Forensic Science International: Genetics 25 (2016) 198–209

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research paper

Massively parallel sequencing of 68 insertion/deletion markers identifies novel microhaplotypes for utility in human identity testing

Frank R. Wendt, BS[a,*], David H. Warshauer, PhD[b], Xiangpei Zeng, MD, PhD[a], Jennifer D. Churchill, PhD[a], Nicole M.M. Novroski, MS[a], Bing Song, BS[a], Jonathan L. King, MS[a], Bobby L. LaRue, PhD[a,c], Bruce Budowle, PhD[a,d]

[a] Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX, 76107, USA
[b] Promega Corporation, 2800 Woods Hollow Rd, Madison, WI, 53711, USA
[c] Department of Forensic Science, College of Criminal Justice, Sam Houston State University, 1003 Bowers Blvd., Huntsville, TX, 77341, USA
[d] Center of Excellence in Genomic Medicine (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 28 May 2016
Received in revised form 1 August 2016
Accepted 19 September 2016
Available online 20 September 2016

Keywords:
INDELs
Population genetics
SNPs
Microhaplotypes
Massively parallel sequencing
MiSeq™
STRait Razor

ABSTRACT

Short tandem repeat (STR) loci are the traditional markers used for kinship, missing persons, and direct comparison human identity testing. These markers hold considerable value due to their highly polymorphic nature, amplicon size, and ability to be multiplexed. However, many STRs are still too large for use in analysis of highly degraded DNA. Small bi-allelic polymorphisms, such as insertions/deletions (INDELs), may be better suited for analyzing compromised samples, and their allele size differences are amenable to analysis by capillary electrophoresis. The INDEL marker allelic states range in size from 2 to 6 base pairs, enabling small amplicon size. In addition, heterozygote balance may be increased by minimizing preferential amplification of the smaller allele, as is more common with STR markers. Multiplexing a large number of INDELs allows for generating panels with high discrimination power. The Nextera™ Rapid Capture Custom Enrichment Kit (Illumina, Inc., San Diego, CA) and massively parallel sequencing (MPS) on the Illumina MiSeq were used to sequence 68 well-characterized INDELs in four major US population groups. In addition, the STR Allele Identification Tool: Razor (STRait Razor) was used in a novel way to analyze INDEL sequences and detect adjacent single nucleotide polymorphisms (SNPs) and other polymorphisms. This application enabled the discovery of unique allelic variants, which increased the discrimination power and decreased the single-locus random match probabilities (RMPs) of 22 of these well-characterized INDELs which can be considered as microhaplotypes. These findings suggest that additional microhaplotypes containing human identification (HID) INDELs may exist elsewhere in the genome.

1. Introduction

Short bi-allelic insertion/deletion (INDEL) polymorphisms are the second most abundant polymorphism, discovered to date, in humans and have been demonstrated as a useful alternative to traditional short tandem repeat (STR) testing in forensic genetics [1–6]. Due to their small amplicon size, INDELs are more advantageous than STRs for typing compromised DNA samples. The small difference in allele size potentially minimizes preferential amplification of smaller size alleles of a heterozygote, a more common occurrence with traditional STR testing. INDELs also have relatively low mutation rates and do not generate stutter products during PCR amplification. Lastly, the ease of multiplexing INDELs enables the development of panels with relatively low random match probabilities (RMPs) for human identity (HID) testing [7,8].

Massively parallel sequencing (MPS, also referred to as next generation sequencing (NGS)) is capable of targeting many loci, including those of forensic relevance, across the genomes of multiple samples simultaneously with relatively high sequence coverage [9–15]. With sequencing, it is possible to define INDELs better and potentially identify proximal single nucleotide polymorphisms (SNPs) that can increase the discrimination power of

* Corresponding author at: Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd, CBH–250, Fort Worth, TX,76107, USA.
E-mail address: Frank.Wendt@my.unthsc.edu (F.R. Wendt).

http://dx.doi.org/10.1016/j.fsigen.2016.09.005
1872-4973/

currently defined INDELs, i.e., by identifying INDEL-containing microhaplotypes. Herein, the Nextera™ Rapid Capture Custom Enrichment Kit was used to prepare DNA libraries that were sequenced on the Illumina MiSeq to generate sequence data for 68 well-described forensically relevant HID INDELs in four major US population groups. In addition, the STR Allele Identification Tool: Razor (STRait Razor) software [16] was used in a novel way to analyze INDEL sequences and detect adjacent SNPs. This application has enabled the discovery of unique allelic variation, which increases the discrimination power and decreases the single-locus random match probabilities of 22 of the INDELs. The results presented here demonstrate the utility of MPS for typing INDEL flanking regions to increase the discrimination power of current bi-allelic markers for HID testing.

2. Materials and methods

2.1. Samples and DNA extraction

DNA was extracted from whole blood and saliva samples obtained from 190 unrelated individuals following the University of North Texas Health Science Center Institutional Review Board Approval. The sample set represented unrelated individuals of four major U.S. population groups with 49 Caucasians (CAU), 49 African Americans (AFA), 49 Hispanics (HIS), and 43 Asians (ASA). DNA extraction was performed using the Qiagen® QIAamp™ DNA Blood Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol [17].

2.2. Library preparation and massively parallel sequencing

Libraries were generated using a custom designed Nextera™ Rapid Capture Enrichment panel (Illumina, Inc., San Diego, CA) using the Illumina Design Studio, as described by Zeng, et al. [18] and Warshauer, et al. [19]. Capture probe sequences will be made available upon request. The HID INDELs for this study were selected based on the results described by LaRue, et al. [3] and Pereira, et al. [6]. INDEL rs number, location, flanking region, and probe design are listed in Supplemental Table 1 [20]. 50 ng of genomic DNA were used as input for each library preparation reaction. Each sample library was diluted to 2 nM and paired-end sequencing (12 pooled libraries per run) was performed on the Illumina MiSeq according to the manufacturer's recommended protocol with a read length of 250 bases [21].

2.3. STRait Razor design

A configuration file was created for use with STRait Razor v2.5 (Supplemental Fig. 1 and Supplemental Table 2) [16]. To create the file, locus coordinates for each INDEL were located on the hg19 human reference genome using the Integrative Genomics Viewer (IGV) [22,23]. STRait Razor flanking regions up and downstream of the INDEL motif, and the complementary sequences, were recorded. The average size of the STRait Razor flank, used to mine sequence data for regions of interest, was 24 bases ± 0.10. The bases between STRait Razor flanks contained the INDEL motif and approximately 50 bases on either end. The STRait Razor flanks

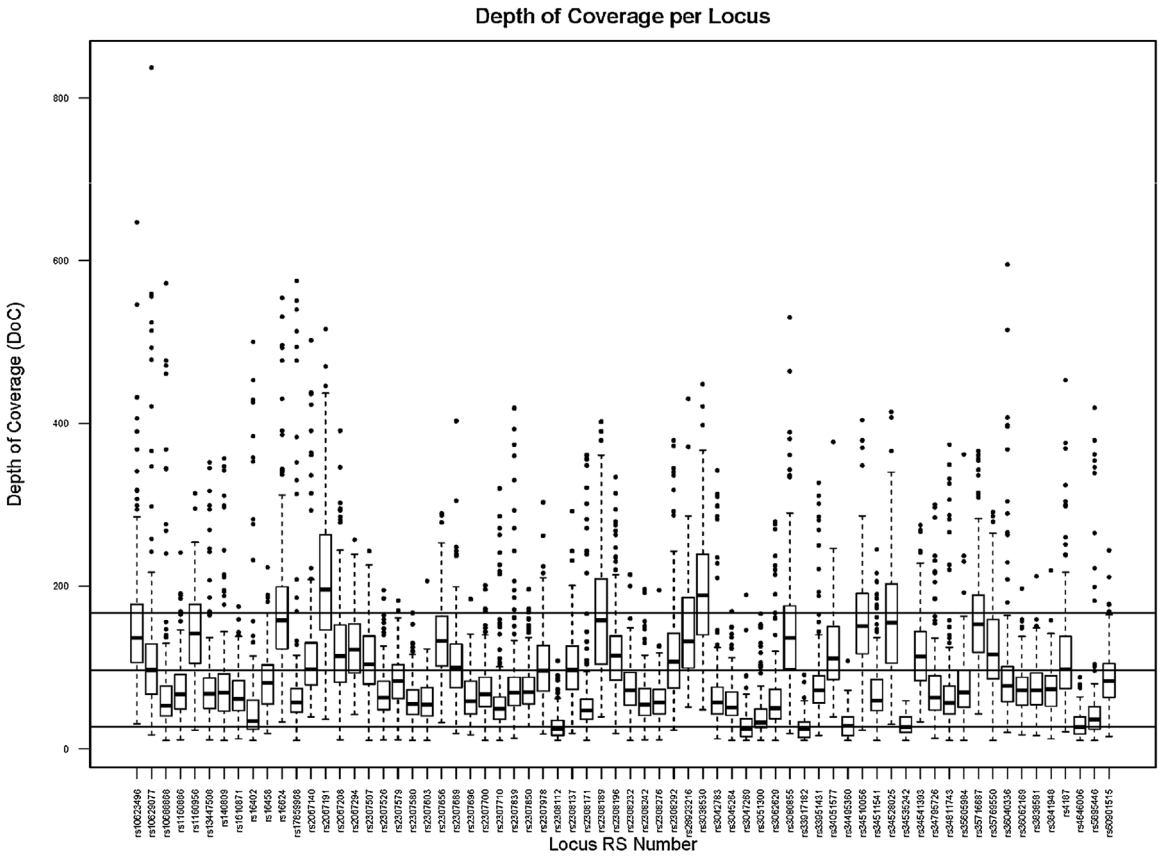

Fig. 1. Depth of coverage (DoC) values for 68 human identity INDELs using the Nextera™ Rapid Capture Enrichment kit and the Illumina MiSeq. Each box plot represents a single locus; the center horizontal line represents the median, the lower and upper boundaries of each box represent the first and third quartiles, respectively, the vertical dashed lines indicate plus/minus three times the interquartile range, and closed circles indicate outliers. The center horizontal line indicates the mean across 68 loci and the top and bottom horizontal lines indicate plus and minus one standard deviation, respectively for all loci combined.

were designed to capture sequence variation in the flanking regions adjacent to the target INDEL (Supplemental Tables 3 and 4) while keeping total target size relatively small. The average length of this region (target INDEL plus approximately 50 bases on either end) was 99 bases ± 4 and 102 bases ± 4 for the deletion and insertion alleles, respectively. Lastly, a relatively short sequence between the STRait Razor flanks, but unique relative to the INDEL motif, was recorded; the average length of these sequences was 12 bases ± 0.15. Analysis of the resulting data was performed using the STRait Razor Sequence Analysis toolkit to assign genotypes to each sample and compile depth of coverage (DoC) and allele coverage ratio (ACR) data. ACRs were calculated by dividing the DoC of one allele by the DoC of the second allele. An ACR of 1.0 was considered perfectly balanced. It should be noted that the hg19 reference genome was used to design the STRait Razor configuration file, however, the sequences within the file are identical to those in the hg38 reference genome.

2.4. Analysis Concordance

Sixty-nine of the samples were analyzed manually by a second reviewer to confirm STRait Razor allele calls. Fastq files were aligned using Burrows-Wheeler Aligner (BWA) and Sequence Alignment/Map Tools (SAMtools) [24–26]. The resulting binary alignment/map (.bam) files were used as input for the Genome Analysis Toolkit (GATK) [27]. The resulting variant call format (.vcf) files were analyzed using an in-house Excel-based workbook. The workbook assigned genotypes and compiled DoC and ACR data for each sample.

2.5. Population statistical analyses

Length-based and sequence-based allele frequencies, observed and expected heterozygosities, and testing for departures from Hardy-Weinberg Equilibrium (HWE) and linkage disequilibrium (LD) assessments were performed using Genetic Data Analysis (GDA) [28]. An in-house Excel-based workbook was used to generate power of discrimination values and single-locus and combined RMPs.

3. Results and discussion

A total of 190 samples were sequenced. One run, containing 11 African American samples and one Asian sample, performed poorly with insufficient sequencing Q scores (between 10 and 20) for all of read 2 and part of read 1. This run was removed from analysis due to poor sequence quality. Ultimately, 178 samples were analyzed, consisting of 48 Caucasians, 38 African Americans, 49 Hispanics, and 43 Asians.

3.1. Locus performance

Analysis of the resulting data was performed using operationally selected DoC and ACR thresholds of 10 x and 0.20, respectively. Mean profile completion was 96.3% ± 0.108, ranging from 44.1% to 100% for the 178 samples. Full HID INDEL profiles were obtained for 70 samples. The average DoC and ACR for 68 HID INDELs was 96.9x ± 69.9 and 0.727 ± 0.182, respectively (Figs. 1 and 2). One locus, rs33917182, fell below one standard deviation from the

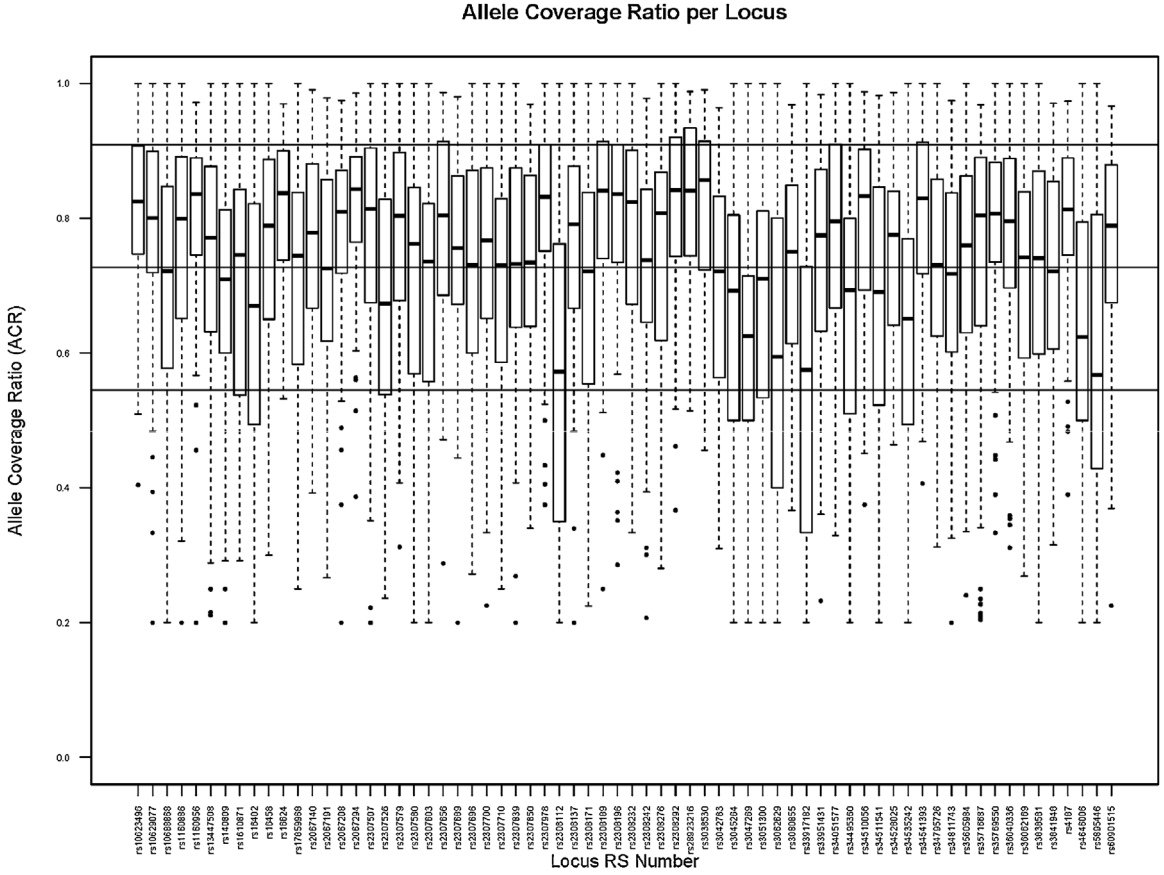

Fig. 2. Allele coverage ratio (ACR) values for 68 human identity INDELs using the Nextera™ Rapid Capture Enrichment kit and the Illumina MiSeq. Each box plot represents a single locus; the center horizontal line represents the median, the lower and upper boundaries of each box represent the first and third quartiles, respectively, the vertical dashed lines indicate plus/minus three times the interquartile range, and closed circles indicate outliers. The center horizontal line indicates the mean across 68 loci and the top and bottom horizontal lines indicate plus and minus one standard deviation, respectively for all loci combined.

