Title

Optimal processing for proteomic genotyping of single human hairs

Authors

Goecker, Zachary C
Salemi, Michelle R
Karim, Noreen
et al.

Publication Date

2020-07-01

DOI

10.1016/j.fsigen.2020.102314

Peer reviewed

incl Acknowledgements

Title:
Optimal Processing for Proteomic Genotyping of Single Human Hairs

Authors:
Zachary C. Goecker[1], Michelle R. Salemi[2], Noreen Karim[1], Brett S. Phinney[2], Robert H. Rice[1], Glendon J. Parker[1,*]

[1] Department of Environmental Toxicology, University of California, Davis, CA, USA
[2] Proteomics Core Facility, University of California, Davis, CA, USA

[*] Corresponding author.
Glendon Parker PhD
Department of Environmental Toxicology
University of California – Davis
One Shields Ave. Davis, California 95616
p. (530) 752 9870
e. gjparker@ucdavis.edu

Funding

This project was supported by Award No. 2015-DN-BX-K065, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

Disclaimer

The authors have declared no conflict of interest, with the exception of GJP who has a patent based on the 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). 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.

Acknowledgments

The authors thank Dr. Blythe Durbin-Johnson from UC Davis and Dr. Susan Walsh of Indiana University – Purdue University Institute, for their advice. Sorenson forensics also played a crucial role for their assistance in tissue procurement. This publication was made possible, in part, with support from the UC Davis Genome Center Bioinformatics Core Facility. The sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. We specifically acknowledge the assistance of Jie Li, Emily Kumimoto, Siranoosh Ashtari, Vanessa K Rashbrook, and Lutz Froenicke.

Title:

Optimal Processing for Proteomic Genotyping of Single Human Hairs

Authors:

Highlights:

- Development of an optimized proteomic processing method to maximize yield of genetically variant peptides
- Genetically variant peptide analysis of single scalp hairs
- Discovery of genetically variant peptides

Abstract:

The use of hair evidence for human identification is undergoing considerable improvement through the adoption of proteomic genotyping. Unlike traditional microscopic comparisons, protein sequencing provides quantitative and empirically based estimates for random match probability. Non-synonymous SNPs are translated as single amino acid polymorphisms and result in genetically variant peptides. Using high resolution mass spectrometry, these peptides can be detected in hair shaft proteins and used to infer the genotypes of corresponding SNP alleles. We describe experiments to optimize the proteomic genotyping approach to individual identification from a single human scalp hair 2 cm in length (~100 µg). This is a necessary step to develop a protocol that will be useful to forensic investigators. To increase peptide yield from hair, and to maximize genetically variant peptide and ancestral information, we examined the conditions for reduction, alkylation, and protein digestion that specifically address the

distinctive chemistry of the hair shaft. Results indicate that optimal conditions for proteomic analysis of a single human hair include 6 hrs of reduction with 100 mM dithiothreitol at room temperature, alkylation with 200 mM iodoacetamide for 45 min, and 6 hrs of digestion with two 1:50 (enzyme:protein) additions of stabilized trypsin at room temperature, with stirring incorporated into all three steps. Our final conditions using optimized temperatures and incubation times increased the average number of genetically variant peptides from 20 ± 5 to 73 ± 5 ($p = 1 \times 10^{-13}$), excluding intractable hair samples. Random match probabilities reached up to 1 in 620 million from a single hair with a median value of 1 in 1.1 million, compared to a maximum random match probability of 1 in 1380 and a median value of 1 in 24 for the original hair protein extraction method. Ancestral information was also present in the data. While the number of genetically variant peptides detected were equivalent for both European and African subjects, the estimated random match probabilities for inferred genotypes of European subjects were considerably smaller in African reference populations and *vice versa*, resulting in a difference in likelihood ratios of 6.8 orders of magnitude. This research will assure uniformity in results across different biogeographic backgrounds and enhance the use of novel peptide analysis in forensic science by helping to optimize genetically variant peptide yields and discovery. This work also introduces two algorithms, GVP Finder and GVP Scout, which facilitate searches, calculate random match probabilities, and aid in discovery of genetically variant peptides.

Abbreviations: ABC, ammonium bicarbonate; DTE, dithioerythritol; DTT, dithiothreitol; GVP, genetically variant peptide; IA, iodoacetamide; KAP, keratin-associated protein; MAF, minor allele frequency; RMP, random match probability; RMT, reductively methylated trypsin; SD, sodium dodecanoate; SNP, single nucleotide polymorphism; TFA, trifluoroacetic acid.

Keywords: Proteomic Genotyping; Hair Shafts; Hair Chemistry; Genetically Variant Peptides; Proteomics; Human Identification

1. Introduction

Hair is a ubiquitous biological material that is shed from the human body at a rate of about 100 to 150 scalp hair shafts per day [1]. Because hair is a complex biological material, it contains information that can potentially be exploited to provide a link between an individual and a location [2-4]. Forensic hair analysis for identification of individuals, ancestry and species has historically been conducted using morphologic hair comparison, which is now considered controversial [5-12]. Hair shaft protein was recently demonstrated to be a carrier of genetic information in the form of genetically variant peptides (GVPs) [13]. These peptides contain single amino acid polymorphisms, the result of non-synonymous SNPs. Detection of these peptides allows for the inference of the corresponding SNP genotypes [13]. Like any DNA genotype, these can be used to estimate random match probability (RMP) and to statistically associate an individual with a given hair shaft [13]. However, in order to be useful to the forensic science community, several technical issues must be addressed. Primary among these is the need to obtain forensically relevant RMPs from a single human hair [2, 13-16].

Hair is a challenging substrate. The bulk of hair consists of highly structured keratin intermediate filaments that are stabilized by a range of covalent bonds that result in a physically robust and chemically resistant tissue [17-19]. These covalent bonds consist of isopeptide bonds, the result of transglutaminase reactions, and particularly high levels of disulfide bonds. Keratins and particularly keratin-associated proteins (KAPs) are cysteine-rich, resulting in a highly cross-linked tissue matrix [18, 20]. Hair remains an underutilized forensic substrate that contains important

biological information from mitochondrial and fragmented nuclear DNA, proteins, and other, small molecules. Any protocol development would need to balance the chemical fragility of the target molecule against the conditions required to thoroughly decontaminate the hair surface or open up the hair matrix for proteolytic release of internal biomolecules. An ideal processing protocol would efficiently and consistently release informative molecules from the matrix with minimal introduction of analytical biases, regardless of hair biology or human behavior. The starting point for any such protocol should be based on the biochemical and biophysical nature of the hair shaft.

This project is a systematic evaluation of chemical treatments of hair shafts from the scalp to maximize the proteomic yield of GVPs using subjects of European or African ancestry. Present work reaches a counter-intuitive finding that milder conditions result in maximal detection and identification of target GVPs. A significant increase in the amount of DTT reductant, up to 100 mM, maintains the gentle conditions while also opening up the keratin matrix to increase the release of peptides from keratin-associated and other proteins. These optimizations, when applied to single hairs, increase proteolytic release of KAPs and detection of GVPs. A single 2 cm hair shaft resulted in detection of up to 80 GVPs with an RMP of up to 1 in 620 million, a three-fold increase of GVP detection and an average increase in RMP of 4 orders of magnitude compared to earlier findings. Tools have also been developed to more efficiently identify and discover GVPs in proteomic data and are hereby made available to the forensic community.

2. Materials and Methods

2.1 Hair Collection and Preparation

Reference hair and matching DNA were collected from 3 self-described African subjects (XXXXXX) and 3 self-described European subjects (XXXXXX) using IRB compliant protocols (IRB# XXXXXX).

Only two biogeographic groups were studied in this work to demonstrate a proof of concept of the work. These two groups were chosen to represent the two largest demographic groups in the United States. The average length of hair on the head before cutting was 10 cm. Hair roots were purposefully excluded from the processing. Hairs were collected by cutting a few inches inward from the distal end. Hair shafts were either weighed to give 4 mg of hair per subject per replicate, or cut to 2 cm in length with no regard to distal or proximal orientation.

All hair shafts were washed three times in 1 mL of 2% (w/v) sodium dodecanoate (SD) (Sigma-Aldrich, St. Louis, MO) in 50 mM ammonium bicarbonate (ABC) (Honeywell, Muskegon, MI) to minimize contamination from exogenous materials, such as environmental epidermal corneocytes. Samples were vortexed for 10 sec with each wash, and the wash eluent was discarded. For single hair analysis, a 2 cm length was cut into 10 separate 2 mm segments and placed in a protein LoBind tube (Eppendorf, Hamburg, Germany) with the entire hair shaft submerged in solution. Hair samples of 4 mg were left intact and not cut into segments. All reagent solutions were passed through solid-phase extraction filtration with the exception of the reductively methylated trypsin (RMT) [21] and SD, as these would bind to the stationary phase of the cartridge. This step was applied to minimize contamination by exogenous organic material.

2.2 Chemical Processing Optimization

The starting chemistry for proteomic processing of human hair was obtained from an NCJRS report [22] and related publications [23, 24]. This method, referred to as the original processing method, employed overnight incubation at a high temperature for disulfide reduction and 3 days of digestion. In this method, 400 µL of a solution of 2% SD + 50 mM ABC and 50 mM dithioerythritol (DTE) (Sigma-Aldrich, St Louis, MO) was added to the LoBind tube with 4 mg of prepared hair. A cleaned magnetic stir flea (Sigma-Aldrich) was added to the tube and stirred at medium speed for 1 hr at room temperature

before incubation in an oven with no agitation at 70°C for 18 hrs. Samples were again stirred at medium speed at room temperature for 1 hr. Free thiols were alkylated with the addition of iodoacetamide (IA) (Sigma-Aldrich) to give a final concentration of 100 mM. The hair-containing solution was stirred in the dark for 45 min. The sample was then acidified (pH ~2) with 8 µL of trifluoroacetic acid (TFA) (ThermoFisher, Chicago, IL) to precipitate the detergent. Detergent extraction was achieved using three consecutive additions of 700 µL ethyl acetate (Sigma-Aldrich). For each extraction, the sample was vortexed and then centrifuged for 3 min at 14000 relative centrifugal force (rcf). The organic (upper) phase was removed by pipetting with care not to disturb the interphase containing denatured protein and/or fragmented hair. The pH was then adjusted to ~8 using 2.5 µL of ammonium hydroxide (Fisher Scientific) and 25 µL of 1 M ABC. Three 1:50 (enzyme:protein) additions of RMT were added to the sample, with one addition per day for three days [21]. Digests were then centrifuged at 14000 rcf for 15 min, and the supernatant was collected for mass spectral analysis. The only modifications to this protocol were made during the final optimization comparison, where the volume of reagents was reduced by 75%, for a final volume of ~160 µL, and 2 cm of a hair shaft was used instead of 4 mg.

The resulting chemistry for proteomic processing of human hair, referred to as the optimized processing method, employs a 14 hr protocol. In this method, 100 µL of a solution of 2% SD + 50 mM ABC + 100 mM dithiothreitol (DTT) (Invitrogen, Carlsbad, CA) was added to each LoBind tube with 2 cm of prepared and cut hair. A cleaned magnetic stir flea was added to the tube and stirred at medium speed for 6 hrs at room temperature. Free thiols were then alkylated with the addition of IA to a final concentration of 200 mM, and the solution was stirred in the dark for 45 min. The sample was then acidified to a pH of ~2 using 2 µL of TFA to precipitate the detergent. Detergent extraction was achieved using three consecutive additions of 175 µL of EtOAc. For each extraction, the sample was vortexed and then centrifuged for 3 min at 14000 rcf to minimize the interphase containing denatured protein and/or fragmented hair before pipetting off the upper organic phase. The pH was then adjusted to ~8 using 6.3

µL of 1 M ABC and 0.6 µL of ammonium hydroxide. Two 1:50 (enzyme:protein) additions of RMT were added to the sample, with one addition every 3 hrs, for a total digestion time of 6 hrs. Digests were then centrifuged at 14000 rcf for 15 min, and the supernatant was collected for mass spectral analysis.

2.3 Peptide Quantification

Digestion efficiency was quantified by reaction of insoluble protein with ninhydrin after hydrolysis with 10% sulfuric acid [25, 26]. Samples were analyzed based on A570 and compared to a standard curve of hydrolyzed bovine serum albumin. The percentage (w/w) of hair that was in the insoluble fraction was then calculated using the mass of the insoluble pellet divided by the total hair mass, which was usually 4 mg for initial experiments. Before instrumental analysis, solubilized tryptic peptides were quantified using the Pierce™ Quantitative Fluorometric Peptide Assay (ThermoFisher) after 1:10 dilution. Fluorescence was measured using a Synergy H1 hybrid multi-mode reader (BioTek, Winooski, VT).

2.4 Data Acquisition

Samples were analyzed using a ThermoScientific Q-Exactive Plus Orbitrap mass spectrometer with built in Proxeon nanospray and Proxeon Easy-nLC II HPLC. A sample (10 µL) containing 0.75 µg of digested peptide material was loaded on a 100 µm × 25 mm Magic C18 100 Å 5 U reverse phase trap, desalted online and separated over 140 min gradient using a 75 µm × 150 mm Magic C18 200 Å 3 U reverse phase column at 300 nL/min flow rate [27]. The solvent gradient for the elution of peptides began with 5% acetonitrile (ACN) and increased linearly to 20% ACN at 92 minutes, 32% ACN at 112 minutes, and 80% ACN at 119 minutes. The 80% ACN solvent ratio was maintained for 10 minutes, reduced to 5% at 130 minutes, and held for 10 minutes. MS survey was conducted at the m/z range of 350-1600, and the 15 most abundant ions from the spectra were subjected to higher-energy C-trap

dissociation (HCD) to fragment the precursor peptides and obtain MS/MS spectra [28]. Precursor ions selected in a 1.6 m/z isolation mass window were fragmented via 27% normalized collision energy. A 20 s duration was used for dynamic exclusion.

2.5 Data Analysis

Raw data files were converted into mzML format using MsConvert GUI software (Proteowizard 2.1, http://proteowizard.sourceforge.net). Files were converted using numpress linear compression and numpress short logged float compression along with peak picking with vendor algorithm for all mass spectrometry levels. These mzML files were then analyzed using GPM Fury software (X!Tandem Alanine (2016.10.15.2)) using the advanced search option. Default search settings were chosen except for exclusion of prokaryotes and viruses in the taxon heading, peptide and protein log(e) score minimum of -1 and -1 respectively, fragment mass error of 20 ppm, parent mass error of ± 100 ppm, and inclusion of point mutations under the refinement specification heading [27]. Post-search filtering based on specific transition levels was manually applied to GVP spectra to account for broad mass error filtering. The output from X!Tandem in the Global Proteome Machine environment included the annotation of single amino acid variants, that were genetic or chemical in origin. These annotations form the basis of subsequent analyses of GVP discovery, detection and post-translational modifications.

A spreadsheet, termed GVP Finder (v1.1), was created to search for GVPs and calculate random match probabilities (RMPs). This spreadsheet can be obtained from the resources menu of XXXXXX. In short, previously identified GVPs were searched for by exporting each sample peptide spreadsheet in the GPM Fury software and then were bioinformatically extracted from the list of total identified peptide spectral matches. These GVPs were prescreened to eliminate those that were not unique, defined as sharing the amino acid sequence from another gene product in the human proteome including variants. Unique sequences that correspond to GVPs were searched for, along with chemical

modifications or single amino acid polymorphisms. False positive rates, due to errors in peptide spectral matching or errors in software or spreadsheet analysis, were not able to be measured when used in isolation. GVP detection required subsequent validation through DNA genotyping of matching DNA samples. Genotypic frequencies from the European and African reference populations of the 1000 Genomes Consortium were consulted to calculate RMP [29]. When combining datasets from three biological replicates of a sample, presence of a GVP was determined by detection in any of the datasets, with no additional weighting for the second identification. RMPs from combined datasets are reported as averaged and not a cumulative probability with higher discrimination.

2.6 Calculation of Random Match Probability

RMP was calculated using the product rule [13, 30] with genotypic frequencies from the 1000 Genomes Project (https://www.internationalgenome.org) from five populations; African, European, East Asian, South Asian, and American [29]. Complete linkage for GVPs shared within an open reading frame was assumed as well as no linkage between open reading frames of different genes. For GVPs that were determined to be genetically linked within an open reading frame, a cumulative genotypic frequency was estimated using summation of all potential diplotype combinations. Sensitivity was calculated as the true positive rate divided by the sum of true positives and false negatives. Homozygosity was not assumed when only one allele was detected from a locus. Instead, the estimated genotype frequency ($gf_p = p2 + 2pq$) from the reference population was substituted [27]. To avoid a null value, each genotypic frequency was expressed as $(x + ½)/(n + 1)$, where x is the number of individuals with a given SNP, or combination of SNPs, in the sample population [31, 32].

