isolation of DNA and preparation of samples for analysis

DNA was isolated from buccal swabs using either the AutoMate Express® (Life Technologies, Carlsbad, CA) or the QiaAMP DNA Investigator® Kit (Qiagen, Hilden, Germany) according to the manufacturers' recommendations. The blood samples were purified on an Applied Biosystems 6100 Nucleic Acid Prep Station (Life Technologies). The quantity of DNA was determined by qPCR using the Quantifiler® Quantification Kit and 7500 Real-Time PCR® System (Life Technologies). Samples were normalized to 500 pg/μL and stored at either −20°C or −40°C until amplification.

## 2.5 Amplification and Analysis of the 114 INDELs

Samples containing 500 pg of DNA were analyzed. Each of four preliminary multiplexed primer sets using a Geneamp 9700 (Life Technologies) were amplified with an initial step at 95$^{o}$C for 11 minutes followed by 28 cycles of 20 seconds at 94$^{o}$C for denaturation and 3 minutes at 59$^{o}$C for annealing/extension. A final extension step of 60$^{o}$C for 60 minutes was employed to promote terminal adenylation.

Each sample was prepared immediately prior to electrophoretic analysis and run on a 3500xl Genetic Analyzer® (Life Technologies) with an injection time of 10 seconds and an injection voltage of 3kV. Electrophoretic data were analyzed using Genemapper IDX® (Life Technologies).

### 2.6 Statistical analyses

Allele frequencies were determined by the gene counting method. Population genetic parameters were analyzed by using either Genetic Data Analysis software [21-22] or in-house developed software. Departures from Hardy–Weinberg equilibrium (HWE) and linkage equilibrium were tested using Fisher's exact test. Bonferroni correction for multiple comparisons and population substructure parameter ($F_{ST}$) was estimated by the methods described in Weir and Cockerham [21, 23-24].

## 3. Results and Discussion

### 3.1 Location and Description of the Markers

The 114 INDELs reside in non-coding regions and are distributed among the non-coding regions of chromosomes 1 through 22. The size of the insertion ranged from two to nine nucleotides in all populations assayed (Table 1).  Sample electropherograms of these four preliminary multiplexes are shown in Supplementary Figures 1-4. While an initial criterion was to select indels with at least 3 bp in size for the polymorphism for long term multiplex design, a few dinucleotide indels were included as they were reported in the 38-plex by Pereira et al.[10]. Although not a final construct for a validated multiplex, the amplicons of all INDELs were less than 180bp. Small size amplicons which tend to be more robust for a PCR could be more effective for analysis of degraded DNA samples. The size of each amplicon, although, is not set as the purpose of this study was to determine the subset of indels that would be well-suited for HID; once selected the primers can be redesigned to generate smaller length amplicons.

### 3.2 Population Data

The 114 INDELs were typed in four major populations: Asian (n=287), Southwest Hispanic (n=253), Caucasian (n=264), and African American (n=246). All loci were polymorphic (Table1). Three loci, I-15, I-43, and I-93 displayed departures from Hardy-Weinberg equilibrium in two or more populations and, thus, were not considered for further analyses in this study. For the remaining 111 markers, the numbers of departures from HWE expectations were 9, 4, 2, and 3 in Asian, Southwest Hispanic, Caucasian, and African American populations, respectively. This number of departures is consistent with the number of departures expected by chance (i.e., 5%), except in the Asian population. One explanation for the larger number of departures from HWE in Asians may be that diverse subpopulations might be included in the sample from this group. More studies with subpopulations may provide a better indication for the cause of these departures in the Asian sample population. However, when corrected for multiple comparisons (via the bonferroni correction), none of the 111 INDELs departed significantly from HWE in any of the four populations (Table 1).

Testing for linkage disequilibrium (LD) was performed using Fisher's exact test, with 10000 shufflings [25]. With 111 INDELS there were 6105 pairwise comparisons performed per population sample. A total of 928 (15.2%), 245 (4.0%), 457 (7.5%), and 204 (3.3%) pairs displayed detectable LD at the 0.05 level in the Asian, African American, Southwest Hispanic, and Caucasian populations, respectively. The percentage of pairs displaying significant LD that were observed in African American and Caucasian were fewer than the expected number by random chance (approximately 305 of the 6105 tests per population). However, the number of significant LDs in Hispanic and Asian populations was greater than expected by chance alone.

Upon closer examination, there were 21, 7, 20, and 6 syntenic loci pairs (i.e., only those on the same arm of a chromosome) out of a total of 180 syntenic comparisons in the Asian, African American, Southwest Hispanic, and Caucasian populations, respectively, that displayed significant LDs. Once again, the number of pairs for the African American and Caucasian were fewer and the Hispanic and Asian populations were greater than would be expected due to random chance alone (i.e., approximately 9 pairs).

Among non-syntenic pairs, LD was observed in 907, 238, 437, and 198 pairs out of a total of 5925 comparisons in the Asian, African American, Southwestern Hispanic, and Caucasian populations, respectively. The number of pairs displaying LD would be expected to be approximately 290 pairs if the departures were attributable to chance alone. The same trends were observed as in the overall and syntenic pairs (i.e., fewer than expected in Caucasian and African American, and greater than expected in Hispanic and Asian populations) (Supplementary Table 1).

One plausible explanation for these departures, as stated above could be the construct of the Hispanic and Asian sample populations studied. Another explanation is that the greater than expected numbers of pairs exhibiting LD could be associated with loci which showed departures from HWE, as previously described in the literature [25]. The Asian and Southwestern Hispanic populations had 9 and 4 such loci, respectively, with departures from HWE at a 0.05 level of significance. While not meeting the HWE criterion for elimination (i.e., departing from HWE at 0.05 level in more than one population), these loci may have distorted the LD analysis, exhibiting apparent linkage with other loci in a greater number of instances than would be expected due to chance alone. For example, among non-syntenic loci pairs, the same loci that showed slight departures from HWE were overrepresented as exhibiting LD in comparison to what would be attributable to random chance alone. 220 of 437 (50.3%) pairs with significant LD in Hispanics and 531 of 907 (58.5%) pairs in Asians contained at least one locus that had a departure from HWE. These loci represented 3.6% and 8.1% of the total markers analyzed for LD in Hispanics and Asians, respectively. The fact that these loci are overrepresented in loci pairs demonstrating LD lends supports that these loci may be distorting the LD analysis. In a similar fashion, these same loci accounted for 12 of 20 (60.0%) Hispanic and 11 of 21 (52.4%) Asian syntenic loci pairs exhibiting LD. They represented 16% in Hispanic and 36% of the Asian loci involved in pairs exhibiting LD. If these loci were removed from the LD analysis, the Hispanic and Asian syntenic pairs exhibiting LD would be either slightly lower (Hispanic) or much closer (Asian) to what would be expected due to chance alone (i.e., approximately 9 pairs). These observations further support that these loci may have distorted the LD analysis. When corrected for multiple comparisons (via the Bonferroni correction) [23-24], however, only one of the pairs of syntenic loci (markers I-113 and I-114 in the Southwest Hispanics and about 10 Mbp distant) (Supplemental Table 1). and 19 non-syntenic pairs in the Asian population still demonstrated significant LD (Supplemental Table 2)

To determine the effects of substructure among the four tested major population groups, Wright's $F_{ST}$ was estimated [24]. The global $F_{ST}$ value for the set of 111 INDELs was 0.06. Some markers, such as I-51, I-64, I-79, I-92 and I-109 had individual $F_{ST}$ values greater than 0.20 and thus contributed to elevating the overall $F_{ST}$ value (Table 1). Clearly a subset of the 111 INDELs could be selected that would display a much lower overall $F_{ST}$ and be a desirable candidate panel for HID (see below).