Table 1
Length-based (LBAF) and sequence-based (SBAF) allele frequencies by population for 68 insertion/deletion (INDEL) markers. Target INDEL motifs are underlined and flanking region variants bolded.

INDEL RS Number	Flanking RS Number (s) and hg19 Reference Allele	Length (bp)	Sequence	AFA (N = 38)		ASA (N = 43)		CAU (N = 48)		HIS (N = 49)	
				LBAF	SBAF	LBAF	SBAF	LBAF	SBAF	LBAF	SBAF
rs10623496	chr8:123945676 T[b]	100	TTAAACATTTGATAGTGCCTTATTTATTTATGTATGTGACACATAAAACCATGATTGTTTCTTCTTTCTGTCTTAGCAAGA**T**TTTTTTTTCTGCTTTCAG	0.3816	0.3816	0.3372	0.3256	0.3438	0.3438	0.3061	0.3061
		100	TTAAACATTTGATAGTGCCTTATTTATTTATGTATGTGACACATAAAACCATGATTGTTTCTTCTTTCTGTCTTAGCAAGA**C**TTTTTTTTCTGCTTTCAG		0		0.0116		0		0
		104	TTAAACATTTGATAGTGCCTTATTTATTTATGTATGTGACACATAAAACCAGAATTGATTGTTTCTTCTTTCTGTCTTAGCAAGA**T**TTTTTTTTCTGCTTTCAG	0.6184	0.6184	0.6628	0.6628	0.6563	0.6563	0.6939	0.6939
rs10629077	rs537464320C	100	TTGGTTATCTCTGTCTAATCTA**C**TCTCTTGTGCCCACTTTTACTTACTCA**C**GTCTTTTCCCAACAGCATTCTGTCACACCTCCTAATAGTTTTGCTCATC	0.2632	0.25	0.2791	0.2791	0.1667	0.1563	0.2755	0.2755
	rs201421087C	100	TTGGTTATCTCTGTCTAATCTA**C**TCTCTTGTGCCCACTTTTACTTACTCA**T**GTCTTTTCCCAACAGCATTCTGTCACACCTCCTAATAGTTTTGCTCATC		0.0132		0		0		0
		100	TTGGTTATCTCTGTCTAATCTA**T**TCTCTTGTGCCCACTTTTACTTACTCA**C**GTCTTTTCCCAACAGCATTCTGTCACACCTCCTAATAGTTTTGCTCATC		0		0		0.0104		0
		102	TTGGTTATCTCTGTCTAATCTA**C**TCTCTTGTGCCCACTTTTACTTACTCA**C**ATGTCTTTTCCCAACAGCATTCTGTCACACCTCCTAATAGTTTTGCTCATC	0.7368	0.7368	0.7209	0.7209	0.8333	0.8333	0.7245	0.7245
rs10688868[c]	rs142634555C	100	CCTGTTCCTCGCTCTAGTTCCCCACTTCCATCCCTCCTCTTG**C**CTCAGCCCTT**C**TCATCTCACACGCCACATGGGATCCACCCTTTTTA**T**GCATGTGCAG	0.1974	0.0658	0.5698	0.4767	0.3229	0.0729	0.3673	0.1531
	rs56780729C	100	CCTGTTCCTCGCTCTAGTTCCCCACTTCCATCCCTCCTCTTG**C**CTCAGCCCTT**C**TCATCTCACACGCCACATGGGATCCACCCTTTTTA**C**GCATGTGCAG		0.1316		0.093		0.25		0.2143
	rs10902117C	102	CCTGTTCCTCGCTCTAGTTCCCCACTTCCATCCCTCCTCTTG**C**CTCAGCCCCTTT**C**TCATCTCACACGCCACATGGGATCCACCCTTTTTA**T**GCATGTGCAG	0.8026	0.4342	0.4302	0.4186	0.6771	0.5	0.6327	0.4082
		102	CCTGTTCCTCGCTCTAGTTCCCCACTTCCATCCCTCCTCTTG**C**CTCAGCCCCTTTT**T**CATCTCACACGCCACATGGGATCCACCCTTTTTA**C**GCATGTGCAG		0.3553		0.0116		0.1667		0.2245
		102	CCTGTTCCTCGCTCTAGTTCCCCACTTCCATCCCTCCTCTTG**T**CTCAGCCCCTTTT**T**CATCTCACACGCCACATGGGATCCACCCTTTTTA**C**GCATGTGCAG		0		0		0.0104		0
		102	CCTGTTCCTCGCTCTAGTTCCCCACTTCCATCCCTCCTCTTG**C**CTCAGCCCCTTT**C**TCATCTCACACGCCACATGGGATCCACCCTTTTTA**C**GCATGTGCAG		0.0132		0		0		0
rs1160886[a]	–	97	CAACTCTATTCCTTTTCCCATTGTGCTTAAAACTCCTTTGGAATAGTACTGTTTTCTCTATACTTGGAATACATTTGGGTAGACTGGACTATTTCTC	0.3947		0.4535		0.4063		0.3372	
		100	CAACTCTATTCCTTTTCCCATTGTGCTTAAAACTCCTTTGGAATAGTACTACTGTTTTCTCTATACTTGGAATACATTTGGGTAGACTGGACTATTTCTC	0.6053		0.5465		0.5938		0.6628	
rs1160956	rs138536239T	101	CAAATTGTGTTCCTCCAAGGTATAGCTTTAGAAGGATCTTCTTT**C**AGTTGCTCTTTCAAACTTTTCCTTGTTAGAAAAGAAAAACTAATACAAATTGGTTA	0.5526	0.5526	0.6047	0.6047	0.8125	0.8125	0.6125	0.6125
		104	CAAATTGTGTTCCTCCAAGGTATAGCTTTAGAAGGATCTTCTTT**C**AGTTGCTCTTTCTTTCAAACTTTTCCTTGTTAGAAAAGAAAAACTAATACAAATTGGTTA	0.4474	0.4342	0.3953	0.3953	0.1875	0.1875	0.3875	0.3875
		104	CAAATTGTGTTCCTCCAAGGTATAGCTTTAGAAGGATCTTCTT**G**CAGTTGCTCTTCTTTCAAACTTTTCCTTGTTAGAAAAGAAAAACTAATACAAATTGGTTA		0.0132		0		0		0
rs13447508	rs13447507A	91	AATGTACATTATTAGATGTACT**A**TGGTTCAGTGGAAATAGCATGAACTAACAGAGTC**T**AATTTTTGACCTTAGACATGTCTCTTATCTCTG		0		0.0116		0		0
	rs201219895 DEL	94	AATGTACATTATTAGATGTACT**A**TGGTTCAGTGGAAATAGCATGAACTAACAGAGTC**TAA**TAATTTTTGACCTTAGACATGTCTCTTATCTCTG	0.3553	0.3553	0.5116	0.5	0.2917	0.2917	0.3854	0.3854
		100	AATGTACATTATTAGATGTACT**A**TGGTTCAGTGGAAATAGCATGAACTAACAGACTTAGAGTC**TAA**TAATTTTTGACCTTAGACATGTCTCTTATCTCTG	0.6447	0.6184	0.4884	0.4884	0.7083	0.7083	0.6146	0.6146
		100	AATGTACATTATTAGATGTACT**G**TGGTTCAGTGGAAATAGCATGAACTAACAGACTTAGAGTC**TAA**TAATTTTTGACCTTAGACATGTCTCTTATCTCTG		0.0263		0		0		0
rs140809[c]	rs10905513A	97	ATGTTCTTAGCCATGGA**A**TTCTTTAGGCTTAATTTTACTTCCAGTTAAATGCAGC**TGCTGT**GGTCACTCAGGAGGGGGATGGGCACCCAGAGTTCCT	0.4079	0.3421	0.3605	0.186	0.4788	0.4043	0.1633	0.0816
	rs687805C	97	ATGTTCTTAGCCATGGA**A**TTCTTTAGGCTTAATTTTACTTCCAGTTAAATGCAGC**C**GCTGTGGTCACTCAGGAGGGGGATGGGCACCCAGAGTTCCT		0.0526		0.1628		0.0745		0.0612
		97	ATGTTCTTAGCCATGG**AG**TTCTTTAGGCTTAATTTTACTTCCAGTTAAATGCAGC**T**GCTGTGGTCACTCAGGAGGGGGATGGGCACCCAGAGTTCCT		0.0132		0.0116		0		0.0204
		100	ATGTTCTTAGCCATGGA**A**TTCTTTAGGCTTAATTTTACTTCCAGTTAACAAATGCAGC**C**GCTGTGGTCACTCAGGAGGGGGATGGGCACCCAGAGTTCCT	0.5921	0.4474	0.6395	0.4651	0.5212	0.4574	0.8367	0.7347
		100	ATGTTCTTAGCCATGGA**G**TTCTTTAGGCTTAATTTTACTTCCAGTTAACAAATGCAGC**C**GCTGTGGTCACTCAGGAGGGGGATGGGCACCCAGAGTTCCT		0.0263		0.1628		0.0638		0.0918
		100	ATGTTCTTAGCCATGG**AG**TTCTTTAGGCTTAATTTTACTTCCAGTTAACAAATGCAGC**T**GCTGTGGTCACTCAGGAGGGGGATGGGCACCCAGAGTTCCT		0.1053		0		0		0.0102
		100	ATGTTCTTAGCCATGGA**A**TTCTTTAGGCTTAATTTTACTTCCAGTTAACAAATGCAGC**T**GCTGTGGTCACTCAGGAGGGGGATGGGCACCCAGAGTTCCT		0.0132		0.0116		0		0
rs1610871[c]	rs111817892G	96	TGCAGATAGCCTCACCTTCCTCCCAGTAACCATCAA**G**CCCCATGAAGAAGGAGTAGCAGGGGAAATGGAGTCCACTAAAAGG**CC**GAAGCCC**T**CGGC	0.5132	0.5132	0.3721	0.3721	0.5833	0.5833	0.4744	0.4744
	rs75866020C	100	TGCAGATAGCCTCACCTTCCTCCCAGTAACCATCAA**G**CCCCATGAAGTAGGAAGGAGTAGCAGGGGAAATGGAGTCCACTAAAAGGC**C**GAAGCCC**T**CGGC	0.4868	0.3158	0.6279	0.593	0.4167	0.4167	0.5256	0.5128
	chr5:171088015 T[b]	100	TGCAGATAGCCTCACCTTCCTCCCAGTAACCATCAA**G**CCCCATGAAGTAGGAAGGAGTAGCAGGGGAAATGGAGTCCACTAAAAGGC**G**GAAGCCC**T**CGGC		0.1053		0.0349		0		0.0128
		100	TGCAGATAGCCTCACCTTCCTCCCAGTAACCATCAA**A**CCCCATGAAGTAGGAAGGAGTAGCAGGGGAAATGGAGTCCACTAAAAGGC**C**GAAGCCC**T**CGGC		0.0526		0		0		0
		100	TGCAGATAGCCTCACCTTCCTCCCAGTAACCATCAA**G**CCCCATGAAGTAGGAAGGAGTAGCAGGGGAAATGGAGTCCACTAAAAGGC**C**GAAGCCC**C**CGGC		0.0132		0		0		0
rs16402	–	100	GTTAATGCGCCTTTTTGGTTTTGGTGAAAATTTCAGACAGTGTGACATTAATTATTTGTCTTTGCACAGAATGATGATGCTAAAGCCAGCCCCATACAGT	0.3158		0.2326		0.25		0.2396	
		104	GTTAATGCGCCTTTTTGGTTTTGGTGAAAATTTCAGACAGTGTGACATTAATTATTTATTTGTCTTTGCACAGAATGATGATGCTAAAGCCAGCCCCATACAGT	0.6842		0.7674		0.75		0.7604	
rs16458	–	100	GGTTAATCTTCCCCCAAAAACCTACCAATTAACAGTCTCAAAGTTTTACAATTCCTTCATTTCCAAACTCTCTTATGAGTAATAATCAAAGTATACATAT	0.5263		0.6047		0.6875		0.5	
		104	GGTTAATCTTCCCCCAAAAACCTACCAATTAACAGTCTCAAAGTTTTACAATTCCTTCCTTCATTTCCAAACTCTCTTATGAGTAATAATCAAAGTATACATAT	0.4737		0.3953		0.3125		0.5	
rs16624	rs146701576C	100	TCCCCTTTCCACCACAGTTACCTTTCAGGCTCTTGGGTTCTGGAAAGATGTTTTTGACATTGCAACTTGGAAGTCACA**AT**G**T**ATTTTTTCACAAAACAGTG	0.25	0.25	0.4419	0.4302	0.7812	0.7708	0.4744	0.4286
	rs140263477A	100	TCCCCTTTCCAC**T**ACAGTTACCTTTCAGGCTCTTGGGTTCTGGAAAGATGTTTTTGACATTGCAACTTGGAAGTCACA**ATG**TATTTTTCACAAAACAGTG		0		0		0.0104		0.0102
	rs250921T	100	TCCCCTTTCCAC**C**ACAGTTACCTTTCAGGCTCTTGGGTTCTGGAAAGATGTTTTTGACATTGCAACTTGGAAGTCACA**ATG**CATTTTTCACAAAACAGTG		0		0.0116		0		0
		102	TCCCCTTTCCAC**C**ACAGTTACCTTTCAGGCTCTTGGGTTCTGGAAAGATGTGTTTTTGACATTGCAACTTGGAAGTCACA**ATG**CATTTTTCACAAAACAGTG	0.75	0.7237	0.5581	0.5581	0.2188	0.2188	0.5256	0.5612
		102	TCCCCTTTCCA**C**CACAGTTACCTTTCAGGCTCTTGGGTTCTGGAAAGATGTGTTTTTGACATTGCAACTTGGAAGTCACA**G**TGC**A**TTTTTCACAAAACAGTG		0.0263		0		0		0
rs17859968[a,c]	rs9923304C	103	ATTCAGAGTGCATGCTGTCTTA**A**AAAGATTGTGGGGATTAAATTACAAGGCACATAAAGCACATGGCATTTAGTAGGACCCTCAATAAATGATAACTCTTACTC	0.3421	0.2237	0.3488	0.2674	0.4375	0.4375	0.4184	0.3878
	rs16955268A	103	ATCCAGAGTGCATGCTGTCTTA**A**AAAGATTGTGGGGATTAAATTACAAGGCACATAAAGCACATGGCATTTAGTAGGACCCTCAATAAATGATAACTCTTACTC		0.1184		0.0814		0		0.0306
		107	ATCCAGAGTGCATGCTGTCTTA**A**AAAGATTGTGGGGATTAAATTACAAGGCACATAAATAAAGCACATGGCATTTAGTAGGACCCTCAATAAATGATAACTCTTACTC	0.6579	0.6316	0.6512	0.6512	0.5625	0.5625	0.5816	0.5816
		107	ATCCAGAGTGCATGCTGTCTT**G**AAAGATTGTGGGGATTAAATTACAAGGCACATAAATAAAGCACATGGCATTTAGTAGGACCCTCAATAAATGATAACTCTTACTC		0.0263		0		0		0
rs2067140[c]	rs192851878T	96	AAAGGAAAATACAGGCA**T**GCCAATCACTACCCACCAAATGTC**A**CTGACACCAAAGAATCACTGATTAACCTGGAGAGCACTGCAAGGTGAGCTATA	0.2237	0.2237	0.6512	0.6512	0.6146	0.6146	0.6224	0.6224
	rs254233C	100	AAAGGAAAATACAGGCA**T**GCCAATCACTACCCACCAAATGTC**CC**TGACACCAGTCAAAGAATCACTGATTAACCTGGAGAGCACTGCAAGGTGAGCTATA	0.7763	0.4605	0.3488	0.1395	0.3854	0.2292	0.3776	0.1939
		100	AAAGGAAAATACAGGCA**T**GCCAATCACTACCCACCAAATGTC**A**CTGACACCAGTCAAAGAATCACTGATTAACCTGGAGAGCACTGCAAGGTGAGCTATA		0.3158		0.1977		0.1563		0.1837
		100	AAAGGAAAATACAGGCA**AG**CCAATCACTACCCACCAAATGTC**A**CTGACACCAGTCAAAGAATCACTGATTAACCTGGAGAGCACTGCAAGGTGAGCTATA		0		0.0116		0		0
rs2067191	rs115923419C	91	GGATTAGAGTAATGTAAGTAATCTGATGAAATTTACCACTTCTAGTTATTTCCTTGTTTATGAACTACAGAATC**C**GATTACCCTGAAATTC	0.4211	0.4211	0.4302	0.4302	0.4896	0.4896	0.3523	0.3523
		95	GGATTAGAGTAATGTAAGTAATCTGATGAAATTTACCACTTCTAGATAGTTATTTCCTTGTTTATGAACTACAGAATC**C**GATTACCCTGAAATTC	0.5789	0.5658	0.5698	0.5698	0.5104	0.5104	0.6477	0.6477
		95	GGATTAGAGTAATGTAAGTAATCTGATGAAATTTACCACTTCTAGATAGTTATTTCCTTGTTTATGAACTACAGAATC**T**GATTACCCTGAAATTC		0.0131		0		0		0