2.7 Genetic Validation of Variant Peptides

Matching genomic DNA was extracted from buccal cells and saliva obtained from a mouthwash and isolated using Gentra Puregene Tissue Kit from Qiagen Inc. (European samples) or from buffy coat using an in-house phenol/chloroform protocol by Sorenson Forensics LLC, Salt Lake City, UT (African samples). Exome sequencing data obtained using the DNA Technologies core and Bioinformatics core facilities in the Genome Center at the University of California, Davis [27]. Barcode-indexed sequencing libraries were generated from genomic DNA samples (1000 ng) sheared on an E220 Focused Ultrasonicator (Covaris, Woburn, MA). The sonicated DNA was size selected with KAPA Pure beads to obtain fragments of about 300bp. Size selected DNA (30 ng) were used for library preparations with the KAPA Hyper DNA library kit, according to the manufacturer's instructions. Ten cycles of PCR were conducted to amplify the libraries. Each library (500 ng) was pooled for exome capture using the IDT xGen® hybridization capture protocol according to the manufacturer's instructions. Seven cycles of PCR were conducted to amplify the library that was analyzed with a Bioanalyzer 2100 instrument (Agilent, Santa Clara, CA), quantified by fluorometry on a Qubit instrument (LifeTechnologies, Carlsbad, CA), and combined in two pools at equimolar ratios. The pools were quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems-Roche) and each pool was sequenced on one lane of an Illumina Nova Seq (Illumina, San Diego, CA) with paired-end 150 bp reads. Raw Illumina paired-end 151 bp reads were first subjected to quality control. Adapters were removed from the sequencing reads using scythe (https://github.com/vsbuffalo/scythe, version 0.994 beta). Base quality was controlled using a window-based method, sickle (https://github.com/najoshi/sickle, version 1.33), with the cutoff set at 30. Reads less than 30 bp in length were discarded. Reads that passed the quality control were mapped to hg19 reference genome using parameter -M for downstream analysis compatibility [33]. PCR duplicates were removed using Picard tools (http://broadinstitute.github.io/picard/, version 2.18.4). Variants were identified using HaplotypeCaller function in GATK (version 4.0.5.2), followed by variant recalibration

using the recommendations from GATK developers [34]. Genotypes for the six subjects used in this research are available in Table S1.

2.8 Discovery of New Genetically Variant Peptides

A spreadsheet, termed GVP Scout (v1.1), was created to search for putative GVPs in proteomic datasets. This spreadsheet can be obtained from the resources menu of XXXXXX. In short, identified single amino acid variants from GPM software were screened and variant peptides with matching common (>0.5% global minor allele frequency) putative non-synonymous SNP alleles were identified and subsequently filtered manually based on exclusionary characteristics such as unique sequence, minor allele frequency, and mass shift. To prevent the inclusion of peptide with more than one genomic address, all peptide sequences were submitted to PROWL (prowl.rockefeller.edu/prowl/proteininfo) and searched against the IPI human (2010-02- 01) database. Peptides with no match or represented by a single point in the genome were considered unique and included in the study.

The putative list of GVPs was assembled based on hair proteomes using samples from the six individuals in this manuscript (Table S2). Putative GVPs were not held to stringent quality standards and were confirmed using matching the mass spectral data. Transitions ideally flanked the single amino acid variant in question. The quality of the whole spectrum was also assessed. However, proteomes that differed based on the processing or analysis methods contained different members in the detected protein population that introduced additional GVPs with MAF > 0.5%. Putative GVPs that were identified in this manner underwent further standards of confirmation steps such as ensuring that the tryptic sequence was unique, the RSID corresponded to a missense mutation, and the mass shift was not due to a chemical modification. Resulting candidate GVPs underwent additional screening via DNA genotyping to become a validated GVP.

2.9 Data Reporting and Availability

African hair sample A1 (D1.0007) was left out of most calculations and was considered an outlier, due to its chemical intractability. Therefore, results which are reported for African samples only are reported as X ± Y, where X is the average and Y is the variance. All other error values (Y) are reported as standard deviation. Reported P-values also exclude the intractable hair sample. All RAW data files and spreadsheets of detected peptides and proteins from hair digests mentioned in this work, including from the supplemental section, are publicly available on ProteomeXchange (PDX016155) [35]. The folder also includes post-analysis using Global Proteome Machine, such as peptide and protein spreadsheets. See Table S3 for a complete list of data available.

3. Results

3.1 Time and Temperature of Reduction with Detergent Treatment

Since proteins undergo chemical modifications when treated with high temperature for long time periods [36], the first optimized parameters for proteomic processing were the duration and temperature for disulfide reduction that was conducted in the presence of detergent. For this experiment, hair samples were reduced for 18 h with 50 mM dithioerythritol (DTE) without agitation at either room temperature or in an oven at 70°C before three days of digestion. Hair processing was assessed by quantification of the trypsin-insoluble material using ninhydrin as well as proteomic analysis. An initial prediction would be that increased solubilization of hair matrix would result in increased release, and subsequent detection, of hair shaft peptides. Indeed, lower incubation temperatures resulted in more insoluble material (Figure 1A, S1). Insolubility was especially evident with the African hair sample that exhibited only 35% ± 7% solubilization (65% insoluble material) relative to 67% ± 1% solubilization (33% insoluble material) when treated at 70^{o}C ($p = 0.03$, Figure 1A). However,

283 the number of unique peptides actually improved under lower temperatures, increasing from 1840 ±
284 260 to 2570 ± 60 ($p = 0.02$) (Figure 1C). This apparent contradiction indicated that solubilization alone is
285 not a reliable indicator of peptide release and identification from the hair matrix. An insight into the
286 chemical mechanisms at play in the heated sample was provided by deamidation data. Reduction at
287 room temperature decreased the deamidation ratio, defined as the number of peptides containing
288 deamidation divided by the total number of peptides, from 0.19 ± 0.06 to 0.05 ± 0.01 ($p = 0.007$) (Figure
289 1B). This demonstrated that higher temperatures were increasing conformational mobility of the
290 peptide and facilitating chemical modifications that change the peptide mass and result in a dilution of
291 the initially-released peptide.
292 The reduction time was then assessed by comparing the 18 hrs 70°C static reduction with a 6 hrs
293 23°C reduction that incorporated stirring at medium speed (Figure 1D). The reduction with stirring,
294 shorter incubation time, and lower incubation temperature yielded an increase in the number of unique
295 peptides from 2060 ± 50 to 2830 ± 70 ($p = 4 \times 10^{-4}$), compared to samples that were held static for 18 hrs
296 at 70°C. This suggests that shorter durations of reduction at room temperature are beneficial for
297 proteome coverage and maximizing useful peptides for GVP analysis.
298
299 *3.2 Trypsin Time-Course*
300 The second parameter to be optimized was the time required for trypsin proteolysis. The initial
301 condition was for three days with one 1:50 addition each day. A time-course experiment was conducted,
302 where a single 1:50 addition of reductively methylated trypsin (RMT) was made to 4 mg of hair for one
303 subject of European ancestry and one subject of African ancestry. Digestion was stopped by freezing at
304 either 1, 3, 6, or 24 hrs. Figure 2 demonstrates the effect digestion had on the number of unique
305 peptides and the number of genetically variant peptides (GVPs) detected. After 6 hrs of digestion, both
306 European and African hair values reached a plateau. However, the African hair samples yielded fewer

unique peptides (2590 ± 10 compared to 2890 ± 50, p = 0.01) and fewer GVPs (38 ± 1 compared to 46 ± 3, p = 0.02) compared to the European samples at 6 hrs of digestion. This difference is primarily due to the concentration of reducing agent, as mentioned in the next section. The data suggested that there was no advantage in longer incubation times beyond the 6 hr digestion period. Likewise, there were no advantages in terms of time of digestion for the detection of proteins of interest such as keratin associated proteins (KAPs) (Figure S2A).

3.3 Concentration of Reducing Agent

Hair shafts have high levels of disulfide bonds that result in extensive protein-to-protein cross-linking and subsequent tissue rigidity and robustness. This makes disulfide bonds an attractive target for opening up the keratin matrix to increase access to internal biomolecules in a way that avoids harsh chemistries. Accordingly, a European and an African hair sample were reduced using DTE concentrations of 25 mM, 50 mM, 75 mM, and 100 mM in biological triplicates (Figure 3, S3, & Table S4). After trypsin digestion and proteomic mass spectrometry, resulting datasets were analyzed for protein coverage (Figure 3A, Table S4). Higher levels of DTE increased coverage of detected proteins. At 100 mM DTE, protein coverage improved to the point that 37 of the 427 proteins had 100% coverage and 76 proteins had 50% or more coverage, compared to that at 25 mM DTE, which had 6 of the 656 proteins at 100% coverage and 53 proteins with over 50% coverage. The initial processing conditions for hair processing used 50 mM reductant [22-24], and at this level only 8 of 475 proteins had full coverage and 50 had 50% coverage or greater.

Part of the increase in protein coverage can be attributed to an increased number of identified KAPs in both the European and African hair samples (Figure S2B). This diverse family of small proteins can contain up to 36% of their amino acids as cysteine [37]. In terms of KAPs, the African hair increased from 8 ± 1 to 38 ± 2 ($p = 7 \times 10^{-4}$) and the European hair increased from 31 ± 3 to 47 ± 2 (p = 0.009) going

from 25 to 100 mM DTE. There was also an increase in the number of detected KAPs for the European sample after reducing time and temperature during reduction and also reducing digestion time (Figure 3B). The numbers of KAPs detected were similar between the modified method (M+100) and a previously reported urea-based method (P+100). 47 KAPs were detected using the reduction-optimized method and 48 KAPs were detected using the urea-based method for the European sample, and 38 versus 36 KAPs for the African sample. This increase was not observed for the African hair sample until modifying the concentrations of reducing agent.

With higher levels of reductant, access to the relaxed keratin matrix facilitates the release of genetically variant peptides from other proteins (Figure 3C). The African hair increased in GVP number from 50 ± 1 to 70 ± 8 ($p = 0.02$) and the European hair increased from 66 ± 2 to 83 ± 4 ($p = 0.01$) going from 25 to 100 mM DTE. Some GVPs were identified more frequently in non-KAP proteins when using higher concentrations of reducing agent such as those derived from SNPs rs9916724, rs9916484, and rs9916475 in KRT37. Both groups yielded the most GVPs at 100 mM DTE, which was taken as the optimum for subsequent analysis.

3.4 Comparing the Finalized and Original Chemistries

A comparison was made between the original processing chemistry and the optimized processing chemistry for 2 cm of reference hair from six subjects (Figures 4 and 5). Three subjects were of African ancestry and three subjects were of European ancestry. All subjects had three replicates for each condition (original and optimized) that were separately digested and analyzed. The resulting profiles of detected GVPs, as illustrated in the insert for Figure 4 (Gene, rsID, SAP and sequence), gave inferred profile of non-synonymous SNP alleles that were directly compared with whole exome sequencing from the same individuals. Four performance outcomes for each inference (TP, true positive, blue; FP, false positive, red; TN, true negative, white; FN, false negative, green) were indicated for each

broad protein class in hair shafts, keratins, KAPs and other proteins. The rate (%) of each outcome is indicated. The most noticeable improvement in true positive inference is the detection of GVPs in KAPs. The intractable hair sample was especially lacking in this protein class with only 1 GVP identified, a clear outlier. Because of this we did not include results from this sample in overall comparisons outlined below. This is primarily due to an overall loss in KAPs from family 4, 5, and 9 (Table S5). Overall sensitivity of the analysis (TP/(TP+FN)) improved 3-fold from 11% to 34%, without altering instrumental parameters. The improved sensitivity was attributed mostly to GVPs in KAPs, increasing from 0 to 49. However, more GVPs were identified and detected in all protein categories, indicating that cleavage of disulfide bonds resulted in opening up the keratin matrix and increased overall protein digestion and release of peptides from the matrix. The total identified GVPs increased from 45 to 127 for the optimized processing method (Figure 4 & S4). The false positive rate (TP/(TP+FP)) did not change with the use of optimized chemistry.

Results indicate that the optimized processing method outperformed the original processing method except with an intractable hair sample from one subject (A1) (Figure 5). Optimization of processing increased the number of unique peptides 1.7-fold from 1590 ± 160 to 2700 ± 230 ($p = 5 \times 10^{-13}$) (Figure 5A). The average number of genetically variant peptides detected increased 3.7-fold from 20 ± 5 to 73 ± 5 ($p = 1 \times 10^{-13}$) after optimization (Figure 5B). RMP increased from a maximum of 1 in 1400 and a median value of 1 in 24 for the original processing method to up to 1 in 620 million from a single hair with a median value of 1 in 1.1 million after chemical processing optimization ($p = 4 \times 10^{-7}$) (Figure 5C). Likewise, median RMPs for the African samples increased from 1 in 5.1×10^{1} to 1 in 1.5×10^{8}, and European samples increased from 1 in 1.3×10^{1} to 1 in 2.2×10^{3}. While the numbers of unique peptides and GVPs were similar between the European and African subjects, calculated RMPs were higher (1.5×10^{8} vs 2.2×10^{3}) in African subjects due to the differences in the genotype frequency of inferred loci in each reference population.

RMPs calculated using genotype frequencies from different reference populations (1000 Genomes Project) were compared using a likelihood ratio (LR) defined as the RMP calculated from the African population divided by the RMP calculated from the European population (LR = Pr(GVP profile|AFR) / Pr(GVP profile|EUR)) (Figure 5D). With optimization and increased GVP detection, the likelihood ratio for European samples decreased by 0.94 ± 0.39 orders of magnitude ($p = 1 \times 10^{-4}$), while the African samples increased by 3.90 ± 0.32 orders of magnitude ($p = 5 \times 10^{-4}$). The GVP profiles from African subjects were therefore considerably less frequent in European populations than in African ones and *vice versa*. Final likelihood ratio estimates averaged 4.1 ± 0.6 orders of magnitude for the two tractable African samples, and negative 2.7 ± 1.3 orders of magnitude (average ± standard deviation, of log transformed values) for the European samples ($p = 0.008$, using log transformed values) a difference of 6.8 orders of magnitude. These effects reflect differences in the structure of the respective reference populations. The use of LR values for ancestral characterization may be further explored with a larger cohort of Europeans and African samples.

3.5 Newly Discovered Genetically Variant Peptides

In summary, using the discovery protocols described in the Methods section, a total of 125 non-synonymous SNP loci were discovered and 152 GVPs were confirmed proteomically and subsequently validated by direct comparison with DNA sequenced genotypes (Tables S1, S6, and S7). To make these discoveries, the GVP Scout spreadsheet was used and the peptides filtered for uniqueness. Non-synonymous SNP loci were identified in the genes, described in more detail in Tables S6 and S2. Of the 125 SNPs, 59 have not been reported in other forensic proteomic literature. Of these 59, six are in KRT genes and 19 are in KRTAP genes. Of particular interest are common SNPs that have a global minor allele frequency above 0.30 (rs58001094, rs2037912, rs4818950, rs2074285, rs688906, rs537301040, rs9897031, and rs238239). These loci are expected to be observed as heterozygote genotypes more

403 frequently resulting in higher discriminatory power. A comprehensive description of the chemical and
404 genetic properties of all GVPs used in this study is included in the Supplemental section (Tables S1 and
405 S7).
406
407 **4. Discussion**
408
409 Forensically-applicable proteomic genotyping requires sample workflows to be developed that
410 are sensitive enough to extract the necessary genetic information from the minimum of material, in this
411 case a fraction of a single hair shaft. This development project optimized the sensitivity of hair
412 proteomic genotyping by focusing on two factors: milder chemical conditions and sulfur chemistry. The
413 milder conditions were assisted by the use of sodium dodecanoate that is strongly amphipathic and an
414 effective denaturant, while also being relatively easy to remove through brief acidification and organic
415 extraction [38]. Mild chemistries, such as lower temperatures and shorter incubation times, decreased
416 the soluble fraction after digestion and yet increased the number of unique peptides, most likely due to
417 the reduced level of post-digestion peptide modification. The modification that best illustrates this is
418 deamidation (Figure 1B), but other modifications would also be present (data not shown). Therefore, an
419 increase in solubilization of hair protein did not necessarily equate to better proteomic data. The overall
420 result of using mild processing chemistries is an improvement in digestion efficiency that increased the
421 number of unique peptides, genetically variant peptides (GVPs), and resulting random match
422 probabilities (RMPs) from human hair. The data from 2 cm of a hair shaft is now equivalent in yield to
423 that previously obtained from 4 mg [39] or even 10 mg [13] of hair tested. The focus on mild chemistries
424 has the additional benefit of reduced processing times, that are currently only 14 hours.
425 Hair has distinctively high levels of disulfide chemistry and so higher levels of reductant allowed
426 the keratin matrix to open up further to promote hair protein proteolysis and release keratin-associated

and other proteins for subsequent analysis. To optimize detection, a target peptide needs to have a maximal concentration in a sample and have minimal modifications so that signal was focused into a single mass. This requires a balance between the release of a peptide into the sample from the keratin matrix with a reduction in subsequent down-stream chemistries that will change the mass of the peptide through chemical modification, or miscleavage [40]. The chemistry required to maximize the release of target peptides from the keratin matrix also acts to modify the peptides and spread the signal across a range of masses resulting in a lower yield of unique peptides and GVPs with a single mass [26, 41]. This project shifts the balance point between these two opposing factors by using high levels of reductant, as much as 100 mM dithiothreitol (DTT), and a strong detergent that opens up the keratin matrix releasing proteins and peptides without resorting to harsher chemistries. The evidence of this is the increased presence of keratin associated proteins (KAPs) in the samples, along with their GVPs (Figure 3 & 4). Increased levels of reductant have previously been shown to be critical to releasing KAPs in wool and textiles [20, 42].

Earlier reports on forensic proteomics that focused on hair shaft protein used high amounts of hair, 4 or 10 mg (Table S8), since they were focused on either basic science questions, such as protein profiles, or discovery of genetically variant peptides for proteomic genotyping [13, 39, 43]. Naturally, development of a forensically useful hair proteomic protocol would focus on a method that required only a fraction of a single hair shaft that would be the limit of material obtained through casework [44-47]. This study has been an open part of this process [48-51]. Over that time period other single hair methods for proteomics and proteomic genotyping have also been reported, and like this study also demonstrate high levels of protein detection and/or discrimination with 1 mm to 20 or 25 mm of a single hair shaft [44-46, 52, 53]. Some of the chemistry in this project is similar to that reported, but not fully documented, by other protocols [44].