Using the four major population groups to derive a $F_{ST}$ value provided an indication of an upper bound of the effects of population substructure. For HID purposes the degree of substructure within a major population may have more practical application. Since the major population group samples were collected from two geographically distinct areas, substructure within the United States, $F_{ST}$ for geographically different populations was estimated. The overall $F_{ST}$ for each major population group was approximately $6.57 x 10^{-4}$, $1.0 x 10^{-5}$, $3.52 x 10^{-3}$, and $1.65 x 10^{-4}$ in the Asian, African American, Southwest Hispanic, and Caucasian populations, respectively. These data indicate that the effects of substructure within a major United States population group may be nominal. More subgroup data from within major population groups from around the world would be necessary to define better the effects.

The cumulative random match probability (RMP) for all 111 INDELs assuming independence and no effects of substructure approached $10^{-42}$ in all populations. The RMP is provided as a guide only (Table 1).This could be an overestimation of the RMP as the assumption of independence may not hold for all loci.

Selecting a robust HID candidate INDEL panel is desirable. This panel should be one that is effective across major populations and thus should exhibit low effects of substructure. With low substructure effects not as many population databases may need to be generated for use across the HID laboratories and a maximized discrimination power can be obtained. In addition, those pairs of loci that do not demonstrate detectable LD or are sufficiently separated physically on the chromosomes are desirable for simplifying estimation of the RMP. To identify a set of INDELs that could be included in a potential panel the following criteria were used:  minor allele frequencies greater than 0.20 in all four populations; $F_{ST}$ values per locus approximately or less than 0.06 (similarly set for SNPs by Kidd et al [11, 26]); physical distance greater than 40 Mbp between markers or for a larger alternative set that includes the 40 Mbp set and additional INDELs that are at least 20 Mbp distant on the same chromosome.

Given that Pereira et al. and others [2-3, 6, 8, 10] already described a multiplex INDEL panel, some of these markers were given preference for compatibility or data sharing if the INDEL met the above criteria in all four sample populations when a similarly performing alternative INDEL was less than 40 Mbp or 20 Mbp for each panel set. The frequency of alleles observed at each locus in the individual populations generally were similar between the same population groups described herein and those described by Fondevilla et al [3]. The few discrepancies observed were in the US Asian populations. Again a likely reason is that the broad category of Asian samples may be composed of notably different subpopulations; further studies are needed with better defined Asian population categories.

16 of the markers from Pereira et al [10] met the above criteria and were included in the initial panel of INDELs separated by at least 40 Mbp (Table 1). The primary panel that met the selection criteria contains 38 INDELs (Table 1). The RMP assuming independence approached $10^{-16}$ for each population group. The overall $F_{ST}$ value for this primary panel was approximately 0.023 which was less than those from the Pereira et al [10] (Table 1).

If the physical distance criterion was relaxed to approximately 20 Mbp, the number of INDELs included in the secondary panel increased to 49. The RMP for the secondary panel, assuming independence, increased to $10^{-19}$ and the overall $F_{ST}$ value was similar to that of the primary panel (Table 1).

## 4. Conclusions

Using the criteria of HWE, allele frequency distribution, physical location, population substructure, lack of detectable LD, and conformity with the assumption of mutual independence, a candidate INDEL panel set of 38 or 49 markers (the latter if the physical distance criterion is relaxed) has been identified. The $F_{ST}$ value across these major populations is relatively low (i.e., 0.023) and will be lower if calculated for each population instead of combining the major population groups. More subpopulation data are needed to define better major population-specific $F_{ST}$ values. These INDELs should be good candidates for development of an HID panel.

## 5. Acknowledgements

This work was funded in part by the Intelligence Community Postdoctoral Research Fellowship program.

## 6. Conflict of Interest

R. Lagacé, C-W Chang, A. Holt, and L. Hennessy were/are employees of Life Technologies, Foster City, CA.  BL. LaRue, J. Ge, JL. King, R. Chakraborty, and B. Budowle have no conflict of interest.

## 7. Protection of Human Subjects

All protocols have been approved by the UNTHSC Institutional Review Board to ensure the ethical protection of human subjects.

## 8. References

1. Budowle B, van Daal A. Forensically relevant SNP classes. BioTechniques2008.
2. Edelmann J, Hering S, Augustin C, Szibor R. Indel polymorphisms--An additional set of markers on the X-chromosome. Forensic Science International: Genetics Supplement Series. [doi: 10.1016/j.fsigss.2009.08.148]. 2009;2(1):510-2.
3. Fondevila M, Phillips C, Santos C, Pereira R, Gusmão L, Carracedo A, Butler JM, Lareu MV, Vallone PM. Forensic performance of two insertion–deletion marker assays. International journal of legal medicine2012 2012/09/01;126(5):725-37.
4. Francez PAdC, Ribeiro-Rodrigues EM, dos Santos SEB. Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon. Forensic Science International: Genetics2012;6(1):132-5.
5. Friis SL, Børsting C, Rockenbauer E, Poulsen L, Fredslund SF, Tomas C, Morling N. Typing of 30 insertion/deletions in Danes using the first commercial indel kit—Mentype® DIPplex. Forensic Science International: Genetics. [doi: 10.1016/j.fsigen.2011.08.002]. (0).
6. LaRue BL, Ge J, King JL, Budowle B. A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification. International Journal of Legal Medicine2012:1-8.
7. Li C, Zhao S, Zhang S, Li L, Liu Y, Chen J, Xue J. Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population. Forensic Science International: Genetics. [doi: 10.1016/j.fsigen.2010.03.004]. 2011;5(1):e27-e30.
8. Pereira R, Phillips C, Alves C, Amorim A, Carracedo Á, Gusmão L. Insertion/deletion polymorphisms: A multiplex assay and forensic applications. Forensic Science International: Genetics Supplement Series. [doi: 10.1016/j.fsigss.2009.09.005]. 2009;2(1):513-5.
9. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G. Human Diallelic Insertion/Deletion Polymorphisms. The American Journal of Human Genetics. [doi: 10.1086/342727]. 2002;71(4):854-62.
10. Pereira R, Phillips C, Alves C, Amorim A, Carracedo Á, Gusmão L. A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis2009;30(21):3682-90.
11. Pakstis A, Speed W, Kidd J, Kidd K. Candidate SNPs for a universal individual identification panel. Human Genetics2007;121(3):305-17.
12. Pena HB, Pena SDJ. Automated genotyping of a highly informative panel of 40 short insertion-deletion polymorphisms resolved in polyacrylamide gels for forensic identification and kinship analysis. Transfusion Medicine and Hemotherapy2012;39(3):211.

13. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research2001 January 1, 2001;29(1):308-11.

14. Kolpakov R, Bana G, Kucherov G. mreps: efficient and flexible detection of tandem repeats in DNA. Nucleic Acids Research2003;31(13):3672-8.

15. Ndifon W, Nkwanta A, Hill D. Identifying Nonrandom Occurrences of Simple Sequence Repeats in Genomic DNA Sequences. ETHNICITY AND DISEASE2005;15(4):5.

16. Leclercq S, Rivals E, Jarne P. DNA slippage occurs at microsatellite loci without minimal threshold length in humans: a comparative genomic approach. Genome Biology and Evolution2010;2:325.

17. Dieringer D, Schlötterer C. Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. Genome Research2003;13(10):2242-51.

18. Messer PW, Arndt PF. The majority of recent short DNA insertions in the human genome are tandem duplications. Molecular Biology and Evolution2007;24(5):1190-7.

19. Amos W. Mutation biases and mutation rate variation around very short human microsatellites revealed by human–chimpanzee–orangutan genomic sequence alignments. Journal of Molecular Evolution2010;71(3):192-201.

20. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol2000;132(3):365-86.

21. Lewis P, Zaykin D. GDA: software for the analysis of discrete genetic data. Free computer program distributed by the authors at: http://hydrodictyoneebuconnedu/people/plewis/softwarephp1999.

