Title

Elucidation of familial relationships using hair shaft proteomics

Authors
Karim, Noreen
Plott, Tempest J
Durbin-Johnson, Blythe P
et al.

Publication Date
2021-09-01

DOI
10.1016/j.fsigen.2021.102564

Peer reviewed

Elucidation of Familial Relationships Using Hair Shaft Proteomics

Noreen Karim[a], Tempest J. Plott[a,b], Blythe P. Durbin-Johnson[c], David M. Rocke[c], Michelle Salemi[d], Brett S. Phinney[d], Zachary C. Goecker[a], Marc J. M. Pieterse[a], Glendon J. Parker[a,b], Robert H. Rice[a,b,]*

[a]Department of Environmental Toxicology, University of California, Davis, USA

[b]Forensic Science Program, University of California, Davis, USA

[c]Division of Biostatistics, Department of Public Health Sciences, Clinical and Translational, Science Center Biostatistics Core, University of California, Davis, USA

[d]Proteomics Core Facility, University of California, Davis, USA

*Corresponding author

Keywords

Proteomic profiling, genetically variant peptides, human hair, forensic investigation, relationship testing

Proteomics repository files

The proteomics data for samples P, M, S1-S4 are available on the MassIVE repository (https://massive.ucsd.edu) MassIVE # MSV000086665 (reviewer password "Hair Shaft"), ProteomeExchange # = PXD23446. Proteomics data for samples A-H are available in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD016169.

Acknowledgments

We thank Dr. Elizabeth Heller for microscopic inspection of the hair shaft samples. This work was supported by NIH grant UL1 TR001860 from the National Center for Advancing Translational Sciences, NIJ grants 2011-DN-BX-K543 and 2015-DN-BX-K065, NIJ fellowship 2019-R2-CX-0051 and USDA (NIFA)/University of California Agricultural Experiment Station project CA-D-ETX-2152-H. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests

The authors declare no conflict of interest, with the exception of GJP, who has a patent based on use of genetically variant peptides for human identification (US 8,877,455 B2, Australian Patent 2011229918, Canadian Patent CA 2794248, and European Patent EP11759843.3, GJP inventor). The patent is owned by Parker Proteomics LLC. Protein-Based Identification Technologies LLC (PBIT) has an exclusive license to develop the intellectual property and is co-owned by Utah Valley University and GJP. This ownership of PBIT and associated intellectual property does not alter policies on sharing data and materials. These financial conflicts of interest are administered by the Research Integrity and Compliance Office, Office of Research at the University of California, Davis to ensure compliance with University of California Policy.

Elucidation of Familial Relationships Using Hair Shaft Proteomics

Abstract

This study examines the potential of hair shaft proteomic analysis to delineate genetic relatedness. Proteomic profiling and amino acid sequence analysis provide information for quantitative and statistically-based analysis of individualization and sample similarity. Protein expression levels are a function of cell-specific transcriptional and translational programs. These programs are greatly influenced by an individual's genetic background, and are therefore influenced by familial relatedness as well as ancestry and genetic disease. Proteomic profiles should therefore be more similar among related individuals than unrelated individuals. Likewise, profiles of genetically variant peptides that contain single amino acid polymorphisms, the result of non-synonymous SNP alleles, should behave similarly. The proteomically-inferred SNP alleles should also provide a basis for calculation of combined paternity and sibship indices. We test these hypotheses using matching proteomic and genetic datasets from a family of two adults and four siblings, one of which has a genetic condition that perturbs hair structure and properties. We demonstrate that related individuals, compared to those who are unrelated, have more similar proteomic profiles, profiles of genetically variant peptides and higher combined paternity indices and combined sibship indices. This study builds on previous analyses of hair shaft protein profiling and genetically variant peptide profiles in different real-world scenarios including different human hair shaft body locations and pigmentation status. It also validates the inclusion of proteomic information with other biomolecular substrates in forensic hair shaft analysis, including mitochondrial and nuclear DNA.

Introduction

Hair shafts are a common component of crime scenes and are currently underutilized forensically. Use of morphological patterns in hair shafts is currently considered controversial in forensic science due to the intrinsically subjective nature of pattern matching (Council, 2009). Additionally, nuclear DNA is degraded in hair shafts as part of the natural cornification process (Linch et al, 2001; McNevin et al, 2005). This effectively eliminates the possibility of routinely obtaining identifying STR genotypes. Since the abundant mitochondrial DNA, unlike nuclear DNA, persists in the hair shaft, its matrilineal haplotype analysis is the current best practice for obtaining identifying genetic information from the hair shaft. Recent research has demonstrated that hair shaft protein may also provide forensically relevant identifying information in the form of genetically variant peptides (GVPs) (Goecker et al, 2020; Parker et al, 2016). The forensic utility and scope of proteomic genotyping continues to be extended and demonstrated to be unaffected in forensically relevant, real-world contexts including hair from different body locations (Chu et al, 2019; Milan et al, 2019), different pigmentation states (Franklin et al, 2020), from long term storage (Plott et al, 2020), and even in hair from experimental explosive devices (Chu et al, 2020). This study examines whether the proteomic information in hair shafts is able to delineate familial relationships.

Proteomic information in forensic genetics consists of two basic forms, the amino acid sequences themselves and the relative profile of protein expression. The profile, a lineup of the many proteins in the sample and their relative levels of expression, is a function of cell-specific transcriptional and translational programming. In addition to a myriad of physiological, anatomical and biochemical contexts, the genetic background of each individual would also play a significant role. Previous findings with mice (Rice et al, 2012) and humans (Wu et al, 2017) indicate that protein expression levels in the hair shaft are largely genetically determined. However, wide variation is observed among hair samples from individuals in the outbred human population (Laatsch et al, 2014), likely arising from sequence variations in noncoding regions of the genome (Hindorff et al, 2009; Martin-Trujillo et al, 2020), including gene promoters and miRNA binding sites that affect transcription factor binding sites or chromatin accessibility. This background of variation would be predicted to be lower in genetically related individuals, and the proteomic profiles of related individuals would therefore be predicted to be more similar to each other than to those of unrelated individuals (Wu et al, 2017). Since children would be expected to inherit determinants of individual hair protein expression level from each parent, their individual hair protein levels would be expected to mimic those of either parent or to be intermediate between them. Based on this expectation, we test the hypothesis that hair protein profiles in a family are more similar in two-way comparisons between a parent and individual children than between the parents. The family studied in this case has three unaffected offspring and one diagnosed with a rare genetic condition where the hair is brittle and has an unusual protein/lipid ratio (Alsop et al, 2016). This happenstance has permitted the opportunity to determine whether a hair sample appears abnormal within the context of a family.

In addition to providing information on protein expression levels, hair shaft proteomic digests also permit analysis of GVPs within those proteins and the development of a proteomically-inferred genotype of non-synonymous single nucleotide polymorphism (SNP) alleles (Parker et al, 2016). This manifestation of allelic differences permits inference of corresponding SNPs in the genomic DNA of hair donors. Although hair protein profiling may have utility in distinguishing individuals, GVPs are more robust and offer a greater power of discrimination. Like any genotype marker system, these profiles would be predicted to be more similar in related individuals, and therefore have the potential also to be exploited to develop measures of genetic relatedness. The present study offers an opportunity to determine kinship indices by analysis of hair shaft digests from a single family compared to nine unrelated individuals.

Materials and Methods

Sample Collection and Processing

For the current study, six family members of European ancestry were enrolled after obtaining written informed consent either from the individuals or from the parents in the case of minors <18 years of age. The study was conducted in accordance with protocols and procedures approved by the Institutional Review Board of the University of California Davis. The enrolled individuals included mother (M), father (P) and their four children, two sons (S1 and S2) and two daughters (S3 and S4). Hair shafts were collected from each enrolled individual. Abnormalities in hair shaft structure were not visible by light microscopy. For the proteomic analysis three replicates of hair samples from each individual except P and S2 (four and six replicates, respectively) were processed as previously described (Plott et al, 2020), and the randomized protein digests were subjected to LC-MS/MS using a Thermo Scientific Q Exactive Plus Orbitrap mass spectrometric analysis (Wu et al, 2017).

Database Searching and Proteomic Profiling

The data files generated by LC-MS/MS were searched against a Uniprot human database appended with a database containing identical but reversed (decoy) peptides and common human contaminants using X!Tandem (2016.10.15.2). The peptide and protein identifications were validated in Scaffold (version 4.8.2, Proteome Software Inc., Portland). Proteins identified at a minimum of 99% probability and represented by at least two peptides identified at 95% probability were included in the analysis (false discovery rate of 0.7% for proteins and <0.1% for peptides). The weighted spectral count data provided by Scaffold were used for the profiling and statistical analyses after confirming protein presence by exclusive spectral counts. To obtain the number of significant differences between profiles, two-way comparisons were conducted, where the weighted spectral counts were compared separately for each protein from the two subjects (Table S10). These differential protein expression analyses were conducted using the limma-voom Bioconductor pipeline (Ritchie et al, 2015), which was originally developed for RNA sequencing data (limma version 3.44.3, edgeR version 3.30.3). Normalization factors were calculated using trimmed mean of M values (Robinson and Oshlack, 2010). P-values were adjusted for multiple testing across proteins (Benjamini and Hochberg, 1995). The model used in limma included effects for individual and batch. Analyses were conducted using R version 4.0.0 Patched (2020-05-18 r78487). The raw data files and scaffold analysis files are available on the MassIVE repository (https://massive.ucsd.edu) MassIVE # MSV000086665 (reviewer password "Hair Shaft"), ProteomeExchange # = PXD023446.

GVP Analysis

To obtain genetically variant peptides (GVPs) profiles, the raw data files for all the samples were first converted by MSConvertGUI (Proteowizard 2.1 http://proteowizard.sourceforge.net) to MzML format and were subsequently searched using X!Tandem peptide spectra matching algorithm (GPM Fury, X!Tandem Alanine 149 v.3.0 (2017-02-01)). Default search parameters were used except that the search was limited to eukaryotic reference libraries, peptide and protein log(e) scores were set to <-1, fragment mass error of 20 ppm, parent mass error of 100 ppm, and point mutations from the refinement specifications were included in the search. The peptides identified by GPM Fury for each sample were subsequently searched for previously identified GVPs using GVP Finder (v 1.2) (https://parkerlab.ucdavis.edu/gvp-finder) where searches for GVPs in the peptide data followed the previously established criteria (Borja et al, 2019; Goecker et al, 2020; Plott et al, 2020). Moreover, the .xml files for all the individuals were also explored using the discovery approach by looking for peptides carrying single amino acid variations with log(e) scores <-2 with no other chemical or genetic modifications and no peaks representing the alternate amino acid (Borja et al, 2019). The single amino acid variations carrying peptides were evaluated against all human protein sequences in the PROWL web portal (prowl.rockefeller.edu/prowl/) for uniqueness to confirm that they were translated from a single site in the genome. Because of the familial structure of the study, the GVPs were not filtered based on their low allele frequencies contrary to earlier studies (Parker et al, 2016; Borja et al, 2020). The obtained GVP profiles of individuals P, M, S1 and S2 were validated from their DNA data. The genetic data of individuals S3 and S4 were not available.

Exome Sequencing Data

Exome DNA sequencing data were provided by the Department of Human Genetics, Radboud University Medical Center, the Netherlands. Data for P, M and S1 were obtained using Illumina HiSeq and those for S2 using SOLiDxl 5500 instrumentation. Genotype information analogous to the detected GVPs were obtained from the exome data for all the four individuals. Data from S2 were not consistent with the M and P at 5 loci encoding GVPs by Mendelian genetics, but proteomic data permitted correction of three of these loci (**Table S1**). The discrepancy reflects the higher error rate in the older SOLiDxl method.

Hierarchical Clustering

Data from a previously published set of unrelated European American individuals (Plott et al, 2020) were merged with the GVP list of the currently studied family. A binary format data matrix was generated with 1 representing a GVP detection and 0 a non-detection (**Table S2**). Each row of the matrix represents the GVP information for each individual with the columns representing SNPs. The matrix was used to calculate Euclidean distance between the rows/samples, based on which agglomerative hierarchical clustering was performed, and a dendrogram was plotted for the clustering using the hclust function of R (Version 3.6.3 (2020-02-29)).

Parentage Index and Sibship Index Calculation

The GVPs detected in the samples were used in kinship calculation (parentage indexes and sibship indexes) that can provide a statistical value for the probability of relationship between samples. Likelihood ratios were calculated using the SNP data obtained from exome sequencing corresponding to all the identified GVPs. Moreover, SNPs were inferred from the GVP profiles for all the studied individuals where each locus was treated as homozygous for an allele if a peptide corresponding to only one allele at the locus was detected and heterozygous if both GVPs were detected in the proteomic data. GVPs from the loci where only one GVP was unique were excluded from this analysis. GVPs from different genes were assumed as completely independent whereas complete linkage between loci within a gene was assumed to account for linkage disequilibrium. In cases of more than two GVPs within one gene, the two with the highest allele frequencies of the minor allele were used. Likelihood ratios were calculated as described (Sozer et al, 2010; Wenk et al, 1996) using the formulae in **Table S3**. Relationship indices for each locus were calculated with allele frequencies from the European population (Consortium et al, 2015). Combined paternity and sibship indices were obtained by taking a product of the respective indices for all the loci included in each analysis.

RESULTS

Proteomic Profiling

The protein levels in hair samples from all six studied individuals were subjected to two way comparisons to evaluate the impact of their genetic relationships. Using the standard significance level of p<0.05 (after correction for multiple testing) showed few protein level differences between the parents (**Table 1A**). While unusual, this degree of similarity is occasionally observed among unrelated individuals (Laatsch et al, 2014). Nevertheless, supporting the original hypothesis, the profiles of three offspring (S1, S3, S4) exhibited few proteins whose levels differed from those in the parental hair samples (0-2) or from each other (0) by this criterion. In contrast, however, the profile of one child (S2) with a rare hair phenotype was quite distinct, showing 0-11 proteins differing in level from those in other family members (**Table 1A**).

A

p<0.05	M	S1	S2	S3	S4
P	1	0	**11**	0	0
M		1	**5**	0	2
S1			**4**	0	0
S2				**2**	**0**
S3					0

B

p<0.1	M	S1	S2	S3	S4
P	13	5 (1)	**24 (2)**	1 (1)	3 (3)
M		1 (1)	**13 (2)**	0	3 (3)
S1			**6**	2	0
S2				**2**	**0**
S3					0

Table 1. Proteins with significant differences in hair protein profiles. Values for two-way comparisons between Father (P), Mother (M) and siblings (S1-S4) are tabulated with p<0.05 (A) or p<0.1 (B). In parentheses (B) are numbers of proteins in each case that match those differing between the parents and thus plausibly result through inheritance from the other parent.