Other hair processing protocols take different approaches. At one extreme a recently published method using heavily alkaline conditions was used to quickly extract around 50% of hair shaft protein [54]. These harsh conditions resulted in poor protein and peptide yields, and presumably would result in chemical degradation within the hair. One of the most widely used protocols, the Shindai method, uses 2.6 M thiourea, 5 M urea and 5% beta-mercaptoethanol at high temperatures (50˚C) for 24 to 72h at pH 8.5 [54-56]. This and related commonly used methods using 8 M urea have the advantage of not relying on detergent that can be difficult to remove prior to mass spectrometry [42, 45, 47, 56, 57]. These often resulted in similar levels of protein and peptide yields [47]. Other research groups remove detergent and desalt using in-gel digestion that has the advantage of further denaturing protein and increasing fractionation [42, 53]. However, in-gel digestion protocols result in sample loss since they do not use insoluble material that are a potentially rich source of proteomic material and are time and resource intensive [39]. The chemistry employed in the initial GVP-demonstration paper used urea and a mass spectrometry-compatible surfactant, along with 100 mM DTT [13]. We did not pursue development of this method, although it also achieves rich proteomic datasets for large quantities of hair, because of the chemical fragility and milder amphipathic character of the acid-labile surfactant [58].

There are still some chemistries that may be incorporated into hair sample processing. We find that 15-20% of hair mass is left insoluble after digestion. We hypothesize that this is due to covalent linkages that would not change when solubilizing in SDS instead of ABC (data not shown) [15]. Improvements in the protocol may focus on stronger detergents, combined use of urea and thio-urea as used in the Shindai method. Other buffers, detergents, enzymes, and alkylating agents could still be tested to further optimize proteomic processing. Further optimization of the timing and combination of the steps employed in this project is still possible.

Intractable hair samples in our hands comprised about 3% of both African or European samples (data not shown). About 50% of intractable hair samples have undergone hair-straightening treatment.

In our analysis of intractable hair, many methods were tested to aid in solubilization. Sonication, high temperatures, freeze-thawing, organic extraction, and increasing the concentration of DTT were all tested, without success. Intractable hair samples were slightly more digested using the original processing method compared to the optimized method. However, intractable hair samples still yield less than 20% of the unique peptides and unique GVPs compared to normal hair samples. The major proteomic difference between normal and intractable hair samples is that they lack peptides from KAPs that are high in cysteine content (Table S5). More effort will be invested in future research to diagnose and mitigate the problems seen with intractable hair samples.

Proteomic datasets should ideally be equivalent in terms of protein, unique peptide, and GVP number between different biogeographic groups, color, and age. Datasets differing in these characteristics may yield a systematic bias in the GVP profiles and in resulting statistical analyses between these groups. For instance, the original processing method had on average 1.4x more GVPs in the European cohort than in the African cohort. This may indicate that certain groups would hold higher evidentiary value of proteomic data. Present research, aiming to reduce statistical bias between a European and African cohort, has decreased the difference in GVP number down to 1.1x between Europeans and Africans. However, RMP calculations will still benefit from the variety and intrinsic distribution of SNPs in the African population that result from its deeper evolutionary history [29].

Future research for the study of genetically variant peptides in human hair may well involve targeted proteomics, ancestral classification, automation in sample processing, scouting and identification of novel GVPs, and developing a genotyping kit for confirmed and validated GVPs. However, the method proposed here is a significant advance and demonstrates a three-fold increase in sensitivity of GVP detection and a three orders of magnitude increase in RMP. This foundation, in addition to being a resource for the field, also allows us now to investigate other areas of development necessary for implementation as a forensic tool. These include investigating different casework

scenarios that would affect data yields or introduce statistical bias into the analysis [43, 59]. Our improvements also provide a foundation for further refinement of downstream mass spectrometry data acquisition and bioinformatics processing protocols.

5. Conclusion

In forensic science it is essential to maximize the extraction of the target biological material. An effective use of human hair in forensic proteomics requires sensitive and efficient sample processing protocols that can be used on a single hair shaft. Maximization of peptide production and minimization of additional chemistries is required to increase the detection of informative peptides. Harsher chemistries are especially problematic because they chemically modify peptides and further dilute the mass signatures. In this research, we combine milder digestion conditions with an increase in reductive compounds, up to 100 mM DTT, to cleave the high levels of disulfide bonds and open up the keratin and keratin-associated protein matrix. This approach should also work for those investigating other chemically fragile biomolecules in the hair shaft, such as mitochondrial DNA and chemically labile small molecules. This optimized method produces more unique peptides, genetically variant peptides, and more discriminatory random match probabilities, particularly through the release of keratin-associated proteins. Random match probability has also improved to 1 in > 600 million for a single hair. The method outlined here produces a similar number of genetically variant peptides between European and African hair digests, and significantly improves the evidentiary value of 2 cm of hair.

Funding
Disclaimers
Acknowledgements

References

1. Kligman, A. M. (1961). Pathologic dynamics of human hair loss: I. Telogen effluvium. *Archives of dermatology*, *83*(2), 175-198.

2. Barthélemy, N. R., Bednarczyk, A., Schaeffer-Reiss, C., Jullien, D., Van Dorsselaer, A., & Cavusoglu, N. (2012). Proteomic tools for the investigation of human hair structural proteins and evidence of weakness sites on hair keratin coil segments. *Analytical biochemistry*, *421*(1), 43-55.
3. Desiere, F., Deutsch, E.W., Nesvizhskii, A.I., Mallick, P., King, N.L., Eng, J.K., Aderem, A., Boyle, R., Brunner, E., Donohoe, S. and Fausto, N. (2005). Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome biology*, *6*(1), 9.
4. Kadiyala, C. S. R., Tomechko, S. E., & Miyagi, M. (2010). Perfluorooctanoic acid for shotgun proteomics. *PLoS One*, *5*(12), e15332.
5. Edwards, H., & Gotsonis, C. (2009). Strengthening forensic science in the United States: A path forward. Statement before the United State Senate Committee on the Judiciary.
6. Houck, M. M., & Budowle, B. (2002). Correlation of microscopic and mitochondrial DNA hair comparisons. *Journal of Forensic Science*, *47*(5), 1-4.
7. Kim, B. J., Na, J. I., Park, W. S., Eun, H. C., & Kwon, O. S. (2006). Hair cuticle differences between Asian and Caucasian females. *International journal of dermatology*, *45*(12), 1435-1437.
8. Ogle, J., & Fox, M. J. (2017). *Atlas of human hair: microscopic characteristics*. Crc Press.
9. Brace, C. L. (1995). Region does not mean "race"—reality versus convention in forensic anthropology. *Journal of Forensic Science*, *40*(2), 171-175.
10. Deedrick, D. W., & Koch, S. L. (2004). Microscopy of hair part 1: a practical guide and manual for human hairs. *Forensic Science Communications*, *6*(1).
11. Koch, S. L., Shriver, M. D., & Jablonski, N. G. (2019). Variation in human hair ultrastructure among three biogeographic populations. *Journal of structural biology*, *205*(1), 60-66.
12. Hicks, J. W. (1977). *Microscopy of hairs: a practical guide and manual*. Federal Bureau of Investigation, FBI Laboratory.
13. Parker, G.J., Leppert, T., Anex, D.S., Hilmer, J.K., Matsunami, N., Baird, L., Stevens, J., Parsawar, K., Durbin-Johnson, B.P., Rocke, D.M. and Nelson, C. (2016). Demonstration of protein-based human identification using the hair shaft proteome. *PloS one*, *11*(9), p.e0160653.
14. Gilbert, M.T.P., Wilson, A.S., Bunce, M., Hansen, A.J., Willerslev, E., Shapiro, B., Higham, T.F., Richards, M.P., O'Connell, T.C., Tobin, D.J. and Janaway, R.C. (2004). Ancient mitochondrial DNA from hair. *Current Biology*, *14*(12), 463-R464.
15. Rice, R. H. (2011). Proteomic analysis of hair shaft and nail plate. *Journal of cosmetic science*, *62*(2), 229.
16. Rice, R. H., Bradshaw, K. M., Durbin-Johnson, B. P., Rocke, D. M., Eigenheer, R. A., Phinney, B. S., & Sundberg, J. P. (2012). Differentiating inbred mouse strains from each other and those with single gene mutations using hair proteomics. *PLoS One*, *7*(12), e51956.
17. Chou, C. C., & Buehler, M. J. (2012). Structure and mechanical properties of human trichocyte keratin intermediate filament protein. Biomacromolecules, 13(11), 3522-3532.
18. Plowman, J. E., Harland, D. P., & Deb-Choudhury, S. (Eds.). (2018). *The Hair Fibre: Proteins, Structure and Development* (Vol. 1054). Springer.
19. Lee, C. H., Kim, M. S., Chung, B. M., Leahy, D. J., & Coulombe, P. A. (2012). Structural basis for heteromeric assembly and perinuclear organization of keratin filaments. *Nature structural & molecular biology*, *19*(7), 707.
20. Solazzo, C., Dyer, J. M., Clerens, S., Plowman, J., Peacock, E. E., & Collins, M. J. (2013). Proteomic evaluation of the biodegradation of wool fabrics in experimental burials. *International Biodeterioration & Biodegradation*, *80*, 48-59.
21. Rice, R. H., Means, G. E., & Brown, W. D. (1977). Stabilization of bovine trypsin by reductive methylation. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *492*(2), 316-321.
22. Rice, R. H., Wu, P. W., & Mann, S. M. (2015). Human Hair Proteomics-Improved Evidence Discrimination. https://www.ncjrs.gov/pdffiles1/nij/grants/249102.pdf

23. Laatsch, C.N., Durbin-Johnson, B.P., Rocke, D.M., Mukwana, S., Newland, A.B., Flagler, M.J., Davis, M.G., Eigenheer, R.A., Phinney, B.S. and Rice, R.H. (2014). Human hair shaft proteomic profiling: individual differences, site specificity and cuticle analysis. *PeerJ*, *2*, p.e506.
24. Wu, P.W., Mason, K.E., Durbin-Johnson, B.P., Salemi, M., Phinney, B.S., Rocke, D.M., Parker, G.J. and Rice, R.H. (2017). Proteomic analysis of hair shafts from monozygotic twins: Expression profiles and genetically variant peptides. *Proteomics*.
25. Moore, S., & Stein, W. H. (1954). A modified ninhydrin reagent for the photometric determination of amino acids and related compounds. *J. biol. Chem*, *211*(2), 907-913.
26. Schiffman, G. (1966) Immunological methods for characterizing polysaccharides. Meth. Enzymol. 8, 79-85
27. Borja, T., Karim, N., Goecker, Z., Salemi, M., Phinney, B., Naeem, M., Rice, R., Parker, G. (2019). Proteomic Genotyping of Fingermark Donors with Genetically Variant Peptides. *Forensic Science International: Genetics* https://doi.org/10.1016/j.fsigen.2019.05.005
28. de Graaf, E.L., Altelaar, A.M., van Breukelen, B., Mohammed, S. and Heck, A.J. (2011). Improving SRM assay development: a global comparison between triple quadrupole, ion trap, and higher energy CID peptide fragmentation spectra. *Journal of proteome research*, *10*(9), 4334-4341.
29. 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68.
30. Evett, I.W. and Weir, B.S. (1998). *Interpreting DNA evidence: statistical genetics for forensic scientists*. Sinauer Associates Sunderland MA.
31. Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *186*(1007), 453-461.
32. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.
33. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
34. Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D. and Shakir, K., 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, p.201178.
35. Perez-Riverol, Y., Xu, Q.W., Wang, R., Uszkoreit, J., Griss, J., Sanchez, A., Reisinger, F., Csordas, A., Ternent, T., del-Toro, N. and Dianes, J.A. (2016). PRIDE inspector toolsuite: moving toward a universal visualization tool for proteomics data standard formats and quality assessment of ProteomeXchange datasets. *Molecular & Cellular Proteomics*, *15*(1), pp.305-317.
36. Hao, P., Ren, Y., Alpert, A. J., & Sze, S. K. (2011). Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Molecular & cellular proteomics*, *10*(10), O111-009381.
37. Rogers, M.A., Langbein, L., Praetzel-Wunder, S., Winter, H. and Schweizer, J. (2006). Human hair keratin-associated proteins (KAPs). *International review of cytology*, *251*, 209-263.
38. Lin, Y., Huo, L., Liu, Z., Li, J., Liu, Y., He, Q., Wang, X. and Liang, S. (2013). Sodium laurate, a novel protease-and mass spectrometry-compatible detergent for mass spectrometry-based membrane proteomics. *PloS one*, *8*(3), p.e59779.
39. Lee, Y. J., Rice, R. H., & Lee, Y. M. (2006). Proteome analysis of human hair shaft: from protein identification to posttranslational modification. *Molecular & cellular proteomics*, *5*(5), 789-800.
40. Šlechtová, T., Gilar, M., Kalíková, K., & Tesařová, E. (2015). Insight into trypsin miscleavage: comparison of kinetic constants of problematic peptide sequences. *Analytical chemistry*, *87*(15), 7636-7643.

41. Robinson, N. E., & Robinson, A. B. (2001). Deamidation of human proteins. *Proceedings of the National Academy of Sciences*, *98*(22), 12409-12413.
42. Plowman, J. E., Deb-Choudhury, S., Thomas, A., Clerens, S., Cornellison, C. D., Grosvenor, A. J., & Dyer, J. M. (2010). Characterisation of low abundance wool proteins through novel differential extraction techniques. *Electrophoresis*, *31*(12), 1937-1946.
43. Milan, J.A., Wu, P.W., Salemi, M.R., Durbin-Johnson, B.P., Rocke, D.M., Phinney, B.S., Rice, R.H. and Parker, G.J. (2019). Comparison of protein expression levels and proteomically-inferred genotypes using human hair from different body sites. *Forensic Science International: Genetics*, *41*, pp.19-23.
44. Mason, K. E., Paul, P. H., Chu, F., Anex, D. S., & Hart, B. R. (2019). Development of a Protein-based Human Identification Capability from a Single Hair. *Journal of forensic sciences*. Doi 10.1111/1556-4029.13995.
45. Carlson, T.L., Moini, M., Eckenrode, B.A., Allred, B.M. and Donfack, J. (2018). Protein extraction from human anagen head hairs 1-millimeter or less in total length. *BioTechniques*, *64*(4), pp.170-176.
46. Lei, F., Li, J., Shan-Fei, L., Jian, Z., Hai-Bo, L., An-Quan, J., Jian, Y., Gui-Qiang, W. and Cai-Xia, L. (2019). Development and Validation of Protein-based Forensic Ancestry Inference Method Using Hair Shaft Proteome. *Progress in Biochemistry and Biophysics*, *46*(1), pp.81-88.
47. Catlin, L.A., Chou, R.M., Goecker, Z.C., Mullins, L.A., Silva, D.S.B.S.S., Spurbeck, R.R., Parker, G.J. and Bartling, C.M. (2019). Demonstration of a mitochondrial DNA-compatible workflow for genetically variant peptide identification from human hair samples. *Forensic Science International: Genetics*, *43*, 102148.
48. Goecker, Z. C., Salemi, S. R., Phinney, B. S., Rice, R. H., Parker, G. J. Optimization of Peptide Generation from Human Hair for Proteomic Analysis. Poster presented at: 65th annual American Society for Mass Spectrometry; 2017 Jun 4-8; Indianapolis, IN.
49. Goecker, Z. C., Salemi, S. R., Phinney, B. S., Rice, R. H., Parker, G. J. Optimization of Human Hair Proteomic Processing to Maximize Total and Genetically-Variant Peptides. Poster presented at: 28th annual International Symposium on Human Identification; 2017 Oct 2-5; Seattle, WA, Poster #19.
50. Goecker, Z. C., Salemi, S. R., Phinney, B. S., Rice, R. H., Parker, G. J. The Optimization of Human Hair Proteomic Processing for Single Hair and Ancestral Analysis. Poster presented at: 70th annual American Academy of Forensic Sciences; 2018 Feb 19-24; Seattle, WA, Poster B57.
51. Parker, G. J. Interplay Between Proteomic and Genomic Datasets: Merging both Information Flows in a Forensic Context. Oral presentation at: Gordon Research Conference, Forensic Analysis of Human DNA; 2018 Jun 17-22; Newry, ME.
52. Jones, K. F., Carlson, T. L., Eckenrode, B. A., & Donfack, J. (2020). Assessing protein sequencing in human single hair shafts of decreasing lengths. *Forensic Science International: Genetics*, *44*, 102145.
53. Zhang, Z., Burke, M.C., Wallace, W.E., Liang, Y., Sheetlin, S.L., Mirokhin, Y.A., Tchekhovskoi, D.V. and Stein, S.E. (2019). Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Keratin. *Journal of forensic sciences*. Doi: 10.1111/1556-4029.14229
54. Wong, S. Y., Lee, C. C., Ashrafzadeh, A., Junit, S. M., Abrahim, N., & Hashim, O. H. (2016). A high-yield two-hour protocol for extraction of human hair shaft proteins. *PloS one*, *11*(10), e0164993.
55. Fujii, T., & Li, D. (2008). Preparation and properties of protein films and particles from chicken feather. *J Biomacromolecules*, *8*, 48-55.
56. Nakamura, A., Arimoto, M., Takeuchi, K., & Fujii, T. (2002). A rapid extraction procedure of human hair proteins and identification of phosphorylated species. *Biological and Pharmaceutical Bulletin*, *25*(5), 569-572.

57. Araki, N., & Moini, M. (2011). Age estimation of museum wool textiles from Ovis aries using deamidation rates utilizing matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, *25*(22), 3396-3400.
58. Hellberg, P. E., Bergström, K., & Holmberg, K. (2000). Cleavable surfactants. *Journal of Surfactants and Detergents*, *3*(1), 81-91.
59. Franklin, R. N., Karim, N., Goecker, Z. C., Durbin-Johnson, B. P., Rice, R. H., & Parker, G. J. (2020). Proteomic Genotyping: Using Mass Spectrometry to Infer SNP Genotypes in Pigmented and Non-Pigmented Hair. *Forensic Science International,* 310, 110200.