Table 1 (*Continued*)

INDEL RS Number	Flanking RS Number (s) and hg19 Reference Allele	Length (bp)	Sequence	AFA (N = 38)		ASA (N = 43)		CAU (N = 48)		HIS (N = 49)	
				LBAF	SBAF	LBAF	SBAF	LBAF	SBAF	LBAF	SBAF
rs2067208[a,c]	rs74531425G	95	GGGGCTCAGGCAGCTGAAGAGAATGTTCTAGAATCCACAAAGAGCCTGGCAGGAGCCTGGGGACCTGGA**G**ATGCAGCCGGGTGCATCTATTCAGG	0.1974	0.1974	0.1744	0.1744	0.3125	0.3125	0.3061	0.3061
		100	GGGGCTCAGGCAGCTGAAGAGAATGTTCTAGAATCCACAAAGAGCCTGGCCTGGCAGGAGCCTGGGGACCTGGA**G**ATGCAGCCGGGTGCATCTATTCAGG	0.8026	0.8026	0.8256	0.7907	0.6875	0.5417	0.6939	0.6429
		100	GGGGCTCAGGCAGCTGAAGAGAATGTTCTAGAATCCACAAAGAGCCTGGCCTGGCAGGAGCCTGGGGACCTGGA**C**ATGCAGCCGGGTGCATCTATTCAGG		0		0.0349		0.1458		0.051
rs2067294	–	100	GAAATGCAGCTTGATTCTCTATGCTATCCCGGCAATTCTCTATTTTTTTGTCCTTTCTCTCCTCCCATTCTGATATTCTCTTGCTTCTTGTTCACTCCTCGA	0.1316		0.2326		0.3333		0.5	
		103	GAAATGCAGCTTGATTCTCTATGCTATCCCGGCAATTCTCTATTTTTTTGTCCTTCTTTCTCTCCTCCATTCTGATATTCTCTTGCTTCTTGTTCACTCCTCGA	0.8684		0.7674		0.6667		0.5	
rs2307507	rs540604306G	95	TTTTAAAATATGTTATATTTTAGGGCTTTGAAATTACT**G**AAAATAATTTAATGTTACATAGTTTTAAATCACTTTCATAGCTTAATAGGGTTTTA	0.1974	0.1842	0.3293	0.3293	0.4271	0.4271	0.4512	0.4512
		95	TTTTAAAATATGTTATATTTTAGGGCTTTGAAATTACT**A**AAAATAATTTAATGTTACATAGTTTTAAATCACTTTCATAGCTTAATAGGGTTTTA		0.0132		0		0		0
		100	TTTTAAAATATGTTATATTTTAGGGCTTTGAAATTACTGAAAATAATTTAATTTTATGTTACATAGTTTAAATCACTTTCATAGCTTAATAGGGTTTTA	0.8026	0.8026	0.6707	0.6707	0.5729	0.5729	0.5488	0.5488
rs2307526[c]	rs814781C	100	TAATGCCAGGAGATAATTTATATACAAAGCAAAGGATGCTCAACAACATACGTAATGGGCTTCCGAATTACAAAGAACAGAATCAGGACT**C**AAGCAGATG	0.3684	0.3684	0.6341	0.6341	0.4688	0.4688	0.2955	0.2955
		104	TAATGCCAGGAGATAATTTATATACAAAGCAAAGGATGCTCAACAACATACGTGTGTAATGGGCTTCCGAATTACAAAGAACAGAATCAGGACT**G**AAGCAGATG	0.6316	0.3684	0.3659	0.3171	0.5313	0.4063	0.7045	0.4886
		104	TAATGCCAGGAGATAATTTATATACAAAGCAAAGGATGCTCAACAACATACGTGTGTAATGGGCTTCCGAATTACAAAGAACAGAATCAGGACT**C**AAGCAGATG		0.2682		0.0488		0.125		0.2159
rs2307579	rs77911955G	100	TAAAATTGTCACTGTACTAAAAATACCATTAATAATAAAGTGTGGAAAGACATATTCAGCTCTCTTCAAAAGTAGCTATTT**G**GTTATTTAAAGTTAGCTC	0.5921	0.5921	0.1744	0.1744	0.4062	0.4062	0.35	0.35
		103	TAAAATTGTCACTGTACTAAAAATACCATTAATAATAAAGTGTGGAAAGACATGATATTCAGCTCTCTTCAAAAGTAGCTATTT**G**GTTATTTAAAGTTAGCTC	0.4079	0.3947	0.8256	0.8256	0.5938	0.5938	0.65	0.65
		103	TAAAATTGTCACTGTACTAAAAATACCATTAATAATAAAGTGTGGAAAGACATGATATTCAGCTCTCTTCAAAAGTAGCTATTT**C**GTTATTTAAAGTTAGCTC		0.0132		0		0		0
rs2307580	–	100	TGGTTCTTAACCTCTGAAATTAATTTTTATATTTCTAAATTGCTTAATTGAATAGTTATGCTATGGTCCTGTTCCATATCTAGATACCTCCCAAGTATTC	0.2632		0.4286		0.4375		0.5513	
		104	TGGTTCTTAACCTCTGAAATTAATTTTTATATTTCTAAATTGCTTAATTAATTGAATAGTTATGCTATGGTCCTGTTCCATATCTAGATACCTCCCAAGTATTC	0.7368		0.5714		0.5625		0.4487	
rs2307603[a]	–	95	GAAAAAAGTCAGTGTGCCTACAGATACCACTTTATTTTGTCCACTTAGAACTAGGTTGCTATCTTTAGCTTAATTCTAGCATATGAGAAATGTGA	0.3947		0.4419		0.6146		0.3846	
		100	GAAAAAAGTCAGTGTGCCTACAGATACCACTTTATTTTGTCCACTTAGATCAGAACTAGGTTGCTATCTTTAGCTTAATTCTAGCATATGAGAAATGTGA	0.6053		0.5581		0.3854		0.6154	
rs2307656[c]	rs13158027T	95	AAATGTTGCTTCCC**T**ATCTACACAACCTGGTTTGCCCACGTGCCTCTTGATAATTAGAATAGAACTCTAGAAGGATATTAGCATGTCATTTTTCA	0.6184	0.6184	0.5581	0.5581	0.5	0.5	0.6429	0.6429
		100	AAATGTTGCTTCCC**C**ATCTACACAACCTGGTTTGCCCACGTGCCTCTTGATAAGTTAATTAGAATAGAACTCTAGAAGGATATTAGCATGTCATTTTTCA	0.3816	0.2237	0.4419	0.1163	0.5	0.3542	0.3571	0.2262
		100	AAATGTTGCTTCCC**T**ATCTACACAACCTGGTTTGCCCACGTGCCTCTTGATAAGTTAATTAGAATAGAACTCTAGAAGGATATTAGCATGTCATTTTTCA		0.1579		0.3256		0.1458		0.131
rs2307689[c]	rs36120065A	86	CTTCCCCAAGGCCAAACCTCCTCCCTCAGAAGAGGGCTGTTCTTCCTCAAGTTACTAAGCCTCCTTCTTCAAACGAGGGG**A**TACCC	0.5526	0.5526	0.2674	0.2674	0.2188	0.2188	0.3605	0.3605
		89	CTTCCCCAAGGCCAAACCTCCTCCCTCAGAAGAGGGCTGTTCTTCTTCCTCAAGTTACTAAGCCTCCTTCTTCAAACGAGGGG**A**TACCC	0.4474	0.2895	0.7326	0.407	0.7812	0.5	0.6395	0.3488
		89	CTTCCCCAAGGCCAAACCTCCTCCCTCAGAAGAGGGCTGTTCTTCTTCCTCAAGTTACTAAGCCTCCTTCTTCAAACGAGGGG**G**TACCC		0.1579		0.3256		0.2813		0.2907
rs2307696	–	96	ACACTACACAACAGAAACAATAACGAATCCCCCCCATCACACACACACAGCGACCTGAGGGTTCTGAAGAGTAAATAAGAGCAGTCAGATGGGCTC	0.5263		0.3023		0.4164		0.4459	
		100	ACACTACACAACAGAAACAATAACGAATCCCCCCCATCACACACACACAGCGACCGACCTGAGGGTTCTGAAGAGTAAATAAGAGCAGTCAGATGGGCTC	0.4737		0.6977		0.5833		0.5541	
rs2307700[c]	rs4239922G	106	AACCTGGGACAGCTGGCAGGGCCAGAGCTCAGCT**G**ATCATTCATGGGCTTGTCAGTCAGTAAAAGAATGGGGATTCCGAGGAAGGACAAACAAAGGAAAGAAAAAA	0.3421	0.3421	0.2674	0.2674	0.5417	0.5417	0.3125	0.3125
		110	AACCTGGGACAGCTGGCAGGGCCAGAGCTCAGCT**C**ATCATTCATGGGCTTGTCACTCAGTCAGTAAAAGAATGGGGATTCCGAGGAAGGACAAACAAAGGAAAGAAAAAA	0.6579	0.3816	0.7326	0.6744	0.4583	0.2396	0.6875	0.5521
		110	AACCTGGGACAGCTGGCAGGGCCAGAGCTCAGCT**G**ATCATTCATGGGCTTGTCACTCAGTCAGTAAAAGAATGGGGATTCCGAGGAAGGACAAACAAAGGAAAGAAAAAA		0.2763		0.0581		0.2188		0.1354
rs2307710	–	100	CATTTGCTAAATTTGCAGAGCCCATACCTACTGTGACCAGAGAAGGAAGGAGCCTGAGATGTCATTAATTGAATTCATTCATTTTGTAGACAAGCCTAAC	0.4474		0.2674		0.375		0.2041	
		104	CATTTGCTAAATTTGCAGAGCCCATACCTACTGTGACCAGAGAAGGAAGGAAGGAGCCTGAGATGTCATTAATTGAATTCATTCATTTTGTAGACAAGCCTAAC	0.5526		0.7326		0.625		0.7959	
rs2307839	–	100	CTAAGAAATGTTTTAAGAAAAATTGTCAGCTAATAATGGTTATTCTTATTTGGAAATATATATACATAGATAAATACGTAAAGAAATGCACAGAGAAAGACAC	0.1892		0.4302		0.2396		0.2755	
		102	CTAAGAAATGTTTTAAGAAAAATTGTCAGCTAATAATGGTTATTCTTATTTCTGGAAATATATATACATAGATAAATACGTAAAGAAATGCACAGAGAAAGACAC	0.8108		0.5698		0.7604		0.7245	
rs2307850[a]	–	96	TCTGGAAAAGCTTTTACTGAGATGCCCAACCTGCGTGGAAAGCCTCCACCTTGCAGCATAAGTGCGGAGGAAACGGTGACAAAATCATTGCCCCGC	0.4211		0.2907		0.3438		0.3889	
		100	TCTGGAAAAGCTTTTACTGAGATGCCCAACCTGCGTGGAAAGCCTCCACCCACCTTGCAGCATAAGTGCGGAGGAAACGGTGACAAAATCATTGCCCCGC	0.5789		0.7093		0.6563		0.6111	
rs2307978[c]	rs188547G	105	CTAAGAAATGTTTTAAGAAAAATTGTCAGCTAATAATGGTTATTCTTATTTGGAAATATATATACATAGATAAATACGTAAAGAAATGCACAGAGAAAGACA**C**	0.3947	0.3947	0.407	0.407	0.1667	0.1667	0.3553	0.3553
		107	CTAAGAAATGTTTTAAGAAAAATTGTCAGCTAATAATGGTTATTCTTATTTCTGGAAATATATATACATAGATAAATACGTAAAGAAATGCACAGAGAAAGACA**C**	0.6053	0.3421	0.593	0.3256	0.8333	0.4792	0.6447	0.4079
		107	CTAAGAAATGTTTTAAGAAAAATTGTCAGCTAATAATGGTTATTCTTATTTCTGGAAATATATATACATAGATAAATACGTAAAGAAATGCACAGAGAAAGACA**G**		0.2632		0.2674		0.3542		0.2368
rs2308112	–	95	ATCCACAGAAGAGGCGGTGCTGATGGACTGGACCCCAGAGAGTCCTGCTTGGTGCCATACCTTGGGTCCCTCCAGATCAGGATGCTCCCAGCAAC	0.5263		0.5		0.6146		0.5256	
		100	ATCCACAGAAGAGGCGGTGCTGATGGACTGGACCCCAGAGAGTCCACACCTGCTTGGTGCCATACCTTGGGTCCCTCCAGATCAGGATGCTCCCAGCAAC	0.4737		0.5		0.3854		0.4744	
rs2308137	–	98	GGCAAGTTGGGAGTTTCCCGTTGAGCTTTGCTTCCCGTGAAAACAACCATGAGAACTCCAGAAATGGCCTTCTCAGGCCAGGCTTCGTGACTTTCAGG	0.6974		0.4535		0.3021		0.2692	
		100	GGCAAGTTGGGAGTTTCCCGTTGAGCTTTGCTTCCCGTGAAAACAACCATGAGAGAACTCCAGAAATGGCCTTCTCAGGCCAGGCTTCGTGACTTTCAGG	0.3026		0.5465		0.6979		0.7308	
rs2308171[a]	–	100	GGTAAGTGTTGCAACATCTGCCACAATTTAAAACACTTACTCAGGTGCCTGCTCAGTGCCTATGCACTAAGTCAAAGAGGCACAGGCCAGCTGTCCAGAA	0.4737		0.0814		0.2234		0.1771	
		104	GGTAAGTGTTGCAACATCTGCCACAATTTAAAACACTTACTCAGGTGCCTGTCTGCTCAGTGCCTATGCACTAAGTCAAAGAGGCACAGGCCAGCTGTCCAGAA	0.5263		0.9186		0.7766		0.8229	
rs2308189[a,c]	rs176295C	99	**T**ACATCCTGCTCCTTTCACCTGGAACCTGATTCA**C**GCTGCTGGAATCATTAGTTGTTTCGAATACTGTTTTCAATTTTGTTCTTGTCGTTCCTTTGTTT	0.5395	0.1711	0.4643	0.4643	0.3511	0.3511	0.5541	0.5405
	rs176294T	99	**C**ACATCCTGCTCCTTTCACCTGGAACCTGATTCA**C**GCTGCTGGAATCATTAGTTGTTTCGAATACTGTTTTCAATTTTGTTCTTGTCGTTCCTTTGTTT		0.3158		0		0		0.0135
		99	**T**ACATCCTGCTCCTTTCACCTGGAACCTGATTCA**T**GCTGCTGGAATCATTAGTTGTTTCGAATACTGTTTTCAATTTTGTTCTTGTCGTTCCTTTGTTT		0.0526		0		0		0
		104	**C**ACATCCTGCTCCTTTCACCTGGAACCTGATTCA**C**GCTGCTGGAATCATTAGTTTAGTTGTTTCGAATACTGTTTTCAATTTTGTTCTTGTCGTTCCTTTGTTT	0.4605	0.1447	0.5357	0.3333	0.6489	0.2872	0.4459	0.2973
		104	**C**ACATCCTGCTCCTTTCACCTGGAACCTGATTCA**T**GCTGCTGGAATCATTAGTTTAGTTGTTTCGAATACTGTTTTCAATTTTGTTCTTGTCGTTCCTTTGTTT		0.3158		0.2024		0.3617		0.1486
rs2308196	–	100	TTGAGAGATAAGTGTCTTGTAAAAATGATTTTGTGTTTCTGCAAACTTGTTACTGGAGACCATGGAATTTCCTTTTTCTTTTCTTTCACTGTGGTGGAAAA	0.6842		0.6512		0.5625		0.6531	
		104	TTGAGAGATAAGTGTCTTGTAAAAATGATTTTGTGTTTCTGCAAACTTGATTGTTACTGGAGACCATGGAATTTCCTTTTTCTTTTCTTTCACTGTGGTGGAAAA	0.3158		0.3488		0.4375		0.3469	
rs2308232[c]	rs1093240C	100	TTAATCTAAATGGTGAATGATTGATGCAATCTCACACTAC**C**TTTGAAAAGCTATTCGTCTAGCTCTTGATCAAATTTAAGATCATATAATACGAAGCAAG	0.25	0.2368	0.2683	0.0976	0.2979	0.2021	0.2895	0.0921
		100	TTAATCTAAATGGTGAATGATTGATGCAATCTCACACTAC**T**TTTGAAAAGCTATTCGTCTAGCTCTTGATCAAATTTAAGATCATATAATACGAAGCAAG		0.0132		0.1707		0.0957		0.1974
		106	TTAATCTAAATGGTGAATGATTGATGCAATCTCACACTAC**T**TTTGAAAAGCTAAACTTATTCGTCTAGCTCTTGATCAAATTTAAGATCATATAATACGAAGCAAG	0.75	0.75	0.7317	0.7317	0.7021	0.7021	0.7105	0.7105
rs2308242	chr3:8616681 T[b]	100	AGGAGAGCTCTGCGGAGTTCATT**T**TCTGCTGATAAGGGGACAGGGAGAAAGCAGCACGGGCAGTCATCCTTGGCCCACACAAGGGGACCATCCAGTGCCAA	0.3289	0.3289	0.2674	0.2674	0.2128	0.2128	0.225	0.225
		102	AGGAGAGCTCTGCGGAGTTCATT**T**TCTGCTGATAAGGGGACAGGGAGAAAGAGCAGACGGGCAGTCATCCTTGGCCCACACAAGGGGACCATCCAGTGCCAA	0.6711	0.6711	0.7326	0.7326	0.7872	0.7766	0.775	0.775
		102	AGGAGAGCTCTGCGGAGTTCATCTTCTGCTGATAAGGGGACAGGGAGAAAGAGCAGACGGGCAGTCATCCTTGGCCCACACAAGGGGACCATCCAGTGCCAA		0		0		0.0106		0
rs2308276[a,c]	rs10209911T	100	CAAGTATATACCATTTCATTGTGACTCAATACTTTATAAAGATGAATTTAAATTTTGAAGTAAACTT**T**TTTTGTCTCAAATGGAGAAATTCCATGCAATA	0.5526	0.5526	0.407	0.407	0.4792	0.4792	0.4487	0.4487
		105	CAAGTATATACCATTTCATTGTGACTCAATACTTTATAAAGATGAATTTAATTTAAATTTTGAAGTAAACTT**T**TTTTGTCTCAAATGGAGAAATTCCATGCAATA	0.4474	0.4079	0.593	0.1977	0.5208	0.5208	0.5513	0.5128
		105	CAAGTATATACCATTTCATTGTGACTCAATACTTTATAAAGATGAATTTAATTTAAATTTTGAAGTAAACTT**C**TTTTGTCTCAAATGGAGAAATTCCATGCAATA		0.0395		0.3953		0		0.0385
rs2308292[c]		95	GACCATGCTTTATATATTCTTA**A**AATTA**T**TGCAAACATTATATTTACTT**T**AAAGTTTCTGTGACACAGTG**T**GCCTAACAGATAGTGGGAATTTTA	0.5395	0.5395	0.4419	0.4419	0.2917	0.2813	0.3163	0.3163
		95	GACCATGCTTTATATATTCTTA**A**AATTA**T**TGCAAACATTATATTTACTTTAAAGTTTCTGTGACACAGTG**C**GCCTAACAGATAGTGGGAATTTTA		0		0		0.0104		0