To obtain a more expansive view of the proteomic relationships, differences were analyzed at a less stringent significance level of p<0.1. As shown in **Table 1B**, the profiles of mother (M) and father (P) exhibited differences in 13 proteins. Hair from three siblings (S1, S3, S4) exhibited few differences with each parent (0-5), and most of the differences (9 of 13) from one parent were shared with the other parent. Samples from the fourth offspring S2 showed a small number of differences from those of the other siblings (0-6). By contrast, samples from this offspring exhibited numerous differences with the parents (13 and 24), most of which were not evident in comparisons of samples from the parents with each other. The identities of the proteins differing among the parents and offspring are shown in Figure S1.

Profile of Proteomically-Inferred SNP Genotypes

Database searching of the samples by GPM Fury identified on average 550 ± 38 proteins with 2390 ± 310 unique peptides per sample (all values given as mean ± std dev), which were then checked for GVPs. A total of 181 GVPs corresponding to 96 loci were identified in datasets of the six studied individuals (**Table S4**). The replicates had on average 52 ± 9 GVPs while the cumulative data of the replicates for each individual showed 75.4 ± 3.6 GVPs. GVPs identified in the individuals P, M, S1 and S2 were validated from the parallel exomic sequencing data, and the GVPs were designated as true positive (TP), true negative (TN), false positive (FP) or false negative (FN) as previously described (Borja et al, 2019; Parker et al, 2016) . The analysis showed a total of 304 (41.7%) TP, 303 (41.6%) TN, 107 (14%) FN, and 14 (1.9%) FP assignments (**Table S4**). The GVPs were also categorized more precisely as undetected when protein regions containing them were not represented due to low yields in the MS run (**Table S5**). Previously such GVPs were assigned to the false negative category (Borja et al, 2019; Parker et al, 2016). This modification avoids assumptions in cases where no data were provided by the MS scan and increased the negative predictive value (TN/(TN+FN) from 73.9% for data in **Table S4** to 92.8% for data in **Table S5**. The positive predictive value (TP/(TP+FP)) for the data was 95.4%. About 94% of the assumptions made were correct ((TP + TN)/(TP + TN + FP +

FN)) when compared to the exomic data. Moreover, because a majority of the homozygous assumptions were made on the major alleles with frequencies >75%, homozygosity was the most conservative assumption.

Hierarchical Clustering

To evaluate the identifying powers of the GVP profiles, the profiles of the 6 studied family members were compared with those of 9 unrelated individuals. Data from a previously published dataset (the 9 unrelated individuals), processed contemporaneously by the same individual using an identical protocol, were merged with the current GVP dataset (Plott et al, 2020). Each GVP detection was assigned a value of 1 and non-detection a value of 0. The file was then imported into R and Euclidean distances between the samples were calculated. Using agglomerative hierarchical clustering, similar profiles were clustered together based on the Euclidean distances. The clustering showed that the GVP profiles of the 6 related individuals were more closely correlated to each other than to GVP profiles of unrelated subjects (**Figure 2**), not likely a manifestation of a batch effect of processing (Plott et al, 2020). This was the case even for the sibling with an RPS23 mutation (S2) who manifested a distinct 'wiry' hair shaft phenotype with low lipid levels. The results indicate a high utility of GVP profiling for forensic identification purposes, especially in cases of mass fatalities when samples from the close family members are available for identification, which would likely increase the power of this approach.

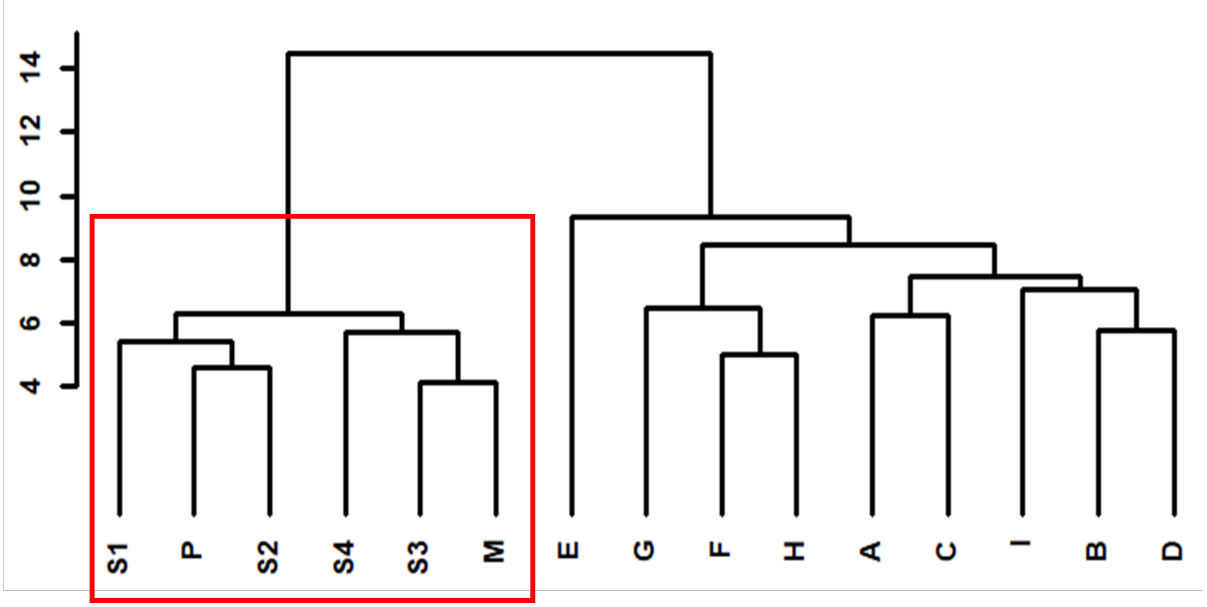


Figure 2. Hierarchical clustering performed using the GVP data of the currently studied family and nine unrelated individuals. The six family members clustered together (boxed), indicating similarity to each other in contrast to unrelated individuals.

Relationship Index Using Genotypic Data Corresponding to the Detected GVPs Acquired from Exome Sequencing

A likelihood ratio (LR) is traditionally used for relationship testing using STRs and/or SNPs where ratios >1 are evidence for individuals to be related, and the higher the LR, the stronger the evidence. However, the value of LR to indicate a relationship conclusively varies among laboratories from 1 to 10 or even 100 (American Association of Blood Banks, 2013; (Ge and Budowle, 2021). The present GVP data were analyzed using several relationship-testing approaches. Initially the corresponding SNP profiles for the GVP profiles of P, M, S1 and S2 were obtained from exome data (**Table S6**) and the profiles of S1 and S2 were tested for the likelihood that they were the offspring of the parents P and M. The likelihood ratios showed combined paternity indexes (CPIs) of 402904 and 5100 and posterior odds of 99.99% and 99.98% calculated with prior probabilities of 0.5 for S1 and S2, respectively. Sibship indexes calculated for the four individuals showed high combined sibship index (CSI) values strongly supporting a relationship for all the genetically related individuals except for M with S2 (7.2) (**Table 2**). This observation reflects a lower number of minor allelic GVPs shared by the siblings with their mother as compared to the father. At 66 of the 96 studied loci, all four analyzed members of the family were homozygous for the same allele. About 90% of this homozygosity was on the major alleles, and these loci added a CSI of 42.1 to the calculations. On the other 30 loci, S1 shared a minor allele with M and P at 6 and 8 loci while S2 shared 4 and 6 such alleles with M and P respectively.

Table 2: Combined sibship index values calculated using the genotype data for the GVP loci obtained only from the exome sequencing. P: father, M: mother, S1: sibling 1, S2: sibling 2.

	M	**S1**	**S2**
P	<0.01	161.30	150.97
M		17.39	7.22
S1			47.09

Relationship Index using proteomically-inferred genotypes

The evaluation of the genetic data was proceeded by the same analyses for SNPs inferred from the proteomic data. In this analysis, data from 8 European-American individuals in a previously published cohort (Plott 2020) were included to expand the GVP data from the currently studied family. Loci were assumed heterozygous if peptides encoded by both alleles and homozygous if peptides encoded by only one allele were seen in the proteomic data. Loci where none of the peptides was detected in any of the replicate samples of an individual were called as undetected or uninformative. GVPs for which the frequency of minor allele in European population was 0 were excluded from this analysis. Parentage indexes (ratios based on trio models) using P and M as parents and sibship indexes (ratios based on duo sibling models) for every possible pair of the individuals (from the family and from the additional subjects) were then calculated. The calculations were performed both including **(Table S7 and S8**) and excluding (with the rationale to not include the frequently observed false positive GVPs in real world practices) (**Table 3 and 4)** the false positive GVPs identified in the data. A locus was included in the calculation only if genotype information for that locus could be inferred for both (sibship calculation) or all three (parentage indexes calculation) individuals. Only the four actual children of the couple P and M showed CPIs and posterior probabilities that support the relationship (**Table 3**). The one locus at which an obligate allele was not found in S2 was rs1455555 in SERPINB5, a false negative (**Table S5**) which was kept out of the CPI calculation for S2 owing to the very low mutation rate per nucleotide (1-2 x 10^{-8}) (Kong et al, 2012). However, allele dropouts due to technical reasons (e.g. low volatilization of some peptides) that are much more likely in MS based proteomic analyses were taken into account by excluding peptides with a history of false negative detections. The unrelated individuals had at least three loci at which the obligate allele was not present either in the parents or the tested sample except for two (G and H) with two such loci. However, for these individuals the number of loci at which genotype information could be inferred was lower, 21 and 20, respectively.

Table 3: Combined paternity indexes and posterior probabilities calculated using the prior odds for the four true offspring and eight random individuals. (For each individual, the chance of being an offspring of the given parents is 4 of a total of 12 individuals or 4/12.) Using P and M as father and mother, the profiles of the 12 individuals were compared. The loci column

represents the number of genes used for each analysis with the values in parenthesis indicating numbers of genes with two loci. CPI: combined paternity index.

CPI when both parents are available				
Individual	Loci used in the calculation	Loci with no obligate allele	CPI	Posterior Probability (%)
S3	24(9)	0	1286.03	99.92
S1	25(8)	0	1676.36	99.94
S4	22(8)	0	258.23	99.61
S2	28(9)	1	1380.90	99.92
A	23(9)	3	--	--
B	23(9)	4	--	--
C	21(8)	5	--	--
D	20(9)	3	--	--
F	23(8)	4	--	--
G	21(7)	2	--	--
H	20(8)	2	--	--
I	25(7)	3	--	--

Sibship indexes were also calculated for each possible pair of the siblings and eight unrelated individuals belonging to the same population. A total of 91 comparisons were made. The calculated values were >10 for 13 of 14 true sibling pairs. The pair M-S1 was the only one with CSI value <10 (9.75) (**Table 4**). It was observed in the hierarchical clustering that the profiles of S1 and S2 were closer to the father than the mother. The low number of minor alleles shared with the mother could account for the lower relationship index value for this pair. Of the 77 unrelated pairs, only two pairs (A-C and D-I) had CSIs >10 falsely supporting a relationship. Consistent with the above calculations, the CPIs including FP-GVPs were more accurate compared to CSIs because of lower genetic similarity in siblings based on Mendelian inheritance patterns (1/4th chance of no allele identical by descent at a given locus) and the inclusion of two profiles (M and P) for comparison in CPI compared to one in CSI (Table S7 and S8). Even though a majority of the calculations were appropriate with a threshold CSI of 10 or greater, a certain threshold for inclusion or exclusion of sibship could not be established in the present study. (However, including more loci to the analysis in the future using optimized techniques (Goecker et al, 2020) should overcome this problem.) The number of loci at which each analysis was made are presented in **Table S9**. Nonetheless, the present findings support the usefulness of GVP profiles in statistically differentiating between related and unrelated individuals.

Table 4: Combined sibship index values calculated for the family members and 8 unrelated individuals. The CSI values higher than 10 for unrelated individuals or lower than 10 for true siblings are shown in bold, and the ones that support the relationship in the cases of true relationships are bold italicized.

	P	**S3**	**S1**	**S4**	**S2**	A	B	C	D	F	G	H	I
M	3.21	***198.13***	**9.75**	***1383.55***	***10.45***	0.51	5.93	1.39	6.08	0.06	8.96	0.09	2.70
P		***15.28***	***1203.59***	***11.04***	***287.77***	1.76	0.15	0.02	0.05	5.54	0.08	2.20	0.18
S3			***565.55***	***42.90***	***24.15***	1.41	0.77	0.34	2.67	0.25	0.45	0.07	2.83
S1				***15.03***	***125.22***	0.13	0.08	0.03	0.07	0.43	0.03	0.36	1.12
S4					***35.17***	0.66	0.86	0.06	2.32	0.07	1.18	0.15	4.19
S2						0.17	0.02	0.00	0.79	1.96	0.73	7.39	0.09
A							0.83	**235.11**	1.07	0.14	2.19	4.99	5.78
B								7.87	0.68	0.02	0.25	0.04	0.02
C									8.77	0.02	5.27	0.95	0.23
D										0.00	3.27	0.13	**16.74**
F											0.24	1.71	0.01
G												0.19	4.95
H													0.03
I													

DISCUSSION

This study investigates the potential for using proteomic variation, both protein abundance and amino acid sequence information, to compare measures of relatedness within and beyond the family unit. When investigating hair from unidentified remains, reference DNA may be difficult to obtain, while potentially related individuals may be available to investigators. Similarity in protein profiling, or calculations of relationship indices, may be all that can be obtained by investigators. Accordingly, different approaches to measuring relatedness were tested and compared: two way comparison of the proteomic profiles, measurement of correlation distancing using hierarchical clustering of GVP profiles, and indices of relationship using DNA and proteomic genotyping data. Like transcriptional analysis, proteomic profiling is the product of the transcriptional and translation program of each cell. A genetic role in modulating the relative expression and increased similarity in proteomic profiles within the family unit was observed. This was also true for GVP content even though one of the siblings had a genetic condition that affected the hair phenotype and the protein profile. Likewise, paternity and sibling indices using proteomically-inferred SNP allele genotypes showed elevated scores for related compared to unrelated individuals. This demonstrates the potential for using protein levels and sequences to assist in identification of unidentified remains.