Figures

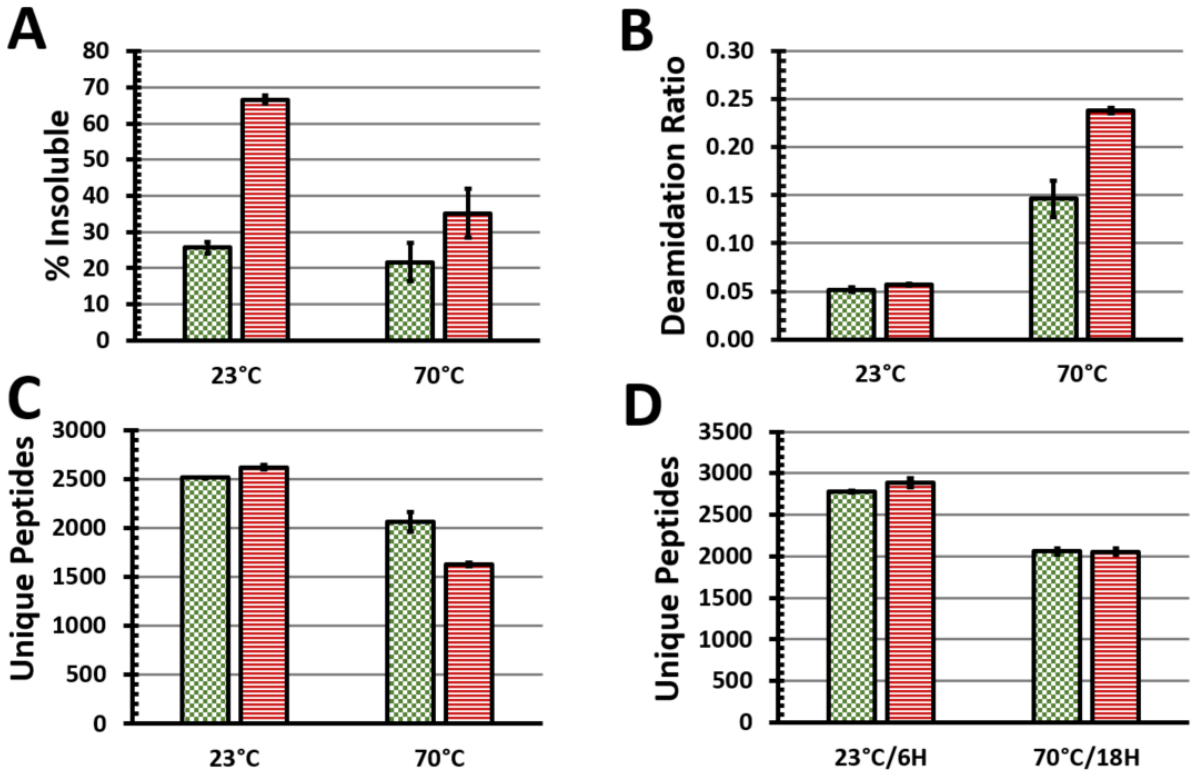


Figure 1. **Effect of temperature and time during disulfide reduction**. Hair samples (4 mg) from European (green) and African (red) subjects. **A)** % Protein (w/w) remaining insoluble after digestion of samples reduced at room temperature using the original processing method or at 70°C. **B)** Deamidation ratio (number of deamidations divided by the total number of peptides) as a function of incubation temperature. Conditions are the same as Figure 1A. **C)** The numbers of unique peptides from the original processing method. **D)** Numbers of unique peptides compiled from the original processing method (70°C/18H) or at 23°C for 6 hrs (23°C/6H).

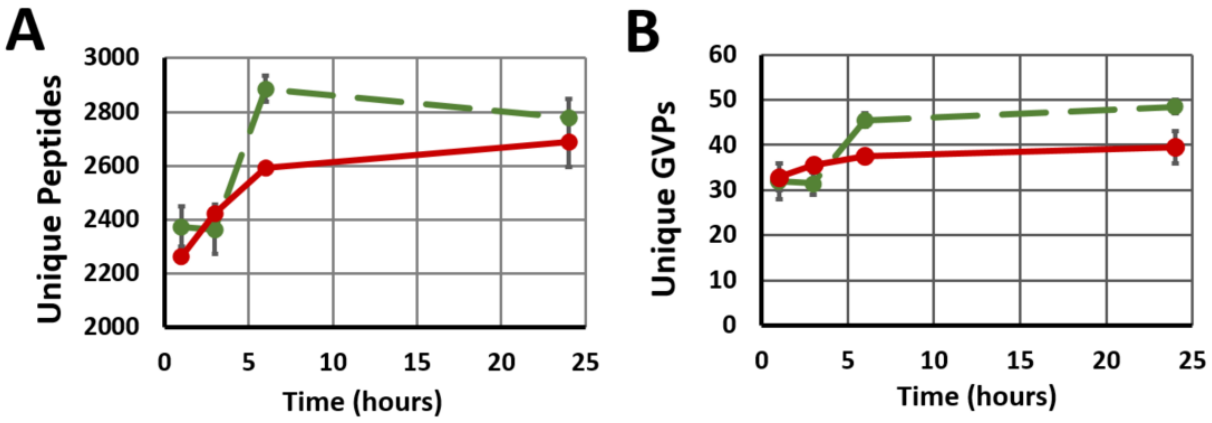


Figure 2. **Time course of hair protein digestion**. Production of unique peptides (A) and unique GVPs (B) after a single 1:50 (enzyme:protein) addition of trypsin in samples from European (green-dashed line) and African (red-full line) subjects.

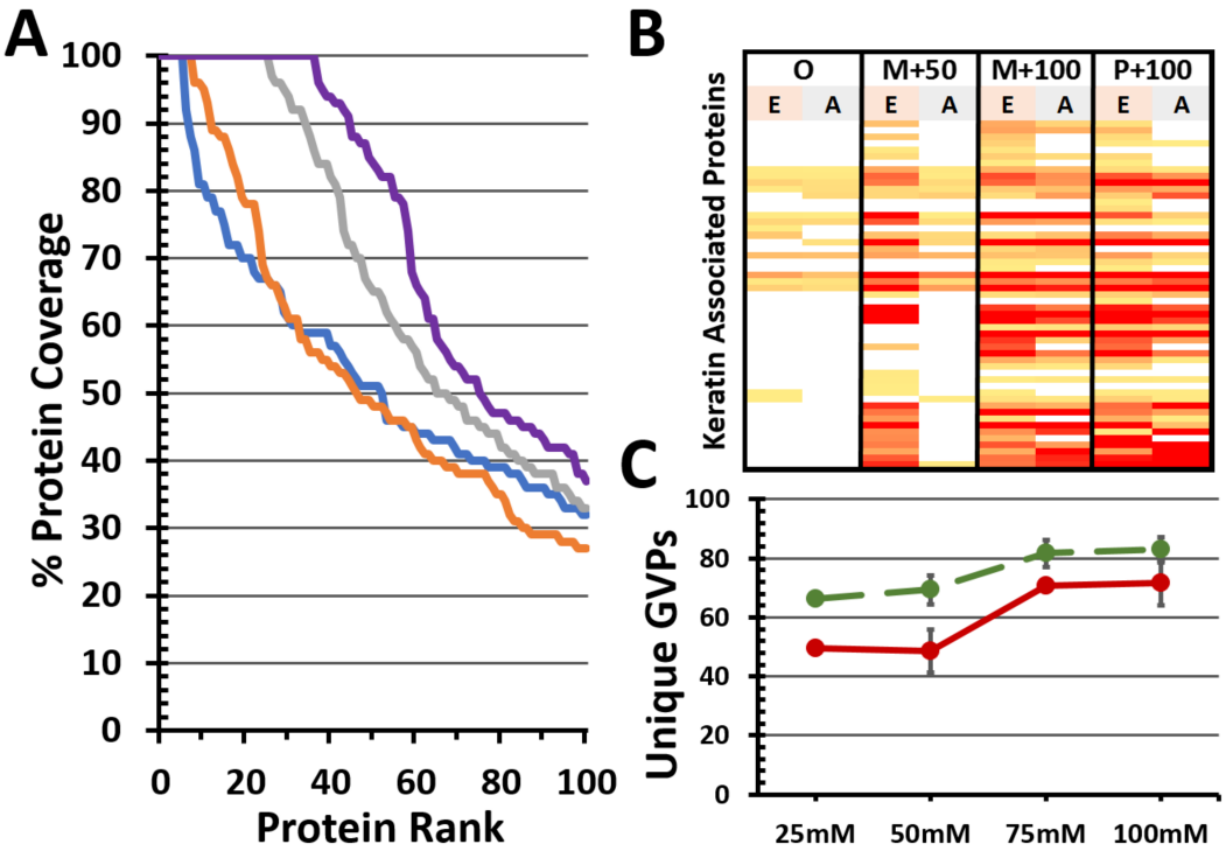


Figure 3. **Concentration of reducing agent using 4mg of hair. A)** Protein coverage from an African sample with different concentrations of reducing agent. Blue represents 25 mM DTE, orange represents 50 mM DTE, grey represents 75 mM DTE, and purple represents 100 mM DTE. Proteins are ranked based on coverage and only 100 proteins of the highest coverage are included. See Table S4 for more details. **B)** A heatmap of keratin associated proteins comparing a subject of European (E) and African (A) ancestry. White denotes no protein detected and red indicates a high level of protein detected (over 100 peptides). The abbreviation "O" indicates original method while "M" indicates use of the optimized method, "+50" and "+100" indicate using 50 mM and 100 mM DTE, respectively. The abbreviation "P+100" indicates a method of hair processing described by Parker et al [13] where Protease-Max and urea were used. **C)** Unique GVPs detected in samples from a subject of European (green) and African (red) ancestries processed using the optimized processing method.

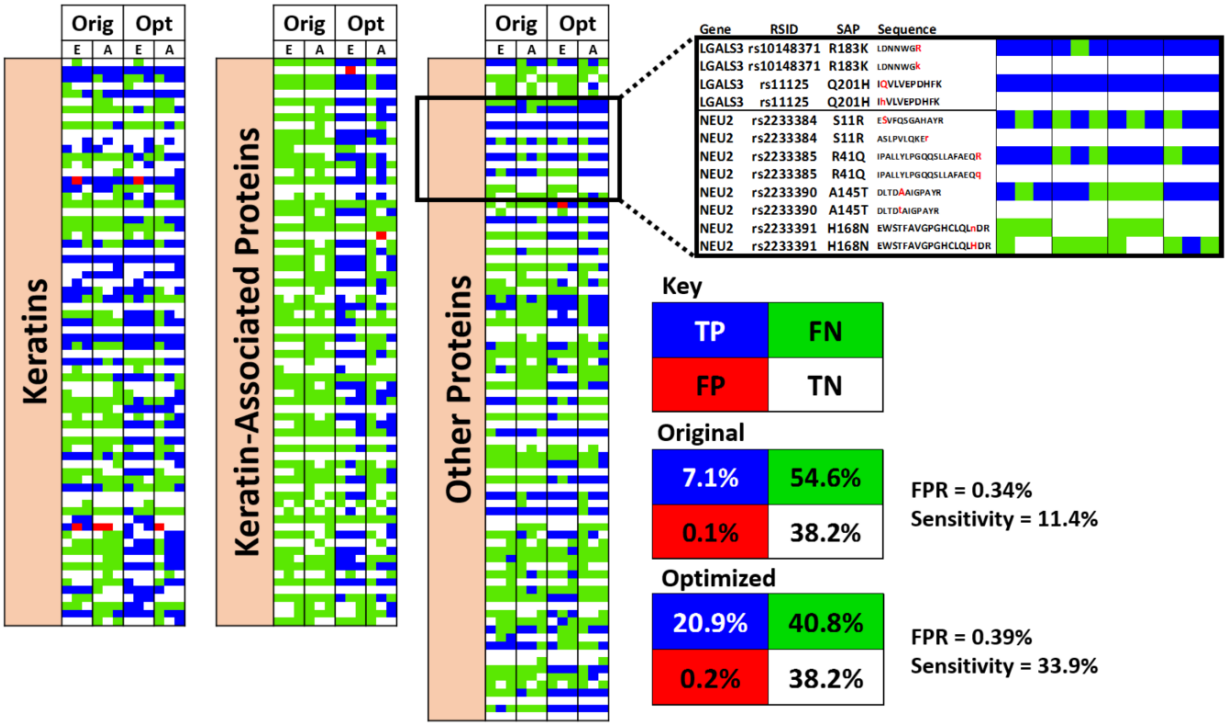


Figure 4. **GVP matrix comparing original and optimized processing methods from single hairs.** This matrix represents GVPs that have been verified via whole exome sequencing. As indicated by the zoomed-in insert in the top right corner, each row is a variant peptide. Each column is an accumulated GVP profile from three replicates. Orig, original processing method; Opt, optimized processing method; E, three European subjects; A, three African subjects; TP, true positive; FN, false negative; FP, false negative; TN, true negative; FPR, false positive rate. See figure S4 for more details.

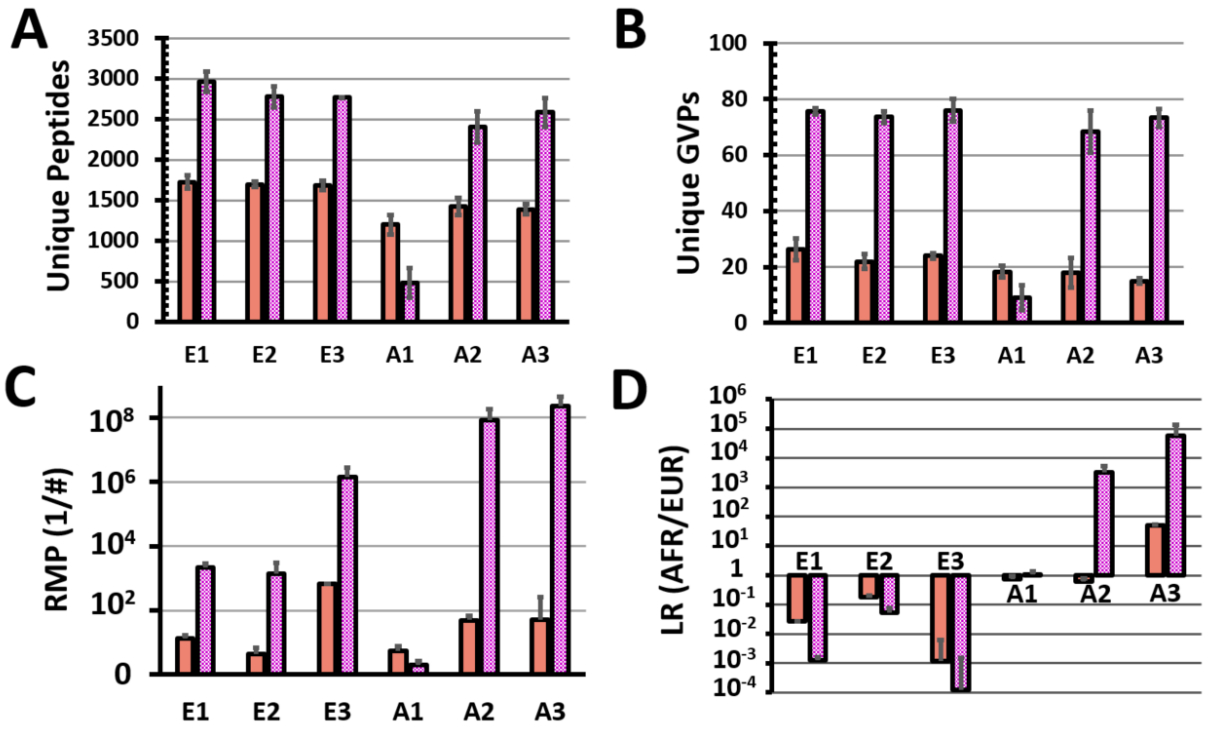


Figure 5. **Results from single hairs.** Comparisons of original (salmon) and optimized (purple) methods of hair processing are shown. **A)** Numbers of unique peptides; **B)** Numbers of GVPs; **C)** Random match probabilities; **D)** Likelihood ratios from three subjects of European (E) ancestry and three subjects of African (A) ancestry.