	rs147933644A	100	GACCATGCTTTATATATTCTTA**A**AATT**A**TTGCAAACATTATATTTACTTTTAAGTAAAGTTTCTGTGACACAGTG**C**GCCTAACAGATAGTGGGAATTTTA	0.4605	0.3421	0.5581	0.1977	0.7083	0.3958	0.6837	0.2653
	rs140159023A	100	GACCATGCTTTATATATTCTTA**A**AATT**A**TTGCAAACATTATATTTACTTTTAAGTAAAGTTTCTGTGACACAGTG**T**GCCTAACAGATAGTGGGAATTTTA		0.1053		0.3605		0.2604		0.398
	rs396196C	100	GACCATGCTTTATATATTCTTA**A**AATT**T**TTGCAAACATTATATTTACTTTTAAGTAAAGTTTCTGTGACACAGTG**C**GCCTAACAGATAGTGGGAATTTTA		0		0		0.0417		0.0204
		100	GACCATGCTTTATATATTCTTA**C**AATT**A**TTGCAAACATTATATTTACTTTTAAGTAAAGTTTCTGTGACACAGTG**T**GCCTAACAGATAGTGGGAATTTTA		0.0132		0		0.0104		0
rs28923216	rs184394929T	100	GTTTATAGTTTTGAAAGTGAATTGATCACTTTGTTTCTTGCTCTGAATCTTTAATTTTTTTTCCTAAAGGAAAAAGATAATTTAC**T**TTTTATAGAGCAAAA	0.6842	0.6842	0.4405	0.4405	0.5312	0.5312	0.5417	0.5417
		105	GTTTATAGTTTTGAAAGTGAATTGATCACTTTGTTTCTTGCTCTGAATCTTTTGTAAATTTTTTTTCCTAAAGGAAAAAGATAATTTAC**T**TTTTATAGAGCAAAA	0.3158	0.2895	0.5595	0.5595	0.4688	0.4688	0.4583	0.4583
		105	GTTTATAGTTTTGAAAGTGAATTGATCACTTTGTTTCTTGCTCTGAATCTTTTGTAAATTTTTTTTCCTAAAGGAAAAAGATAATTTAC**G**TTTTATAGAGCAAAA		0.0263		0		0		0
rs3038530[c]	rs1923740C	104	**A**C**T**CTGGCAATTACTTAAGGCTATCATTCATTCTG**C**AGAAATCGCTTTGTAAATCAGCTTATTTGGCTCAGTTTATATTTGAAAATTTTGATATGGAGGAATTC	0.3421	0.3421	0.3605	0.3605	0.3438	0.3438	0.4615	0.4615
	rs115288378C	108	**C**C**T**CTGGCAATTACTTAAGGCTATCATTCATTCTGCAGAAATCGCTTTGTAAATCAGC**T**TGATTATTTGGCTCAGTTTATATTTGAAAATTTTGATATGGAGGAATTC	0.6579	0.5395	0.6395	0.4186	0.6562	0.2812	0.5385	0.2308
		108	**A**C**T**CTGGCAATTACTTAAGGCTATCATTCATTCTGCAGAAATCGCTTTGTAAATCAGC**T**TGATTATTTGGCTCAGTTTATATTTGAAAATTTTGATATGGAGGAATTC		0.0921		0.2209		0.375		0.3077
		108	**C**C**T**TTGGCAATTACTTAAGGCTATCATTCATTCTGCAGAAATCGCTTTGTAAATCAGC**T**TGATTATTTGGCTCAGTTTATATTTGAAAATTTTGATATGGAGGAATTC		0.0263		0		0		0
rs3042783[c]	chr2:222160737 T[b]	100	TTCTGTTGTGGTTAGGAGGGATATTGACC**T**GAAGGCA**T**ACTGTTCCCTCTTAAAGAAGCAAGATCAACATTCAGAATCGCTCAGTGACAAATCTCAATTG	0.7763	0.1053	0.6163	0.5698	0.7292	0.6146	0.4898	0.3878
	rs3943815C	100	TTCTGTTGTGGTTAGGAGGGATATTGACC**T**GAAGGC**A**CACTGTTCCCTCTTAAAGAAGCAAGATCAACATTCAGAATCGCTCAGTGACAAATCTCAATTG		0.6711		0.0465		0.1146		0.102
		105	TTCTGTTGTGGTTAGGAGGGATATTGACC**T**GAAGGC**A**CACTGTTCCCTCTTAACTCAAAGAAGCAAGATCAACATTCAGAATCGCTCAGTGACAAATCTCAATTG	0.2237	0.2237	0.3837	0.3837	0.2708	0.2604	0.5102	0.5102
		105	TTCTGTTGTGGTTAGGAGGGATATTGACC**C**GAAGGC**A**CACTGTTCCCTCTTAACTCAAAGAAGCAAGATCAACATTCAGAATCGCTCAGTGACAAATCTCAATTG		0		0		0.0104		0
rs3045264	rs183114846G	100	AACAAAGTTTCCAAGGGCTCTTACCTACGTGGTTGGTGACCTTTTGGGGCTAATAATTAAGAAGGATGGGTATATATGAAATTATTTAAATCTCCATC**G**G	0.6351	0.2763	0.3571	0.3571	0.3438	0.3438	0.3293	0.3293
		104	AACAAAGTTTCCAAGGGCTCTTACCTACGTGGTTGGTGACCTTTTGGGGCTGTCTAATAATTAAGAAGGATGGGTATATATGAAATTATTTAAATCTCCATC**G**G	0.3649	0.7237	0.6429	0.6429	0.6562	0.6458	0.6707	0.6707
		104	AACAAAGTTTCCAAGGGCTCTTACCTACGTGGTTGGTGACCTTTTGGGGCTGTCTAATAATTAAGAAGGATGGGTATATATGAAATTATTTAAATCTCCATC**A**G		0		0		0.0104		0
rs3047269[a]	–	100	CACAGAGAGGCGGTGGGAGACAGGCACACCAGCATGCAAGTACAGTGCACTGGGATCTCTCTCACTAGATACAAGGAAGGTTTGGCAAATTAGTTTGTTT	0.625		0.6429		0.4362		0.55	
		104	CACAGAGAGGCGGTGGGAGACAGGCACACCAGCATGCAAGTACAGTGCACTGACTGGGATCTCTCTCACTAGATACAAGGAAGGTTTGGCAAATTAGTTTGTTT	0.375		0.3571		0.5638		0.45	
rs3051300[a]	rs186936660A	100	TTCTATTTAGAATTGGAAGATCTGGGTAGAAGTCCTATCTAGTCCATGTATACATTGTGTGAGACTGGGCAATATTGAATCTCCTAAATGTGGACAC**A**AA	0.1842	0.1842	0.3837	0.3837	0.5319	0.5319	0.3617	0.3617
		104	TTCTATTTAGAATTGGAAGATCTGGGTAGAAGTCCTATCTAGTCCATGTATGTATACATTGTGTGAGACTGGGCAATATTGAATCTCCTAAATGTGGACAC**A**AA	0.8158	0.8158	0.6163	0.6163	0.4681	0.4681	0.6383	0.6277
		104	TTCTATTTAGAATTGGAAGATCTGGGTAGAAGTCCTATCTAGTCCATGTATGTATACATTGTGTGAGACTGGGCAATATTGAATCTCCTAAATGTGGACAC**G**AA		0		0		0		0.0106
rs3062629	–	100	AATTCACCAGAATTTAAAATACTGCTGGGTATGATGTCACATGTCTGTAGTCCCTAGCTACTTGAGAGGCTGAGATGGGAGGATTCCTTGAGTCTAGGAA	0.3158		0.6977		0.3936		0.6531	
		105	AATTCACCAGAATTTAAAATACTGCTGGGTATGATGTCACATGTCTGTACTGTAGTCCCTAGCTACTTGAGAGGCTGAGATGGGAGGATTCCTTGAGTCTAGGAA	0.6842		0.3023		0.6064		0.3469	
rs3080855	–	100	GTATATTTATAAATTTCATGTAAGTAGATATTCTAAATACTAGTAAAGCTTGTTTTCATTTTGTGTTTTAAAAAAGAATTATGATATTTTTCTCCATGCC	0.25		0.4651		0.375		0.3776	
		104	GTATATTTATAAATTTCATGTAAGTAGATATTCTAAATACTAGTAAAGCTTAATTGTTTTCATTTTGTGTTTTAAAAAAGAATTATGATATTTTTCTCCATGCC	0.75		0.5349		0.625		0.6224	
rs33917182	–	100	GCTGTGCAGAGAGAGAGTAGGGGGAGGAGGTGGAGAACCCGGAAGAGCAGCTCTGAGCAGATTTCACACGATTAATAGGGGGTTTGTTGGCTGGTGGACT	0.5263		0.5465		0.6064		0.7105	
		102	GCTGTGCAGAGAGAGAGTAGGGGGAGGAGGTGGAGAACCCGGAAGAGCAGCCATCTGAGCAGATTTCACACGATTAATAGGGGGTTTGTTGGCTGGTGGACT	0.4737		0.4535		0.3936		0.2895	
rs33951431[c]	rs4741748G	96	GGGACATAACAAAGCCTCGGCGATAGACAGCATTTGTGAGTTACATCACATGTTAGAACTTAATAGTTCTCCCACATGCTGTGA**A**TAAACAGGGTT	0.6842	0.3421	0.6163	0.5698	0.5729	0.5729	0.6224	0.5816
		96	GGGACATAACAAAGCCTCGGCGATAGACAGCATTTGTGAGTTACATCACATGTTAGAACTTAATAGTTCTCCCACATGCTGTGA**G**TAAACAGGGTT		0.3421		0.0465		0		0.0408
		100	GGGACATAACAAAGCCTCGGCGATAGACAGCATTTGTGAGTTACATCACAAGTTTGTTAGAACTTAATAGTTCTCCCACATGCTGTGA**A**TAAACAGGGTT	0.3158	0.3158	0.3837	0.3605	0.4271	0.4271	0.3776	0.3776
		100	GGGACATAACAAAGCCTCGGCGATAGACAGCATTTGTGAGTTACATCACAAGTTTGTTAGAACTTAATAGTTCTCCCACATGCTGTGA**G**TAAACAGGGTT		0		0.0233		0		0