Genetic match is the strongest and most widely accepted evidence for identification of tissues procured from crime scenes, resolving relationship conflicts and/or identification of remains in mass fatalities. To this end, probabilities for marker profiles and relationship indexes from the corresponding population genetics data can be calculated based on laws of Mendelian genetics, hence assigning a statistical value for the degree of match between profiles. As manifestations of allelic differences, permitting inference of corresponding single nucleotide polymorphisms in the genomic DNA, GVP proteomic data permit judging match or mismatch like other genetic marker systems. With random match probabilities as low as one in 640 million (Goecker et al, 2020), they have a greater power of discrimination than protein profiling among related and unrelated samples regardless of the age of individuals/hair samples, anatomic collection sites, chemical treatment and exposure of hairs to extreme conditions (Chu et al, 2020; Franklin et al, 2020; Plott et al, 2020). Thus, proteomics may provide useful information in cases where DNA evidence is insufficient due to age or suboptimal storage.

The combined sibship index values are defined as the likelihood of obtaining the genetic data when the two individuals are related versus unrelated; therefore, a higher LR value supports the relationship and vice versa (Ge and Budowle, 2021). However, the likelihood threshold values for inclusion, exclusion and inconclusive results vary among laboratories. According to American Association of Blood Banks nearly 6% of the laboratories use a LR threshold of 1 and a similar number of laboratories use 10 for inclusion in testing full siblings vs unrelated, while about 20% of the laboratories use 100 (Unit, 2013). In the current data, the likelihood threshold if kept at 10 supported all the relationships, but there were two unrelated pairs with CSI values >10. On the other hand, increasing the minimum LR value supporting a relationship to 100 eliminates the false positives in the data but brings 8 of the true relations into the uncertain range (1 < LR < 100), although not excluding them completely (<1). However, it should be noted that STRs traditionally used for such analyses often exhibit a greater degree of polymorphism (numbers of tandem repeats) than SNP loci, and the number of GVPs used in the current study were lower than the number of SNP loci used earlier in similar testing (Yousefi et al, 2018).

Moreover, the detection of GVPs limited the sensitivity of the SNP panel rather than the discriminatory powers of SNPs in a population.

Present GVP analysis provided promising results for relationship testing even using only 29-35 SNP marker loci for trio parent-child analyses **(Table 3)** and 30-49 loci in 20-28 genes for duo sibship analyses. The former, including data from two individuals (mother and father) for comparison and the obligation for certain alleles to be present, successfully identified the true relationships. Sibship analysis, on the other hand, was less discriminatory as has been seen for DNA analyses of 'duo sibship' cases. The 25% chance that two siblings will have no allele identical by descent at a given diploid locus leads to difficulty solving such identification cases (Lee et al, 2012). Therefore, increasing the amount of data used for the calculation both in terms of GVPs and individual profiles, e.g., comparing the profile of a subject to those of two known true siblings, can better discriminate among the related and unrelated individuals (Lee et al, 2012). In the case of STR markers, because of the higher degree of polymorphism at each locus, the number of markers sufficient to discriminate successfully between individuals is relatively low, 13-17, and ~30-40 for resolving second degree relationship status (Fimmers et al, 2008; Presciuttini et al, 2004). This number, due to the low mutation rate and polymorphism, is far higher for SNPs, ~50-150, where including a higher number of markers provides a higher power of discrimination (Chang et al, 2015; Phillips et al, 2008). The same holds true for GVPs in kinship analyses, since GVPs are the expressed manifestation of SNPs in the studied proteomes. Recently published hair sample processing procedures improve the number of GVP identifications by several fold, which will allow for more confident assignment of GVP heterozygosity and will result in higher discrimination and higher indices of relatedness (Goecker et al, 2020).

An obvious limitation of the current study is the inference of homozygosity at loci where the alternate allele was not detected. Detection of a peptide encoded by only one allele of a SNP locus provides half the number of markers on which the probability of match/randomness of a profile are calculated provided by DNA sequencing analyses. An intrinsic limitation of GVP detection when using shotgun proteomics is that the presence of an allele can be inferred but no claim can be made concerning alternate alleles. This is currently addressed using genotype frequencies instead of potential homozygotic frequencies for calculating random match probabilities but could lead to inaccuracy if the peptide representing another allele is not detected due, for example, to low volatility. This limitation will be alleviated when GVP quantitation becomes more precise using targeted mass spectrometry. This is a significant issue in analyzing relationships, as the kinship/relationship indexes calculations require data from both alleles at a locus. However, the negative predictive value obtained using the present categorization scheme improved by 20% the value for such data using an earlier approach (Borja et al, 2019, Parker, 2016 #2247), thereby increasing confidence in the inferences of SNP alleles. This study makes two assumptions that may change with more study and investigation. This study assumes homozygosity when only one GVP at a locus is detected. Of all the assumptions made for the four individuals whose GVP profiles could be validated, 92.6% were correct when compared with the exome data, whereas 7.2% were incorrect (3.6% were less conservative ($fa^2 < f2ab$) and 3.6% were more conservative ($fa^2 > f2ab$), a balanced outcome). In the future, as genotyping for GVP-inferred loci improves based on proteomic workflows and instruments that are more sensitive and quantitative, this assumption will become moot since the status of the alternate allele based on GVP quantitation could be inferred directly from proteomic detection. The second assumption is that GVP-inferred SNP loci were statistically independent unless they fell in the same gene. Observed SNP locus combinations within a gene were counted in the European population of the 1000 genome project to determine the genotype frequency of the diplotype, consistent with previous studies (Parker et al, 2016). This assumes that linkage disequilibrium dissipates beyond the gene boundary. Although this is worthy of revisiting in the future, it had little impact on final paternity index values in this instance. When calculations were made using both models, treating inferred SNP loci independently or expanding the boundaries of the locus to incorporate the entire reading frame, there was no change in the median sibship indices, and only 2 of 91 changes in concluded relationships (i.e,. sibship index < 10, data not shown). Even so, peptides that are consistently or frequently undetected using a certain protocol should be noted and not included in the analysis. Examples in the current dataset are rs3744786_T in KRT32, rs17843021_A in KRT39, rs2852464_C in KRT83, rs951773_A in KRT 84, rs9636845_T in KRTAP11, and rs13070515_A in LRRC15 **(Table S5)**. Moreover, employing different hair processing protocols, MS instruments or data acquisition strategies will lead to detection of a different set of peptides and proteins, thereby affecting the GVPs detected downstream. Nonetheless, the current study provides a basis and demonstrates feasibility for the use of GVPs in analyzing relationship status.

Proteomic profiling, as applied in this study, used label free quantification. Differential protein expression analysis was based on weighted spectral counts obtained from the Scaffold software. This type of label free quantitation is commonly used in proteomics for judging variation in a given protein's level among parallel samples (Dowle et al, 2016; Liu et al, 2004). Consistent with the expected correlation of hair protein profiles within the family, the profiles from three offspring were intermediate between those of the parents in two-way comparisons. Samples from the fourth offspring were distinctly different from both parents, however. The latter finding can possibly be attributed to departure of the hair from an unaffected phenotype due to a *de novo* heterozygous mutation (c.200G>A) in the ribosomal protein RPS23 (Paolini et al, 2017), although a connection to the observed perturbation of hair shaft protein levels in offspring S2 is not obvious. The genetic bases for numerous hair abnormalities are known, and others remain to be discovered (Duverger and Morasso, 2014). We speculate this example could illustrate how a genetic defect could result in an unusual phenotype due to loss of a critical protein or to perturbation of expression levels of a group of proteins in an intracellular signaling pathway. Proteomic analysis could potentially assist in diagnoses or help connect genotype and phenotype if the abnormalities manifested characteristic protein profiles.

CONCLUSION

The major significance of the present work for forensic casework is that GVP analysis of hair evidence offers a viable approach to testing familial relationships. The results obtained complement, and can be combined with, those from mitochondrial DNA analysis. Results from protein profiling, although not readily applicable to calculating random match probabilities, would be expected to support the outcomes of GVP analysis. Discrepancies in protein expression level that do not fit expectation within a family could be indicative of genetic differences not evident by GVP analysis. Such cases may be useful in discovery and characterization of genetic hair abnormalities.

Supplementary File Legends

Figure S1. Two way comparisons of hair protein levels among offspring and parents. Each Venn diagram shows the number of significant differences in samples from the father (P) and mother (M) with each other and with one sibling (S1-S4). Proteins in blue are those significantly different in amount from the mother in samples from sibling and father, while those in red are

those different from the father in samples from the mother and sibling. The two way differences between the family members are tabulated in the inset. Note S2 exhibited many more differences than the other siblings with P and M.

Table S1. Loci at which the genotype obtained from exome data of S2 was not consistent with the parents P and M. Assignments consistent with proteomic data are listed as "corrected".

Table S2. GVP data matrix used for hierarchical clustering. Each GVP detection was assigned a value of 1 and a non-detection of 0.

Table S3. Formulae to calculate paternity indices and sibship indices. Capital letters indicate alleles whereas lower case letters indicate the allele frequencies from 1000 Genome Project (Consortium et al, 2015).

Table S4. Cumulative GVP profiles identified in the six members of the family. The GVPs from P, M, S1 and S2 were validated from the corresponding genomic data. True positive identifications are highlighted in blue, true negative as white, false positive as red and false negative as green.

Table S5. GVPs identified in the six members of the family. The GVPs from P, M, S1 and S2 were validated from the corresponding genomic data. True positive identifications are highlighted in blue, true negative as white, false positive as red and false negative as green. GVPs present in the protein regions that were not sequenced in the MS runs were called as undetected and highlighted as grey.

Table S6. Genotypes of individual P, M, S1 and S2 for the identified genetically variant peptides. Genotypes at the five dubious loci are highlighted in bold italic.

Table S7. Combined paternity indexes and posterior probabilities calculated using all the detected GVPs including false positives. The posterior probabilities were obtained using the prior odds of 4/12 for the four true offspring and eight random individuals. Using P and M as father and mother, the profiles of the 12 individuals were compared. CPI: combined paternity index.

Table S8. Combined sibship index values calculated using all the detected GVPs including false positives for the family members and eight unrelated individuals. The CSI values higher than 10 for unrelated individuals or lower than 10 for true siblings are shown in bold, and the ones that support the relationship in the cases of true relationships are bold italicized.

Table S9. Number of loci at which each comparison was based for CSI calculations. The numbers inside the parentheses represent the genes with two loci included.

Table S10. Pairwise comparisons of protein levels in samples from the parents and siblings. Shown are the fold difference (FC) for each protein, calculated p values before (P.Value) and after (adj.P.Val) correction for multiple testing, identified proteins, accession numbers in the Uniprot human database and the protein molecular weight for each (MW).

References

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. Nature 526:68-74. doi: 10.1038/nature15393

Alsop RJ, Soomro A, Zhang Y, Pieterse M, Fatona A, Dej K, Rheinstädter MC (2016) Structural abnormalities in the hair of a patient with a novel ribosomopathy. PLoS One 11(3):e0149619. doi: 10.1371/journal.pone.0149619

American Association of Blood Banks (2013) Annual Report Summary for Testing in 2013. Relationship Testing Program Unit. https://www.aabb.org/docs/default-source/default-document-library/accreditation/2013-relationship-testing-summary-report.pdf?sfvrsn=da4315b2_2

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Royal Stat Soc, Ser B 57:289-300. doi: 10.1016/0306-9877(95)90228-7

Borja T, Karim N, Goecker Z, Salemi M, Phinney BS, Naeem M, Rice RH, Parker GJ (2019) Proteomic genotyping of fingermark donors with genetically variant peptides. Foren Sci Int: Genet 42:21-30. doi: 10.1016/j.fsigen.2019.05.005

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7. doi: 10.1186/s13742-015-0047-8

Chu F, Mason KE, Anex DS, Jones AD, Hart BR (2019) Hair proteome variation at different body locations on genetically variant peptide detection for protein-based human identification. Sci Rep 9(1):7641. doi: 10.1038/s41598-019-44007-7

Chu F, Mason KE, Anex DS, Jones AD, Hart BR (2020) Proteomic characterization of damaged single hairs recovered after an explosion for protein-based human identification. J Proteome Res 19:3088-3099. doi: 10.1021/acs.jproteome.0c00102

Dowle AA, Wilson J, Thomas JR (2016) Comparing the diagnostic classification accuracy of iTRAQ, peak-area, spectral-counting, and emPAI methods for relative quantification in expression proteomics. J Proteome Res 15:3550-3562. doi: 10.1021/acs.jproteome.6b00308

Duverger O, Morasso MI (2014) To grow or not to grow: Hair morphogenesis and human genetic hair disorders. Seminars Cell Develop Biol 25-26:22-33. doi: 10.1016/j.semcdb.2013.12.006

Fimmers R, Baur M, Rabold U, Seifried E, Seidl C (2008) STR-profiling for the differentiation between related and unrelated individuals in cases of citizen rights. For Sci Intl: Genet Suppl Ser 1:510-513. doi: 10.1016/j.fsigss.2008.01.002

Franklin RN, Karim N, Goecker ZC, Durbin-Johnson BP, Rice RH P, G J (2020) Proteomic genotyping: Using mass spectrometry to infer SNP genotypes in pigmented and non-pigmented hair. Foren Sci Int 310:110200. doi: 10.1016/j.forsciint.2020.110200

Ge J, Budowle B (2021) Forensic investigation approaches of searching relatives in DNA databases. J Forensic Sci 66:430-443. doi: 10.1111/1556-4029.14615

Goecker ZC, Salemi MR, Karim N, Phinney BS, Rice RH, Parker GJ (2020) Optimal processing for proteomic genotyping of single human hairs. Foren Sci Int: Genet 47:102314. doi: 10.1016/j.fsigen.2020.102314

Hindorff LA, Sethupathy P, Junkinsa HA, Ramosa EM, Mehtac JP, Collins FS, Manolio TA (2009) Potential etiologic and functional implications ofgenome-wide association loci for human diseasesand traits. Proc Natl Acad Sci USA 106:9362-9367. doi: 10.1073pnas.0903103106

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, Wong W, Sigurdsson G, Walters GB, Steinberg S, Helgason H, Thorleifsson G, Gudbjartsson DF, Helgason A, Magnusson OT, Thorsteinsdottir U, Stefansson K (2012) Rate of de novo mutations and the importance of father's age to disease risk. Nature 488:471-475. doi: 10.1038/nature11396

Laatsch CN, Durbin-Johnson BP, Rocke DM, Mukwana S, Newland AB, Flagler MJ, Davis MG, Eigenheer RA, Phinney BS, Rice RH (2014) Human hair shaft proteomic profiling: individual differences, site specificity and cuticle analysis. PeerJ 2:e506. doi: 10.7717/peerj.506