22. Weir BS. Genetic data analysis II: Sinauer Associates; 1996.

23. Dunn OJ. Multiple comparisons among means. Journal of the American Statistical Association1961;56(293):52-64.

24. Weir B, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution1984:1358-70.

25. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics2003;164(4):1567.

26. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SLB, Karoma NJ, Kungulilo S, Kim J-J, Lu R-B, Odunsi A, Okonofua F, Parnas J, Schulz LO, Zhukova OV, Kidd JR. Developing a SNP panel for forensic identification of individuals. Forensic Science International. [doi: 10.1016/j.forsciint.2005.11.017]. 2006;164(1):20-32.

Supplementary Figure 1. A sample electropherogram of preliminary multiplex assay 1.

375 Supplementary Figure 2. A sample electropherogram of preliminary multiplex assay 2.

376 Supplementary Figure 3. A sample electropherogram of preliminary multiplex assay 3.

Supplementary Figure 4. A sample electropherogram of preliminary multiplex assay 4.

**Table 1. A Description, Location, and Distribution of 114 Small INDELs In Four North American Populations**

| | | | | | **Asian** *(n=287)* | | | | **Southwestern Hispanic** *(n=253)* | | | | **Caucasian** *(n=264)* | | | | **African American** *(n=246)* | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Marker | RS Number[a] | Alleles[a] | Chr[a] | Location[a] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | $F_{ST}$[d] |
| I-1[f,g] | 4646006 | -/CTCA | 1 | 15845022 | 0.3739 | 0.4435 | 0.4906 | 0.3924 | 0.5456 | 0.4606 | 0.3016 | 0.3771 | 0.4494 | 0.5253 | 0.3791 | 0.3776 | 0.2913 | 0.4008 | 0.6418 | 0.4299 | 0.046 |
| I-2[f,g] | 13447508 | -/CTTAGA | 1 | 91977954 | 0.4087 | 0.3913 | 0.0046* | 0.3838 | 0.3498 | 0.4609 | 0.8845 | 0.4006 | 0.2763 | 0.4436 | 0.0910 | 0.4401 | 0.2901 | 0.4156 | 1.0000 | 0.4307 | 0.014 |
| I-3[e,f,g] | 3047269 | -/CTGA | 1 | 162810828 | 0.5660 | 0.4340 | 0.0825 | 0.3795 | 0.4717 | 0.5348 | 0.3005 | 0.3758 | 0.4283 | 0.4303 | 0.0638 | 0.3803 | 0.6763 | 0.4730 | 0.2388 | 0.4119 | 0.047 |
| I-4[g] | 2307507 | -/ATTTT | 1 | 190257015 | 0.4022 | 0.4913 | 0.7847 | 0.3851 | 0.4897 | 0.5413 | 0.2414 | 0.3751 | 0.4494 | 0.4553 | 0.2061 | 0.3776 | 0.2789 | 0.4008 | 1.0000 | 0.4382 | 0.032 |
| I-5[e,f,g] | 2307579 | -/ATG | 1 | 247812083 | 0.2863 | 0.4103 | 1.0000 | 0.4332 | 0.3092 | 0.3991 | 0.3505 | 0.4194 | 0.5082 | 0.5082 | 0.9002 | 0.3751 | 0.4979 | 0.5228 | 0.5216 | 0.3750 | 0.056 |
| I-6[f,g] | 3838581 | -/TAAC | 2 | 100050427 | 0.3500 | 0.4478 | 0.8865 | 0.4005 | 0.5806 | 0.4917 | 1.0000 | 0.3817 | 0.4319 | 0.4903 | 1.0000 | 0.3798 | 0.3905 | 0.5083 | 0.3398 | 0.3879 | 0.038 |
| I-7[f,g] | 2308276 | -/TTTAA | 2 | 172915805 | 0.3630 | 0.3957 | 0.0317* | 0.3959 | 0.4835 | 0.5391 | 0.2474 | 0.3753 | 0.3716 | 0.4397 | 0.3487 | 0.3931 | 0.5434 | 0.5000 | 1.0000 | 0.3769 | 0.029 |
| I-8[f,g] | 3042783 | -/GAGTT | 2 | 222160758 | 0.6325 | 0.4056 | 0.0540 | 0.3944 | 0.6319 | 0.4676 | 1.0000 | 0.3942 | 0.6745 | 0.3843 | 0.0464* | 0.4110 | 0.7522 | 0.3973 | 0.3777 | 0.4629 | 0.012 |
| I-9[e] | 16624 | -/GT | 2 | 235016391 | 0.4100 | 0.3800 | 0.0118* | 0.3835 | 0.5063 | 0.5125 | 0.8762 | 0.3750 | 0.7805 | 0.3659 | 0.4930 | 0.4908 | 0.2870 | 0.3889 | 0.5686 | 0.4327 | 0.170 |
| I-10[e] | 2308242 | -/CT | 3 | 8616709 | 0.2702 | 0.3872 | 0.7430 | 0.4445 | 0.1787 | 0.2979 | 1.0000 | 0.5421 | 0.2070 | 0.3074 | 0.3215 | 0.5051 | 0.3402 | 0.4481 | 1.0000 | 0.4044 | 0.025 |
| I-11[f,g] | 3841948 | -/ATTTA | 3 | 30715071 | 0.4239 | 0.4913 | 1.0000 | 0.3810 | 0.3704 | 0.4856 | 0.5888 | 0.3935 | 0.4261 | 0.4786 | 0.7946 | 0.3806 | 0.3107 | 0.4321 | 1.0000 | 0.4185 | 0.011 |
| I-12[f,g] | 35716687 | -/TTAA | 3 | 112650221 | 0.6565 | 0.4609 | 0.8852 | 0.4031 | 0.5123 | 0.5391 | 0.2468 | 0.3752 | 0.5467 | 0.5019 | 0.8994 | 0.3772 | 0.7190 | 0.4132 | 0.8815 | 0.4368 | 0.036 |
| I-13[f,g] | 2307603 | -/GATCT | 3 | 153886702 | 0.5341 | 0.5542 | 0.0964 | 0.3762 | 0.4731 | 0.5165 | 0.6040 | 0.3757 | 0.5725 | 0.4627 | 0.3711 | 0.3804 | 0.4489 | 0.4723 | 0.5029 | 0.3776 | 0.011 |
| I-14 | 3057785 | -/ATTTG | 3 | 188417221 | 0.2848 | 0.4043 | 0.8728 | 0.4342 | 0.2984 | 0.4403 | 0.5376 | 0.4256 | 0.3852 | 0.5136 | 0.1947 | 0.3892 | 0.0926 | 0.1687 | 1.0000 | 0.7063 | 0.076 |
| I-15[i] | 17131840 | -/CCGCCCTGC | 4 | 1283077 | 0.7730 | 0.0000 | <0.0001* | 0.4829 | 0.8031 | 0.0438 | <0.0001* | 0.5175 | 0.5667 | 0.4424 | 0.2091 | 0.3796 | 0.7184 | 0.0190 | <0.0001* | 0.4363 | 0.050 |
| I-16[f,g] | 60901515 | -/AAGT | 4 | 23792754 | 0.6225 | 0.4739 | 1.0000 | 0.3914 | 0.6058 | 0.4813 | 1.0000 | 0.3869 | 0.6569 | 0.4510 | 1.0000 | 0.4032 | 0.6170 | 0.4340 | 0.2153 | 0.3898 | 0.000 |
| I-17[f,g] | 2308292 | -/TAAGT | 4 | 107889773 | 0.5109 | 0.5000 | 1.0000 | 0.3751 | 0.3180 | 0.4184 | 0.5460 | 0.4147 | 0.3366 | 0.4764 | 0.3249 | 0.4060 | 0.4195 | 0.4915 | 1.0000 | 0.3817 | 0.030 |