Supplemental Results

Optimization of processing chemistry

As discussed previously, agitation with stirring at room temperature improved proteomic results. However, other types of agitation are possible and need to be compared to ensure appropriate optimization with multiple physical factors in chemical processing. Experiments were conducted comparing the insoluble pellets of digested material after stirring, mixing, and static conditions, as illustrated in Figures S1A and S1C. Stirring and static conditions at room temperature were first compared to assess solubilization of hair mass among three biogeographic backgrounds (Figure S1A). Results indicate that improved solubilization occurred for two of the three biogeographic groups with shorter periods of reduction with stirring at room temperature, although results are only statistically significant for the African cohort. The insoluble pellet decreased for the East Asian sample from 28% ± 7% to 15% ± 5% ($p = 0.21$), European samples maintained the level of insolubility from 22% ± 8% to 21% ± 3% ($p = 0.43$), and African samples decreased from 37% ± 3% to 29% ± 4% ($p = 0.002$). However, conducting the reduction step with shaking agitation, where LoBind tubes were put directly in a thermomixer, did not produce more solubilization compared with that of static conditions (Figure S1C). Among three subjects at three different temperatures, shaking produced an insoluble pellet of 48% ± 9% of the hair mass compared to 22% ± 6% at static conditions ($p = 9 \times 10^{-12}$). Heating during reduction also resulted in more solubilization of hair shaft protein for the African subject, decreasing the insoluble pellet from 46% ± 3% to 33% ± 2% going from 23°C to 70°C ($p = 0.03$, Figure S1B). However, the European subject had a higher mass of insoluble pellet when heated during reduction, going from 18% ± 2% to 29% ± 4% ($p = 0.01$). For three subjects, heating at 70°C compared to 37°C during reduction without shaking decreased the insoluble pellet from 24% ± 7% to 17% ± 5% ($p = 0.09$) and with shaking

776 from 58% ± 5% to 38% ± 5% ($p = 9 \times 10^{-4}$), Figure S1C). Therefore, agitation with stirring at room
777 temperature during reduction, alkylation, and digestion improved the yield of proteomic information.
778 The concentration of reducing agent was paramount in improving protein coverage, keratin
779 associated protein (KAP) detection, and GVP detection (Figures S2B and Table S4). Hair shaft material (4
780 mg) was processed using 25-100mM dithioerythritol (DTE). The reducing agent was later switched to
781 dithiothreitol (DTT), a chemical with similar reducing properties and more water solubility, in the fully
782 optimized protocol to ensure no precipitation of reagents during alkylation. KAPs increased for the
783 European subject from 31 ± 3 to 47 ± 2 ($p = 0.009$) and for the African subject from 8 ± 1 to 38 ± 2 ($p = 8$
784 $\times 10^{-4}$) going from 25mM DTE to 100mM (Figure S2B). While the number of KAPs after optimization is
785 not fully equivalent between the two biogeographic populations, the gap is smaller (9 versus 23) and
786 may be due to natural biological differences. Other proteins that were represented by increased release
787 of genetically variant peptides, due to an increase in reducing agent, include KRT31, KRT37, KRT39,
788 KRT82, KRT83, KRT84, and S100A3 (Figure S3).
789 Longer digestion times for any protein matrix, regardless of chemical stability, may lead to semi-
790 tryptic and non-tryptic peptides due to minor contributions from the inherent low chymotrypsin-like
791 activity of trypsin. However, shorter digestion times may lead to reduced cleavage. This issue was
792 mitigated with 2 additions of the enzyme. A comparison of one and two additions of trypsin at a
793 concentration of 0.02% of the total protein present was conducted (Figure S1D), where the degree of
794 digestion was measured from the fraction of protein in the insoluble pellet (i.e. tryptic core) after
795 centrifugation. In the experiment, one addition of RMT (1:50) was made at T = 0. Four hair digests were
796 stirred for 6 hrs and four more were stirred for 3 hrs before a second addition of 1:50 RMT. Results
797 indicated that the mass of the insoluble pellet was 13% lower (from 33% ± 3% to 20% ± 4%, $p = 0.06$) for
798 European digests and 4% lower (from 34% ± 4% to 30% ± 2%, $p = 0.36$) for African digests for samples
799 digested with two rounds of RMT. Therefore, two additions of trypsin over a 6 hrs period of digestion

were selected for subsequent experiments. A time-course experiment of trypsin digestion indicated that the time of digestion does not assist or inhibit KAP detection, going from 10 ± 1 to 15 ± 2 ($p = 0.12$) for the European from 1 to 24 hours and from 7 ± 2 to 8 ± 2 ($p = 0.10$) for the African (Figure S2A). While KAPs were detected at 1 and 3 hrs of digestion, they were detected at lower levels.

Final optimization parameters involve 6 hrs of reduction at room temperature using 100mM DTT while stirring, followed by alkylation using 200mM IA in the dark for 45 minutes, and digestion for 6 hrs at room temperature with two 1:50 additions of trypsin. Total results for genetically variant peptides detected before and after optimization can be found in Figure S4. GVPs that were not detected in the original processing method and were detected in the optimized method include all GVPs from proteins GSTP1, KRT9, KRT32, KRT39, S100A3, VSIG8 and most KAPs.

Intractable hair samples

Chemically intractable hair is defined as a hair sample that does not swell or break apart after 6 hours of reduction at room temperature using 100 mM DTT. This included the African sample A1 (D1.0007) in Figure 4 and 5. After composing a GVP profile of the six individuals, sample A1 was lacking all GVPs from the KAP protein class (Figure S4). A visual examination was conducted to compare proteins that were present and missing in the intractable hair proteome compared to six normal sample digest proteomes. The most discriminating difference was the lack of KAPs with high sulfur content. These include KAP family 4, 5, and 9. Average cysteine content for these missing KAPs is 36%, whereas the average cysteine content for other KAPs that were detected is ~14% [37]. These proteins were either left in the insoluble pellet due to the failure of the reducing agent to penetrate the hair shaft, these proteins were not present in the hair shaft, or they were present but were incorporated into a covalent matrix of isopeptide bonds.

Attempts to mitigate intractability of hair samples included preparatory chemistries to break apart the structure of the hair before reduction. Treatment of hair with sonication before and after reduction, lipid extraction with organics such as methanol and hexane, freeze-thawing with dry ice and acetone, and reduction with high concentrations of DTT (500 mM) were all conducted on three intractable hair samples. Results for the higher DTT concentration can be found in Table S8, and other results are not shown. Only one condition showed visual improvement on hair solubilization for one hair sample, that is incubation in methanol overnight before processing. This condition did not show improvement on any other intractable hair samples. The use of heating is avoided here to avoid potential modifications altering peptide chemistry.

Alternative chemistries

Four alternative hair processing chemistries were performed to compare metadata with the optimized processing method (Table S8). The urea-based method developed by Parker et al [13] was replicated at UC Davis. This method uses 10 mg of hair shaft, about 100x the mass of a single hair, and yielded a large number of unique peptides (3990 ± 679 versus 2571 ± 318, $p = 0.03$) and GVPs (89 ± 5 versus 63 ± 8, $p = 0.01$) compared to the optimized chemistry. The second and third chemistries tested were two modified versions of the urea-based method, using stirring and different durations of incubation. These produced fewer unique and genetically variant peptides compared to the optimized protocol. Unique peptides were 1937 ± 44 and 2036 ± 98 and GVPs were 42 ± 3 and 36 ± 4 for the modified urea-based chemistries, whereas unique peptides were 2571 ± 318 and GVPs were 63 ± 8 for the optimized chemistry. The final chemistry tested brought the concentration of reducing agent up to 500 mM DTT to attempt to remedy the intractable hair samples. This strategy did not succeed, with only 503 ± 79 unique peptides and 10 ± 2 GVPs. The urea-based method did produce acceptable results for the intractable hair samples (4937 ± 911 unique peptides and 54 ± 12 GVPs), but this used 10 mg of hair

versus the 20 mm of hair for the 500 mM DTT samples in order to be consistent with the literature. Overall, random match probabilities ranged from 1 in 1 to 1 in 27 billion. Not all GVP profiles here have been validated genetically and therefore these RMP estimates are expected to become more conservative to account for false positives.

Identifying and validating GVPs

All GVPs that were detected as a part of this manuscript are reported in Tables S1-S3, S6, and S7. Table S1 reports genotypes for all identified GVPs for the six subjects used in this manuscript. These were used to validate detected GVPs. A table was also created to give examples of each GVP that was detected in the datasets (Table S6). Accumulating mass spectral data of each GVP can further help in identifying false positive data based on ion ratio and intensities. RSIDs were assembled from the GVP list provided and searched for in whole exome sequencing data. A comprehensive list of putative GVPs was compiled using Ensembl Biomart (Table S2). GVPs were included if they reside within genes from a compiled proteome, were above a global MAF of 0.5%, and were classified as missense SNPs. The list is theoretical and has not gone through additional filters, such as uniqueness of sequence or mass shift confirmation. Validated GVPs which have undergone filtering at both the mass spectral level and genetic level are listed in Table S7. This list provides basic chemical information on each GVP, such as sequence, precursor mass, charge state, and mass shifts as well as genetic information such as genotype frequency for different biogeographic populations. The final list that is included is a directory of all samples cited in this manuscript, with corresponding information such as protein and csv file name, raw file name, where they are found in the paper, and what conditions were used to process each sample (Table S3). This table can be used to help navigate file downloads from proteomeXchange (PDX016155).

Supplemental Figures

Table S1. **Genotypes for GVP-associated SNPs from whole exome sequencing**. Each box represents a genotype, and only genotypes of interest are included. D1 series donors are of African ancestry, and U1 series donors are of European ancestry. D1.0007, A1; D1.0017, A2; D1.0020, A3; U1.0001, E1, U1.0003, E2; U1.0005, E3.

Table S2. **Putative GVPs obtained from the human hair proteome.** Putative GVPs from common hair proteins of at least 0.5% MAF that have been characterized and annotated in Ensembl. Not all of these GVP-containing peptides have been identified or confirmed proteomically.

Table S3. **Sample master list.** This table includes all samples that were mentioned in this research. All samples used for this manuscript were compiled and uploaded to proteome exchange (PDX016155) and are available online.

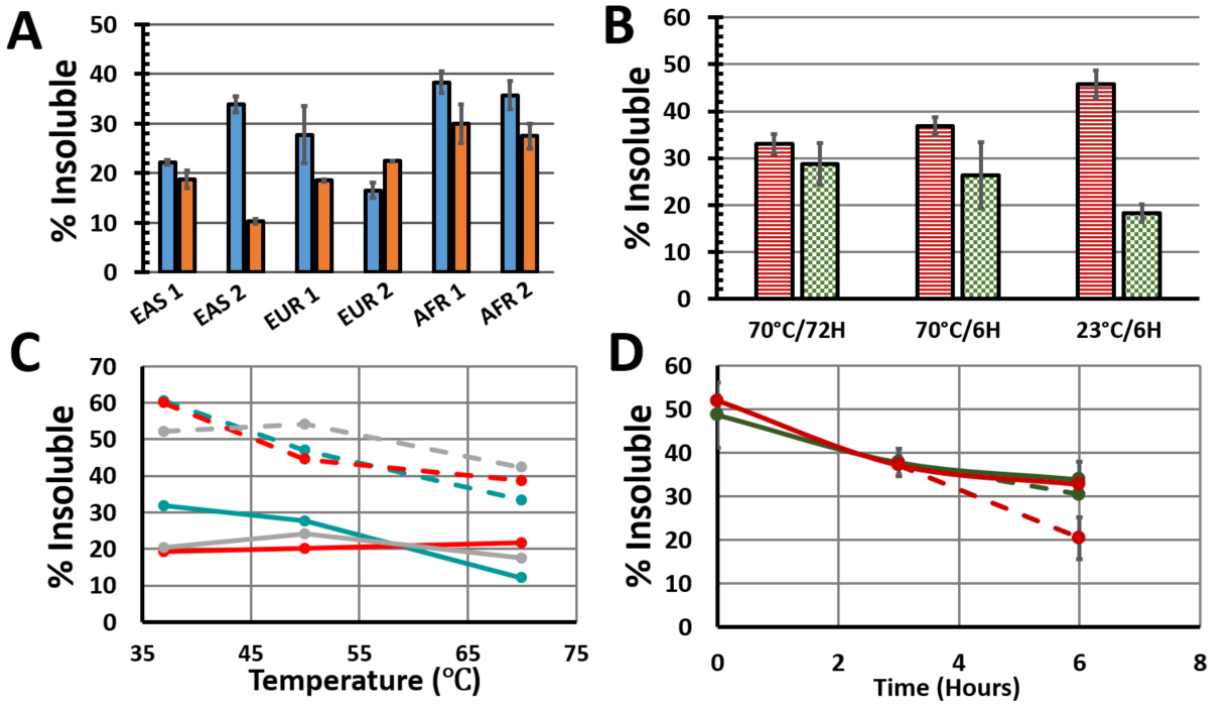


Figure S1. **Processing efficiency.** Shown are the % w/w of the original hair mass in insoluble pellets after trypsin digestion, based in reaction with ninhydrin reagent. **A)** Introducing agitation during reduction. Blue, stirred for 6 hrs; orange, held static for 18 hrs. **B)** Hair samples were processed using the optimized method (70°C/72H), heated at 70°C for 18 hrs with 6 hrs of digestion (70°C/6H), or processed at room temperature while stirring for 6 hrs along with 6 hrs of digestion (23°C/6H). Green, subject of European ancestry; red, subject of African ancestry. **C)** Effects of temperature and agitation combined. Solid lines, no agitation during disulfide reduction; dashed lines, mixing in a thermomixer during reduction. Teal lines, subject of East Asian ancestry; red lines, subject of European ancestry; grey lines, subject of European ancestry. **D)** Number of trypsin additions. One addition of 1:50 trypsin:protein was made at the 0 hr mark and another addition was made at the 3 hr mark for the samples with 2 additions. Dashed lines, 2-additions; green lines, European subject; red lines, African subject.

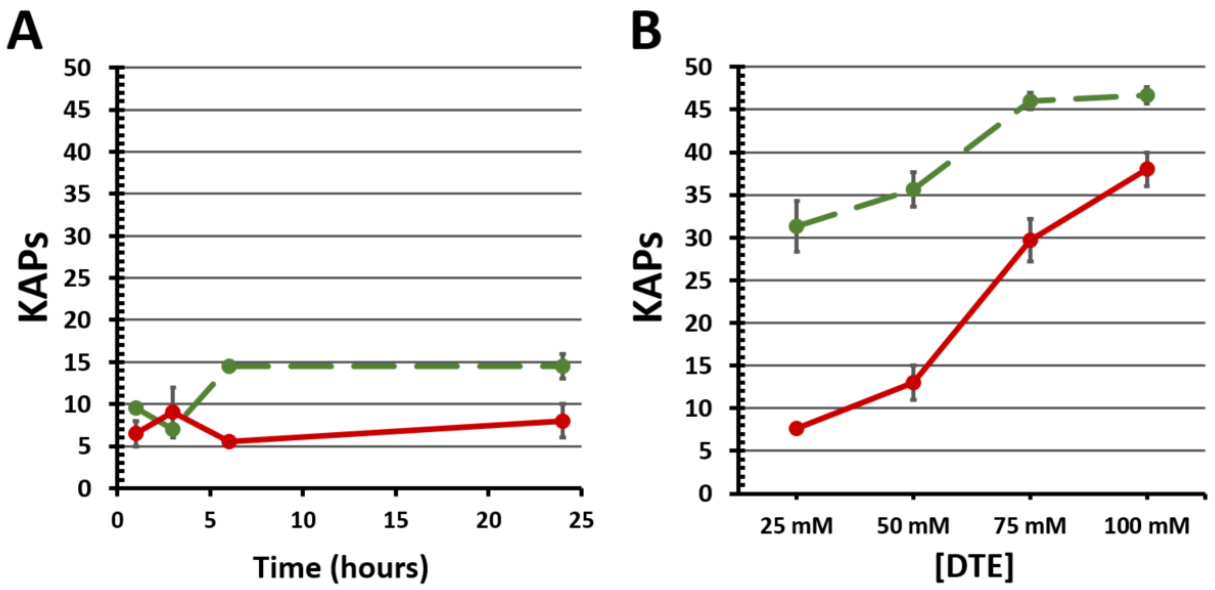


Figure S2. **Number of KAPs identified as a function of digestion time and DTE concentration.** Each data point represents an individual sample and not an aliquot from the same sample. 4mg of hair was processed. The dashed green line represents a European subject, and the solid red line represents an African subject. Samples were processed using the optimized processing method described in the methods section. **A)** A single addition of trypsin was made at 0 hrs, and samples were frozen at 3, 6, and 24 hrs. **B)** 25, 50, 75, and 100 mM of DTE was added to African and European hair.

Figure S3. **Detailed GVP matrix for varying concentrations of reducing agent.** This figure provides the data supporting figure 3. The matrix represents genetically variant peptides that have been verified via whole exome sequencing. Blue indicates a true positive result, green indicates a false negative result, white indicates a true negative result, and red indicates a false positive result. All columns are individual samples, and all GVP rows are included in this matrix with corresponding information included. Concentrations of DTE ranged from 25 – 100 mM. SAP, single amino acid polymorphism; RSID, reference SNP number; nuc, nucleotide.

Table S4. **Detailed protein coverage information for varying concentrations of reducing agent.** Protein name and protein coverage, which was corrected for unlikely peptides, were plotted and sorted by protein coverage for each sample. The first replicate of three was chosen to represent each reduction concentration. The sample is from one African donor.

Missing KAPs		
Family 4	Family 5	Family 9
34 - 37% Cys	29 - 36% Cys	31 - 33% Cys
KAP 4-2	KAP 5-1	KAP 9-1
KAP 4-5	KAP 5-2	KAP 9-2
KAP 4-6	KAP 5-3	KAP 9-4
KAP 4-7		KAP 9-6
KAP 4-9		KAP 9-7
KAP 4-11		KAP 9-8
KAP 4-12		
KAP 4-16		

Table S5. **KAPs missing from intractable hair digests**. Proteomes were compared between normal sample digests and intractable sample digests. Unique peptides from the listed proteins were not present in the intractable hair sample and are typically present in all other hair digests using the optimized processing protocol.

Figure S4. **Detailed GVP matrix comparing original and optimized processing methods for a single hair.** This figure provides data for figure 4. The matrix represents genetically variant peptides that have been verified via whole exome sequencing. Blue indicates a true positive result, green indicates a false negative result, white indicates a true negative result, and red indicates a false positive result. All columns are individual samples and all GVP rows are included in this matrix with corresponding information included. E represents the European cohort, A represents the African cohort, and B represents two blank samples with trypsin added.