Lee JC, Lin YY, Tsai LC, Lin CY, Huang TY, Chu PC, Yu YJ, Linacre A, Hsieh HM (2012) A novel strategy for sibship determination in trio sibling model. Croat Med J 53:336-342. doi: 10.3325/cmj.2012.53.336

Linch CA, Whiting DA, Holland MM (2001) Human hair histogenesis for the mitochondrial DNA forensic scientist. J Forensic Sci 46:844-853. doi: 10.1520/JFS15056J

Liu H, Sadygov RG, Yates JRI (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Analyt Chem 76:4193-4201. doi: 10.1021/ac0498563

Martin-Trujillo A, Patel N, Felix Richter F, Bharati Jadhav B, Garg P, Morton SU, McKean DM, R DS, Goldmuntz E, Gruber D, Kim BR, Jane W. Newburger JW, Porter GAJ, Alessandro Giardini A, Bernstein D, Tristani-Firouzi M, Seidman JG, Seidman CE, Chung WK, Gelb BD, Sharp AJ (2020) Rare genetic variation at transcription factor binding sites modulates local DNA methylation profiles. PLoS Genetics 16(11):e1009189. doi: 10.1371/journal.pgen.1009189

McNevin D, Wilson-Wilde L, Robertson J, Kyda J, Lennard C (2005) Short tandem repeat (STR) genotyping of keratinised hair. Part 1. Review of current status and knowledge gaps. Foren Sci Int 153:237-246. doi: 10.1016/j.forsciint.2005.05.005

Milan J, Wu P-W, Salemi M, Durbin-Johnson B, Rocke DM, Phinney BS, Rice RH, Parker GJ (2019) Comparison of protein expression levels and proteomically-inferred genotypes using human hair from different body sites. Foren Sci Int: Genet 41:19-23. doi: 10.1016/j.fsigen.2019.03.009

National Research Council (2009) Strenghtening Forensic Science in the United States: A Path Forward. The National Academies Press, pp 155-161. doi: 10.17226/12589

Paolini NA, Attwood M, Sondalle SB, dos Santos Vieira CM, van Adrichem AM, di Summa FM, O'Donohue M-F, Gleizes P-E, Rachuri S, Briggs JW, Fischer R, Ratcliffe PJ, Wlodarski MW, Houtkooper RH, von Lindern M, Kuijpers TW, Dinman JD, Baserga SJ, Cockman ME, MacInnes AW (2017) A ribosomopathy reveals decoding defective ribosomes driving human dysmorphism. Am J Hum Genet 100:506-522. doi: 10.1016/j.ajhg.2017.01.034

Parker GJ, Leppert T, Anex DS, Hilmer JK, Matsunami N, Baird L, Stevens J, Parsawar K, Durbin-Johnson BP, Rocke DM, Nelson C, Fairbanks DJ, Wilson AS, Rice RH, Woodward SR, Bothner B, Hart H, Leppert M (2016) Demonstration of protein-based human identification using the hair shaft proteome. PLoS One 11(9):e0160653. doi: 10.1371/journal.pone.0160653

Phillips C, Fondevila M, García-Magariños M, Rodriguez A, Salas A, Carracedo A, Lareu MV (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. Foren Sci Int: Genet 2:198-204. doi: 10.1016/j.fsigen.2008.02.002

Plott TJ, Karim N, Durbin-Johnson BP, Swift DP, Youngquist RS, Salemi M, Phinney BS, Rocke DM, Davis MG, Parker GJ, Rice RH (2020) Age-related changes in hair shaft protein profiling and genetically variant peptides Foren Sci Int: Genet 47:102309. doi: 10.1016/j.fsigen.2020.102309

Presciuttini S, Toni C, Marronia F, Spinetti I, Bailey-Wilson JE, Domenici R (2004) The number of STR markers necessary to resolve relationships in deficiency paternity cases. Int Congr Ser 1261:541-543. doi: doi.org/10.1016/S0531-5131(03)01664-9

Rice RH, Bradshaw KM, Durbin-Johnson BP, Rocke DM, Eigenheer RA, Phinney BS, Sundberg JP (2012) Differentiating inbred mouse strains from each other and those with single gene mutations using hair proteomics. PLoS One 7:e51956. doi: 10.1371/journal.pone.0051956

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucl Acids Res 43(7):e47. doi: 10.1093/nar/gkv007

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25. doi: 10.1186/gb-2010-11-3-r25

Sozer A, Baird M, Beckwith M, Harmon B, Lee D, Riley G, Schmitt S (2010) Guidelines for mass fatality DNA identification operations. American Association of Blood Banks. https://www.aabb.org/docs/default-source/default-document-library/about/guidelines-for-mass-fatality-dna-identification-operations.pdf?sfvrsn=af1c96a9_0

Wenk RE, Traver M, Chiafari FA (1996) Determination of sibship in any two persons. Transfusion 36:259-262. doi: 10.1046/j.1537-2995.1996.36396182146.x

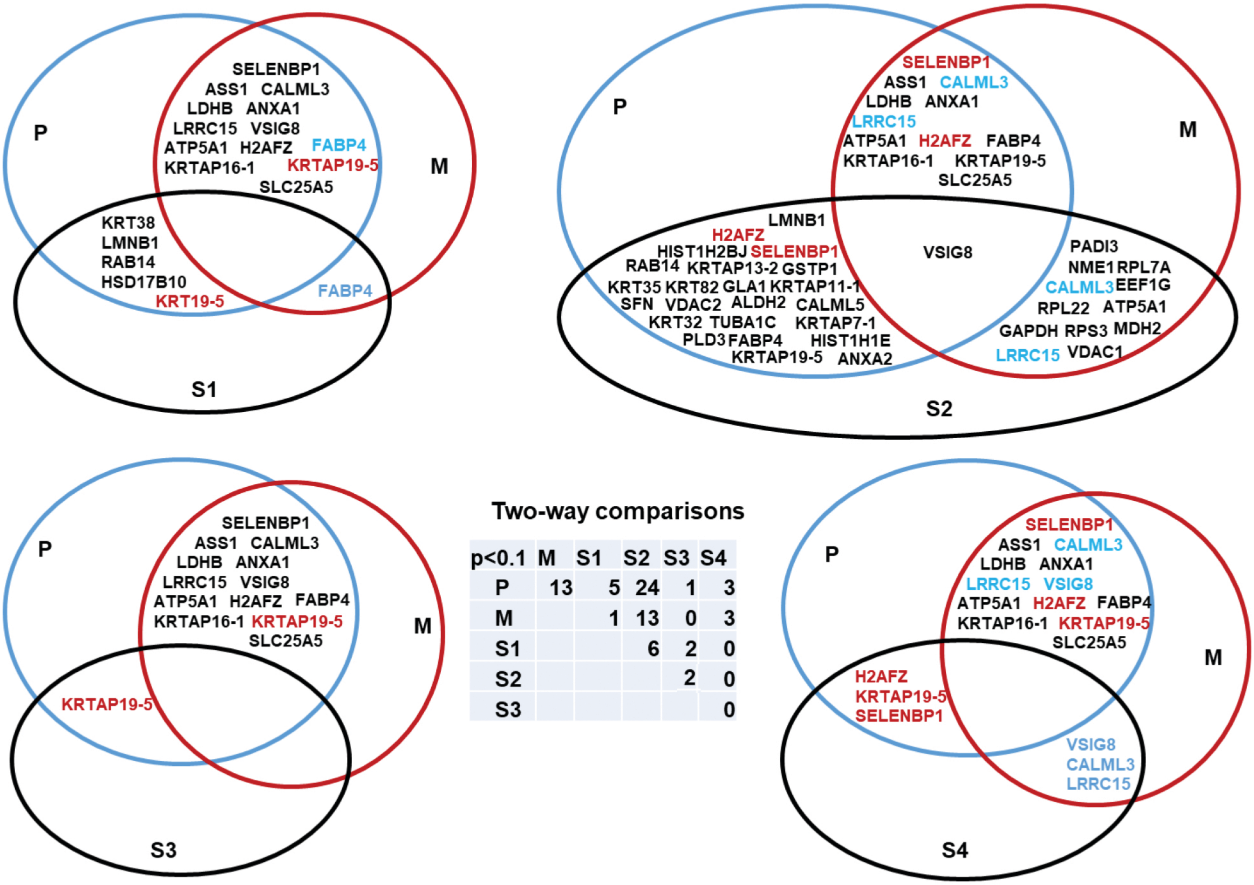
Wu P-W, Mason KE, Durbin-Johnson BP, Salemi M, Phinney BS, Rocke DM, Parker GJ, Rice RH (2017) Proteomic analysis of hair shafts from monozygotic twins: Expression profiles and genetically variant peptides. Proteomics 17:13-14, 1600462. doi: 10.1002/pmic.201600462

Yousefi S, Abbassi-Daloii T, Kraaijenbrink T, Vermaat M, Mei H, van 't Hof P, van Iterson M, Zhernakova DV, Claringbould A, Lude Franke L, 't Hart LM, Slieker RC, van der Heijden A, de Knijff P, consortium B, 't Hoen PAC (2018) A SNP panel for identification of DNA and RNA specimens. BMC Genomics 19(1):90. doi: 10.1186/s12864-018-4482-7

Elucidation of Familial Relationships Using Hair Shaft Proteomics

Supplementary Material

(Figure S1, Tables S1-S9)



Two-way comparisons

p<0.1	M	S1	S2	S3	S4
P	13	5	24	1	3
M		1	13	0	3
S1			6	2	0
S2				2	0
S3					0

Figure S1. Two way comparisons of hair protein levels among offspring and parents. Each Venn diagram shows the number of significant differences in samples from the father (P) and mother (M) with each other and with one sibling (S1-S4). Proteins in blue are those significantly different in amount from the mother in samples from sibling and father, while those in red are those different from the father in samples from the mother and sibling. The two way differences between the family members are tabulated in the inset. Note S2 exhibited many more differences than the other siblings with P and M.

Table S1 Loci at which the genotype obtained from exome data of S2 was not consistent with the parents P and M. Assignments consistent with proteomic data are listed as “corrected”.

Gene Name	rs#	Reference	P	M	S1	S2	S2 corrected
KRT32	rs2604953	G	TT	TT	TT	GG	TT
KRTAP4-1	rs398825	C	TT	CT	TT	CC	??
KRTAP4-9	rs113059833	A	AA	AA	AA	AT	AA
KRTAP9-2	rs9902235	G	GG	GG	GG	CC	**
KRTAP10-6	rs465279	G	GG	AA	GA	GG	AG

?? no peptides

** false positives

Analysis of the DNA profile of S2 revealed 5 loci at which the genotype was not consistent with those of the parents. These included rs2604953 (KRT32), rs398825 (KRTAP4-1), rs113059833 (KRTAP 4-9), rs9902235 (KRTAP9-2) and rs465279 (KRTAP10-6) (Table S5). This problem was attributed to the exome analysis of S2 being performed using an older technique and at a separate time from those of P, M and S1. However, the data obtained from the proteomic analysis at these loci were consistent with the parental genotypes. According to DNA sequencing, both parents were homozygous TT for rs2604953, but S2 was homozygous GG at that position. The proteomic data showed peptides supporting only the T allele in S2, consistent with the parental genotypes. Similarly, for rs113059833 the DNA data of S2 showed a heterozygous AT genotype, but the parents were homozygous for A at that position. The proteomic data for S2 showed translation products only of an A as expected from the genotypes of the parents. For rs465279, the parental genotypes were AA and GG, inconsistent with the sequence of S2 as GG whereas, in the proteomic data, peptides for both the alleles were seen (Table S3). There was no proteomic information at the locus for rs398825 in any of the S2 replicates. The GVP corresponding to rs9902235 in KRTAP9-2 seemed unreliable since the other members of the family had it as a false positive in their GVP profiles.

Note: Table S2 is at the end of the file after Table S9.

Table S3: Formulae to calculate paternity indices and sibship indices. Capital letters indicate alleles whereas lower case letters indicate the allele frequencies from 1000 genome project (1000 Genomes Project Consortium et al, 2015).

Paternity Indices Calculations

Parent 1	Parent 2	Subject	Formula
AA	AA	AA	$1/a^2$
AA	BB	AB	1/2ab
AA	AB	AA	$1/2a^2$
AA	AB	AB	1/4ab
AB	AB	AB	1/4ab
AB	AB	AA	$1/4a^2$
AA	BC	AB	1/4ab
AB	AC	AA	$1/4a^2$
AB	BC	AB	1/8ab
AB	BC	BC	1/8bc

Sibship Indices Calculations

Subject 1	Subject 2	Formula
AA	AA	$(1+a)^2/(2a)^2$
AA	AB	(1+a)/4a
AB	AB	(1+a+b+2ab)/8ab
AA	BB	1/4
AB	AC	(1+2a)/8a

Table S4: Cumulative GVP profiles identified in the six members of the family. The GVPs from P, M, S1 and S2 were validated from the corresponding genomic data. True positive identifications are highlighted in blue, true negative as white, false positive as red and false negative as green.