| I-18[e,f,g] | 2307526 | -/ACAC | 5 | 5125112 | 0.5723 | 0.4213 | 0.0346* | 0.3804 | 0.3092 | 0.4079 | 0.5394 | 0.4194 | 0.3648 | 0.4918 | 0.4086 | 0.3953 | 0.4066 | 0.5145 | 0.3409 | 0.3842 | 0.049 |
| I-19 | 2308240 | -/AGAA | 5 | 18217324 | 0.3353 | 0.4538 | 0.8891 | 0.4065 | 0.3471 | 0.4959 | 0.1626 | 0.4017 | 0.4765 | 0.4824 | 0.6203 | 0.3756 | 0.1723 | 0.2851 | 1.0000 | 0.5515 | 0.066 |
| I-20[g] | 2307656 | -/TAAGT | 5 | 34844425 | 0.4217 | 0.4435 | 0.1775 | 0.3813 | 0.5926 | 0.4774 | 0.8934 | 0.3840 | 0.4903 | 0.4591 | 0.2107 | 0.3751 | 0.5494 | 0.5226 | 0.4397 | 0.3775 | 0.019 |
| I-21 | 2307661 | -/TTCT | 5 | 34893909 | 0.4197 | 0.5422 | 0.0870 | 0.3817 | 0.3777 | 0.4378 | 0.3237 | 0.3913 | 0.4294 | 0.4745 | 0.6041 | 0.3801 | 0.5022 | 0.4848 | 0.6858 | 0.3750 | 0.008 |
| I-22 | 2307848 | -/AAGTGCACG | 5 | 36819396 | 0.4618 | 0.4980 | 1.0000 | 0.3765 | 0.3954 | 0.4393 | 0.2235 | 0.3867 | 0.2680 | 0.3760 | 0.5158 | 0.4462 | 0.6474 | 0.4231 | 0.2509 | 0.3996 | 0.096 |
| I-23[e] | 1160956 | -/AGA | 5 | 65378460 | 0.5830 | 0.5191 | 0.3486 | 0.3822 | 0.6346 | 0.4316 | 0.3186 | 0.3951 | 0.8689 | 0.2213 | 0.5840 | 0.6221 | 0.5602 | 0.4896 | 0.8960 | 0.3787 | 0.087 |
| I-24[f,g] | 2308196 | -/ATTG | 5 | 73798863 | 0.6871 | 0.3681 | 0.0636 | 0.4174 | 0.6350 | 0.4479 | 0.7326 | 0.3952 | 0.5904 | 0.4819 | 1.0000 | 0.3836 | 0.6656 | 0.4110 | 0.3771 | 0.4070 | 0.004 |
| I-25 | 1610959 | -/CTTA | 5 | 76003944 | 0.4799 | 0.5261 | 0.4420 | 0.3754 | 0.3947 | 0.4825 | 1.0000 | 0.3868 | 0.4255 | 0.4667 | 0.5182 | 0.3807 | 0.6498 | 0.4185 | 0.2449 | 0.4005 | 0.047 |
| I-26 | 10590424 | -/AATAA | 5 | 79347159 | 0.5060 | 0.5329 | 0.4432 | 0.3750 | 0.5274 | 0.4939 | 0.8746 | 0.3758 | 0.5215 | 0.5399 | 0.3449 | 0.3755 | 0.6182 | 0.4606 | 0.7358 | 0.3901 | 0.007 |
| I-27 | 35864678 | -/GTAACTAC | 5 | 100097302 | 0.8213 | 0.3012 | 0.8349 | 0.5422 | 0.5885 | 0.4425 | 0.2152 | 0.3832 | 0.6569 | 0.4667 | 0.6790 | 0.4032 | 0.5826 | 0.4420 | 0.1668 | 0.3821 | 0.053 |
| I-28 | 1160936 | -/ATTTA | 5 | 115787453 | 0.2043 | 0.2870 | 0.1028 | 0.5083 | 0.4318 | 0.4835 | 0.7890 | 0.3798 | 0.5233 | 0.4786 | 0.5327 | 0.3755 | 0.1723 | 0.2941 | 0.8179 | 0.5516 | 0.127 |
| I-29[f,g] | 2067140 | -/CAGT | 5 | 115887784 | 0.5982 | 0.4356 | 0.2542 | 0.3852 | 0.5215 | 0.3804 | 0.0027* | 0.3755 | 0.5873 | 0.5000 | 0.7479 | 0.3830 | 0.3282 | 0.4110 | 0.3930 | 0.4097 | 0.058 |
| I-30[g] | 2067191 | -/TCTA | 5 | 135274588 | 0.5141 | 0.5060 | 0.8965 | 0.3752 | 0.4539 | 0.5044 | 0.8916 | 0.3771 | 0.5216 | 0.4784 | 0.5336 | 0.3755 | 0.4408 | 0.5044 | 0.7867 | 0.3786 | 0.005 |
| I-31 | 2307687 | -/TTGT | 5 | 144002681 | 0.2952 | 0.3655 | 0.0651 | 0.4275 | 0.2438 | 0.3625 | 0.8633 | 0.4665 | 0.2157 | 0.3765 | 0.0920 | 0.4950 | 0.1603 | 0.2607 | 0.6371 | 0.5704 | 0.016 |
| I-32 | 1160941 | -/AAAAGC | 5 | 156621965 | 0.9960 | 0.0080 | 1.0000 | 0.9841 | 0.9871 | 0.0258 | 1.0000 | 0.9501 | 0.9882 | 0.0235 | 1.0000 | 0.9543 | 0.9957 | 0.0087 | 1.0000 | 0.9828 | 0.001 |
| I-33[e,f,g] | 1610871 | -/TAGG | 5 | 171087970 | 0.4120 | 0.4869 | 1.0000 | 0.3831 | 0.4153 | 0.4587 | 0.4287 | 0.3825 | 0.4789 | 0.4981 | 1.0000 | 0.3754 | 0.4551 | 0.4939 | 1.0000 | 0.3770 | 0.002 |
| I-34 | 2307680 | -/CAAA | 6 | 10020397 | 0.1109 | 0.1522 | 0.002* | 0.6640 | 0.3128 | 0.4198 | 0.7646 | 0.4174 | 0.3794 | 0.4553 | 0.5974 | 0.3908 | 0.2078 | 0.3086 | 0.3324 | 0.5041 | 0.069 |
| I-35[e,f,g] | 2307710 | -/AGGA | 6 | 47821263 | 0.2128 | 0.3149 | 0.3357 | 0.4983 | 0.2830 | 0.4043 | 1.0000 | 0.4354 | 0.3053 | 0.4631 | 0.1806 | 0.4215 | 0.4419 | 0.4855 | 0.7899 | 0.3784 | 0.040 |
| I-36 | 2307938 | -/CCCA | 6 | 79100815 | 0.2490 | 0.3293 | 0.0663 | 0.4618 | 0.5473 | 0.5021 | 0.8973 | 0.3773 | 0.4863 | 0.5098 | 0.7991 | 0.3752 | 0.6857 | 0.4346 | 1.0000 | 0.4166 | 0.126 |
| I-37 | 2308231 | -/GACAAA | 6 | 116436397 | 0.6152 | 0.5087 | 0.3338 | 0.3893 | 0.4136 | 0.4650 | 0.5126 | 0.3828 | 0.4086 | 0.4514 | 0.3023 | 0.3838 | 0.0885 | 0.1687 | 0.7049 | 0.7164 | 0.186 |
| I-38[e,g] | 2307839 | -/GA | 6 | 117093558 | 0.4295 | 0.5427 | 0.1114 | 0.3801 | 0.2467 | 0.3511 | 0.4733 | 0.4639 | 0.2500 | 0.4098 | 0.1758 | 0.4609 | 0.2261 | 0.3361 | 0.5739 | 0.4837 | 0.041 |
| I-39[e] | 2308137 | -/GA | 6 | 149614198 | 0.4340 | 0.4851 | 0.8908 | 0.3795 | 0.2983 | 0.4163 | 1.0000 | 0.4256 | 0.3115 | 0.4262 | 0.8838 | 0.4181 | 0.6058 | 0.4647 | 0.6951 | 0.3869 | 0.081 |
| I-40[f,g] | 34510056 | -/CTTTA | 6 | 153353935 | 0.6696 | 0.4174 | 0.3739 | 0.4087 | 0.5864 | 0.4403 | 0.1490 | 0.3828 | 0.5233 | 0.5331 | 0.3155 | 0.3755 | 0.5165 | 0.5537 | 0.1181 | 0.3753 | 0.018 |
| I-41 | 1160847 | -/TAAAA | 7 | 11562255 | 0.8609 | 0.2609 | 0.2631 | 0.6070 | 0.7798 | 0.3333 | 0.7149 | 0.4901 | 0.8288 | 0.2879 | 1.0000 | 0.5532 | 0.6322 | 0.4959 | 0.3380 | 0.3943 | 0.055 |
| I-42 | 1611033 | -/GAAA | 7 | 70068205 | 0.1978 | 0.3000 | 0.3954 | 0.5163 | 0.3244 | 0.4174 | 0.4671 | 0.4115 | 0.4942 | 0.4903 | 0.8033 | 0.3750 | 0.4104 | 0.5375 | 0.1061 | 0.3834 | 0.066 |