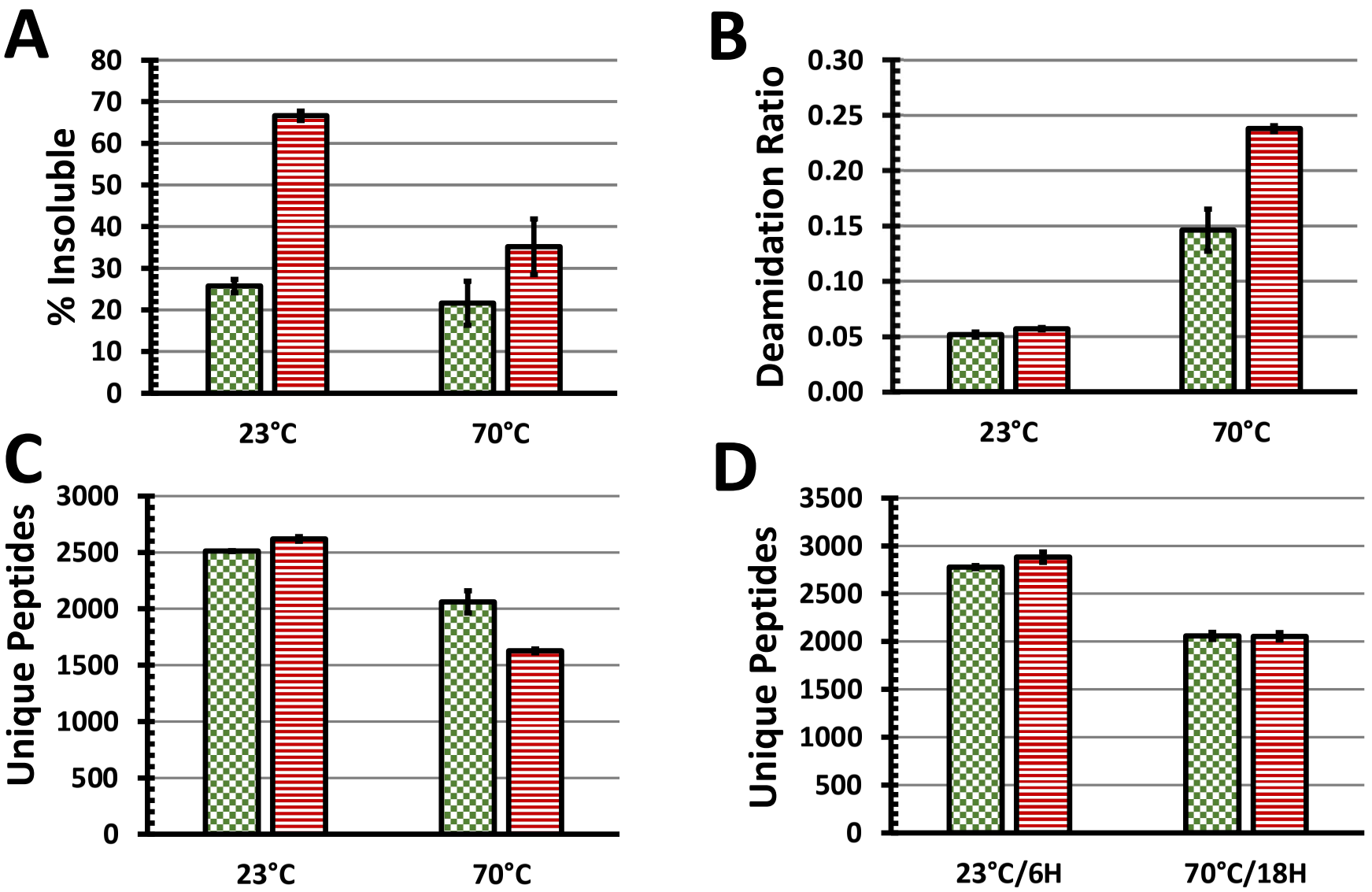
Table S6. **Samples that contain examples of validated GVPs.** GVPs found in this report have been validated through whole exome sequencing. GVPs that are not included in this list and are included in the detailed GVP matrix have been validated in other work not shown here. GVPs may be found in samples other than those indicated.

Table S7. **Validated GVP masses and modifications.** This table lists tryptic GVPs with corresponding genotypic frequencies, masses, charge state, and chemical modifications that have been discovered in all samples for this report. Red text indicates the location of the GVP, with lower-case lettering representing the minor allele, and capital lettering representing the major allele. Green text indicates that the variant SAP of interest is an R or K and that the tryptic sequence which is observed is from the downstream peptide after cleavage at the R or K of interest. GN, gene; rs#, reference SNP number; SAP, single amino acid polymorphism; gf, genotypic frequency; M+H, precursor mass; Z=, charge.

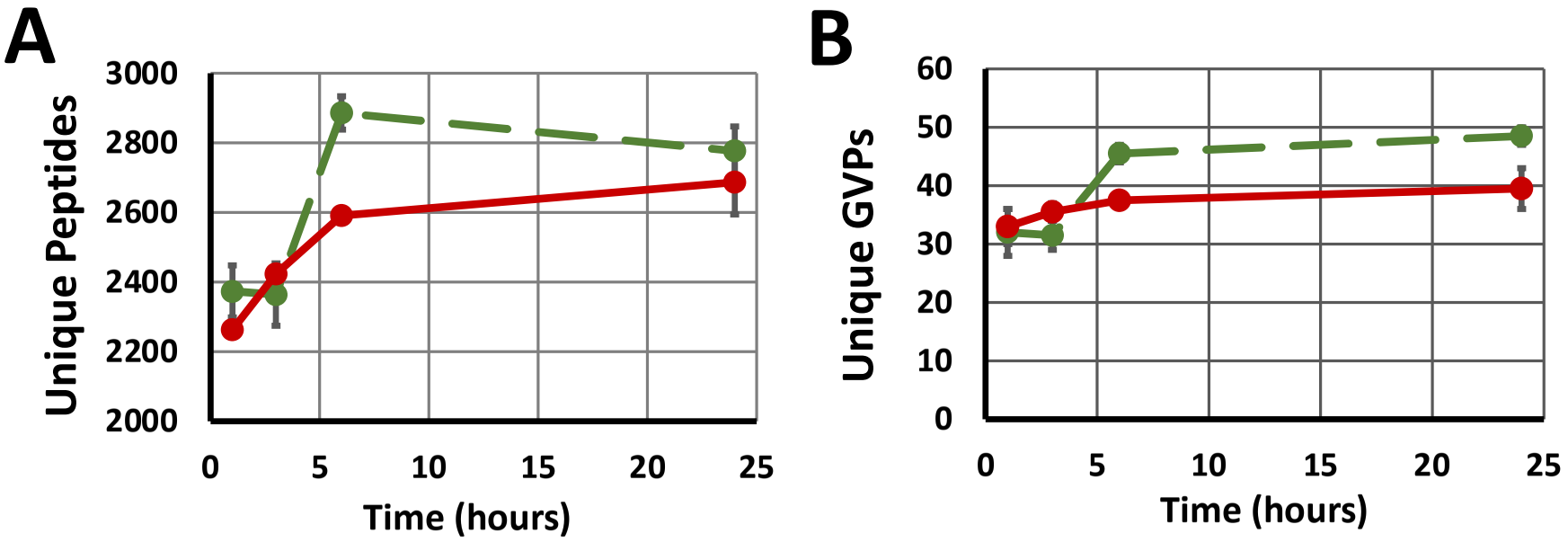
Sample Name	Sample ID	Method	Number of Unique Peptides	Number of GVPs	Random Match Probability
Urea-Based Method Results With 10 mg of Hair					
European U1.0003	ZG149	Urea-based	3504	87	2.27E-04
European U1.0003	ZG150	Urea-based	3558	84	3.13E-04
European U1.0003	ZG151	Urea-based	3367	83	5.00E-04
African D1.0017	ZG152	Urea-based	3944	92	1.00E-10
African D1.0017	ZG153	Urea-based	5122	96	3.70E-11
African D1.0017	ZG154	Urea-based	4445	91	6.00E-10
Blank with Trypsin	ZG155	Urea-based	150	0	1.00E+00
Comparing Three Scaled-Back Methods for Single Hair Analysis					
European U1.0003	ZG83	Optimized (1/4 vol)	2507	63	7.00E-10
European U1.0003	ZG84	Optimized (1/4 vol)	3034	73	1.69E-05
African D1.0017	ZG85	Optimized (1/4 vol)	2327	56	6.25E-08
African D1.0017	ZG86	Optimized (1/4 vol)	2415	58	1.41E-08
Blank with Trypsin	ZG87	Optimized (1/4 vol)	387	0	1.00E+00
European U1.0003	ZG88	Urea Mod. 1 (1/4 vol)	1989	43	1.35E-10
European U1.0003	ZG89	Urea Mod. 1 (1/4 vol)	1954	37	2.63E-05
African D1.0017	ZG90	Urea Mod. 1 (1/4 vol)	1887	43	1.20E-07
African D1.0017	ZG91	Urea Mod. 1 (1/4 vol)	1919	43	7.69E-11
Blank with Trypsin	ZG92	Urea Mod. 1 (1/4 vol)	296	0	1.00E+00
European U1.0003	ZG93	Urea Mod. 2 (1/4 vol)	2038	39	1.14E-08
European U1.0003	ZG94	Urea Mod. 2 (1/4 vol)	2091	37	1.10E-05
African D1.0017	ZG95	Urea Mod. 2 (1/4 vol)	1898	30	5.00E-07
African D1.0017	ZG96	Urea Mod. 2 (1/4 vol)	2117	37	2.50E-10
Blank with Trypsin	ZG97	Urea Mod. 2 (1/4 vol)	237	0	1.00E+00
Comparing Two Methods for Intractable Hair Processing					
African D1.0001	ZG156	Urea-based	5848	71*	1.10E-07
African D1.0007	ZG157	Urea-based	4026	47	4.76E-05
African D1.0017	ZG158	Urea-based	3536	63	2.86E-07
African D1.0001	ZG159	Optimized + 500 mM DTT	581	12*	2.70E-01
African D1.0007	ZG160	Optimized + 500 mM DTT	424	8	4.76E-01
African D1.0017	ZG161	Optimized + 500 mM DTT	2920	55	6.67E-05

Table S8. **Proteomic metadata for alternative processing chemistries**. 28 samples were run for three experiments to compare urea-based methods, single hair methods, and methods to address intractable hair samples. Random match probabilities ranged from 1 in 1 to 1 in 27 billion. Not all GVP profiles have been validated genetically.

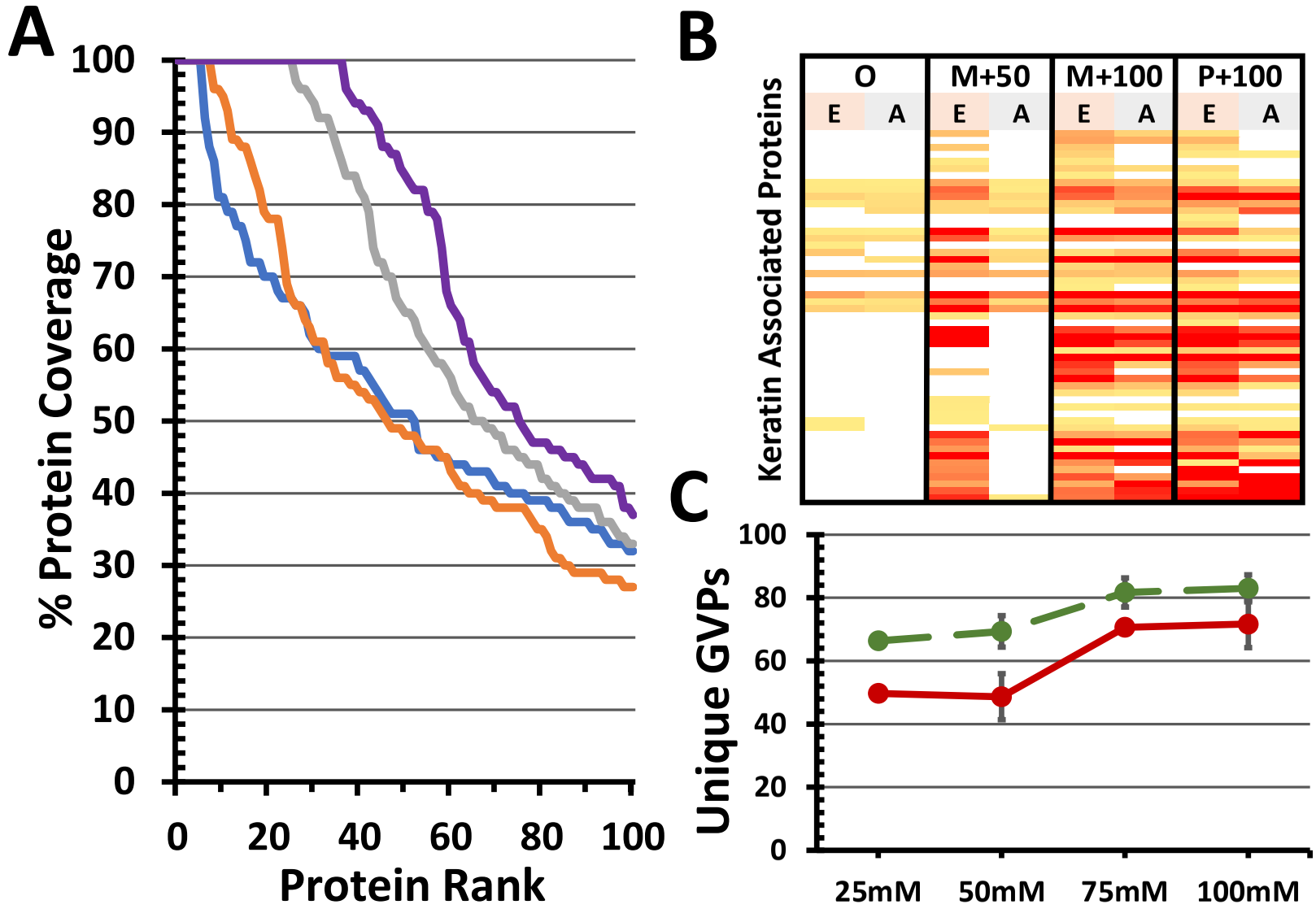
Figure


A
% Insoluble
80
70
60
50
40
30
20
10
0
23°C
70°C
B
Deamidation Ratio
0.30
0.25
0.20
0.15
0.10
0.05
0.00
23°C
70°C
C
Unique Peptides
3000
2500
2000
1500
1000
500
0
23°C
70°C
D
Unique Peptides
3500
3000
2500
2000
1500
1000
500
0
23°C/6H
70°C/18H

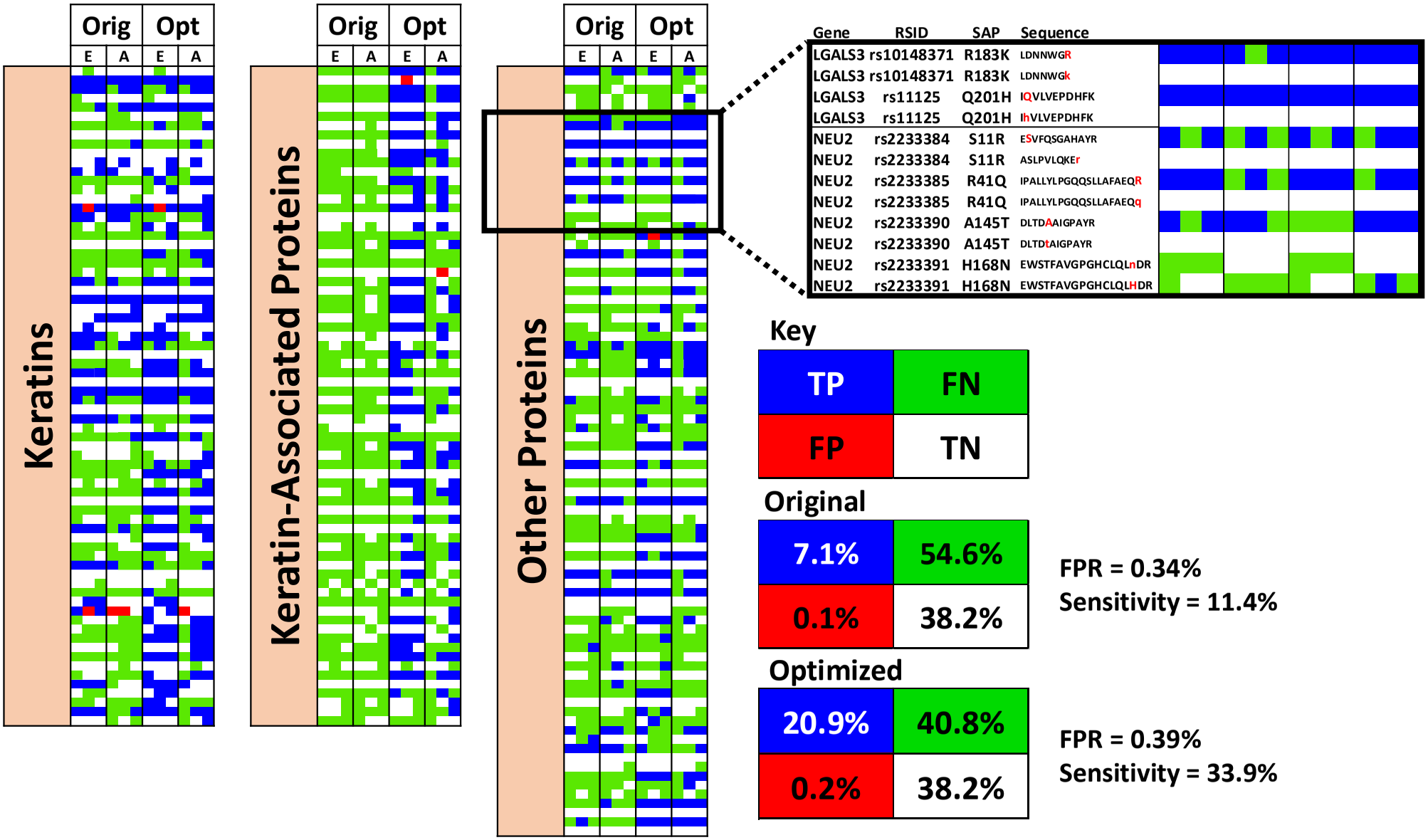
Figure


A
Unique Peptides
3000
2800
2600
2400
2200
2000
0
5
10
15
20
25
Time (hours)
B
Unique GVPs
60
50
40
30
20
10
0
0
5
10
15
20
25
Time (hours)