Gene Name	rs#_nucleotide	SAP	peptide sequence	P	M	S1	S2	S3	S4
ALDH2	rs671_G	E504K	ELGEYGLQAYTEVK	🟦	🟦	🟦	🟦	1	1
ALDH2	rs671_A	E504K	ELGEYGLQAYTk						
ATG9B	rs7804893_T	N493S	HFNELPHELR	🟦	🟦	🟦	🟦	1	
ATG9B	rs7804893_C	N493S	HFsELPHELR						
ATP5A1	rs79011243_C	A32S	VLSIGDGIAR	🟦	🟦	🟦	🟦	1	1
ATP5A1	rs79011243_A	A32S	VLSIGDGIsR						
CSRP1	rs3738283_T	K108I	HEEAPGHRPTTNPNASK	🟦	🟦	🟦	🟦	1	1
CSRP1	rs3738283_A	K108I	HEEAPGHRPTTNPNASiFAQK						
DSC3	rs276937_A	S78T	VLNDGSVYTAR		🟦	🟩	🟦		
DSC3	rs276937_T	S78T	VLNDGtVYTAR	🟩		🟩	🟩		
DSC3	rs35296997_T	K180Q	GVDKEPLNLFYIER	🟩	🟦	🟦	🟦	1	
DSC3	rs35296997_G	K180Q	GVDqEPLNLFYIER						
DSP	rs80325569_G	G939S	NLHSEISGK	🟦	🟦	🟦	🟦	1	
DSP	rs80325569_A	G939S	NLHSEISsK						
DSP	rs2076299_A	Y1512C	VQYDLQK	🟩	🟩	🟩	🟦		1
DSP	rs2076299_G	Y1512C	VQcDLQK	🟩					
DSP	rs28763966_C	N1526K	ANSSATETINK	🟩	🟩	🟩	🟦	1	
DSP	rs28763966_A	N1526K	ANSSATETIk						
DSP	rs6929069_A	R1738Q	GqSEADSDKNATILELR						
DSP	rs6929069_G	R1738Q	GRSEADSDKNATILELR/SEADSDKNATILELR	🟦	🟦	🟦	🟦	1	1
DSP	rs28763967_C	R1537C	VQEQELTR	🟦	🟦	🟦	🟦	1	1
DSP	rs28763967_T	R1537C	VQEQELTcLR						
FAM83H	rs9969600-C	Q201H	VNLQHVDFLR						
FAM83H	rs9969600-A/G	Q201H	VNLhHVDFLR	🟩	🟩	🟦	🟩		
GSDMA	rs3894194_A	R18Q	QLNPqGDLTPLDSLIDFK						
GSDMA	rs3894194_G	R18Q	QLNPR/GDLTPLDSLIDFK	🟦	🟦	🟦	🟦	1	1
GSDMA	rs7212938_G	V128L	ALETVQER						
GSDMA	rs7212938_T	V128L	ALETlQER	🟦	🟩	🟩	🟩		1
GSDMA	rs56030650_A	T314N	GHEVnLEALPK	🟩					
GSDMA	rs56030650_C	T314N	GHEVTLEALPK	🟩	🟩	🟦	🟩		
GSTP1	rs1138272_C	A114V	YISLIYTNYEAGKDDYVK	🟦	🟩	🟦	🟦	1	1
GSTP1	rs1138272_T	A114V	YISLIYTNYEvGKDDYVK						
GSTP1	rs1695_A	I105V	YISLIYTNYEAGKDDYVK	🟦	🟩	🟦	🟦		
GSTP1	rs1695_G	I105V	YvSLIYTNYEAGKDDYVK		🟩		🟦		
HEXB	rs10805890_A	I207V	GILIDTSR	🟦	🟦	🟦	🟦	1	1
HEXB	rs10805890_G	I207V	GILvDTSR						
HEXB	rs77499935_A	I420V	LAPGTIVEVWKDSAYPEELSR/LAPGTIVEVWK	🟦	🟦	🟦	🟦	1	1
HEXB	rs77499935_G	I420V	LAPGTvVEVWKDSAYPEELSR						
JUP	rs41283425_C	R142H	SAIVHLINYQDDAELATR	🟦	🟦	🟦	🟦	1	1
JUP	rs41283425_T	R142H	SAIVHLINYQDDAELAThALPELTK						
JUP	rs143043662_C	V648I	NEGTATYAAAVLFR	🟦	🟦	🟦	🟦	1	1
JUP	rs143043662_T	V648I	NEGTATYAAAiLFR						
KRT1	rs17678945_A	A454S	NKLNDLEDALQQsKEDLAR/LNDLEDALQQsK						
KRT1	rs17678945_C	A454S	NKLNDLEDALQQAKEDLAR/LNDLEDALQQAK	🟦	🟩	🟩	🟦	1	1
KRT31	rs6503627_A	A82V	DNvELENLIR/QLERDNvELENLIR		🟦		🟥	1	1

KRT32	rs2071561_T	S222Y	ADLEAQVEyLK	🟦		🟦	🟦	1	
KRT32	rs72830046_C/rs	R280H/C	CQYEAMVEANRR	🟦	🟦	🟦	🟦	1	1
KRT32	rs72830046_T	R280H	CQYEAMVEANhR	🟦		🟦		1	
KRT32	rs2604953_G	P427T	SLLENEDCKLPCNPCSTPSCTTCVPSPCVPR/LPCNPCSTPSCTTCVPSPCVPR				🟩		
KRT32	rs2604953_T	P427T	SLLENEDCKLPCNPCSTPSCTTCVPSPCVtR/LPCNPCSTPSCTTCVPSPCVtR	🟩	🟩	🟩	🟥	1	
KRT32	rs3744786_T	Q72R	TYLSSSCQAASGISGSMGPGSWYSEGAFNGNEK	🟩	🟩	🟩	🟩		
KRT32	rs3744786_C	Q72R	TYLSSSCr	🟦		🟦	🟦	1	
KRT32	rs2071560_A	I171T	MVVNIDNAK	🟦	🟦	🟦	🟦	1	1
KRT32	rs2071560_G	I171T	MVVNtDNAK						
KRT32	rs146792525_C	A255T	LNIEVDAAPPVDLTR	🟦	🟦	🟦	🟦	1	1
KRT32	rs146792525_T	A255T	LNIEVDtAPPVDLTR						
KRT33A	rs373657561_C	G33R	PCVPPSCHGCTLPGACNIPANVSNCNWFCEGSFNGSEK	🟦	🟦	🟦	🟦	1	1
KRT33A	rs373657561_T	G33R	PCVPPSCHGCTLPr						
KRT33A	rs12937519_A	A270V	QVVSSSEQLQSYQvEIIELR		🟦			1	1
KRT34	rs2071599_T	H348R	DSLENTLTESEAHYSSQLSQVQSLITNVESQLAEIR	🟦	🟦	🟦	🟦	1	1
KRT35	rs743686_A	S36P	VSAMYSSSSCKLPSLSPVAR		🟦			1	1
KRT35	rs743686_G	S36P	VSAMYSSSpCKLPSLSPVAR	🟦	🟦	🟦	🟦	1	1
KRT35	rs200355130-C	E141D	LVVEIDNAK	🟦	🟦	🟦	🟦	1	1
KRT35	rs200355130-A	E141D	LVVdIDNAK	🟥	🟥		🟥	1	1
KRT35	rs138303882_A	R163W	YETEVSLwQLVESDINGLR						
KRT35	rs138303882_G	R163W	YETEVSLRQLVESDINGLR/QLVESDINGLR	🟦	🟦	🟦	🟦	1	1
KRT36	rs75790652_G	A202G	CQLGDRLNVEVDAAPPVDLNK/LNVEVDAAPPVDLNK	🟦	🟦	🟦	🟦	1	1
KRT36	rs75790652_C	A202G	CQLGDRLNVEVDgAPPVDLNK						
KRT36	rs11657323_T	N357T	YSSQLAQMQCLISNVEAQLSEIR	🟩	🟩	🟩	🟦		
KRT36	rs11657323_G	N357T	YSSQLAQMQCLIStVEAQLSEIR	🟩	🟩	🟩	🟩		
KRT36	rs9904102_G	R277C	CQYEALVENNR	🟩	🟩	🟦	🟩		
KRT36	rs9904102_A	R277C	CQYEALVENNcR						
KRT39	rs17843021_G	T341M	DSQECILTETEAR	🟦	🟦	🟦	🟦	1	1
KRT39	rs17843021_A	T341M	DSQECILmETEAR	🟩					
KRT39	rs7213256_C	R456Q	SGAIESTAPACTSSSPCSLKEHCSACGPLSR/EHCSACGPLSR/EHCSACGPLSRILVK	🟦	🟦	🟦	🟦	1	1
KRT39	rs7213256_T	R456Q	SGAIESTAPACTSSSPCSLKEHCSACGPLSqLLVK/EHCSACGPLSqLLVK/EHCSACGPLSqILVK						
KRT39	rs17843023_G	L383M	QNQEYEILLDVK	🟦	🟦	🟦	🟦	1	1
KRT39	rs17843023_T	L383M	QNQEYEILmDVK						
KRT39	rs112557906_G	S423F	CEPSPWTSCK	🟦	🟦	🟦	🟦	1	1
KRT39	rs112557906_A	S423F	CEPSPWTfCK						
KRT39	rs142154718_C	S86N	FSLDDCSWYGEGINSNEK	🟦	🟦	🟦	🟦	1	1
KRT39	rs142154718_T	S86N	FSLDDCnWYGEGINSNEK						
KRT40	rs2010027_C	R235H	NHEEEVNLLREQLGDR/NHEEEVNLLR	🟦	🟦	🟦	🟦	1	1
KRT40	rs2010027_T	R235H	NHEEEVNLLhEQLGDR	🟦					
KRT40	rs140634473_C	R108H	R.SLEETNAELESR	🟦	🟦	🟦	🟦	1	1
KRT40	s140634473_T	R108H	*V h SLEETNAELESR*						
KRT7	rs6580870_A	H186R	NKYEDEINHRTAAENEFVVLK	🟩			🟩		

KRT7	rs6580870_G	H186R	NKYEDEINrRTAAENEFVVLK	🟩	🟩	🟦	🟩		
KRT81	rs6580873_A	L248R	LYEEEILILQSHISDTSVVVK	🟦	🟦	🟦	🟥	1	1
KRT82	rs2658658_A	T458M	GAFLYEPCGVSmPVLSTGVLR	🟦	🟦		🟦		1
KRT82	rs2658658_G	T458M	GAFLYEPCGVSTPVLSTGVLR	🟦	🟦	🟦		1	1
KRT82	rs1732263_C	E452D	GAFLYEPCGVSTPVLSTGVLR	🟦	🟩	🟩	🟦		
KRT82	rs1732263_G	E452D	GAFLYdPCGVSTPVLSTGVLR						
KRT82	rs1791634_C	E219Q	KYEEELSLRPCVENEFVALK	🟦	🟦	🟦	🟦	1	1
KRT82	rs1791634_G	E219Q	KYEEELSLRPCVqNEFVALK						
KRT83	rs61485872_A	C23G	PGNFSCVSACGPR	🟦	🟦	🟦	🟦		1
KRT83	rs61485872_C	C23G	PGNFSCVSAgGPR						
KRT83	rs2852464_C	I279M	DLNMDCmVAEIK	🟩			🟩		
KRT83	rs2852464_G	I279M	DLNMDCIVAEIK	🟦	🟦	🟦	🟦	1	1
KRT83	rs2857663_G	R149C	LQFYQNR.ECCQSNLEPLFAGYIETLR/LQFYQNR	🟦	🟦	🟦	🟦	1	1
KRT83	rs2857663_A	R149C	LQFYQNcECCQSNLEPLFAGYIETLR						
KRT84	RS951773_A	C446R	CEYQELMNAKLGLDIEIATYR		🟩	🟩			
KRT84	RS951773_G	C446R	QLrEYQELMNAKLGLDIEIATYR	🟦	🟦	🟦	🟦	1	1
KRT85	rs2852471-C	W155L	WQFYQNQR	🟦	🟦	🟦	🟦	1	1
KRT85	rs2852471-A	W155L	lQFYQNQR	🟥	🟥	🟥	🟥	1	1
KRTAP1-5	rs148449559_G	T32S	TCCQTSFCGYPSFSISGTCGSSCCQPSCCETSCCQPR	🟩	🟩	🟦	🟩		
KRTAP1-5	rs148449559_C	T32S	TCCQTSFCGYPSFSISGTCGSSCCQPSCCEsSCCQPR						
KRTAP1-5	rs62623375_C	C35Y	MTCCQTSFCGYPSFSISGTCGSSCCQPSCCETSCCQPR	🟩	🟦	🟦	🟦	1	
KRTAP1-5	rs62623375_T	C35Y	MTCCQTSFCGYPSFSISGTCGSSCCQPSCCETSCyQPR						
KRTAP3-2	rs9897046_T	S8G	MDCCASRSCSVPTGPATTICSSDKSCR	🟦	🟦	🟦	🟦	1	1
KRTAP3-2	rs3829598_G	R27C	SCSVPTGPATTICSSDKSCR	🟦	🟦	🟦	🟦		
KRTAP3-2	rs3829598_A	R27C	K.SCCCGVCLPSTCPHTVWLLEPTCCDNCPPPCHIPQPCVPTCFLLNSCQPTPGLETLNLTTFTQPCCEPCLPR.G						
KRTAP3-2	rs3813050_A	I46T	CGVCLPSTCPHTVWLLEPICCDN	🟦	🟩	🟦	🟦		
KRTAP4-1	rs398825_C	T134A	TTCCRPSCCGSSC-		🟦		🟩	1	1
KRTAP4-1	rs398825_T	T134A	aTCCRPSCCGSSC-	🟩	🟦	🟩		1	1
KRTAP4-3	rs428371_G	P152S	PACCISSCCHPSCCVSSCR	🟦	🟦	🟦	🟦		1
KRTAP4-3	rs428371_A	P152S	sACCISSCCHPSCCVSSCR						
KRTAP4-4	rs366700_C	R154S	TTCCRPSCCVSRCYR/TTCCRPSCCVSR/TTCCRPSCCVSRCYRPHCGQSLCC-	🟦	🟦	🟦	🟦	1	1
KRTAP4-4	rs366700_G	R154S	TTCCRPSCCVSsCYR/TTCCRPSCCVSsCYRPHCGQSLCC-						
KRTAP4-4	rs385055_T	Y25C	VNSCCGSVCSDQGCGLENCCRPSYCQTTCCR	🟦	🟦	🟦	🟦	1	1
KRTAP4-4	rs75030409_T	Q109R	TTCCRPSCCRPQCC	🟦	🟦	🟦	🟦		1
KRTAP4-4	rs75030409_C	Q109R	TTCCRPSCCRPr						
KRTAP4-6	rs73983172_G	P63S	R.TTCCRPSCCVSSCCRPQCCQSVCCQPTCCRPSCCPSCCQTTCCR.T	🟦	🟩	🟩	🟦		1
KRTAP4-9	rs149483591_G	R26H	VSSCCGSVCSDQGCGQDLCQETCCR	🟩	🟩	🟩	🟦		1
KRTAP4-9	rs149483591_A	R26H	VSSCCGSVCSDQGCGQDLCQETCChPSCCETTCCR						
KRTAP4-9	rs113059833_A	D18V	VSSCCGSVCSDQGCGQDLCQETCCRPSCCETTCCR	🟩	🟩	🟩	🟦		1
KRTAP4-9	rs113059833_T	D18V	VSSCCGSVCSDQGCGQvLCQETCCRPSCCETTCCR				🟩		
KRTAP5-2	rs35925287_C	G29R	GCGSGCGGCGSSCGGCGSGCGGCGSGR	🟦	🟦	🟦	🟦		1
KRTAP5-2	rs35925287_T	G29R	GCGSGCGGCGSSCGGCGSGCr						
KRTAP9-2	rs9902235_C	C56S	CRPTsCQNTCCR	🟥	🟥	🟥	🟦	1	1