**Table 1. (Cont)**

| Marker | RS Number[a] | Alleles[a] | Chr[a] | Location[a] | Asian *(n=287)* | | | | Southwestern Hispanic *(n=253)* | | | | Caucasian *(n=264)* | | | | African American *(n=246)* | | | | $F_{ST}$[d] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | |
| I-43[i] | 2067151 | -/TATTA | 7 | 78252645 | 0.3635 | 0.3815 | 0.0075* | 0.3957 | 0.6288 | 0.4592 | 0.7840 | 0.3932 | 0.6608 | 0.3569 | 0.0013* | 0.4049 | 0.4539 | 0.5044 | 0.8915 | 0.3771 | 0.079 |
| I-44[e] | 2307978 | -/GA | 7 | 83283913 | 0.3426 | 0.4553 | 1.0000 | 0.4035 | 0.3075 | 0.4027 | 0.4271 | 0.4203 | 0.1557 | 0.2705 | 0.8075 | 0.5778 | 0.3942 | 0.5145 | 0.2727 | 0.3869 | 0.049 |
| I-45 | 1610907 | -/AAAGT | 7 | 110559277 | 0.1891 | 0.3087 | 1.0000 | 0.5277 | 0.5351 | 0.4752 | 0.5151 | 0.3762 | 0.6031 | 0.4825 | 1.0000 | 0.3863 | 0.2128 | 0.3182 | 0.4383 | 0.4983 | 0.183 |
| I-46[f,g] | 16458 | -/TTCC | 7 | 122151327 | 0.6135 | 0.4663 | 0.8708 | 0.3889 | 0.5859 | 0.5460 | 0.1485 | 0.3827 | 0.6536 | 0.4759 | 0.6046 | 0.4019 | 0.4663 | 0.5276 | 0.5369 | 0.3761 | 0.024 |
| I-47 | 2307571 | -/TACTT | 7 | 137050412 | 0.6767 | 0.4297 | 0.7722 | 0.4121 | 0.5932 | 0.4682 | 0.6788 | 0.3841 | 0.6608 | 0.4902 | 0.1681 | 0.4049 | 0.7500 | 0.3584 | 0.4864 | 0.4609 | 0.016 |
| I-48 | 3062629 | -/CTGT | 8 | 10606219 | 0.6145 | 0.5221 | 0.1447 | 0.3891 | 0.4871 | 0.5172 | 0.6979 | 0.3752 | 0.4725 | 0.4980 | 1.0000 | 0.3758 | 0.2978 | 0.4304 | 0.7536 | 0.4259 | 0.064 |
| I-49 | 17515041 | -/CAAGA | 8 | 16855495 | 0.5109 | 0.5174 | 0.7001 | 0.3751 | 0.4877 | 0.4733 | 0.4452 | 0.3752 | 0.4222 | 0.4786 | 0.8042 | 0.3813 | 0.1379 | 0.2263 | 0.4250 | 0.6093 | 0.118 |
| I-50[f,g] | 34535242 | -/GTAG | 8 | 18429416 | 0.5402 | 0.4859 | 0.7975 | 0.3766 | 0.6234 | 0.5105 | 0.2136 | 0.3916 | 0.6608 | 0.4275 | 0.4898 | 0.4049 | 0.5736 | 0.5065 | 0.6867 | 0.3806 | 0.010 |
| I-51 | 2308127 | -/TCAAG | 8 | 24053261 | 0.0000 | 0.0000 | 1.0000 | 1.0000 | 0.0434 | 0.0785 | 0.3661 | 0.8443 | 0.0039 | 0.0078 | 1.0000 | 0.9845 | 0.2827 | 0.4135 | 0.8821 | 0.4356 | 0.227 |
| I-52 | 34293322 | -/ACTC | 8 | 34948880 | 0.6968 | 0.4297 | 0.8765 | 0.4227 | 0.4958 | 0.4768 | 0.5188 | 0.3750 | 0.3706 | 0.4510 | 0.5860 | 0.3934 | 0.3803 | 0.4359 | 0.2671 | 0.3905 | 0.090 |
| I-53 | 10666410 | -/AGTG | 8 | 61190688 | 0.4761 | 0.5000 | 1.0000 | 0.3756 | 0.4918 | 0.5144 | 0.7038 | 0.3751 | 0.5097 | 0.5058 | 0.8983 | 0.3751 | 0.1983 | 0.3140 | 0.8381 | 0.5157 | 0.086 |
| I-54[e] | 35769550 | -/TGAC | 8 | 76518680 | 0.5830 | 0.4681 | 0.5938 | 0.3822 | 0.5085 | 0.4468 | 0.1122 | 0.3751 | 0.3668 | 0.4631 | 1.0000 | 0.3946 | 0.1805 | 0.3029 | 0.8270 | 0.5396 | 0.124 |
| I-55 | 35146764 | -/TCTTA | 8 | 117130337 | 0.6345 | 0.4498 | 0.6824 | 0.3951 | 0.6508 | 0.4587 | 1.0000 | 0.4009 | 0.5882 | 0.4627 | 0.5240 | 0.3831 | 0.3468 | 0.5234 | 0.0213* | 0.4018 | 0.075 |
| I-56[f,g] | 10623496 | -/GAAT | 8 | 123945645 | 0.4237 | 0.4297 | 0.0691 | 0.3810 | 0.2991 | 0.4274 | 0.8823 | 0.4251 | 0.4039 | 0.4706 | 0.7017 | 0.3847 | 0.3565 | 0.5043 | 0.1526 | 0.3981 | 0.011 |
| I-57[e] | 5895447 | -/CA | 8 | 138420594 | 0.2851 | 0.4255 | 0.6341 | 0.4340 | 0.3249 | 0.4009 | 0.2068 | 0.4113 | 0.3402 | 0.4672 | 0.5764 | 0.4045 | 0.2729 | 0.3958 | 1.0000 | 0.4425 | 0.003 |
| I-58[f,g] | 33951431 | -/AGTT | 9 | 2626813 | 0.5913 | 0.4348 | 0.1340 | 0.3838 | 0.6446 | 0.4711 | 0.7825 | 0.3985 | 0.5856 | 0.5175 | 0.3035 | 0.3827 | 0.6379 | 0.3951 | 0.0264* | 0.3962 | 0.002 |
| I-59[e,g] | 16402 | -/TTAT | 9 | 38406788 | 0.2553 | 0.3574 | 0.3843 | 0.4564 | 0.3013 | 0.4330 | 0.7459 | 0.4238 | 0.2930 | 0.4385 | 0.4333 | 0.4288 | 0.3382 | 0.4440 | 0.8859 | 0.4053 | 0.003 |
| I-60[e] | 2067294 | -/CTT | 9 | 71314421 | 0.1936 | 0.3106 | 1.0000 | 0.5217 | 0.3655 | 0.4622 | 1.0000 | 0.3950 | 0.3566 | 0.4426 | 0.5814 | 0.3981 | 0.1722 | 0.2531 | 0.1112 | 0.5517 | 0.051 |
| I-61 | 2308113 | -/TACT | 9 | 98479484 | 0.2028 | 0.3173 | 0.6957 | 0.5101 | 0.2045 | 0.3595 | 0.1180 | 0.5080 | 0.2373 | 0.3412 | 0.3797 | 0.4726 | 0.1097 | 0.1772 | 0.1681 | 0.6666 | 0.017 |
| I-62[f,g] | 2308112 | -/ACACC | 9 | 98574109 | 0.4819 | 0.5141 | 0.7059 | 0.3753 | 0.5917 | 0.5333 | 0.1364 | 0.3838 | 0.5647 | 0.5020 | 0.7994 | 0.3793 | 0.5468 | 0.5149 | 0.5908 | 0.3772 | 0.007 |