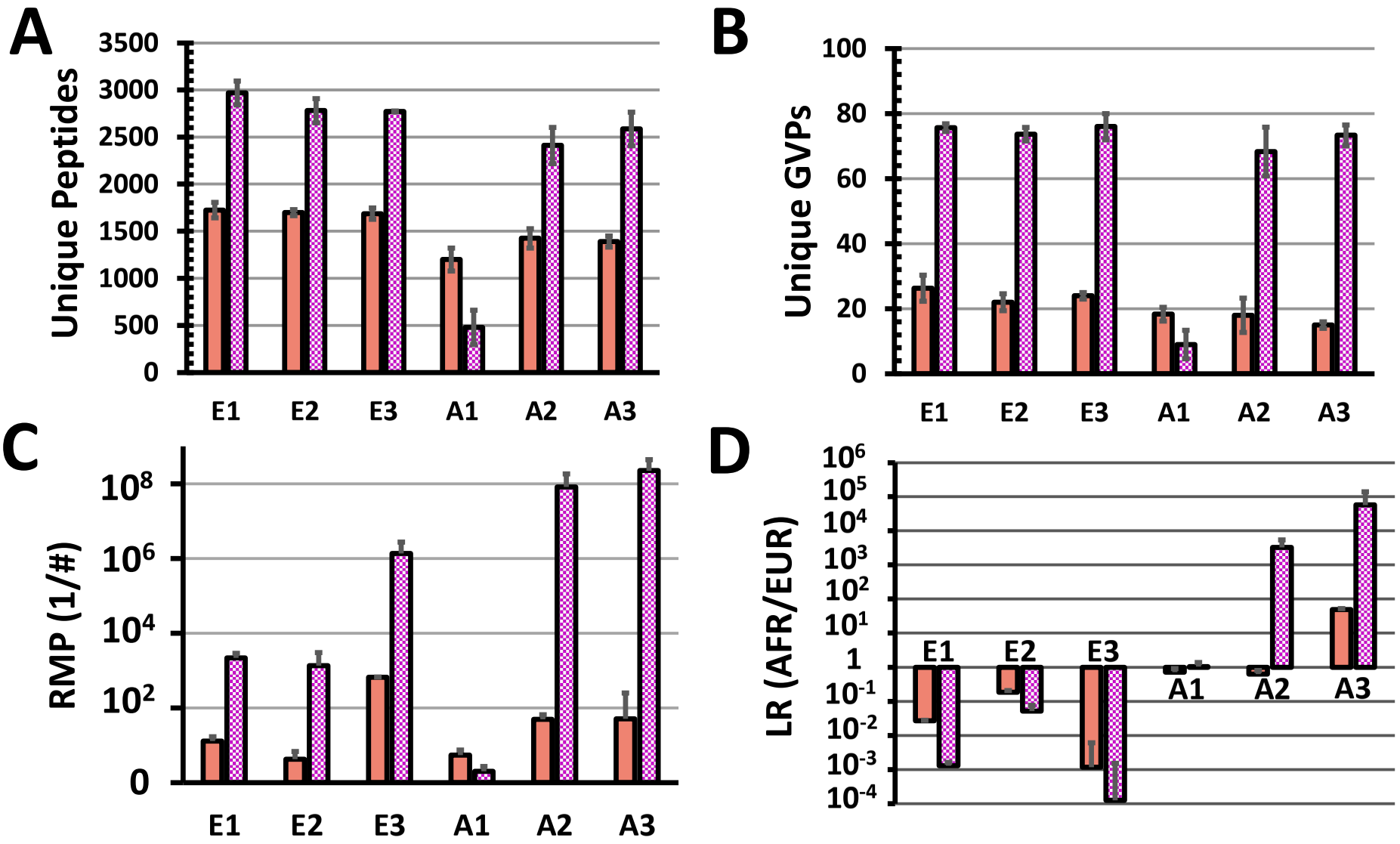
Figure


A
100
90
80
70
60
50
40
30
20
10
0
% Protein Coverage
0
20
40
60
80
100
Protein Rank
B
O
M+50
M+100
P+100
E
A
Keratin Associated Proteins
C
Unique GVPs
25mM
50mM
75mM
100mM

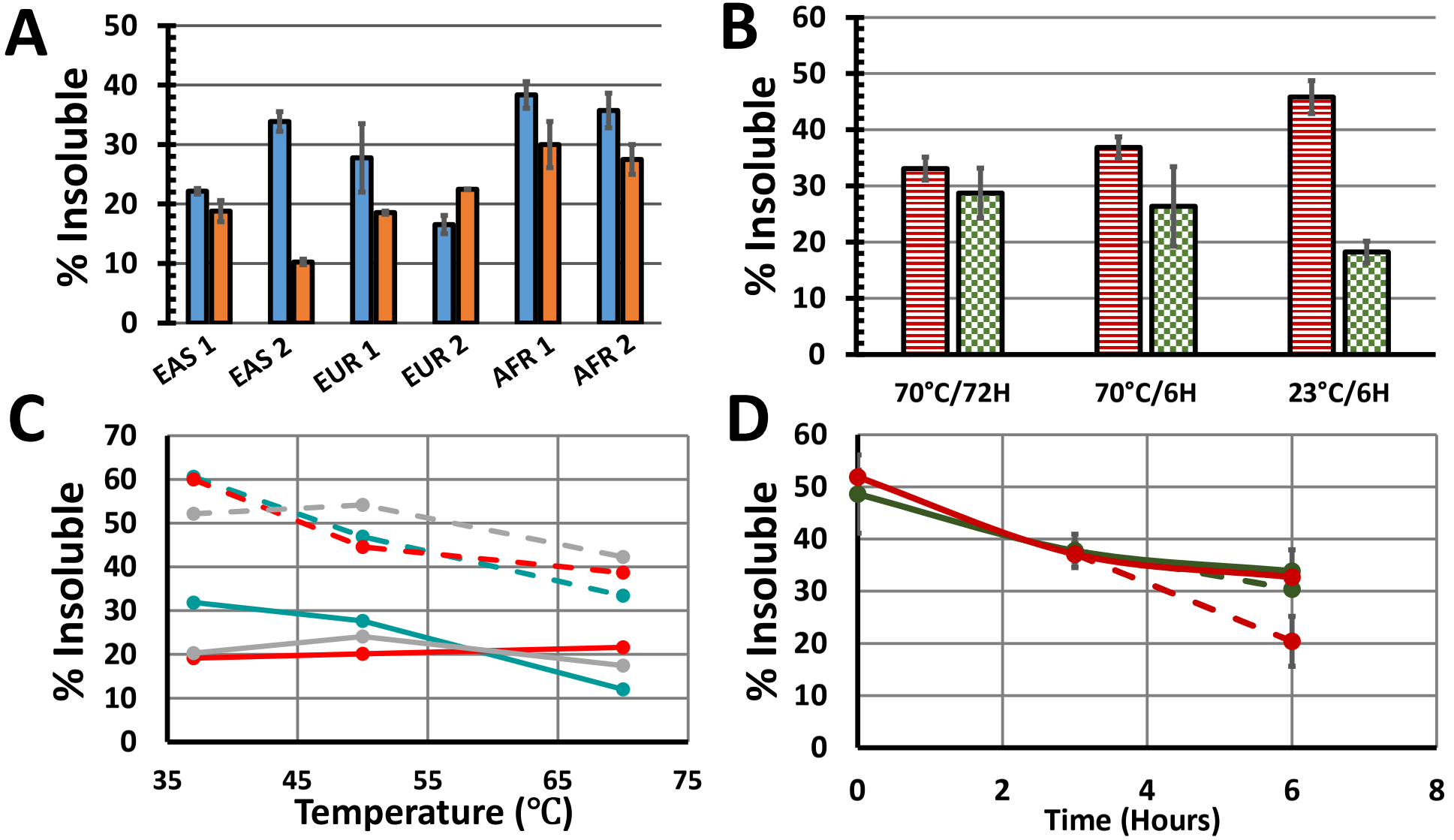
Figure


Orig
Opt
E
A
Keratins
Keratin-Associated Proteins
Other Proteins
Gene RSID SAP Sequence
LGALS3 rs10148371 R183K LDNNWGR
LGALS3 rs10148371 R183K LDNNWGk
LGALS3 rs11125 Q201H IQVLVEPDHFK
LGALS3 rs11125 Q201H IhVLVEPDHFK
NEU2 rs2233384 S11R ESVFQSGAHAYR
NEU2 rs2233384 S11R ASLPVLQKEr
NEU2 rs2233385 R41Q IPALLYLPGQQSLLAFAEQR
NEU2 rs2233385 R41Q IPALLYLPGQQSLLAFAEQq
NEU2 rs2233390 A145T DLTDAAIGPAYR
NEU2 rs2233390 A145T DLTDtAIGPAYR
NEU2 rs2233391 H168N EWSTFAVGPGHCLQLnDR
NEU2 rs2233391 H168N EWSTFAVGPGHCLQLHDR
Key
TP
FN
FP
TN
Original
7.1%
54.6%
0.1%
38.2%
FPR = 0.34%
Sensitivity = 11.4%
Optimized
20.9%
40.8%
0.2%
38.2%
FPR = 0.39%
Sensitivity = 33.9%

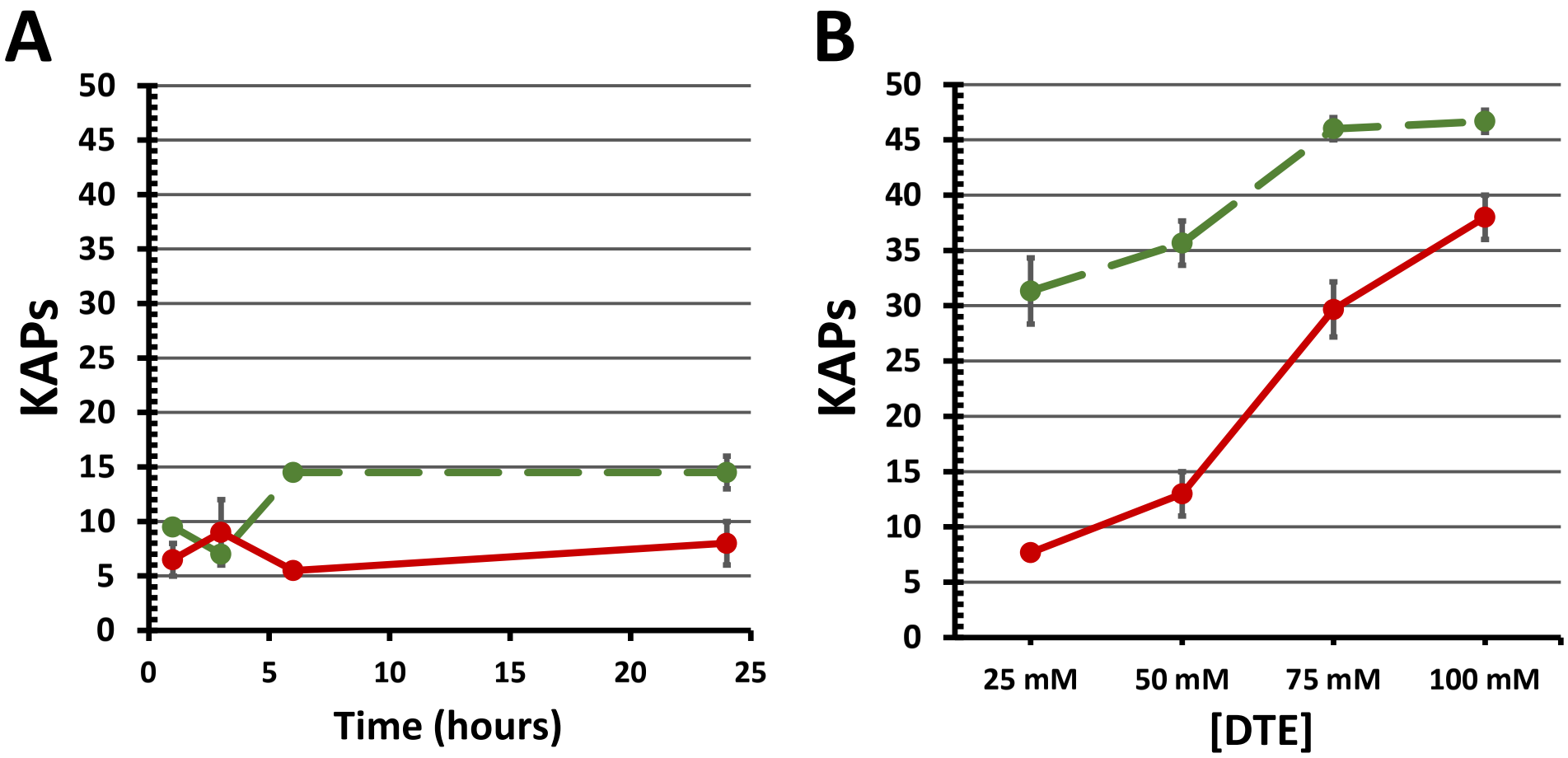
Figure


A
Unique Peptides
3500
3000
2500
2000
1500
1000
500
0
E1
E2
E3
A1
A2
A3
B
Unique GVPs
100
80
60
40
20
0
E1
E2
E3
A1
A2
A3
C
RMP (1/#)
10^8
10^6
10^4
10^2
0
E1
E2
E3
A1
A2
A3
D
LR (AFR/EUR)
10^6
10^5
10^4
10^3
10^2
10^1
1
10^{-1}
10^{-2}
10^{-3}
10^{-4}
E1
E2
E3
A1
A2
A3

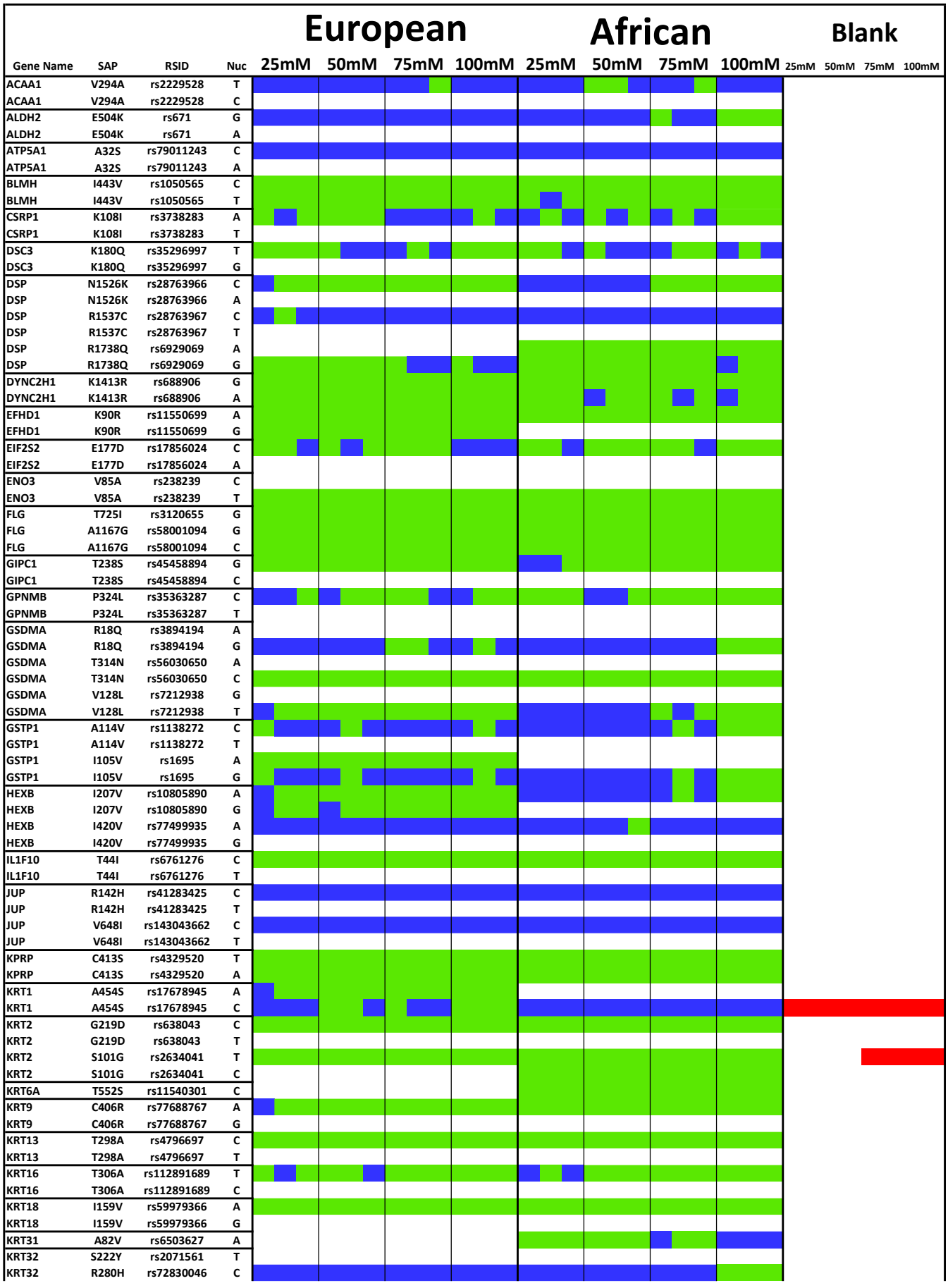
Figure


A
% Insoluble
0
10
20
30
40
50
EAS 1
EAS 2
EUR 1
EUR 2
AFR 1
AFR 2
B
% Insoluble
0
10
20
30
40
50
60
70°C/72H
70°C/6H
23°C/6H
C
% Insoluble
0
10
20
30
40
50
60
70
35
45
55
65
75
Temperature (°C)
D
% Insoluble
0
10
20
30
40
50
60
0
2
4
6
8
Time (Hours)