KRTAP9-2	rs9902235_G	C56S	CRPTCCQNTCCR	🟩	🟦			1	1
KRTAP9-4	rs2191379_A	S146Y	R.TCYYPTTVCLPGCLNQSCGSNCCQPCCRPACCETTCFQPTCVySCCQPFCC-	🟩	🟩	🟩	🟦		
KRTAP9-4	rs2191379_C	S146Y	RTCYYPTTVCLPGCLNQSCGSNCCQPCCRPACCETTCFQPTCVSSCCQPFCC-		🟩				
KRTAP10-6	rs465279_A	S300P	SSSSVSLLCHPVCK		🟦	🟦	🟥	1	
KRTAP10-12	rs61745911_G	C236Y	LASCGSLLCR	🟦	🟦	🟩	🟦	1	
KRTAP10-12	rs61745911_A	C236Y	LASCGSLLyR						
KRTAP10-12	rs34302939_G	G226S	RVPVPSCCVPTSSCQPSCGR/VPVPSCCVPTSSCQPSCGR	🟦	🟦	🟦	🟦	1	
KRTAP10-12	rs34302939_A	G226S	RVPVPSCCVPTSSCQPSCsR						
KRTAP11-1	rs71321355_C	R72Q	CIVPVAQVTTTSTTDADCLGGICLPSSFQTGSWLLDHCQETCCEPTACQPTCYRR/R.RTSCVSNPCQVTCSR	🟦	🟦	🟦	🟦	1	1
KRTAP11-1	rs71321355_T	R72Q	CIVPVAQVTTTSTTDADCLGGICLPSSFQTGSWLLDHCQETCCEPTACQPTCYqR						
KRTAP11-1	rs79258920_G	S78F	TSCVSNPCQVTCSR	🟦	🟦	🟦	🟦	1	1
KRTAP11-1	rs79258920_A	S78F	TSCVfNPCQVTCSR						
KRTAP11-1	rs9636845_A	C111S	QTTCISNPCSTTYSRPLTFVSSGCQPLGGISSVCQPVGGISTVCQPVGGVSTVCQPACGVSR	🟦	🟦	🟦	🟦	1	1
KRTAP11-1	rs9636845_T	C111S	QTTCISNPCSTTYSRPLTFVSSGsQPLGGISSVCQPVGGISTVCQPVGGVSTVCQPACGVSR		🟩	🟩			
KRTAP16-1	rs2074285_G	P340R	RCPSVCPEPVSCPSTSCR	🟦	🟦	🟦	🟦	1	1
KRTAP16-1	rs2074285_C	P340R	RCrSVCPEPVSCPSTSCR		🟦			1	
LAMP1	rs9577230_T	I309T	FFLQGIQLNTILPDAR	🟦	🟦	🟦	🟦	1	1
LAMP1	rs9577230_C	I309T	FFLQGIQLNTtLPDAR						
LGALS3	rs10148371_G	R183K	LDNNWGR	🟦	🟦	🟦	🟦	1	1
LGALS3	rs10148371_A	R183K	LDNNWGk						
LGALS3	rs11125_A	Q201H	IQVLVEPDHFK	🟦	🟦	🟦	🟦	1	1
LGALS3	rs11125_T	Q201H	IhVLVEPDHFK						
LRRC15	rs13070515_A	P286L	ELSlGIFGPMPNLR		🟩				
LRRC15	rs13070515_G	P286L	ELSPGIFGPMPNLR	🟦	🟦	🟦	🟦	1	
NEU2	rs2233384_C	S11R	ESVFQSGAHAYR	🟦	🟦	🟦	🟦	1	
NEU2	rs2233384_A	S11R	ASLPVLQKEr						
NEU2	rs2233385_G	R41Q	IPALLYLPGQQSLLAFAEQR	🟩	🟩	🟦	🟩	1	1
NEU2	rs2233385_A	R41Q	IPALLYLPGQQSLLAFAEQq						
NEU2	rs2233390_G	A145T	DLTDAAIGPAYR	🟦	🟦	🟦	🟦	1	1
NEU2	rs2233390_A	A145T	DLTDtAIGPAYR						
PKP1	rs61818256_C	R684W	AAEAARLLLSDMWSSK/LLLSDMWSSK	🟦	🟦	🟦	🟦	1	1
PKP1	rs61818256_T	R684W	AAEAAwLLLSDMWSSK						
PLCD1	rs933135_C	R257H	EEAAGPALALSLIER	🟩	🟩	🟦	🟩		
PLCD1	rs933135_T	R257H	EEAAGPALALSLIEhYEPSETAK						
PPL	rs2037912_C	Q1573E	eNLQLETR	🟩	🟩	🟩	🟩		
PPL	rs2037912_G	Q1573E	QNLQLETR	🟩	🟩	🟦			
PPL	rs143676756_C	R1457Q	VVLQQDPQQAREHALLR	🟩	🟩	🟩	🟦		
PPL	rs143676756_T	R1457Q	VVLQQDPQQAqEHALLR						
S100A3	rs36022742_C	R3K	ARPLEQAVAAIVCTFQEYAGR	🟦	🟦	🟦	🟦	1	1
S100A3	rs36022742_T	R3K	AkPLEQAVAAIVCTFQEYAGR						

SERPINB5	rs1455555_A	I319V	GVALSNVIHK	🟦		🟩	🟦		
SERPINB5	rs1455555_G	I319V	GVALSNVvHK		🟦	🟩	🟩		
SYNGR2	rs142608913_G	A28S	FLTQPQVVAR	🟦	🟦	🟦	🟩	1	
SYNGR2	rs142608913_T	A28S	FLTQPQVVsR						
TCHH	rs2515663_A	L63R	TVDLILELLDLDSNGR						
TCHH	rs2515663_C	L63R	TVDLILELLDr	🟦	🟩	🟩	🟩		
TGM3	rs214814_G	S249N	SWNGSVEILK	🟦	🟦	🟦	🟦	1	1
TGM3	rs214814_A	S249N	nWNGSVEILK						
TRIM29	rs11604169_T	Y544C	GYPSLMR	🟦	🟦	🟩	🟦	1	
TRIM29	rs11604169_C	Y544C	GcPSLMR						
VSIG8	rs62624468_C	V47I	R.LGCPYVLDPEDYGPNGLDIEWMQVNSDPAHHR.E	🟦	🟦	🟩	🟦		
VSIG8	rs62624468_T	V47I	R.LGCPYiLDPEDYGPNGLDIEWMQVNSDPAHHR.E						

# of observations	
Total detectec	235
True positive	235
False Positive	0
True Negative	235
False Negative	16
Undetected	138

Table S5. GVPs used for CPI and CSI calculation in the six members of the family. The GVPs from P, M, S1 and S2 were validated from the corresponding genomic data. True positive identifications are highlighted in blue, true negative as white, and false negative as green. GVPs present in the protein regions that were not sequenced in the MS runs were called as undetected and highlighted as grey.

Gene Name	rs#_nucleotide	SAP	Peptide	P	M	S1	S2	S3	S4
ALDH2	rs671_G	E504K	ELGEYGLQ	🟦	🟦	🟦	🟦	1	1
ALDH2	rs671_A	E504K	ELGEYGLQ	⬜	⬜	⬜	⬜		
ATG9B	rs7804893_T	N493S	HFNELPHEL	🟦	🟦	🟦	🟦	1	
ATG9B	rs7804893_C	N493S	HFsELPHEL	⬜	⬜	⬜	⬜		
ATP5A1	rs79011243_C	A32S	VLSIGDGIA	🟦	🟦	🟦	🟦	1	1
ATP5A1	rs79011243_A	A32S	VLSIGDGIs	⬜	⬜	⬜	⬜		
CSRP1	rs3738283_T	K108I	HEEAPGHR	🟦	🟦	🟦	🟦	1	1
CSRP1	rs3738283_A	K108I	HEEAPGHR	⬜	⬜	⬜	⬜		
DSC3	rs276937_A	S78T	VLNDGSVY	▨	🟦	▨	🟦		
DSC3	rs276937_T	S78T	VLNDGtVYT	▨	⬜	▨	🟩		
DSC3	rs35296997_T	K180Q	GVDKEPLNL	▨	🟦	🟦	🟦	1	
DSC3	rs35296997_G	K180Q	GVDqEPLNL	▨	⬜	⬜	⬜		
DSP	rs80325569_G	G939S	NLHSEISGK	🟦	🟦	🟦	🟦	1	
DSP	rs80325569_A	G939S	NLHSEISsK	⬜	⬜	⬜	⬜		
DSP	rs2076299_A	Y1512C	VQYDLQK	▨	▨	▨	🟦		1
DSP	rs2076299_G	Y1512C	VQcDLQK	▨	▨	▨	⬜		
DSP	rs28763966_C	N1526K	ANSSATETI	▨	▨	▨	🟦	1	
DSP	rs28763966_A	N1526K	ANSSATETI	▨	▨	▨	⬜		
DSP	rs6929069_A	R1738Q	GqSEADSD	⬜	⬜	⬜	⬜		
DSP	rs6929069_G	R1738Q	GRSEADSD	🟦	🟦	🟦	🟦	1	1
DSP	rs28763967_C	R1537C	VQEQELTR	🟦	🟦	🟦	🟦	1	1
DSP	rs28763967_T	R1537C	VQEQELTcL	⬜	⬜	⬜	⬜		
FAM83H	rs9969600-C	Q201H	VNLQHVDF	▨	▨	⬜	▨		
FAM83H	rs9969600-A/G	Q201H	VNLhHVDFL	▨	▨	🟦	▨		
GSDMA	rs3894194_A	R18Q	QLNPqGDLT	⬜	⬜	⬜	⬜		
GSDMA	rs3894194_G	R18Q	QLNPR/GDL	🟦	🟦	🟦	🟦	1	1
GSDMA	rs7212938_G	V128L	ALETVQER	⬜	▨	▨	▨		
GSDMA	rs7212938_T	V128L	ALETIQER	🟦	▨	▨	▨		1
GSDMA	rs56030650_A	T314N	GHEVnLEAL	▨	▨	⬜	▨		
GSDMA	rs56030650_C	T314N	GHEVTLEAL	▨	▨	🟦	▨		
GSTP1	rs1138272_C	A114V	YISLIYTNYE	🟦	▨	🟦	🟦	1	1
GSTP1	rs1138272_T	A114V	YISLIYTNYE	⬜	▨	⬜	⬜		
GSTP1	rs1695_A	I105V	YISLIYTNYE	🟦	▨	🟦	🟦	1	
GSTP1	rs1695_G	I105V	YvSLIYTNYE	⬜	▨	⬜	🟦		1
HEXB	rs10805890_A	I207V	GILIDTSR	🟦	🟦	🟦	🟦	1	1
HEXB	rs10805890_G	I207V	GILvDTSR	⬜	⬜	⬜	⬜		

HEXB	rs77499935_A	I420V	LAPGTIVEV	🟦	🟦	🟦	🟦	1	1
HEXB	rs77499935_G	I420V	LAPGTvVEV						
JUP	rs41283425_C	R142H	SAIVHLINYQ	🟦	🟦	🟦	🟦	1	1
JUP	rs41283425_T	R142H	SAIVHLINYQ						
JUP	rs143043662_C	V648I	NEGTATYA	🟦	🟦	🟦	🟦	1	1
JUP	rs143043662_T	V648I	NEGTATYA						
KRT1	rs17678945_A	A454S	NKLNDLEDA		▨	▨			
KRT1	rs17678945_C	A454S	NKLNDLEDA	🟦	▨	▨	🟦	1	1
KRT32	rs72830046_C/rs	R280H/C	CQYEAMVE	🟦	🟦	🟦	🟦	1	1
KRT32	rs72830046_T	R280H	CQYEAMVE	🟦		🟦		1	
KRT32	rs2604953_G	P427T	SLLENEDCK	▨	▨	▨	🟩		
KRT32	rs3744786_C	Q72R	TYLSSSCr	🟦	▨	🟦	🟦	1	
KRT32	rs2071560_A	I171T	MVVNIDNAK	🟦	🟦	🟦	🟦	1	1
KRT32	rs2071560_G	I171T	MVVNtDNAK						
KRT32	rs146792525_C	A255T	LNIEVDAAP	🟦	🟦	🟦	🟦	1	1
KRT32	rs146792525_T	A255T	LNIEVDtAPP						
KRT33A	rs373657561_C	G33R	PCVPPSCH	🟦	🟦	🟦	🟦	1	1
KRT33A	rs373657561_T	G33R	PCVPPSCH						
KRT35	rs743686_A	S36P	VSAMYSSS		🟦			1	1
KRT35	rs743686_G	S36P	VSAMYSSS	🟦	🟦	🟦	🟦	1	1
KRT35	rs200355130-C	E141D	LVVEIDNAK	🟦	🟦	🟦	🟦	1	1
KRT35	rs138303882_G	R163W	YETEVSLRQ	🟦	🟦	🟦	🟦	1	1
KRT36	rs75790652_G	A202G	CQLGDRLN	🟦	🟦	🟦	🟦	1	1
KRT36	rs75790652_C	A202G	CQLGDRLN						
KRT36	rs11657323_T	N357T	YSSQLAQM	▨	▨	▨	🟦		
KRT36	rs11657323_G	N357T	YSSQLAQM	▨	▨	▨	🟩		
KRT36	rs9904102_G	R277C	CQYEALVE	▨	▨	🟦	▨		
KRT36	rs9904102_A	R277C	CQYEALVE	▨	▨		▨		
KRT39	rs17843021_G	T341M	DSQECILTE	🟦	🟦	🟦	🟦	1	1
KRT39	rs17843021_A	T341M	DSQECILmE	🟩					
KRT39	rs7213256_C	R456Q	SGAIESTAP	🟦	🟦	🟦	🟦	1	1
KRT39	rs7213256_T	R456Q	SGAIESTAP						
KRT39	rs17843023_G	L383M	QNQEYEILL	🟦	🟦	🟦	🟦	1	1
KRT39	rs17843023_T	L383M	QNQEYEILm						
KRT39	rs112557906_G	S423F	CEPSPWTS	🟦	🟦	🟦	🟦	1	1
KRT39	rs112557906_A	S423F	CEPSPWTfC						
KRT39	rs142154718_C	S86N	FSLDDCSW	🟦	🟦	🟦	🟦	1	1
KRT39	rs142154718_T	S86N	FSLDDCnW						
KRT40	rs2010027_C	R235H	NHEEEVNLL	🟦	🟦	🟦	🟦	1	1
KRT40	rs2010027_T	R235H	NHEEEVNLL	🟦					
KRT40	rs140634473_C	R108H	R.SLEETNA	🟦	🟦	🟦	🟦	1	1
KRT40	s140634473_T	R108H	*V h SLEETN*						
KRT7	rs6580870_A	H186R	NKYEDEINH	▨	▨		▨		
KRT7	rs6580870_G	H186R	NKYEDEINr	▨	▨	🟦	▨		