| I-63[e] | 2307580 | -/AATT | 9 | 105586193 | 0.5277 | 0.4426 | 0.0914 | 0.3758 | 0.5498 | 0.4842 | 0.7920 | 0.3775 | 0.4877 | 0.5164 | 0.7006 | 0.3752 | 0.2967 | 0.4108 | 0.8772 | 0.4266 | 0.051 |
| I-64 | 41308024 | -/GTAA | 9 | 123793536 | 0.7217 | 0.4087 | 0.8710 | 0.4387 | 0.6770 | 0.4074 | 0.3079 | 0.4122 | 0.5875 | 0.4825 | 1.0000 | 0.3830 | 0.2087 | 0.3430 | 0.6955 | 0.5031 | 0.205 |
| I-65[g] | 2307850 | -/GGTG | 9 | 135380186 | 0.3649 | 0.5040 | 0.2192 | 0.3952 | 0.3347 | 0.4628 | 0.6652 | 0.4068 | 0.3000 | 0.4353 | 0.6453 | 0.4246 | 0.4703 | 0.5424 | 0.1968 | 0.3759 | 0.021 |
| I-66[e,f,g] | 140809 | -/CAA | 10 | 5987163 | 0.2830 | 0.3787 | 0.3256 | 0.4354 | 0.2863 | 0.3965 | 0.6224 | 0.4332 | 0.3668 | 0.5205 | 0.0701 | 0.3946 | 0.4461 | 0.5021 | 0.8990 | 0.3780 | 0.024 |
| I-67[e,f,g] | 1160886 | -/ACT | 10 | 54442386 | 0.5043 | 0.4979 | 1.0000 | 0.3750 | 0.3796 | 0.5185 | 0.1544 | 0.3907 | 0.3689 | 0.5082 | 0.1716 | 0.3940 | 0.3174 | 0.4108 | 0.4626 | 0.4150 | 0.025 |
| I-68[g] | 34051577 | -/TCTTA | 10 | 89690955 | 0.5823 | 0.5221 | 0.2955 | 0.3821 | 0.5833 | 0.4907 | 1.0000 | 0.3822 | 0.6686 | 0.4431 | 1.0000 | 0.4083 | 0.5045 | 0.5336 | 0.3586 | 0.3750 | 0.017 |
| I-69[e,f,g] | 10688868 | -/CT | 11 | 268180 | 0.4511 | 0.4426 | 0.1127 | 0.3774 | 0.3109 | 0.3782 | 0.0710 | 0.4184 | 0.3033 | 0.4344 | 0.7676 | 0.4227 | 0.2614 | 0.3817 | 0.8575 | 0.4514 | 0.028 |
| I-70 | 34823526 | -/AAGT | 11 | 14200361 | 0.4785 | 0.4417 | 0.1618 | 0.3755 | 0.3497 | 0.4540 | 1.0000 | 0.4007 | 0.4277 | 0.4940 | 1.0000 | 0.3804 | 0.5307 | 0.4601 | 0.3485 | 0.3759 | 0.021 |
| I-71[e,g] | 34811743 | -/TG | 11 | 30177690 | 0.6936 | 0.4170 | 0.7513 | 0.4209 | 0.7577 | 0.3260 | 0.1056 | 0.4679 | 0.6393 | 0.4590 | 1.0000 | 0.3967 | 0.6328 | 0.4855 | 0.5812 | 0.3945 | 0.013 |
| I-72 | 2307666 | -/GTTAC | 11 | 64729920 | 0.2087 | 0.2957 | 0.1140 | 0.5031 | 0.3663 | 0.4691 | 1.0000 | 0.3948 | 0.5798 | 0.4669 | 0.5302 | 0.3816 | 0.2284 | 0.3827 | 0.2070 | 0.4814 | 0.125 |
| I-73[f,g] | 2307696 | -/CGAC | 11 | 70595112 | 0.3414 | 0.4659 | 0.6782 | 0.4040 | 0.4661 | 0.4746 | 0.5140 | 0.3762 | 0.4118 | 0.4627 | 0.5176 | 0.3831 | 0.5043 | 0.4224 | 0.0162* | 0.3750 | 0.018 |
| I-74[g] | 34528025 | -/GAGT | 11 | 99514962 | 0.4819 | 0.5382 | 0.2575 | 0.3753 | 0.3548 | 0.4855 | 0.4003 | 0.3988 | 0.2863 | 0.3843 | 0.3622 | 0.4332 | 0.5601 | 0.4764 | 0.5971 | 0.3787 | 0.059 |
| I-75 | 11281892 | -/GTCAT | 11 | 124644227 | 0.7783 | 0.2957 | 0.0356* | 0.4884 | 0.8189 | 0.2716 | 0.1976 | 0.5388 | 0.8327 | 0.2568 | 0.2583 | 0.5592 | 0.2593 | 0.3868 | 1.0000 | 0.4531 | 0.319 |
| I-76 | 2307805 | -/CCATAAACC | 12 | 67705010 | 0.6064 | 0.5141 | 0.2868 | 0.3871 | 0.4219 | 0.5063 | 0.5995 | 0.3813 | 0.3627 | 0.4588 | 0.8938 | 0.3960 | 0.5590 | 0.4716 | 0.5047 | 0.3785 | 0.051 |
| I-77[f,g] | 3045264 | -/GTCT | 12 | 77216833 | 0.3715 | 0.4297 | 0.2243 | 0.3932 | 0.3130 | 0.4328 | 1.0000 | 0.4173 | 0.3863 | 0.4039 | 0.0197* | 0.3889 | 0.4095 | 0.4655 | 0.5864 | 0.3836 | 0.005 |
| I-78[g] | 2308232 | -/AGTTTA | 12 | 96991884 | 0.3000 | 0.4348 | 0.6513 | 0.4246 | 0.2794 | 0.3908 | 0.6345 | 0.4379 | 0.3541 | 0.4981 | 0.1734 | 0.3990 | 0.2748 | 0.4421 | 0.1090 | 0.4412 | 0.005 |
| I-79[e] | 2308171 | -/TCTG | 13 | 44880155 | 0.0759 | 0.1296 | 0.1791 | 0.7489 | 0.1687 | 0.2651 | 0.3622 | 0.5571 | 0.2015 | 0.3118 | 0.5725 | 0.5117 | 0.5490 | 0.5020 | 0.9004 | 0.3774 | 0.214 |
| I-80[f,g] | 4187 | -/TAAAGA | 13 | 50106333 | 0.5153 | 0.5276 | 0.5277 | 0.3752 | 0.4294 | 0.5153 | 0.6308 | 0.3801 | 0.5181 | 0.5181 | 0.7519 | 0.3753 | 0.6288 | 0.4847 | 0.7417 | 0.3933 | 0.024 |
| I-81 | 2308057 | -/AATAA | 13 | 110810568 | 0.4261 | 0.4783 | 0.7788 | 0.3806 | 0.2181 | 0.2716 | 0.0029* | 0.4924 | 0.2140 | 0.3502 | 0.5809 | 0.4969 | 0.4339 | 0.4793 | 0.6909 | 0.3795 | 0.067 |
| I-82[f,g] | 3038530 | -/TCAA | 13 | 112546924 | 0.4558 | 0.5100 | 0.7074 | 0.3770 | 0.4635 | 0.4721 | 0.4339 | 0.3763 | 0.4020 | 0.4431 | 0.2486 | 0.3852 | 0.2651 | 0.4353 | 0.0927 | 0.4485 | 0.032 |
| I-83[e,f,g] | 2308189 | -/AACTA | 14 | 29036757 | 0.4043 | 0.5447 | 0.0567 | 0.3847 | 0.4908 | 0.5229 | 0.5902 | 0.3751 | 0.4221 | 0.4836 | 0.8980 | 0.3813 | 0.5041 | 0.5519 | 0.1262 | 0.3750 | 0.008 |