rs34051577[a]	–	100	CTGTCATAACAATAATGAGTCATCCAGATTATCGAGTGAGATACATATTTAAGAATTATCTTTAAAAATTTCAAAAATTTTAATTTTACTGTTGTGTTTT	0.4211		0.3837		0.6875		0.5	
		105	CTGTCATAACAATAATGAGTCATCCAGATTATCGAGTGAGATACATATTTATCTTAAGAATTATCTTTAAAAATTTCAAAAATTTTAATTTTACTGTTGTGTTTT	0.5789		0.6163		0.3125		0.5	
rs34495360	–	100	CCAGTTCTGTGGTTTGGTCTCAGTACTTGTTCTGTGCAGTAGTTAAGTCCTTCCAGCCATCTTGCCACTCAGCTTAGAAAAATACTCTCCAAAAACATGT	0.6757		0.5349		0.6354		0.4474	
		105	CCAGTTCTGTGGTTTGGTCTCAGTACTTGTTCTGTGCAGTAGTTAAGTCAAGTCCTTCCAGCCATCTTGCCACTCAGCTTAGAAAAATACTCTCCAAAAACATGT	0.3243		0.4651		0.3646		0.5526	
rs34510056[a]	–	95	ACAAAATATTCTGAATAGATCCCGGCCCAAAGTCATTTGATTTGGGAATAGTCTTAAAAACAGGCAGGCATACTGTTATTAACATTGTCATTATC	0.4079		0.8023		0.4896		0.4796	
		100	ACAAAATATTCTGAATAGATCCCGGCCCAAAGTCATTTGATTTGGGAATACTTTAGTCTTAAAAACAGGCAGGCATACTAGTTATTAACATTGTCATTATC	0.5921		0.1977		0.5104		0.5204	
rs34511541[c]	rs57941925T	95	AAGCTTATGAGATTTGGAGGACTTTAGTAGAAGAGGAAAATACCACATTTATTATGAGTGTCT**T**TGAACCTAAGAAGGGTCTCATTTGTATACA	0.4079	0.4079	0.5	0.5	0.3229	0.3229	0.5102	0.5102
		100	AAGCTTATGAGATTTGGAGGACTTTAGTAGAAGAGGAAAATACCACATTTCTCTTATTTATGAGTGTCT**T**TGAACCTAAGAAGGGTCTCATTTGTATACA	0.5921	0.4737	0.5	0.3953	0.6771	0.4271	0.4898	0.3673
		100	AAGCTTATGAGATTTGGAGGACTTTAGTAGAAGAGGAAAATACCACATTTCTCTTATTTATGAGTGTCT**C**TGAACCTAAGAAGGGTCTCATTTGTATACA		0.1184		0.1047		0.25		0.1224
rs34528025[a]	rs34247791 DEL	100	GTCTCTAGCGTAGAAAGAGGAAATTTGACCCATGTCTTGGAGAGGAGTCAAATCAGAAACTCCTCCTATTAC**G**TCTTTTTCCTTTGCTCGTTTTGCTTTG	0.4474	0.4474	0.5581	0.5581	0.3333	0.3333	0.3163	0.3163
	rs202051643G	106	GTCTCT**CT**AGCGTAGAAAGAGGAAATTTGACCCATGTCTTGGAGAGGAGTCAAATCAGAAACTCCTCCTGAGTATTAC**G**TCTTTTTCCTTTGCTCGTTTTGCTTTG	0.5526	0.5526	0.4419	0.4419	0.6667	0.6667	0.6837	0.5816
		106	GTCTCT**CT**AGCGTAGAAAGAGGAAATTTGACCCATGTCTTGGAGAGGAGTCAAATCAGAAACTCCTCCTGAGTATTAC**A**TCTTTTTCCTTTGCTCGTTTTGCTTTG		0		0		0		0.102
rs34535242[a]	–	100	AGGGGGTACTACAGACAGGTTTAAAATGAGCAAACCTAGCTGGTAGGTAGTGTCCTTAGAAGAGTTTTAAGTGAAAAAGGACATGATAAAATATGGCTTT	0.5789		0.4651		0.6149		0.5395	
		104	AGGGGGTACTACAGACAGGTTTAAAATGAGCAAACCTAGCTGGTAGGTAGGTAGTGTCCTTAGAAGAGTTTTAAGTGAAAAAGGACATGATAAAATATGGCTTT	0.4211		0.5349		0.3854		0.4605	
rs34541393	–	96	TACATTTCTAGATGTGTCAGGAGTCTAGAAACTTCAGTTTGGAGAATAACTACTTCCCTCACATCATTGTTCATACTGTTTTGGTTTTTATTATAA	0.4342		0.7442		0.4167		0.5395	
		100	TACATTTCTAGATGTGTCAGGAGTCTAGAAACTTCAGTTTGGAGAATAAAACTCTACTTCCCTCACATCATTGTTCATACTGTTTTGGTTTTTATTATAA	0.5658		0.2558		0.5833		0.4605	
rs34795726	rs189603436G	100	ATAATGTAGAGTTATTCAAAAAAAAG**GT**CTTTTAGAAATTCTTTTTAAATTATTTGCTACCTATCCATGTTTTCTCCAAATCTATCAGCAGCACAGAGT**G**	0.5921	0.5789	0.6977	0.6977	0.4583	0.4479	0.6042	0.6042
	rs4646566G	100	ATAATGTAGAGTTATTCAAAAAAAAG**TT**CTTTTAGAAATTCTTTTTAAATTATTTGCTACCTATCCATGTTTTCTCCAAATCTATCAGCAGCACAGAGT**G**		0		0		0.0104		0
		100	ATAATGTAGAGTTATTCAAAAAAAAG**GT**CTTTTAGAAATTCTTTTTAAATTATTTGCTACCTATCCATGTTTTCTCCAAATCTATCAGCAGCACAGAGT**A**		0.0132		0		0		0
		104	ATAATGTAGAGTTATTCAAAAAAAAG**GT**CTTTTAGAAATTCTTTTTAAATTATTTGCTACCTATCCATGTTTTCTCCAAATCTATCAGCAGCACAGAGT**A**	0.4079	0.4079	0.3023	0.3023	0.5417	0.5417	0.3958	0.3958
rs34811743[c]	rs532272C	100	TATGTCTCTACATCCCACCCCAA**C**TACAA**C**ACTTCGTACCCAGGATGCAACAGATCAAAGTAGTTGCTTACTATGGGTTGAACAAAAAGGAGAGGCACAC	0.5921	0.5263	0.8095	0.8095	0.6667	0.6667	0.7604	0.7292
		100	TATGTCTCTACATCCCACCCCAA**A**TACAA**C**ACTTCGTACCCAGGATGCAACAGATCAAAGTAGTTGCTTACTATGGGTTGAACAAAAAGGAGAGGCACAC		0.0658		0		0		0.0313
		102	TATGTCTCTACATCCCACCCCAA**C**TACAA**C**ACTTCGTACCCAGGATGCAACTGAGATCAAAGTAGTTGCTTACTATGGGTTGAACAAAAAGGAGAGGCACAC	0.4079	0.4079	0.1905	0.1905	0.3333	0.3333	0.2396	0.2396
rs35605984	rs150571926C	100	TATGTCATAGTAAAAA**C**TTGGAAATAATAAGATGTTGAATAATTGACATTA**T**TAAATTATGCTACATTAGCATAATAAAATATTAGGTAGTTATTTTTAA	0.5526	0.5526	0.4167	0.4167	0.5208	0.5208	0.5816	0.5816
		105	TATGTCATAGTAAAAA**C**TTGGAAATAATAAGATGTTGAATAATTGACATTACTTTATTAAATTATGCTACATTAGCATAATAAAATATTAGGTAGTTATTTTTAA	0.4474	0.4474	0.5833	0.5833	0.4792	0.4688	0.4184	0.4184
		105	TATGTCATAGTAAAAA**AT**TGGAAATAATAAGATGTTGAATAATTGACATTACTTTATTAAATTATGCTACATTAGCATAATAAAATATTAGGTAGTTATTTTTAA		0		0		0.0104		0
rs35716687[a]	–	97	GTCATGCCATCATTAGGGGACTAAATGTGTTAATATCCTGAAAATTATAAGTAATCAATAATTTCTCTTTCGTGATACACCTTGTTTTGAAATATTT	0.6447		0.6395		0.5938		0.602	
		101	GTCATGCCATCATTAGGGGACTAAATGTGTTAATATCCTGAAAATTATACTTAAGTAATCAATAATTTCTCTTTCGTGATACACCTTGTTTTGAAATATTT	0.3553		0.3605		0.4062		0.398	
rs35769550[a]	rs1449554A	100	**A**CTGCGTTTCTGTAGAGGAGTAAATGTACTAAGACTATTAAATAACTTACACCTTAACTAAAACTTTTAGGTGGAAACAAAAGACTGGTTAGAAAAAATG	0.1053	0.1053	0.4767	0.4767	0.4688	0.4688	0.641	0.641
		104	**G**CTGCGTTTCTGTAGAGGAGTAAATGTACTAAGACTATTAAATAACTTACTGACACCTTAACTAAAACTTTTAGGTGGAAACAAAAGACTGGTTAGAAAAAATG	0.8947	0.8815	0.5233	0.5233	0.5312	0.5312	0.359	0.359
		104	**A**CTGCGTTTCTGTAGAGGAGTAAATGTACTAAGACTATTAAATAACTTACTGACACCTTAACTAAAACTTTTAGGTGGAAACAAAAGACTGGTTAGAAAAAATG		0.0132		0		0		0
rs36040336	–	88	CACGGGTTAACAGATGCAGTTATTATGCCCATTTAACACGAGGGAAACTGAGCCCAGAGAGGTTGAGGTTCACAGGTTGCAGCAGGG	0.4079		0.8023		0.8021		0.6735	
		90	CACGGGTTAACAGATGCAGTTATTATGCCCATTTAACACATGAGGGAAACTGAGGCCCAGAGAGGTTGAGGTTCACAGGTTGCAGCAGGG	0.5921		0.1977		0.1979		0.3265	
rs36062169[a]	rs114264449C	94	TTAGGGTTTCCTGTCAACTATTCTACTGCCATTTACCACAGGGTCACCACATTCTAATAAGTCCATCCTT**C**TGAGATATCCTCTTCCTAACATG	0.5132	0.5132	0.3023	0.3023	0.5833	0.5833	0.5513	0.5513