KRT82	rs2658658_A	T458M	GAFLYEPC	🟦	🟦		🟦		1
KRT82	rs2658658_G	T458M	GAFLYEPC	🟦	🟦	🟦		1	1
KRT82	rs1732263_C	E452D	GAFLYEPC	🟦	▨	▨	🟦		
KRT82	rs1732263_G	E452D	GAFLYdPC		▨	▨			
KRT82	rs1791634_C	E219Q	KYEEELSLR	🟦	🟦	🟦	🟦	1	1
KRT82	rs1791634_G	E219Q	KYEEELSLR						
KRT83	rs61485872_A	C23G	PGNFSCVS	🟦	🟦	🟦	🟦		1
KRT83	rs61485872_C	C23G	PGNFSCVS						
KRT83	rs2852464_C	I279M	DLNMDCmV	🟩			🟩		
KRT83	rs2852464_G	I279M	DLNMDCIVA	🟦	🟦	🟦	🟦	1	1
KRT83	rs2857663_G	R149C	LQFYQNR.E	🟦	🟦	🟦	🟦	1	1
KRT83	rs2857663_A	R149C	LQFYQNcE						
KRT84	RS951773_A	C446R	CEYQELMN		🟩	🟩			
KRT84	RS951773_G	C446R	QLrEYQELM	🟦	🟦	🟦	🟦	1	1
KRTAP1-5	rs148449559_G	T32S	TCCQTSFC	▨	▨	🟦	▨		
KRTAP1-5	rs148449559_**C**	T32S	TCCQTSFC	▨	▨		▨		
KRTAP1-5	rs62623375_C	C35Y	MTCCQTSF	▨	🟦	🟦	🟦	1	
KRTAP1-5	rs62623375_T	C35Y	MTCCQTSF	▨					
KRTAP3-2	rs3829598_G	R27C	SCSVPTGP	🟦	🟦	🟦	🟦		
KRTAP3-2	rs3829598_A	R27C	K.SCCCGV						
KRTAP4-1	rs398825_**C**	T134A	TTCCRPSC	▨	🟦		▨	1	1
KRTAP4-1	rs398825_**T**	T134A	aTCCRPSC	▨	🟦		▨	1	1
KRTAP4-3	rs428371_G	P152S	PACCISSCC	🟦	🟦	🟦	🟦		1
KRTAP4-3	rs428371_A	P152S	sACCISSCC						
KRTAP4-4	rs366700_C	R154S	TTCCRPSC	🟦	🟦	🟦	🟦	1	1
KRTAP4-4	rs366700_G	R154S	TTCCRPSC						
KRTAP4-4	rs75030409_T	Q109R	TTCCRPSC	🟦	🟦	🟦	🟦		1
KRTAP4-4	rs75030409_C	Q109R	TTCCRPSC						
KRTAP4-9	rs149483591_G	R26H	VSSCCGSV	▨	▨	▨	🟦		1
KRTAP4-9	rs149483591_A	R26H	VSSCCGSV	▨	▨	▨			
KRTAP4-9	rs113059833_A	D18V	VSSCCGSV	▨	▨	▨	🟦		1
KRTAP4-9	rs113059833_T	D18V	VSSCCGSV	▨	▨	▨	🟩		
KRTAP5-2	rs35925287_C	G29R	GCGSGCG	🟦	🟦	🟦	🟦		1
KRTAP5-2	rs35925287_T	G29R	GCGSGCG						
KRTAP9-4	rs2191379_A	S146Y	R.TCYYPTT	▨	▨	▨	🟦		
KRTAP9-4	rs2191379_C	S146Y	RTCYYPTTV	▨	▨	▨			
KRTAP10-12	rs61745911_G	C236Y	LASCGSLLC	🟦	🟦	▨	🟦	1	
KRTAP10-12	rs61745911_A	C236Y	LASCGSLLy			▨			
KRTAP10-12	rs34302939_G	G226S	RVPVPSCC	🟦	🟦	🟦	🟦	1	
KRTAP10-12	rs34302939_A	G226S	RVPVPSCC						
KRTAP11-1	rs71321355_C	R72Q	CIVPVAQVT	🟦	🟦	🟦	🟦	1	1
KRTAP11-1	rs71321355_T	R72Q	CIVPVAQVT						
KRTAP11-1	rs79258920_G	S78F	TSCVSNPC	🟦	🟦	🟦	🟦	1	1
KRTAP11-1	rs79258920_A	S78F	TSCVfNPCQ						

KRTAP11-1	rs9636845_A	C111S	QTTCISNPC	🟦	🟦	🟦	🟦	1	1
KRTAP11-1	rs9636845_T	C111S	QTTCISNPC		🟩	🟩			
KRTAP16-1	rs2074285_G	P340R	RCPSVCPE	🟦	🟦	🟦	🟦	1	1
KRTAP16-1	rs2074285_C	P340R	RCrSVCPEP		🟦			1	
LAMP1	rs9577230_T	I309T	FFLQGIQLN	🟦	🟦	🟦	🟦	1	1
LAMP1	rs9577230_C	I309T	FFLQGIQLN						
LGALS3	rs10148371_G	R183K	LDNNWGR	🟦	🟦	🟦	🟦	1	1
LGALS3	rs10148371_A	R183K	LDNNWGk						
LGALS3	rs11125_A	Q201H	IQVLVEPDH	🟦	🟦	🟦	🟦	1	1
LGALS3	rs11125_T	Q201H	IhVLVEPDH						
LRRC15	rs13070515_A	P286L	ELSlGIFGP		🟩				
LRRC15	rs13070515_G	P286L	ELSPGIFGP	🟦	🟦	🟦	🟦	1	
NEU2	rs2233384_C	S11R	ESVFQSGA	🟦	🟦	🟦	🟦	1	
NEU2	rs2233384_A	S11R	ASLPVLQKE						
NEU2	rs2233385_G	R41Q	IPALLYLPG	▒	▒	🟦	▒	1	1
NEU2	rs2233385_A	R41Q	IPALLYLPG	▒	▒		▒		
NEU2	rs2233390_G	A145T	DLTDAAIGP	🟦	🟦	🟦	🟦	1	1
NEU2	rs2233390_A	A145T	DLTDtAIGPA						
PKP1	rs61818256_C	R684W	AAEAARLLL	🟦	🟦	🟦	🟦	1	1
PKP1	rs61818256_T	R684W	AAEAAwLLL						
PLCD1	rs933135_C	R257H	EEAAGPALA	▒	▒	🟦	▒		
PLCD1	rs933135_T	R257H	EEAAGPALA	▒	▒		▒		
PPL	rs2037912_C	Q1573E	eNLQLETR	▒	▒	🟩	▒		
PPL	rs2037912_G	Q1573E	QNLQLETR	▒	▒	🟦	▒		
PPL	rs143676756_C	R1457Q	VVLQQDPQ	▒	▒	▒	🟦		
PPL	rs143676756_T	R1457Q	VVLQQDPQ	▒	▒	▒			
S100A3	rs36022742_C	R3K	ARPLEQAV	🟦	🟦	🟦	🟦	1	1
S100A3	rs36022742_T	R3K	AkPLEQAVA						
SERPINB5	rs1455555_A	I319V	GVALSNVIH	🟦		▒	🟦		
SERPINB5	rs1455555_G	I319V	GVALSNVvH		🟦	▒	🟩	1	
SYNGR2	rs142608913_G	A28S	FLTQPQVV	🟦	🟦	🟦	▒	1	
SYNGR2	rs142608913_T	A28S	FLTQPQVVs				▒		
TCHH	rs2515663_A	L63R	TVDLILELLD		▒	▒	▒		
TCHH	rs2515663_C	L63R	TVDLILELLD	🟦	▒	▒	▒		
TGM3	rs214814_G	S249N	SWNGSVEIL	🟦	🟦	🟦	🟦	1	1
TGM3	rs214814_A	S249N	nWNGSVEIL						
TRIM29	rs11604169_T	Y544C	GYPSLMR	🟦	🟦	▒	🟦	1	
TRIM29	rs11604169_C	Y544C	GcPSLMR			▒			
VSIG8	rs62624468_C	V47I	R.LGCPYVL	🟦	🟦	▒	🟦		
VSIG8	rs62624468_T	V47I	R.LGCPYiL			▒			

Table S6: Genotypes of individual P, M, S1 and S2 for the identified genetically variant peptides. Genotypes at the five dubious loci are highlighted in bold italic.

Serial No.	**Gene Name**	**rs#**	**Reference**	**P**	**M**	**S1**	**S2**
1	**ALDH2**	rs671	G	GG	GG	GG	GG
2	**ATG9B**	rs7804893	T	TT	TT	TT	TT
3	**ATP5A1**	rs79011243	C	CC	CC	CC	CC
4	**CSRP1**	rs3738283	T	TT	TT	TT	TT
5	**DSC3**	rs276937	A	TT	AA	AT	AT
6	**DSC3**	rs35296997	T	TT	TT	TT	TT
7	**DSP**	rs2076299	A	AG	AA	AA	AA
8	**DSP**	rs28763966	C	CC	CC	CC	CC
9	**DSP**	rs28763967	C	CC	CC	CC	CC
10	**DSP**	rs80325569	G	GG	GG	GG	GG
11	**DSP**	rs6929069	G	GG	GG	GG	GG
12	**FAM83H**	rs9969600	C	GG	GG	GG	GG
13	**GSDMA**	rs56030650	C	CA	CC	CC	CC
14	**GSDMA**	rs3894194	G	GG	GG	GG	GG
15	**GSDMA**	rs7212938	G	TT	TT	TT	TT
16	**GSTP1**	rs1695	A	AA	AG	AA	AG
17	**GSTP1**	rs1138272	C	CC	CC	CC	CC
18	**HEXB**	rs10805890	A	AA	AA	AA	AA
19	**HEXB**	rs77499935	A	AA	AA	AA	AA
20	**JUP**	rs41283425	C	CC	CC	CC	CC
21	**JUP**	rs143043662	C	CC	CC	CC	CC
22	**KRT1**	rs17678945	C	CC	CC	CC	CC
23	**KRT31**	rs6503627	G	GG	AG	GG	GG
24	**KRT32**	rs2071560	A	AA	AA	AA	AA
25	**KRT32**	rs72830046	C	CT	CC	CT	CC
26	**KRT32**	rs146792525	C	CC	CC	CC	CC
27	**KRT32**	rs2071561	G	GT	GG	GT	GT
28	**KRT32**	rs3744786	T	TC	TT	TC	TC
29	**KRT32**	rs2604953	G	***TT***	***TT***	***TT***	***GG***
30	**KRT33A**	rs373657561	C	CC	CC	CC	CC
31	**KRT33A**	rs12937519	G	GG	GA	GG	GG
32	**KRT34**	rs2071599	T	TT	TC	TT	TT
33	**KRT35**	rs743686	A	GG	AG	GG	GG
34	**KRT35**	rs200355130	C	CC	CC	CC	CC
35	**KRT35**	rs138303882	G	GG	GG	GG	GG
36	**KRT36**	rs75790652	G	GG	GG	GG	GG
37	**KRT36**	rs9904102	G	GG	GG	GG	GG
38	**KRT36**	rs11657323	T	TG	TG	TG	TG
39	**KRT39**	rs7213256	C	CC	CC	CC	CC
40	**KRT39**	rs142154718	C	CC	CC	CC	CC
41	**KRT39**	rs17843021	G	GA	GG	GG	GG
42	**KRT39**	rs17843023	G	GG	GG	GG	GG

43	**KRT39**	rs112557906	G	GG	GG	GG	GG
44	**KRT40**	rs2010027	C	CT	CC	CC	CC
45	**KRT40**	rs140634473	C	CC	CC	CC	CC
46	**KRT7**	rs6580870	A	AG	GG	GG	AG
47	**KRT81**	rs6580873	A	AC	AC	AA	CC
48	**KRT82**	rs1732263	C	CC	CC	CC	CC
49	**KRT82**	rs1791634	C	CC	CC	CC	CC
50	**KRT82**	rs2658658	G	GA	GA	GG	AA
51	**KRT83**	rs61485872	A	AA	AA	AA	AA
52	**KRT83**	rs2852464	G	GC	GG	GG	GC
53	**KRT83**	rs2857663	G	GG	GG	GG	GG
54	**KRT84**	RS951773	A	GG	AG	AG	GG
55	**KRT85**	rs2852471	C	CC	CC	CC	CC
56	**KRTAP10-12**	rs61745911	G	GG	GG	GG	GG
57	**KRTAP10-12**	rs34302939	G	GG	GG	GG	GG
58	**KRTAP10-6**	rs465279	G	***GG***	***AA***	***GA***	***GG***
59	**KRTAP11-1**	rs9636845	A	AA	AT	AT	AA
60	**KRTAP11-1**	rs71321355	C	CC	CC	CC	CC
61	**KRTAP11-1**	rs79258920	G	GG	GG	GG	GG
62	**KRTAP1-5**	rs62623375	C	CC	CC	CC	CC
63	**KRTAP1-5**	rs148449559	G	GG	GG	GG	GG
64	**KRTAP16-1**	rs2074285	G	GG	GC	GG	GG
65	**KRTAP3-2**	rs3813050	A	AA	AA	AA	AA
66	**KRTAP3-2**	rs3829598	G	GG	GG	GG	GG
67	**KRTAP3-2**	rs9897046	T	TT	TT	TT	TT
68	**KRTAP4-1**	rs398825	C	***TT***	***CT***	***TT***	***CC***
69	**KRTAP4-3**	rs428371	G	GG	GG	GG	GG
70	**KRTAP4-4**	rs366700	C	CC	CC	CC	CC
71	**KRTAP4-4**	rs385055	T	TT	TT	TT	TT
72	**KRTAP4-4**	rs75030409	T	TT	TT	TT	TT
73	**KRTAP4-6**	rs73983172	G	GG	GG	GG	GG
74	**KRTAP4-9**	rs149483591	G	GG	GG	GG	GG
75	**KRTAP4-9**	rs113059833	A	***AA***	***AA***	***AA***	***AT***
76	**KRTAP5-2**	rs35925287	C	CC	CC	CC	CC
77	**KRTAP9-2**	rs9902235	**G**	***GG***	***GG***	***GG***	***CC***
78	**KRTAP9-4**	rs2191379	C	AA	CA	AA	AA
79	**LAMP1**	rs9577230	T	TT	TT	TT	TT
80	**LGALS3**	rs11125	A	AA	AA	AA	AA
81	**LGALS3**	rs10148371	G	GG	GG	GG	GG
82	**LRRC15**	rs13070515	G	GG	GA	GG	GG
83	**NEU2**	rs2233384	C	CC	CC	CC	CC
84	**NEU2**	rs2233385	G	GG	GG	GG	GG
85	**NEU2**	rs2233390	G	GG	GG	GG	GG

86	**PKP1**	rs61818256	C	CC	CC	CC	CC
87	**PLCD1**	rs933135	C	CC	CC	CC	CC
88	**PPL**	rs143676756	C	CC	CC	CC	CC
89	**PPL**	rs2037912	G	GC	GC	GC	CC
90	**S100A3**	rs36022742	C	CC	CC	CC	CC
91	**SERPINB5**	rs1455555	A	AA	GG	AG	AG
92	**SYNGR2**	rs142608913	G	GG	GG	GG	GG
93	**TCHH**	rs2515663	A	CC	CC	CC	CC
94	**TGM3**	rs214814	G	GG	GG	GG	GG
95	**TRIM29**	rs11604169	T	TT	TT	TT	TT
96	**VSIG8**	rs62624468	C	CC	CC	CC	CC

Table S7: Combined paternity indexes and posterior probabilities calculated using all the detected GVPs including false positives. The posterior probabilities were obtained using the prior odds of 4/12 for the four true offspring of the couple P and M and eight random individuals. CPI: combined paternity index.