**Table 1. (Cont)**

| Marker | RS Number[a] | Alleles[a] | Chr[a] | Location[a] | Asian *(n=287)* |  |  |  | Southwestern Hispanic *(n=253)* |  |  |  | Caucasian *(n=264)* |  |  |  | African American *(n=246)* |  |  |  | $F_{ST}$[d] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  |  |  |  | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] | Frequency of Deletion | $H_o$[c] | HWE[b] *(p-value)* | RMP[c] |  |
| I-84 | 3059434 | -/CTCTT | 14 | 34152270 | 0.9196 | 0.1261 | 0.0461* | 0.7370 | 0.7335 | 0.4174 | 0.3291 | 0.4473 | 0.6654 | 0.4514 | 0.8918 | 0.4068 | 0.9564 | 0.0788 | 0.3674 | 0.8437 | 0.128 |
| I-85[f,g] | 34795726 | -/AAGA | 15 | 58348104 | 0.6466 | 0.4819 | 0.4886 | 0.3993 | 0.6535 | 0.4440 | 0.7814 | 0.4019 | 0.4490 | 0.5294 | 0.3157 | 0.3776 | 0.5214 | 0.4444 | 0.0921 | 0.3755 | 0.039 |
| I-86 | 2307519 | -/TTTCAA | 15 | 64367358 | 0.2239 | 0.3435 | 0.8570 | 0.4861 | 0.3864 | 0.4256 | 0.1315 | 0.3889 | 0.1829 | 0.2879 | 0.5346 | 0.5362 | 0.3601 | 0.4650 | 1.0000 | 0.3969 | 0.047 |
| I-87 | 3029195 | -/ATGGGA | 16 | 7758509 | 0.2370 | 0.3348 | 0.2706 | 0.3911 | 0.3864 | 0.4421 | 0.7014 | 0.3889 | 0.4377 | 0.5097 | 0.6038 | 0.3790 | 0.1364 | 0.2562 | 1.0000 | 0.6121 | 0.088 |
| I-88[f,g] | 17859968 | -/TAAA | 16 | 55530356 | 0.3783 | 0.4609 | 0.7861 | 0.4657 | 0.4669 | 0.5124 | 0.3367 | 0.3761 | 0.4144 | 0.4708 | 0.6616 | 0.3827 | 0.2955 | 0.4174 | 1.0000 | 0.4273 | 0.020 |
| I-89[e] | 2067208 | -/GCCAG | 16 | 84582287 | 0.2447 | 0.3532 | 0.4809 | 0.4657 | 0.2818 | 0.3771 | 0.3367 | 0.4362 | 0.3176 | 0.4221 | 0.6616 | 0.4149 | 0.1577 | 0.2656 | 1.0000 | 0.5746 | 0.023 |
| I-90[e] | 3051300 | -/GTAT | 17 | 10135941 | 0.3489 | 0.4681 | 0.7751 | 0.4009 | 0.3403 | 0.4202 | 0.3192 | 0.4044 | 0.4508 | 0.4754 | 0.5156 | 0.3775 | 0.1950 | 0.3154 | 1.0000 | 0.5199 | 0.048 |
| I-91[f,g] | 28923216 | -/TTGTA | 17 | 12011874 | 0.5500 | 0.4130 | 0.0102* | 0.3775 | 0.5926 | 0.4856 | 1.0000 | 0.3840 | 0.5233 | 0.5019 | 1.0000 | 0.3755 | 0.6694 | 0.4545 | 0.7747 | 0.4086 | 0.015 |
| I-92 | 16715 | -/AAGCTC | 17 | 61393657 | 0.8204 | 0.2635 | 0.1989 | 0.5408 | 0.6677 | 0.3841 | 0.1070 | 0.4079 | 0.6810 | 0.4172 | 0.5977 | 0.4142 | 0.1606 | 0.2970 | 0.2569 | 0.5698 | 0.316 |
| I-93[i] | 16430 | -/CTTTAA | 18 | 673444 | 0.7450 | 0.3092 | 0.0036* | 0.4566 | 0.4464 | 0.2747 | <0.0001* | 0.3779 | 0.3412 | 0.3137 | <0.0001* | 0.4040 | 0.7162 | 0.3231 | 0.0027* | 0.4348 | 0.156 |
| I-94[g] | 36062169 | -/GTACTG | 18 | 8073016 | 0.4261 | 0.4696 | 0.5869 | 0.3806 | 0.5124 | 0.5455 | 0.1984 | 0.3752 | 0.5992 | 0.4747 | 0.9007 | 0.3854 | 0.4772 | 0.4979 | 1.0000 | 0.3755 | 0.019 |
| I-95[e] | 3080855 | -/AATT | 18 | 23253207 | 0.3090 | 0.4382 | 0.7732 | 0.4195 | 0.2745 | 0.3872 | 0.7408 | 0.4414 | 0.2816 | 0.4023 | 1.0000 | 0.4363 | 0.2837 | 0.4122 | 0.8777 | 0.4349 | 0.001 |
| I-96 | 34000371 | -/GTTA | 18 | 27291283 | 0.3655 | 0.4337 | 0.3465 | 0.3951 | 0.5000 | 0.4746 | 0.4402 | 0.3750 | 0.6255 | 0.4588 | 0.7890 | 0.3922 | 0.5396 | 0.4890 | 0.7855 | 0.3766 | 0.046 |
| I-97 | 34999022 | -/TAAAA | 18 | 33050322 | 0.4196 | 0.4739 | 0.6803 | 0.3817 | 0.5658 | 0.4568 | 0.2979 | 0.3794 | 0.6712 | 0.4553 | 0.6732 | 0.4095 | 0.5640 | 0.4917 | 1.0000 | 0.3792 | 0.041 |
| I-98[e,f,g] | 34511541 | -/CTCTT | 18 | 36423040 | 0.3809 | 0.4809 | 0.8898 | 0.3904 | 0.5000 | 0.4498 | 0.1447 | 0.3750 | 0.3586 | 0.4467 | 0.6807 | 0.3974 | 0.3714 | 0.4772 | 0.7864 | 0.3932 | 0.015 |
| I-99 | 4149614 | -/TTAAA | 18 | 56040243 | 0.5978 | 0.5000 | 0.5876 | 0.3851 | 0.5103 | 0.5350 | 0.3113 | 0.3751 | 0.5778 | 0.4942 | 0.9005 | 0.3813 | 0.2695 | 0.4403 | 0.0747 | 0.4450 | 0.087 |
| I-100[e] | 36040336 | -/AT | 19 | 1402662 | 0.7596 | 0.3447 | 0.3725 | 0.4696 | 0.6915 | 0.4298 | 1.0000 | 0.4197 | 0.8115 | 0.3033 | 0.8360 | 0.5285 | 0.4627 | 0.5021 | 1.0000 | 0.3764 | 0.105 |
| I-101 | 34560670 | -/CATAGAG | 19 | 5059801 | 0.8024 | 0.2754 | 0.0889 | 0.5166 | 0.4055 | 0.5427 | 0.1508 | 0.3844 | 0.3865 | 0.4172 | 0.1358 | 0.3889 | 0.3879 | 0.4970 | 0.6263 | 0.3885 | 0.160 |
| I-102 | 34781304 | -/GATAA | 19 | 38094947 | 0.5804 | 0.4652 | 0.5006 | 0.3817 | 0.2531 | 0.3909 | 0.7410 | 0.4583 | 0.2121 | 0.3152 | 0.3629 | 0.4992 | 0.3889 | 0.4156 | 0.0571 | 0.3883 | 0.113 |