Table 1 (*Continued*)

INDEL RS Number	Flanking RS Number (s) and hg19 Reference Allele	Length (bp)	Sequence	AFA (N = 38)		ASA (N = 43)		CAU (N = 48)		HIS (N = 49)	
				LBAF	SBAF	LBAF	SBAF	LBAF	SBAF	LBAF	SBAF
		100	TTAGGGTTTCCTGTCAACTATTCTACTGCCATTTACCACAGGGTCACCACCAGTACATTCTAATAAGTCCATCCTTCTGAGATATCCTCTTCCTAACATG	0.4868	0.4211	0.6977	0.686	0.4167	0.4167	0.4487	0.4487
		100	TTAGGGTTTCCTGTCAACTATTCTACTGCCATTTACCACAGGGTCACCACCAGTACATTCTAATAAGTCCATCCTTATCAGATATCCTCTTCCTAACATG		0.0658		0.0116		0		0
rs3838581[a]	rs371883530C	96	GATTACTGGTGTTTACTTTTAATTCCAATAAATTAAAAGTTCTACTGTTTTTCTACTTCCTCATACAAATCTTGAGCAAGACAAACTTTAACATTC	0.3026	0.2894	0.3372	0.3372	0.3229	0.3229	0.5	0.5
		96	GATTATTGGTGTTTACTTTTAATTCCAATAAATTAAAAGTTCTACTGTTTTTCTACTTCCTCATACAAATCTTGAGCAAGACAAACTTTAACATTC		0.0132		0		0		0
		100	GATTACTGGTGTTTACTTTTAATTCCAATAAATTAAAAGTTCTACTGTTTGTTATTCTACTTCCTCATACAAATCTTGAGCAAGACAAACTTTAACATTC	0.6974	0.6974	0.6628	0.6628	0.6771	0.6771	0.5	0.5
rs3841948[a]	rs76509761G	95	AAGTGATCCAGATTTGGTCTTTTACTGTGAAAATGCTTTTATACAATTAGTACTATATCCTTTGCACACTGGAAT	0.3947	0.3947	0.5116	0.5116	0.375	0.375	0.2949	0.2949
		100	AAGTGATCCAGATTTGGTCTTTTACTGTGAAAATGCTTTTATACAATTTAGTACAGATGTTATGCAATTGTACTATATCCTTTGCACACTGGAAT	0.6053	0.5263	0.4884	0.4884	0.625	0.625	0.7051	0.7051
		100	AAGTGATCCAGATTTGGTCTTTTACTGTGAAAATGCTTTTATACAATTTAGTAGAGATGTTATGCAATTGTACTATATCCTTTGCACACTGGAAT	0.079		0		0		0	
rs4187	–	100	ATGATTAACAAAAAAACAAGTAGAAAAATAAGAGATGTATTTAAAAAAAAATAAAGATAATCAAATGCTTTTGAAAGACCTGTTCTCTTCACTGCCACACATATT	0.6842	0.3158	0.5233	0.4767	0.5208	0.4792	0.5208	0.4792
		106	ATGATTAACAAAAAAACAAGTAGAAAAATAAGAGATGTATTTAAAAAAAAATAAAGATAATCAAATGCTTTTTGAAAGACCTGTTCTCTTCACTGCCACACATATT								
rs4646006	rs562172870G	100	TGTAAGTCTAAACAATCAGGCACGTCGGCAGCAATGCAGCTCGCAGCTGCCATTTACCAGCCTTTGCTGATCTGTTCATTATTTTGCAGGGC	0.1711	0.1711	0.407	0.3953	0.4255	0.4255	0.5366	0.5366
		100	TGTAAGTCTAAACAATCAGGCACGTCGGCAGCAATGCAGCTCGCAGCTGCCATTTACCAGCCTTTGCTGATCTGTTCATTATTTTGCAGGGC		0		0.0116		0		0
		104	TGTAAGTCTAAACAATCAGGCACGTCGGCAGCAATGCAGCTCGCAGCTGCCATTTACCAGCCTTTGCTGATCTGTTCATTATTTTGCAGGGC	0.8289	0.8289	0.593	0.593	0.5745	0.5745	0.4634	0.4634
rs5895446[c]	rs2960102G	108	GGGAGAGATATAGAGTTACTTTGTATCCTGCCACTATCACTGGGAGATATGTTGGAGAGTCTATCGTGCAAAGTTAAGTGAAAGAGGTTCTAAGGAGATTGTTC	0.3289	0.3289	0.3256	0.3256	0.3646	0.3646	0.3265	0.3265
		110	GGGAGAGATATAGAGTTACTTTGTATCCTGCCACTATCACTGGGAGATATGTTGGACACAGAGTTCTATCGTGCAAAGTTAAGTGAAAGAGGTTCTAAGGAGATTGTTC	0.6711	0.579	0.6744	0.2326	0.6354	0.25	0.6735	0.3469
		110	AGGAGAGATATAGAGTTACTTTGTATCCTGCCACTATCACTGGGAGATATGTTGGACACAGAGTTCTATCGTGCAAAGTTAAGTGAAAGAGGTTCTAAGGAGATTGTTC		0.0921		0.4419		0.3854		0.3265
rs60901515[c]	rs9790699C	100	TGCCTATGCAATTTAACCACATAGAAGACAAGTCAGCAACTTATCTATTCGAACTCAGCAGTCGTTGGTATCCACAGTGCTGCAGATAAATTC	0.6486	0.6081	0.686	0.6628	0.6667	0.6667	0.5976	0.5854
		100	TGCCTTATGCAATTTAAGCAACAATAGAAGACAAGTCAGCAACTTATCTATTCGAACTTAGCAGTCGTTGGTATCCACAGTGGCAGATAAATTC		0.0405		0.0233		0		0.0122
		104	TGCCTTATGCAATTTAAGCAACAATAGAAGACAAGTCAGCAACTTACTATTCTATTGAAACTCAGGAGTGCTTGGTATCCACAGTGGCAGATAAATTC	0.3514	0.3514	0.314	0.3333	0.3333	0.4024	0.4024	

[a] Motif different from LaRue, et al., Pereira, et al., and dbSNP. Sequences confirmed with IGV [3,6,20,22,23].
[b] Due to lack of RS number for the observed SNP, the hg19 locus coordinates are provided.
[c] One of twenty-two INDELs, with substantial sequence variation, that are recommended for future HID INDEL panels.

overall DoC mean with an average DoC of 26.5x ± 15.7 and overall ACR mean with an average of 0.542 ± 0.224. While DoC and ACR values were sufficient for analysis, the rs33917182 locus was typed successfully in only 41.6% of the samples (74/178) after application of the DoC and ACR thresholds. This locus was identified by Pereira, et al. [6] as a valuable HID marker with expected heterozygosity of 0.501 and discrimination power of 0.618; LaRue, et al. [3] reported an F_{st} of 0.0145 and observed heterozygosities ranging from 0.451 to 0.542 in four global populations for the same INDEL. Removal of this locus due to poor success resulted in full HID INDEL profiles for 155 samples, increasing the overall mean profile completion to 97.1% ± 0.110.

3.2. Sequence variation

Using STRait Razor, sequence data were obtained for each INDEL motif and approximately 50 bases on either side of the motif (Table 1). Based on 1000 Genomes Project Phase 3 data [29], 100 known polymorphisms (94 SNPs and 6 non-HID INDELs) exist within 50 bases of the target HID INDELs (Supplemental Tables 3 and 4). Twenty-five and seventy-five of these polymorphisms have global allele frequencies (GAFs) ≥ 0.02 and < 0.02, respectively. The average distance of these polymorphisms from the target INDEL was 27 bases ± 13. All 25 flanking region polymorphisms with GAFs ≥ 0.02 were observed in the population data for four major US populations. Only 18/75 polymorphisms with GAFs < 0.02 were observed as would be expected due to sampling or being private variants.

In all 178 samples, 19 INDEL motifs had different sequences than previously reported, and these sequences were consistent among the samples studied herein [3,6,20]. Eighteen of these motif sequence differences were the result of alignment and were consistent with previous reports after manual analysis in IGV [22,23]. One locus, rs35716687, has been reported as a TTAA deletion but the marker was identified as a TACT deletion. Fifteen markers were associated with a repeat motif; the initial INDEL selection criteria by LaRue, et al. [3] had sought to avoid such structures by excluding loci with three or more repeats. Four of the 15 markers contained three copies or repeats. The remaining 11 loci contained two copies. These motifs range in size from di- to penta-nucleotides (Table 2). While the number of repeats is limited, STR motifs may become problematic if stutter-type artifacts can be generated. Thus, special attention during validation studies should be paid for potential stutter product generation. Though possible, STRs with only a few repeat motifs are less subject to such PCR artifacts relative to STRs with several to many repeats [30–32].

Sequence variation was observed in the region adjacent to the INDEL motif at 42 loci, producing 65 novel microhaplotypes (Table 1) [33–35]. Forty HID INDEL loci are part of a microhaplotype containing one or two SNPs. Two INDELs, rs13447508 and rs34528025, are part of microhaplotypes containing the target INDEL, an adjacent SNP (rs13447507 and rs202051643, respectively), and an adjacent flanking-region INDEL (rs201219895 and rs34247791, respectively). Twenty-two loci had sequence variants that account for ≥ 2% of total alleles in two or more populations. For these 22 microhaplotypes, the presence of additional, sequence-based alleles increased the average number of alleles per marker from 2 to 3.82 ± 1.14 with a range from 3 to 7 alleles (rs1408093). The observed heterozygosity for these 22 loci increased by an average of 0.132 ± 0.0957 for AFA, 0.107 ± 0.0824 for ASA, 0.179 ± 0.106 for CAU, and 0.123 ± 0.0959 for HIS (Table 3). All 68 loci were ranked based on length- and sequence-based observed heterozygosity (Table 4). By length, INDELs rs10688868, rs2308189, rs2308276, and rs2308292 ranked 33rd, 8th, 19th, and 35nd in the HIS, AFA, ASA, and CAU populations, respectively.

Table 2
Insertion/deletion loci that are part of a short tandem repeat (STR) motif. The repeat motif for each locus is underlined.

Locus rs#	STRait Razor Sequence for Insertions	Number of Repeat Motifs	Reference
rs1160886	TAGTACTAC	2	[3,6]
rs1160956	AAAGAAGAGCAAC	2	[6]
rs16402	ATTAATTATTTATT	2	[3,6]
rs16458	TTTTACAATTCCTTCCTTC	2	[3]
rs17859968	GGCACATAAATAAA	2	[3]
rs2067208	AAAGAGCCTGGCCTG	2	[6]
rs2307580	TAATTAATTGAATA	2	[6]
rs2307689	GGCTGTTCTTCTTC	3	[6]
rs2307710	CCAGAGAAGGAAGGAAGGA	3	[3,6]
rs2307839	TGAGAGAACAAC	3	[6]
rs2307850	AGCCTCCACCCACC	2	[3]
rs2308276	GATGAATTTAATTTAAA	2	[3]
rs3051300	AGTCCATGTATGTA	2	[6]
rs34535242	TAGCTGGTAGGTAGGTAG	3	[3]
rs3841948	TATACAATTTAATTT	2	[3]

Table 3
Length-based (LB) and sequence-based (SB) observed (H_o) and expected (H_e) heterozygosities in four major US population groups for 42 INDEL loci that exhibited sequence variation. The change in H_o and H_e as a result of utilizing SB alleles is indicated by ΔH_o and ΔH_e, respectively.

Locus	AFA						ASA					
	LB H_e	SB H_e	ΔH_e	LB H_o	SB H_o	ΔH_o	LB H_e	SB H_e	ΔH_e	LB H_o	SB H_o	ΔH_o
rs10623496	0.48	0.48	0.00	0.61	0.61	0.00	0.45	0.46	0.01	0.40	0.40	0.00
rs10629077	0.39	0.40	0.01	0.42	0.42	0.00	0.41	0.41	0.00	0.33	0.33	0.00
rs10688868	0.32	0.67	0.35	0.34	0.71	0.37	0.50	0.60	0.10	0.53	0.65	0.12
rs1160956	0.50	0.51	0.01	0.47	0.50	0.03	0.48	0.48	0.00	0.42	0.42	0.00
rs13447508	0.46	0.50	0.03	0.50	0.55	0.05	0.51	0.51	0.00	0.51	0.51	0.00
rs140809	0.49	0.68	0.19	0.34	0.45	0.11	0.47	0.70	0.24	0.40	0.51	0.12
rs1610871	0.51	0.63	0.12	0.50	0.63	0.13	0.47	0.51	0.04	0.37	0.42	0.05
rs16624	0.38	0.42	0.04	0.29	0.34	0.05	0.50	0.51	0.01	0.42	0.42	0.00
rs17859968	0.46	0.54	0.09	0.47	0.58	0.11	0.46	0.50	0.04	0.51	0.53	0.02
rs2067140	0.35	0.65	0.29	0.39	0.66	0.26	0.46	0.52	0.06	0.42	0.49	0.07
rs2067191	0.49	0.51	0.02	0.53	0.55	0.03	0.50	0.50	0.00	0.58	0.58	0.00
rs2067208	0.32	0.32	0.00	0.34	0.34	0.00	0.29	0.35	0.06	0.26	0.28	0.02
rs2307507	0.32	0.33	0.00	0.29	0.29	0.00	0.45	0.45	0.00	0.41	0.41	0.00
rs2307526	0.47	0.67	0.20	0.53	0.71	0.18	0.47	0.50	0.03	0.54	0.54	0.00
rs2307579	0.49	0.50	0.01	0.50	0.50	0.00	0.29	0.29	0.00	0.30	0.30	0.00
rs2307656	0.48	0.55	0.07	0.45	0.50	0.05	0.50	0.58	0.08	0.47	0.56	0.09
rs2307689	0.50	0.59	0.09	0.58	0.61	0.03	0.40	0.66	0.27	0.40	0.72	0.33
rs2307700	0.46	0.67	0.21	0.42	0.61	0.18	0.40	0.48	0.08	0.35	0.42	0.07
rs2307978	0.48	0.67	0.18	0.47	0.68	0.21	0.49	0.66	0.18	0.44	0.60	0.16
rs2308189	0.50	0.76	0.25	0.55	0.82	0.26	0.50	0.64	0.14	0.60	0.74	0.14
rs2308232	0.38	0.39	0.01	0.39	0.39	0.00	0.40	0.43	0.03	0.39	0.44	0.05
rs2308242	0.45	0.45	0.00	0.29	0.29	0.00	0.40	0.40	0.00	0.49	0.49	0.00
rs2308276	0.50	0.53	0.03	0.53	0.55	0.03	0.49	0.65	0.16	0.49	0.77	0.28
rs2308292	0.50	0.59	0.08	0.50	0.61	0.11	0.50	0.64	0.14	0.47	0.60	0.14
rs28923216	0.44	0.45	0.02	0.37	0.39	0.03	0.50	0.50	0.00	0.50	0.50	0.00
rs3038530	0.46	0.59	0.13	0.47	0.63	0.16	0.47	0.65	0.19	0.58	0.67	0.09
rs3042783	0.35	0.50	0.14	0.34	0.50	0.16	0.48	0.53	0.05	0.53	0.56	0.02
rs3045264	0.41	0.41	0.00	0.50	0.50	0.00	0.46	0.46	0.00	0.43	0.43	0.00
rs3051300	0.30	0.30	0.00	0.26	0.26	0.00	0.48	0.48	0.00	0.44	0.44	0.00
rs33951431	0.44	0.68	0.24	0.37	0.61	0.24	0.48	0.55	0.07	0.49	0.58	0.09
rs34511541	0.49	0.60	0.11	0.34	0.42	0.08	0.51	0.59	0.08	0.58	0.65	0.07
rs34528025	0.50	0.50	0.00	0.42	0.42	0.00	0.50	0.50	0.00	0.51	0.51	0.00
rs34795726	0.49	0.50	0.02	0.29	0.29	0.00	0.43	0.43	0.00	0.37	0.37	0.00
rs34811743	0.49	0.56	0.07	0.55	0.63	0.08	0.31	0.31	0.00	0.14	0.14	0.00
rs35605984	0.50	0.50	0.00	0.63	0.63	0.00	0.49	0.49	0.00	0.45	0.45	0.00
rs35769550	0.19	0.21	0.02	0.21	0.24	0.03	0.50	0.50	0.00	0.49	0.49	0.00
rs36062169	0.51	0.56	0.06	0.66	0.68	0.03	0.43	0.44	0.02	0.47	0.49	0.02
rs3838581	0.43	0.44	0.01	0.39	0.39	0.00	0.45	0.45	0.00	0.49	0.49	0.00
rs3841948	0.48	0.57	0.08	0.32	0.45	0.13	0.51	0.51	0.00	0.60	0.60	0.00
rs4646006	0.29	0.29	0.00	0.29	0.29	0.00	0.49	0.50	0.01	0.44	0.47	0.02
rs5895446	0.45	0.56	0.11	0.50	0.61	0.11	0.44	0.65	0.21	0.42	0.60	0.19
rs60901515	0.46	0.51	0.05	0.38	0.43	0.05	0.44	0.47	0.03	0.44	0.47	0.02