CPI when both parents are available				
Individual	Loci used in the calculation	Loci with no obligate allele	CPI	Posterior Probability (%)
S3	35	0	1696.75	99.94
S1	36	0	1506.12	99.93
S4	33	0	340.70	99.70
S2	39	1	1239.76	99.91
A	34	3	--	--
B	34	4	--	--
C	31	6	--	--
D	31	4	--	--
F	33	4	--	--
G	30	2	--	--
H	30	2	--	--
I	35	3	--	--

Table S8: Combined sibship index values calculated using all the detected GVPs including false positives for the family members and eight unrelated individuals. The CSI values higher than 10 for unrelated individuals or lower than 10 for true siblings are shown in bold, and the ones that support the relationship in the cases of true relationships are bold italicized.

CSI	**P**	**S3**	**S1**	**S4**	**S2**	**A**	**B**	**C**	**D**	**F**	**G**	**H**	**IYO**
M	3.21	***198.13***	**5.71**	***1383.55***	**6.12**	0.51	7.32	2.18	9.54	0.04	14.07	0.05	1.58
P		***6.16***	***1648.28***	***10.83***	***394.09***	1.76	0.06	0.02	0.05	7.58	0.05	3.01	0.24
S3			***330.91***	***67.33***	***14.13***	1.45	1.00	0.56	4.40	0.15	0.71	0.04	1.70
S1				***8.80***	***171.49***	0.07	0.02	0.03	0.04	0.59	0.01	0.50	1.53
S4					***60.96***	0.66	1.07	0.10	3.65	0.04	1.86	0.09	2.45
S2						0.11	0.01	0.00	0.51	2.83	0.45	**10.65**	0.13
A							1.36	**388.00**	1.77	0.08	1.35	2.99	3.46
B								**10.74**	0.93	0.00	0.34	0.01	0.01
C									**15.22**	0.01	9.15	0.59	0.14
D										0.00	5.67	0.08	**10.30**
F											0.15	2.40	0.01
GH												0.12	3.05
H													0.04
IYO													

Table S9: Number of loci at which each comparison was based for CSI calculations. The numbers inside the parentheses represent the genes with two loci included.

CSI	**P**	**S3**	**S1**	**S4**	**S2**	**A**	**B**	**C**	**D**	**F**	**G**	**H**	**I**
M	30(9)	29(9)	28(8)	24(8)	31(9)	25(9)	24(9)	22(8)	21(9)	24(8)	22(7)	21(8)	25(8)
P		27(9)	27(9)	24(8)	31(11)	26(11)	25(11)	22(10)	21(10)	24(10)	22(9)	21(10)	26(10)
S3			27(10)	24(8)	28(9)	24(10)	23(10)	21(8)	21(9)	23(10)	22(9)	21(10)	24(11)
S1				24(8)	30(9)	25(9)	25(10)	22(8)	22(9)	25(9)	25(9)	23(9)	25(9)
S4					25(8)	24(9)	22(8)	19(7)	20(8)	22(7)	22(7)	19(7)	23(8)
S2						27(12)	26(11)	27(10)	22(11)	27(12)	26(11)	24(12)	27(12)
A							29(12)	31(12)	23(9)	25(13)	28(12)	26(13)	26(12)
B								27(13)	26(10)	26(13)	24(12)	25(13)	25(12)
C									25(9)	27(11)	30(10)	30(11)	27(11)
D										22(11)	23(9)	23(11)	24(11)
F											26(13)	27(15)	25(13)
G												29(14))	25(12)
H													25(13)
I													

Table S2. GVP data matrix used for hierarchical clustering. Each GVP detection was assigned a value of 1 and a non-detection of 0.

Individuals	rs2229528_T	rs2229528_C	rs671_G	rs671_A	rs17845226_A	rs7804893_T	rs7804893_C	rs79011243_C	rs79011243_A	rs3738283_T	rs3738283_A	rs276937_A	rs276937_T	rs35296997_T	rs35296997_G	rs80325569_G	rs80325569_A	rs2076299_A	rs2076299_G	rs28763966_C	rs28763966_A	rs28763967_C	rs28763967_T	rs6929069_A	rs6929069_G	rs9969600-C	rs9969600-A/G	rs11550699_A	rs11550699_G	rs17856024_C	rs17856024_A	rs35363287_C	rs35363287_T	rs3894194_A	rs3894194_G	rs7212938_G	rs7212938_T	rs56030650_A
S3	0	0	1	0	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
S4	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0
S1	0	0	1	0	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0
P	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0
M	0	0	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
S2	0	0	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
A	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0	1	1	0	0	0	0	0	0	1	0
B	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0
C	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0
D	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
G	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
H	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
I	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0

rs56030650_C	rs1138272_C	rs1138272_T	rs1695_A	rs1695_G	rs10805890_A	rs10805890_G	rs77499935_A	rs77499935_G	rs6761276_C	rs6761276_T	rs41283425_C	rs41283425_T	rs143043662_C	rs143043662_T	rs17678945_A	rs17678945_C	rs6503627_A	rs2071561_T	rs73983451_G	rs72830046_T	rs2604953_G	rs2604953_T	rs3744786_T	rs3744786_C	rs2071560_A	rs2071560_G	rs146792525_C	rs146792525_T	rs373657561_C	rs373657561_T	rs12937519_A	rs2239710_A	rs2071599_T	rs61740668_G	rs61740668_A	rs743686_A	rs743686_G	rs200355130-C
0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	0	1	1	1	1	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	0	1	1	1
0	1	0	0	1	1	0	1	0	0	0	1	0	1	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	1	1	1
1	1	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	1	1	1	0	0	0	1	1	0	1	0	1	0	0	0	1	0	0	0	1	1
0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	0	1	0	1	1	1	0	0	0	1	1	0	1	0	1	0	0	0	1	0	0	0	1	1
0	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	1	1	1
0	1	0	1	1	1	0	1	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1	0	1	1	0	1	0	1	0	0	0	1	0	0	0	1	1
0	1	0	1	1	0	1	1	0	0	0	1	0	1	0	0	0	0	1	1	1	0	1	0	0	1	0	1	0	0	0	0	1	1	0	0	1	1	0
0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	1	1	0	0	1	0	0	1	0	1	0	0	0	0	1	1	1	1	1	0	0
0	1	0	1	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	1	1	1	0	1	0	0
0	1	0	0	1	1	1	1	0	0	0	1	0	1	0	0	0	0	1	1	0	0	1	0	0	1	0	1	0	0	0	0	1	1	1	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0
0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	1	1	1	0	1	0	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0
0	1	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0	0	1	0	0	1	1	0
0	1	0	1	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1	1	1	0	1	0	0	1	0	1	0	0	0	0	1	1	1	0	0	1	0
0	1	0	0	1	1	1	1	0	0	0	1	0	1	0	0	0	0	1	1	1	0	1	0	0	1	0	1	0	0	0	0	1	1	1	0	1	1	0

rs200355130-A	rs138303882_A	rs138303882_G	rs12451652_C	rs12451652_T	rs2071601_C	rs2071601_G	rs75790652_G	rs75790652_C	rs11657323_T	rs11657323_G	rs9904102_G	rs9904102_A	rs9916484_T	rs9916484_C	rs9916475_T	rs9916475_A	rs16966811_A	rs897416_A	rs897416_G	rs17843021_G	rs17843021_A	rs7213256_C	rs7213256_T	rs17843023_G	rs17843023_T	rs12557906_G	rs12557906_A	rs142154718_C	rs142154718_T	rs2010027_C	rs2010027_T	rs140634473_C	s140634473_T	rs721957_C	rs721957_T	rs6580870_A	rs6580870_G	rs2232393_A
1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0
1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0	0	1	0
1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	1	1	0	0	0	0	0	0
1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0
1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0
0	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0
0	0	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1	0	1	1	0	0	1	0	1	1	1	0	1	0	0	0	0
0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	1	0	1	1	1	0	0	0	0	0	1
0	0	1	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	1	1	0	1	1	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	1	0	1	1	1	0	0	1	0	0	0
0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	1	1	0	1	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0	0

rs2232393_G	rs6580873_A	rs2658658_A	rs2658658_G	rs1732263_C	rs1732263_G	rs1791634_C	rs1791634_G	rs61485872_A	rs61485872_C	rs2852464_C	rs2852464_G	rs2857671_G	rs2857671_A	rs2857663_G	rs2857663_A	RS951773_A	RS951773_G	rs61630004_C	rs61630004_T	rs2852471-C	rs2852471-A	rs58717266_C	rs150218495_G	rs150218495_C	rs138200823_C	rs148449559_G	rs148449559_C	rs62623375_C	rs62623375_T	rs138758776_T	rs138758776_C	rs9897046_T	rs3829598_G	rs3829598_A	rs3813050_A	rs398825_C	rs398825_T	rs62067292_G
0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0
0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0
0	1	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	1	0	1	0	0	0	1	1	0	1	0	0	0
0	1	1	1	1	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0
0	1	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	1	0
0	1	1	0	1	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0
0	0	1	1	1	0	1	0	0	0	0	1	1	0	1	0	0	1	0	1	0	0	0	1	0	1	1	0	1	0	1	0	1	1	0	0	0	0	1
0	0	1	1	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1	0	1	0	1	0	1	1	0	0	0	0	1
0	0	0	1	1	0	1	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	1	0	1	1	0	0	0	0	1
0	0	1	1	0	0	1	0	0	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0	0	0
0	0	1	0	1	0	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	1	1	0	1	1	0	1	0	1	0	1	1	0	1	0	0	0
0	0	1	1	1	0	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	1	1	0	1	1	0	1	0	1	0	1	1	0	1	0	0	0
0	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	1	1	0	1	1	0	1	0	1	0	1	1	0	0	0	0	0
0	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

rs62067292_C	rs389784_T	rs389784_C	rs428371_G	rs428371_A	rs366700_C	rs366700_G	rs385055_T	rs444509_A	rs444509_T	rs75030409_T	rs75030409_C	rs1497383_A	rs73983172_G	rs201814486_G	rs201814486_C	rs138296121_C	rs138296121_T	rs149483591_G	rs149483591_A	rs113059833_A	rs113059833_T	rs9897031_C	rs9897031_T	rs113376601_C	rs35925287_C	rs35925287_T	rs9902235_C	rs9902235_G	rs2191379_A	rs2191379_C	rs12938692_A	rs12938692_G	rs576405629_A	rs576405629_G	rs537301040_T	rs537301040_C	rs12938374_G	rs12938374_A
0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	1
0	0	0	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	1	0	1	1	0	1	0	1	1	1	0	1	0	0	0	1	0	1	0	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	1	0	1	0	0	1
0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0

rs233252_C	rs233252_T	rs464391_C	rs464391_G	rs465279_A	rs111668637_G	rs111668637_A	rs411254_A	rs411254_G	rs9980129_T	rs9980129_C	rs4818950_G	rs4818950_A	rs61745911_G	rs61745911_A	rs34302939_G	rs34302939_A	rs71321355_C	rs71321355_T	rs79258920_G	rs79258920_A	rs9636845_A	rs9636845_T	rs3804010_G	rs3804010_C	rs2074285_G	rs2074285_C	rs9577230_T	rs9577230_C	rs10148371_G	rs10148371_A	rs11125_A	rs11125_T	rs13060627_C	rs13060627_T	rs13070515_A	rs13070515_G	rs2233384_C	rs2233384_A
0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	1	1	1	0	1	0	1	0	0	0	0	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1	0	0	0	0	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1	0	0	0	0	1	1	0
0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	1	1	1	0	1	0	1	0	0	0	0	1	1	0
0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	1	0	0	0	0	1	1	0
0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	1	0	0	0	1	1	1	0	1	0	1	0	0	0	0	1	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	0	0	1	1	0	1	0	1	0	0	0	0	1	1	0
0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	1	0	1	0	0	0	0	0	1	0	1	1	0	0	1	0	1	0	0	0	0	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	0	1	1	0	0	1	0	1	0	1	0	0	1	1	0
0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1	0	1	0	1	1	0	1	1	0
0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	1	0
1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	0	1	1	1	0	1	0	1	1	0	0	0	1	1	0

rs2233385_G	rs2233385_A	rs2233390_G	rs2233390_A	rs2233391_A	rs2233391_C	rs412750_G	rs412750_A	rs61818256_C	rs61818256_T	rs10920171_C	rs10920171_T	rs6753929_G	rs6753929_C	rs933135_C	rs933135_T	rs2037912_C	rs2037912_G	rs143676756_C	rs143676756_T	rs116208483_G	rs36022742_C	rs36022742_T	rs41265164_G	rs41265164_T/C	rs1455555_A	rs1455555_G	rs142608913_G	rs142608913_T	rs2515663_A	rs2515663_C	rs214803_A	rs214803_C	rs214814_G	rs214814_A	rs11604169_T	rs11604169_C	rs62624468_C	rs62624468_T
1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0
1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0
0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	1	0
0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1	0	1	0
0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0
1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
0	0	1	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0
0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0
1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0