| I-103[e] | 2307689 | -/TTC | 19 | 44204340 | 0.1979 | 0.3021 | 0.4100 | 0.5163 | 0.4013 | 0.5084 | 0.4260 | 0.3853 | 0.2418 | 0.3852 | 0.4826 | 0.4683 | 0.4627 | 0.5519 | 0.1194 | 0.3764 | 0.069 |
| I-104[f,g] | 34495360 | -/AAGT | 20 | 4954109 | 0.5848 | 0.5087 | 0.4988 | 0.3825 | 0.5288 | 0.4897 | 0.8055 | 0.3758 | 0.5603 | 0.5370 | 0.1596 | 0.3787 | 0.6049 | 0.5267 | 0.1414 | 0.3867 | 0.002 |
| I-105 | 35149698 | -/CAACTA | 20 | 7672133 | 0.7329 | 0.3735 | 0.5135 | 0.4469 | 0.4535 | 0.4735 | 0.5049 | 0.3772 | 0.5882 | 0.4627 | 0.5176 | 0.3831 | 0.2701 | 0.4241 | 0.3100 | 0.4446 | 0.146 |
| I-106[e] | 33917182 | -/CA | 20 | 11695625 | 0.5043 | 0.4638 | 0.2929 | 0.3750 | 0.6468 | 0.4587 | 1.0000 | 0.3993 | 0.5656 | 0.4508 | 0.1956 | 0.3794 | 0.5125 | 0.5417 | 0.2388 | 0.3752 | 0.015 |
| I-107 | 33921337 | -/GGGGTCTGA | 20 | 24727238 | 0.9202 | 0.1227 | 0.0618 | 0.7387 | 0.6902 | 0.3865 | 0.2680 | 0.4190 | 0.6657 | 0.4639 | 0.7280 | 0.4070 | 0.5736 | 0.5460 | 0.1492 | 0.3806 | 0.100 |
| I-108[e] | 34541393 | -/AACT | 20 | 30701405 | 0.6685 | 0.4307 | 0.6780 | 0.4082 | 0.4693 | 0.4959 | 1.0000 | 0.3760 | 0.4004 | 0.4866 | 0.8959 | 0.3855 | 0.3959 | 0.5061 | 0.4219 | 0.3865 | 0.065 |
| I-109 | 34785121 | -/TGGA | 20 | 58311383 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | 0.9664 | 0.0588 | 0.2287 | 0.8764 | 0.9942 | 0.0117 | 1.0000 | 0.9770 | 0.6996 | 0.4115 | 0.7596 | 0.4244 | 0.250 |
| I-110[e,f,g] | 35605984 | -/TAAAG | 21 | 15634865 | 0.3894 | 0.5234 | 0.1445 | 0.3881 | 0.5446 | 0.4225 | 0.0385* | 0.3770 | 0.4549 | 0.5082 | 0.7926 | 0.3771 | 0.5858 | 0.4686 | 0.5980 | 0.3827 | 0.029 |
| I-111 | 10629864 | -/TTAAT | 21 | 30695351 | 0.1674 | 0.2217 | 0.0042* | 0.5591 | 0.2243 | 0.3498 | 1.0000 | 0.4857 | 0.3891 | 0.5058 | 0.3543 | 0.3882 | 0.1157 | 0.1818 | 0.1109 | 0.6535 | 0.079 |
| I-112[e] | 10629077 | -/AT | 21 | 31372337 | 0.2511 | 0.3319 | 0.0823 | 0.4600 | 0.2478 | 0.3805 | 0.8621 | 0.4629 | 0.1660 | 0.2828 | 1.0000 | 0.5613 | 0.2116 | 0.3402 | 0.8504 | 0.4997 | 0.007 |
| I-113[e,f,g] | 2307700 | -/TCAC | 22 | 26790901 | 0.2907 | 0.3889 | 0.3724 | 0.4303 | 0.3865 | 0.4382 | 0.2291 | 0.3889 | 0.5209 | 0.4487 | 0.1078 | 0.3754 | 0.2469 | 0.3633 | 0.7326 | 0.4636 | 0.061 |
| I-114 | 3218285 | -/AACC | 22 | 37536724 | 0.4719 | 0.5020 | 1.0000 | 0.3758 | 0.5667 | 0.5583 | 0.0497* | 0.3796 | 0.5392 | 0.4745 | 0.5171 | 0.3766 | 0.1897 | 0.3017 | 0.8330 | 0.5270 | 0.112 |
|  |  |  |  |  | **Asian RMP** |  |  |  | **Southwestern Hispanic RMP** |  |  |  | **Caucasian RMP** |  |  |  | **African American RMP** |  |  |  | **$F_{ST}$** |
| **Overall For 111 Markers[h]** |  |  |  |  | 6.53 x$10^{-42}$ |  |  |  | 5.03 x $10^{-44}$ |  |  |  | 1.87 x $10^{-43}$ |  |  |  | 1.15 x $10^{-41}$ |  |  |  | 0.060 |
| **Overall For 33 Markers Described By Pereria et al. [e]** |  |  |  |  | 4.38x$10^{-13}$ |  |  |  | 2.38x$10^{-13}$ |  |  |  | 6.22x$10^{-13}$ |  |  |  | 3.41x$10^{-13}$ |  |  |  | 0.050 |
| **Overall For Suggested Panel 1[f]** |  |  |  |  | 4.27x$10^{-16}$ |  |  |  | 3.68x$10^{-16}$ |  |  |  | 2.43x$10^{-16}$ |  |  |  | 5.79x$10^{-16}$ |  |  |  | 0.023 |
| **Overall For Suggested Panel 2[g]** |  |  |  |  | 2.30x$10^{-19}$ |  |  |  | 2.52x$10^{-19}$ |  |  |  | 1.60x$10^{-19}$ |  |  |  | 3.62x$10^{-19}$ |  |  |  | 0.023 |

a. According to dbSNP [26]

b. * denotes markers that display departueres from HWE at a critical value of .05; α-level of .05 is adjusted from .05 to 0.000431 when corrected for multiple tests (Bonferroni's correction) [32, 33] as calculated by GDA [32]

c. $H_o$ denotes Observed Heterozygosity and RMP denotes Random Match Probability

d. $F_{ST}$ calculated according to Weir and Cockerham [34]

e. 33 Markers also described in Perieria et al. [10]

f. 38 Markers meeting the criteria of no observable departeures from HWE, LD, minor allele frequencies >.20, $F_{ST}$< .062, and >40Mb between markers on the same chromosome

g. Expanded set of 49 Markers (38 markers from f. and 11 additional markers) meeting the criteria of no observable departeures from HWE, LD, minor allele frequencies >.20, FST <.062, and >20Mb between markers on the same chromosome

h. Calculated assuming independence at population level

i. Markers excluded based on departure from HWE in more than one population