Locus	CAU						HIS					
	LB H_e	SB H_e	ΔH_e	LB H_o	SB H_o	ΔH_o	LB H_e	SB H_e	ΔH_e	LB H_o	SB H_o	ΔH_o
rs10623496	0.46	0.46	0.00	0.31	0.31	0.00	0.43	0.43	0.00	0.41	0.41	0.00
rs10629077	0.28	0.28	0.00	0.29	0.29	0.00	0.40	0.40	0.00	0.39	0.39	0.00
rs10688868	0.44	0.66	0.22	0.44	0.65	0.21	0.47	0.72	0.25	0.45	0.76	0.31
rs1160956	0.31	0.31	0.00	0.33	0.33	0.00	0.48	0.48	0.00	0.48	0.48	0.00
rs13447508	0.42	0.42	0.00	0.33	0.33	0.00	0.48	0.48	0.00	0.48	0.48	0.00
rs140809	0.50	0.62	0.12	0.45	0.55	0.11	0.28	0.45	0.17	0.24	0.41	0.16
rs1610871	0.49	0.49	0.00	0.54	0.54	0.00	0.51	0.52	0.01	0.38	0.41	0.03

Table 3 (*Continued*)

Locus	AFA						ASA					
	LB H_e	SB H_e	ΔH_e	LB H_o	SB H_o	ΔH_o	LB H_e	SB H_e	ΔH_e	LB H_o	SB H_o	ΔH_o
rs16624	0.35	0.36	0.02	0.27	0.29	0.02	0.50	0.51	0.01	0.59	0.59	0.00
rs17859968	0.50	0.50	0.00	0.54	0.54	0.00	0.49	0.52	0.02	0.51	0.55	0.04
rs2067140	0.48	0.55	0.07	0.52	0.56	0.04	0.47	0.55	0.07	0.47	0.53	0.06
rs2067191	0.51	0.51	0.00	0.40	0.40	0.00	0.46	0.46	0.00	0.43	0.43	0.00
rs2067208	0.43	0.59	0.16	0.38	0.54	0.17	0.43	0.50	0.07	0.45	0.51	0.06
rs2307507	0.49	0.49	0.00	0.48	0.48	0.00	0.50	0.50	0.00	0.37	0.37	0.00
rs2307526	0.50	0.61	0.10	0.52	0.63	0.10	0.42	0.63	0.21	0.45	0.73	0.27
rs2307579	0.49	0.49	0.00	0.44	0.44	0.00	0.46	0.46	0.00	0.35	0.35	0.00
rs2307656	0.51	0.61	0.10	0.63	0.71	0.08	0.46	0.52	0.06	0.48	0.50	0.02
rs2307689	0.35	0.63	0.28	0.40	0.69	0.29	0.47	0.67	0.21	0.44	0.56	0.12
rs2307700	0.50	0.61	0.11	0.50	0.63	0.13	0.43	0.59	0.15	0.38	0.52	0.15
rs2307978	0.28	0.62	0.34	0.25	0.67	0.42	0.46	0.66	0.20	0.29	0.53	0.24
rs2308189	0.46	0.67	0.21	0.40	0.70	0.30	0.50	0.61	0.10	0.57	0.62	0.05
rs2308232	0.42	0.46	0.04	0.51	0.53	0.02	0.42	0.45	0.04	0.37	0.42	0.05
rs2308242	0.34	0.36	0.02	0.43	0.43	0.00	0.35	0.35	0.00	0.35	0.35	0.00
rs2308276	0.50	0.50	0.00	0.54	0.54	0.00	0.50	0.54	0.04	0.54	0.59	0.05
rs2308292	0.42	0.70	0.28	0.46	0.75	0.29	0.44	0.68	0.24	0.47	0.69	0.22
rs28923216	0.50	0.50	0.00	0.48	0.48	0.00	0.50	0.50	0.00	0.53	0.53	0.00
rs3038530	0.46	0.67	0.21	0.35	0.48	0.13	0.50	0.65	0.14	0.51	0.64	0.13
rs3042783	0.40	0.55	0.15	0.42	0.60	0.19	0.50	0.58	0.08	0.57	0.63	0.06
rs3045264	0.46	0.47	0.01	0.48	0.48	0.00	0.45	0.45	0.00	0.56	0.56	0.00
rs3051300	0.50	0.50	0.00	0.51	0.51	0.00	0.47	0.48	0.01	0.43	0.45	0.02
rs33951431	0.49	0.49	0.00	0.52	0.52	0.00	0.47	0.52	0.05	0.47	0.51	0.04
rs34511541	0.44	0.66	0.22	0.35	0.54	0.19	0.50	0.60	0.09	0.57	0.61	0.04
rs34528025	0.45	0.45	0.00	0.46	0.46	0.00	0.44	0.56	0.12	0.39	0.49	0.10
rs34795726	0.50	0.51	0.01	0.50	0.52	0.02	0.48	0.48	0.00	0.46	0.46	0.00
rs34811743	0.45	0.45	0.00	0.33	0.33	0.00	0.37	0.41	0.05	0.40	0.46	0.06
rs35605984	0.50	0.51	0.01	0.46	0.48	0.02	0.49	0.49	0.00	0.51	0.51	0.00
rs35769550	0.50	0.50	0.00	0.60	0.60	0.00	0.47	0.47	0.00	0.41	0.41	0.00
rs36062169	0.49	0.49	0.00	0.50	0.50	0.00	0.50	0.50	0.00	0.59	0.59	0.00
rs3838581	0.44	0.44	0.00	0.44	0.44	0.00	0.51	0.51	0.00	0.47	0.47	0.00
rs3841948	0.47	0.47	0.00	0.54	0.54	0.00	0.42	0.42	0.00	0.38	0.38	0.00
rs4646006	0.49	0.49	0.00	0.55	0.55	0.00	0.50	0.50	0.00	0.39	0.39	0.00
rs5895446	0.47	0.66	0.19	0.52	0.73	0.21	0.44	0.67	0.23	0.24	0.55	0.31
rs60901515	0.45	0.45	0.00	0.58	0.58	0.00	0.49	0.50	0.01	0.61	0.61	0.00

However, when ranked by sequence-based observed heterozygosity, microhaplotypes containing these four INDELS displayed the highest heterozygosities in the HIS, AFA, ASA, and CAU populations, respectively. The second highest heterozygosity microhaplotypes in the AFA, HIS, ASA, and CAU populations are rs10688868, rs2307526, rs2308189, and rs5895446, respectively. These four loci increased from their length-based ranks of 52st, 32nd, 3rd, and 16th, in AFA, HIS, ASA, and CAU, respectively. Microhaplotypes containing the rs10688868 and rs2308189 INDELs are ranked highest, or second highest, in heterozygosity in the AFA, ASA, and HIS populations, making them far more informative than even the top ranked length-based marker. Single-locus RMPs were decreased by an average of 0.166 ± 0.0816 for AFA, 0.130 ± 0.0661 for ASA, 0.176 ± 0.0837 for CAU, and 0.134 ± 0.0773 for HIS (Supplemental Table 5).

The remaining 20 loci with detectable adjacent sequence variants did not display substantial sequence variation (average frequency of 0.0234 ± 0.0250 across all four populations). It should be noted that there were specific microhaplotypes containing INDELs rs34528025, rs36062169, and rs3841948 with relatively high frequencies: 0.1020 for HIS, 0.0658 for AFA, and 0.07900 for AFA, respectively. However, they were either observed once or not at all in the other population groups. While microhaplotypes containing these three INDELs did not substantially increase the discrimination power across the populations, these alleles may hold value for ancestry apportionment. The low allele frequency of these sequence variants, or lack of sufficient frequency in multiple populations, suggests that these 20 microhaplotypes do not have increased discrimination power over that of the current length-based allele polymorphism (Table 1) for HID applications.

Length-based allele frequencies and observed and expected heterozygosities were similar to those previously reported by LaRue, et al. [3] and Pereira, et al. [6]. Prior to Bonferroni correction, three AFA, three ASA, no CAU, and three HIS length-based loci and four AFA, five ASA, three CAU, and two HIS sequence-based loci deviated significantly from HWE ($p < 0.05$). After Bonferroni correction, there were no significant departures from HWE for length- or sequence-based loci ($p = 0.00074$, Supplemental Table 6). Prior to Bonferroni correction, 185 AFA, 140 ASA, 197 CAU, and 216 HIS length-based and 205 AFA, 186 ASA, 124 CAU, and 186 HIS sequence-based pairwise LDs were observed ($p = 0.05$). Five (AFA), four (ASA), seven (CAU), and five (HIS) length-based and seven (AFA), eight (ASA), nine (CAU), and six (HIS) sequence-based significant pairwise LDs were observed for markers on the same chromosome but not on the same chromosomal arm. After Bonferroni correction, at most two pairwise locus comparisons showed significant LD for length- and sequence-based alleles per population (rs2308112 and rs34795726 in AFA, rs34541393 and rs34811743 in ASA, $p < 0.0000219$). The observed significant pairwise LDs are less than that due to chance alone (~114). Assuming independence, the combined length-based RMPs were 1.36×10^{-26} for AFA, 5.42×10^{-27} for ASA, 2.94×10^{-27} for CAU, 1.33×10^{-27} for HIS and the combined sequence-based RMPs were 3.29×10^{-32} for AFA, 5.92×10^{-31} for ASA, 6.69×10^{-32} for CAU, and 5.67×10^{-32} for HIS for 68 HID INDEL-containing microhaplotypes (Supplemental Table 5).

The combined RMPs, under the assumption of independence, for 22 microhaplotypes (Supplemental Table 5) were 3.84×10^{-14} for AFA, 3.87×10^{-13} for ASA, 7.76×10^{-14} for CAU, and 1.60×10^{-13} for HIS. These values are comparable to those obtained with larger INDEL panels described by Pereira, et al. [6] and LaRue, et al. [3].

Table 4
Length-based (LB) and sequence-based (SB) observed heterozygosity rank (1 = highest) in four major US population groups for 68 INDEL loci.

Locus	AFA		ASA		CAU		HIS	
	LB Rank	SB Rank	LB Rank	SB Rank	LB Rank	SB Rank	LB Rank	SB Rank
rs10623496[a]	5	13	50	55	64	65	44	51
rs10629077[a]	38	46	59	59	65	66	47	54
rs10688868[a,b]	52	2	9	6	45	7	33	1
rs1160886	10	24	10	18	28	39	37	44
rs1160956[a]	24	26	42	47	63	64	23	34
rs13447508[a]	15	21	12	20	62	62	21	33
rs140809[a,b]	53	41	51	21	41	18	67	52
rs1610871[a,b]	16	7	54	48	10	20	50	49
rs16402	11	25	65	65	61	63	58	61
rs16458	25	34	43	49	36	45	24	35
rs16624[a]	58	57	44	50	66	67	5	12
rs17859968[a,b]	26	19	13	19	11	21	16	19
rs2067140[a,b]	42	6	45	25	17	16	26	21
rs2067191[a]	12	22	4	12	52	56	40	46
rs2067208[a,b]	54	58	66	66	56	27	34	25
rs2067294	64	64	46	51	40	49	20	32
rs2307507[a]	59	59	49	54	33	44	57	60
rs2307526[a,b]	13	3	8	17	15	9	32	2
rs2307579[a]	17	27	62	62	42	50	59	62
rs2307580	65	65	23	31	25	34	39	45
rs2307603	27	35	60	60	3	12	10	15
rs2307656[a,b]	35	28	24	14	2	3	22	28
rs2307689[a,b]	6	14	52	3	51	5	38	17
rs2307696	39	47	25	32	55	59	2	7
rs2307700[a,b]	40	15	57	52	23	10	54	24
rs2307710	2	8	30	37	5	14	56	59
rs2307839	63	63	17	26	67	68	61	64
rs2307850	28	36	31	38	43	51	53	58
rs2307978[a,b]	29	4	32	8	68	6	64	23
rs2308112	66	66	1	4	19	31	51	56
rs2308137	43	50	33	39	31	42	62	65
rs2308171	48	55	67	67	57	60	65	67
rs2308189[a,b]	8	1	3	2	49	4	9	8
rs2308196	30	37	47	53	37	46	35	41
rs2308232[a,b]	44	51	53	44	20	28	55	48
rs2308242[a]	60	60	18	27	46	53	60	63
rs2308276[a,b]	14	23	19	1	9	22	13	13
rs2308292[a,b]	18	16	26	9	35	1	27	3
rs28923216[a]	49	52	16	24	32	43	14	22
rs3038530[a,b]	31	9	5	5	58	37	15	4
rs3042783[a,b]	55	29	11	15	48	11	7	5
rs3045264[a]	19	30	40	45	29	40	11	16
rs3047269	7	20	41	46	26	35	4	11
rs3051300[a]	67	67	34	40	21	32	42	43
rs3062629	32	38	27	33	27	36	18	29
rs3080855	36	42	14	22	12	23	28	37
rs33917182	50	56	58	58	50	55	66	68
rs33951431[a,b]	51	17	20	13	18	29	29	26
rs34051577	3	10	35	41	47	54	30	38
rs34495360	46	54	7	16	54	58	63	66
rs34510056	20	31	63	63	1	8	48	55
rs34511541[a,b]	56	48	6	7	59	26	8	9
rs34528025[a]	41	49	15	23	39	48	49	30
rs34535242	33	39	55	56	30	41	12	18
rs34541393	37	43	61	61	13	24	36	42
rs34795726[a]	61	61	56	57	24	30	31	39
rs34811743[a,b]	9	11	68	68	60	61	45	40
rs35605984[a]	4	12	29	36	34	38	17	27
rs35716687	21	32	36	42	7	17	41	47
rs35769550[a]	68	68	21	28	4	13	43	50
rs36040336	22	33	64	64	53	57	19	31
rs36062169[a]	1	5	28	29	22	33	6	14
rs3838581[a]	45	53	22	30	44	52	25	36
rs3841948[a]	57	44	2	10	14	25	52	57
rs4187	34	40	37	43	38	47	1	6
rs4646006[a]	62	62	38	34	8	19	46	53
rs5895446[a,b]	23	18	48	11	16	2	68	20
rs60901515[a,b]	47	45	39	35	6	15	3	10

[a] Marker is part of a microhaplotype observed in these population data.
[b] Marker is recommended for future massively parallel sequencing HID microhaplotype panels.

4. Conclusion

Sixty-eight HID INDELs were characterized further using MPS and a novel application of the STRait Razor software. Fifteen loci were found to be part of an STR, and as such, although unlikely, special attention to potential stutter artifacts should be given with these markers for PCR-based sample preparation on MPS platforms.

The presence of additional, sequence-based alleles in 42 microhaplotypes increased heterozygosities and decreased the single-locus random match probabilities in four major U.S. populations. A subset of 22 sequence-based microhaplotype alleles became substantially more informative for identity testing than solely their length-based equivalents. The remaining 20 loci had less frequent sequence variation: variants were observed at a frequency < 0.02 in one or more populations (10 loci) or at a frequency ≥ 0.02 in one major US population (7 loci). While not increasing discrimination power substantially, the relatively common alleles seen in only one population group may be informative for ancestry determination.

The sample population sizes used for this exploratory study are less than those typically used for STR population studies. However, there are far less alleles per locus for INDELS and INDELs within microhaplotypes. As described previously by Chakraborty [36] economization of sample size for allele characterization can be achieved by focusing on obtaining reliable frequencies of common alleles in a sample, and rare allele frequencies can be approximated by an upper bound. Since microhaplotype alleles with reasonable minimum allele frequencies of, for example $p = 0.05$ with an $\alpha = 0.05$, can be detected in the sample sizes in this study, the population data are reasonable for identifying markers that likely provide increased discrimination power over using just the INDEL itself.

This study focused on markers that could be converted to short amplicons which would be more beneficial for degraded DNA sample typing than would be STRs. The positions of all INDELs are already known and readily accessible. Therefore, STR population data to test for linkage disequilibrium were not considered in the current study. In addition, to perform such a study a larger population sample would be required to accommodate the more polymorphic STRs.

Approximately 50 bases on either side of the INDEL were searched so that microhaplotype amplicon sizes could be designed to be as short as possible. In this dataset, multiple polymorphisms were captured on the same amplicon so future studies may focus on phasing assessments to aid in mixture deconvolution. While a relatively short amplicon is desirable for analyzing challenged samples, it is possible that additional polymorphisms lie beyond the flanking regions analyzed herein.

The Nextera™ Rapid Capture was used to readily identify markers and variants without the demands of primer design associated with PCR-amplicon enrichment. In addition, while 50 ng of genomic DNA is perfectly acceptable to use for exploratory work, this amount of input DNA clearly is far too much for an assay for forensic utility. It is expected that those microhaplotypes of interest will be converted to an assay that is PCR-based and thus requires input DNA of ≤ 1 ng.

CE and MPS platforms are suitable for analysis of INDELs; however, with MPS, 42 markers had increased variation due to closely linked polymorphisms. The panel of 22 INDEL-containing microhaplotypes had increased numbers of alleles, combined RMPs comparable to those provided by larger INDEL sets in LaRue, et al. (38 and 49 INDELs) [3] and Pereira, et al. (38 INDELs) [6], and heterozygosities greater than some low-performing STR markers [37,38].

Conflict of interest

The authors report no conflict of interest.

Acknowledgements

Portions of this project were supported by National Institute of Justice grant award 2013-DN-BX-K036.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2016.09.005.

References

[1] B. Budowle, A. van Daal, Forensically relevant SNP classes, Biotechniques 44 (April (5)) (2008) 603–608 (610. 10.2144/000112806. Review. PubMed PMID: 18474034).
[2] B.L. LaRue, J. Ge, J.L. King, B. Budowle, A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification, Int. J. Legal Med. 126 (4uly (4)) (2012) 533–540, doi:http://dx.doi.org/10.1007/s00414-012-0667-9 (Epub 2012 Jan 15. PubMed PMID: 22249274).
[3] B.L. LaRue, R. Lagacé, C.W. Chang, A. Holt, L. Hennessy, J. Ge, J.L. King, R. Chakraborty, B. Budowle, Characterization of 114 insertion/deletion (INDEL) polymorphisms, and selection for a global INDEL panel for human identification, Legal Med. (Tokyo) 16 (January (1)) (2014) 26–32, doi:http://dx.doi.org/10.1016/j.legalmed.2013.10.006 (Epub 2013 Nov 1. PubMed PMID: 24296037).
[4] J.M. Mullaney, R.E. Mills, W.S. Pittard, S.E. Devine, Small insertions and deletions (INDELs) in human genomes, Hum. Mol. Genet. 19 (October (R2)) (2010) R131–R136, doi:http://dx.doi.org/10.1093/hmg/ddq400 (Epub 2010 Sep 21. Review. PubMed PMID: 20858594 PubMed Central PMCID: PMC2953750).
[5] H.B. Pena, S.D. Pena, Automated genotyping of a highly informative panel of 40 short insertion-deletion polymorphisms resolved in polyacrylamide gels for forensic identification and kinship analysis, Transfus. Med. Hemother. 39 (June (3)) (2012) 211–216 (Epub 2012 May 11. PubMed PMID: 22851937 PubMed Central PMCID: PMC3375136).
[6] R. Pereira, C. Phillips, C. Alves, A. Amorim, A. Carracedo, L. Gusmão, A new multiplex for human identification using insertion/deletion polymorphisms, Electrophoresis 30 (November (21)) (2009) 3682–3690, doi:http://dx.doi.org/10.1002/elps.200900274 (PubMed PMID: 19862748).
[7] K.B. Gettings, K.M. Kiesler, P.M. Vallone, Performance of a next generation sequencing SNP assay on degraded DNA, Forensic Sci. Int. Genet. 19 (November) (2015) 1–9, doi:http://dx.doi.org/10.1016/j.fsigen.2015.04.010 (Epub 2015 May 27. PubMed PMID: 26036183).
[8] R. Pereira, L. Gusmão, Capillary electrophoresis of 38 noncoding biallelic mini-Indels for degraded samples and as complementary tool in paternity testing, Methods Mol. Biol. 830 (2012) 141–157, doi:http://dx.doi.org/10.1007/978-1-61779-461-2_10 (PubMed PMID: 22139658).
[9] F.R. Wendt, X. Zeng, J.D. Churchill, J.L. King, B. Budowle, Analysis of short tandem repeat and single nucleotide polymorphism loci from single-source samples using a custom HaloPlex target enrichment system panel, Am. J. Forensic Med. Pathol. 37 (June (2)) (2016) 99–107, doi:http://dx.doi.org/10.1097/PAF.0000000000000228 (PubMed PMID: 27075592).
[10] F.R. Wendt, J.D. Churchill, N.M.M. Novroski, J.L. King, J. Hg, R.F. Oldt, K.L. McCulloh, J.A. Weise, D.G. Smith, S. Kanthaswamy, B. Budowle, Genetic analysis of the Yavapai Native Americans from West-Central Arizona, Forensic Sci. Int. Genet. 10.1016/j.fsigen.2016.05.008.
[11] M.A. Quail, M. Smith, P. Coupland, et al., A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers, BMC Genomics 24 (July (13)) (2012) 341, doi:http://dx.doi.org/10.1186/1471-2164-13-341 (PubMed PMID: 22827831 PubMed Central PMCID: PMC3431227).
[12] X. Zeng, J.L. King, M. Stoljarova, et al., High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing, Forensic Sci. Int. Genet. 16 (May) (2015) 38–47, doi:http://dx.doi.org/10.1016/j.fsigen.2014.11.022 (Epub 2014 Dec 3. PubMed PMID: 25528025).
[13] J.L. King, B.L. LaRue, N.M. Novroski, et al., High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq, Forensic Sci. Int. Genet. 12 (Sep) (2014) 128–135, doi:http://dx.doi.org/10.1016/j.fsigen.2014.06.001 (Epub 2014 Jun 7. PubMed PMID: 24973578).
[14] S.L. Fordyce, H.S. Mogensen, C. Børsting, et al., Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™ Forensic Sci. Int. Genet. (2015) J 438 an;14 132-40. 10.1016/j.fsigen.2014.09.020 Epub 2014 Oct 5. PubMed PMID: 25450784.
[15] J.D. Churchill, J. Chang, J. Ge, et al., Blind study evaluation illustrates utility of the Ion PGM™ system for use in human identity DNA typing, Croat. Med. J. 56 (June (3)) (2015) 218–229 (PubMed PMID: 26088846).

[16] D.H. Warshauer, J.L. King, B. Budowle, STRait Razor v2.0: the improved STR allele identification tool–razor, Forensic Sci. Int. Genet. 14 (January) (2015) 182–186, doi:http://dx.doi.org/10.1016/j.fsigen.2014.10.011 (Epub 2014 Oct 22. PubMed PMID: 25450790).
[17] QIAamp®, DNA Mini and Blood Mini Handbook, 3rd edition, (2012) . https://www. qiagen.com/us/resources/resourcedetail?id=67893a91-946f-49b5-8033-394fa5d752ea&lang=en.
[18] X. Zeng, D.H. Warshauer, J.L. King, J.D. Churchill, R. Chakraborty, B. Budowle, Empirical testing of a 23-AIMs panel of SNPs for ancestry evaluations in four major US populations, Int. J. Legal Med. 130 (July (4)) (2016) 891–896, doi: http://dx.doi.org/10.1007/s00414-016-1333-4 (Epub 2016 Feb 25. PubMed PMID: 26914801).
[19] D.H. Warshauer, J.D. Churchill, N. Novroski, J.L. King, B. Budowle, Novel Y-chromosome short tandem repeat variants detected through the use of massively parallel sequencing, Genomics Proteomics Bioinform. 13 (August (4)) (2015) 250–257, doi:http://dx.doi.org/10.1016/j.gpb.2015.08.001 (Epub 2015 Sep 21. PubMed PMID: 26391384 PubMed Central PMCID: PMC4610967).
[20] S.T. Sherry, M.H. Ward, M. Kholodov, J. Baker, L. Phan, E.M. Smigielski, K. Sirotkin, dbSNP: the NCBI database of genetic variation, Nucleic Acids Res. 29 (1) (2001 Jan 1) 308–311.
[21] MiSeq System User Guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-system-guide-15027617-o.pdf.
[22] H. Thorvaldsdóttir, J.T. Robinson, J.P. Mesirov, Integrative Genomics Viewer (IGV): high performance genomics data visualization and exploration, Brief. Bioinform. 14 (March (2)) (2013) 178–192, doi:http://dx.doi.org/10.1093/bib/bbs017 (Epub 2012 Apr 19. PubMed PMID: 22517427 PubMed Central PMCID: PMC3603213).
[23] J.T. Robinson, H. Thorvaldsdóttir, W. Winckler, et al., Integrative genomics viewer, Nat. Biotechnol. 29 (January (1)) (2011) 24–26, doi:http://dx.doi.org/10.1038/nbt.1754 (PubMed PMID: 21221095 PubMed Central PMCID: PMC3346182).
[24] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 genome project data processing subgroup the sequence alignment/map format and SAMtools, Bioinformatics 25 (August (16)) (2009) 2078–2079, doi:http://dx.doi.org/10.1093/bioinformatics/btp352 (Epub 2009 Jun 8. PubMed PMID: 19505943 PubMed Central PMCID: PMC2723002).
[25] H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, Bioinformatics 27 (November (21)) (2011) 2987–2993, doi:http://dx.doi.org/10.1093/bioinformatics/btr509 (Epub 2011 Sep 8. PubMed PMID: 21903627 PubMed Central PMCID: PMC3198575).
[26] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (May (14)) (2009) 1754–1760, doi:http://dx.doi.org/10.1093/bioinformatics/btp324 (Epub 2009 May 18. PubMed PMID: 19451168 PubMed Central PMCID: PMC2705234).
[27] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (September (9)) (2010) 1297–1303, doi: http://dx.doi.org/10.1101/gr.107524.110 (Epub 2010 Jul 19. PubMed PMID: 20644199 PubMed Central PMCID: PMC2928508).
[28] Genetic Data Analysis Software, Lewis and Zaykin, 1999.
[29] D. Karolchik, A.S. Hinrichs, W.J. Kent, The UCSC genome browser, Curr. Protoc. Bioinform. (December) (2012) (Chapter 1:Unit1.4. 10.1002/0471250953. bi0104s40. PubMed PMID: 23255150).
[30] S. Leclercq, E. Rivals, P. Jarne, DNA slippage occurs at microsatellite loci without minimal threshold length in humans: a comparative genomic approach, Genome Biol. Evol. 12 (July (2)) (2010) 325–335, doi:http://dx.doi.org/10.1093/gbe/evq023 (PubMed PMID: 20624737 PubMed Central PMCID: PMC2997547).
[31] H. Fan, J.Y. Chu, A brief review of short tandem repeat mutation, Genomics Proteomics Bioinform. 5 (February (1)) (2007) 7–14 (Review. PubMed PMID: 17572359).
[32] R. Chakraborty, M. Kimmel, D.N. Stivers, L.J. Davison, R. Deka, Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci, Proc. Natl. Acad. Sci. U. S. A. 94 (February (3)) (1997) 1041–1046 (PubMed PMID: 9023379 PubMed Central PMCID: PMC19636).
[33] K.K. Kidd, A.J. Pakstis, W.C. Speed, R. Lagacé, J. Chang, S. Wootton, E. Haigh, J.R. Kidd, Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics, Forensic Sci. Int. Genet. 12 (2014) 215–224, doi:http://dx.doi.org/10.1016/j.fsigen.2014.06.014 (Epub 2014 Jul 1. PubMed PMID: 25038325).
[34] K.K. Kidd, W.C. Speed, Criteria for selecting microhaplotypes: mixture detection and deconvolution, Invest. Genet. 6 (January (1)) (2015), doi: http://dx.doi.org/10.1186/s13323-014-0018-3 (eCollection 2015. PubMed PMID: 25750707 PubMed Central PMCID: PMC4351693).
[35] J. Ge, B. Budowle, J.V. Planz, R. Chakraborty, Haplotype block: a new type of forensic DNA markers, Int. J. Legal Med. 124 (September (5)) (2010) 353–361, doi:http://dx.doi.org/10.1007/s00414-009-0400-5 (Epub 2009 Dec 22. PubMed PMID: 20033199).
[36] R. Chakraborty, Sample size requirements for addressing the population genetic issues of forensic use of DNA typing, Hum. Biol. 64 (April (2)) (1992) 141–159 (PubMed PMID: 1559686).
[37] C. Tomas, H.S. Mogensen, S.L. Friis, C. Hallenberg, M.C. Stene, N. Morling, Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex® ESI 17 and AmpFℓSTR® NGM SElect™ in Somalis, Danes and Greenlanders, Forensic Sci. Int. Genet. 11 (2014) e18–e21, doi:http://dx.doi.org/10.1016/j.fsigen.2014.04.004 (Epub 2014 Apr 18. PubMed PMID: 24810256).
[38] S. Turrina, M. Ferrian, S. Caratti, D. De Leo, Evaluation of genetic parameters of 22 autosomal STR loci (PowerPlex® Fusion System) in a population sample from Northern Italy, Int. J. Legal Med. 128 (March (2)) (2014) 281–283, doi: http://dx.doi.org/10.1007/s00414-013-0934-4 (Epub 2013 Nov 2. PubMed PMID: 24185983).