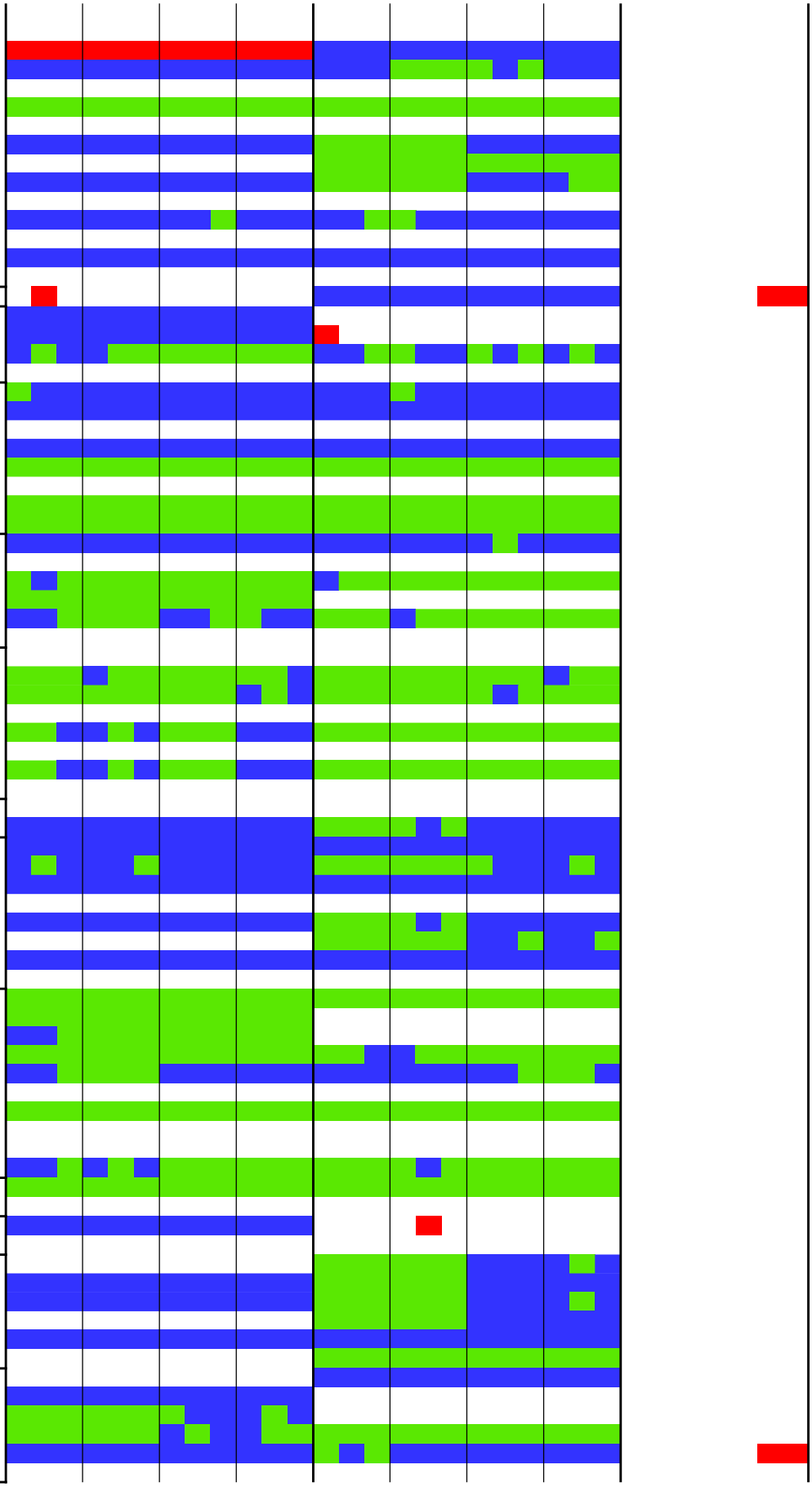
Figure


A
KAPs
50
45
40
35
30
25
20
15
10
5
0
0
5
10
15
20
25
Time (hours)
B
KAPs
50
45
40
35
30
25
20
15
10
5
0
25 mM
50 mM
75 mM
100 mM
[DTE]

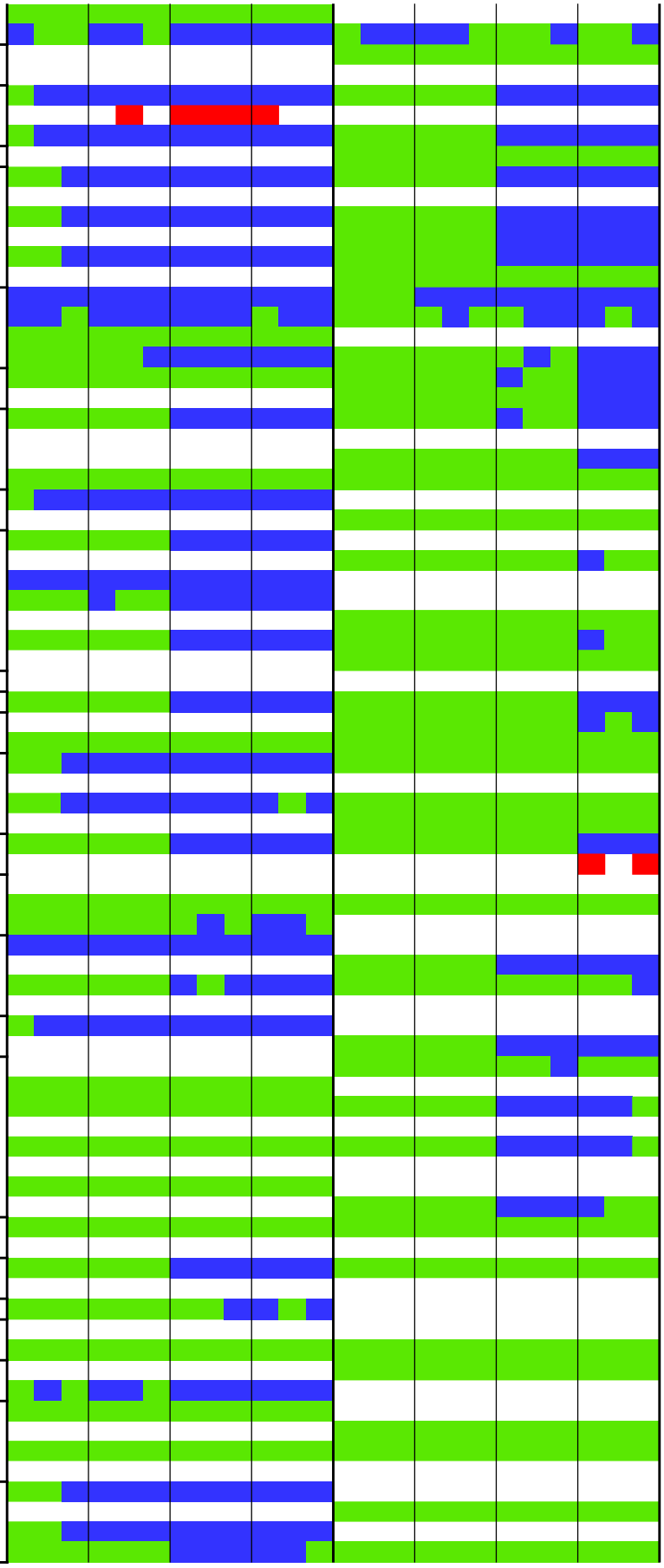
Figure


European
African
Blank
Gene Name
SAP
RSID
Nuc
25mM
50mM
75mM
100mM
25mM
50mM
75mM
100mM
25mM
50mM
75mM
100mM
ACAA1 V294A rs2229528 T
ACAA1 V294A rs2229528 C
ALDH2 E504K rs671 G
ALDH2 E504K rs671 A
ATP5A1 A32S rs79011243 C
ATP5A1 A32S rs79011243 A
BLMH I443V rs1050565 C
BLMH I443V rs1050565 T
CSRP1 K108I rs3738283 A
CSRP1 K108I rs3738283 T
DSC3 K180Q rs35296997 T
DSC3 K180Q rs35296997 G
DSP N1526K rs28763966 C
DSP N1526K rs28763966 A
DSP R1537C rs28763967 C
DSP R1537C rs28763967 T
DSP R1738Q rs6929069 A
DSP R1738Q rs6929069 G
DYNC2H1 K1413R rs688906 G
DYNC2H1 K1413R rs688906 A
EFHD1 K90R rs11550699 A
EFHD1 K90R rs11550699 G
EIF2S2 E177D rs17856024 C
EIF2S2 E177D rs17856024 A
ENO3 V85A rs238239 C
ENO3 V85A rs238239 T
FLG T725I rs3120655 G
FLG A1167G rs58001094 G
FLG A1167G rs58001094 C
GIPC1 T238S rs45458894 G
GIPC1 T238S rs45458894 C
GPNMB P324L rs35363287 C
GPNMB P324L rs35363287 T
GSDMA R18Q rs3894194 A
GSDMA R18Q rs3894194 G
GSDMA T314N rs56030650 A
GSDMA T314N rs56030650 C
GSDMA V128L rs7212938 G
GSDMA V128L rs7212938 T
GSTP1 A114V rs1138272 C
GSTP1 A114V rs1138272 T
GSTP1 I105V rs1695 A
GSTP1 I105V rs1695 G
HEXB I207V rs10805890 A
HEXB I207V rs10805890 G
HEXB I420V rs77499935 A
HEXB I420V rs77499935 G
IL1F10 T44I rs6761276 C
IL1F10 T44I rs6761276 T
JUP R142H rs41283425 C
JUP R142H rs41283425 T
JUP V648I rs143043662 C
JUP V648I rs143043662 T
KPRP C413S rs4329520 T
KPRP C413S rs4329520 A
KRT1 A454S rs17678945 A
KRT1 A454S rs17678945 C
KRT2 G219D rs638043 C
KRT2 G219D rs638043 T
KRT2 S101G rs2634041 T
KRT2 S101G rs2634041 C
KRT6A T552S rs11540301 C
KRT9 C406R rs77688767 A
KRT9 C406R rs77688767 G
KRT13 T298A rs4796697 C
KRT13 T298A rs4796697 T
KRT16 T306A rs112891689 T
KRT16 T306A rs112891689 C
KRT18 I159V rs59979366 A
KRT18 I159V rs59979366 G
KRT31 A82V rs6503627 A
KRT32 S222Y rs2071561 T
KRT32 R280H rs72830046 C

KRT32	R280H	rs72830046	T
KRT32	T339M	rs16966929	A
KRT32	T395M	rs2071563	G
KRT32	T395M	rs2071563	A
KRT32	N402S	rs2604955	T
KRT32	N402S	rs2604955	C
KRT32	P427T	rs2604953	G
KRT32	P427T	rs2604953	T
KRT32	D371E	rs2604956	C
KRT32	Q72R	rs3744786	T
KRT32	Q72R	rs3744786	C
KRT32	I171T	rs2071560	A
KRT32	I171T	rs2071560	G
KRT32	A255T	rs146792525	C
KRT32	A255T	rs146792525	T
KRT33A	A270V	rs12937519	A
KRT34	I238T	rs2239710	G
KRT34	H348R	rs2071599	A
KRT34	R334W	rs61740668	G
KRT34	R334W	rs61740668	A
KRT35	S36P	rs743686	A
KRT35	S36P	rs743686	G
KRT35	R163W	rs138303882	A
KRT35	R163W	rs138303882	G
KRT35	C441Y	rs12451652	C
KRT35	C441Y	rs12451652	T
KRT35	P413A	rs2071601	C
KRT35	P413A	rs2071601	G
KRT36	A202G	rs75790652	G
KRT36	A202G	rs75790652	C
KRT36	N357T	rs11657323	T
KRT36	N357T	rs11657323	G
KRT36	R277C	rs9904102	G
KRT36	R277C	rs9904102	A
KRT37	G13C	rs9910204	A
KRT37	G13C	rs9910204	C
KRT37	N39S	rs9916724	C
KRT37	N39S	rs9916724	A
KRT37	T72A	rs9916484	T
KRT37	T72A	rs9916484	C
KRT37	S73C	rs9916475	T
KRT37	S73C	rs9916475	A
KRT38	S423P	rs897416	A
KRT38	S423P	rs897416	G
KRT39	T341M	rs17843021	G
KRT39	T341M	rs17843021	A
KRT39	S86N	rs142154718	C
KRT39	S86N	rs142154718	T
KRT39	R456Q	rs7213256	C
KRT39	R456Q	rs7213256	T
KRT39	L383M	rs17843023	G
KRT39	L383M	rs17843023	T
KRT40	C349R	rs150812789	A
KRT40	C349R	rs150812789	G
KRT40	R235H	rs2010027	C
KRT40	R235H	rs2010027	T
KRT40	R108H	rs140634473	C
KRT40	R108H	s140634473	T
KRT40	S406L	rs16968862	G
KRT40	S406L	rs16968862	A
KRT40	C265Y	rs721957	C
KRT40	C265Y	rs721957	T
KRT75	E242G	rs2232393	A
KRT75	E242G	rs2232393	G
KRT81	G52R	rs2071588	G
KRT81	R248L	rs6580873	A
KRT82	T458M	rs2658658	A
KRT82	T458M	rs2658658	G
KRT82	E452D	rs1732263	C
KRT82	E452D	rs1732263	G
KRT82	E219Q	rs1791634	C
KRT82	E219Q	rs1791634	G
KRT83	I279M	rs2852464	C
KRT83	I279M	rs2852464	G
KRT83	H493Y	rs2857671	G
KRT83	H493Y	rs2857671	A
KRT83	R149C	rs2857663	G
KRT83	R149C	rs2857663	A

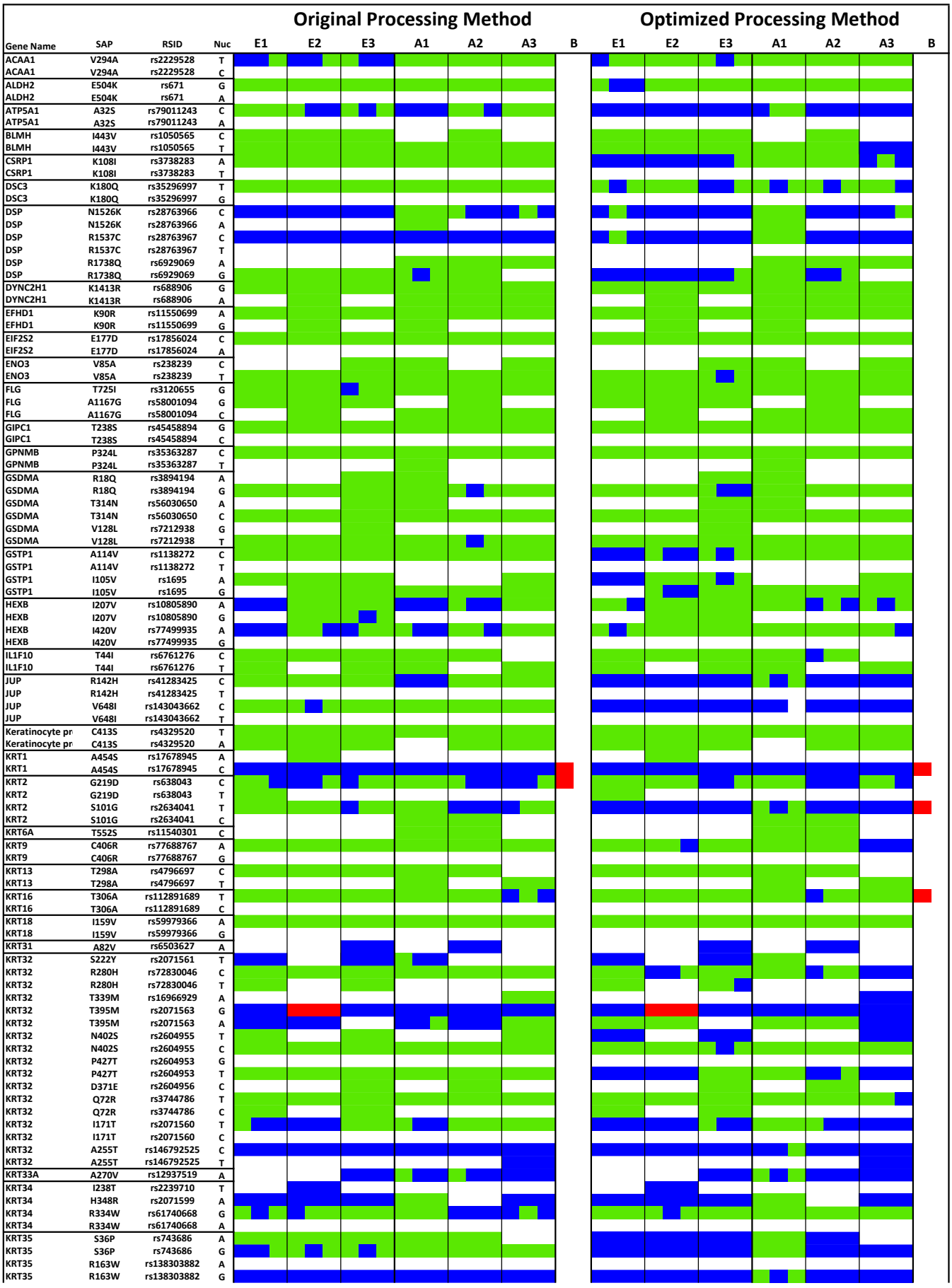


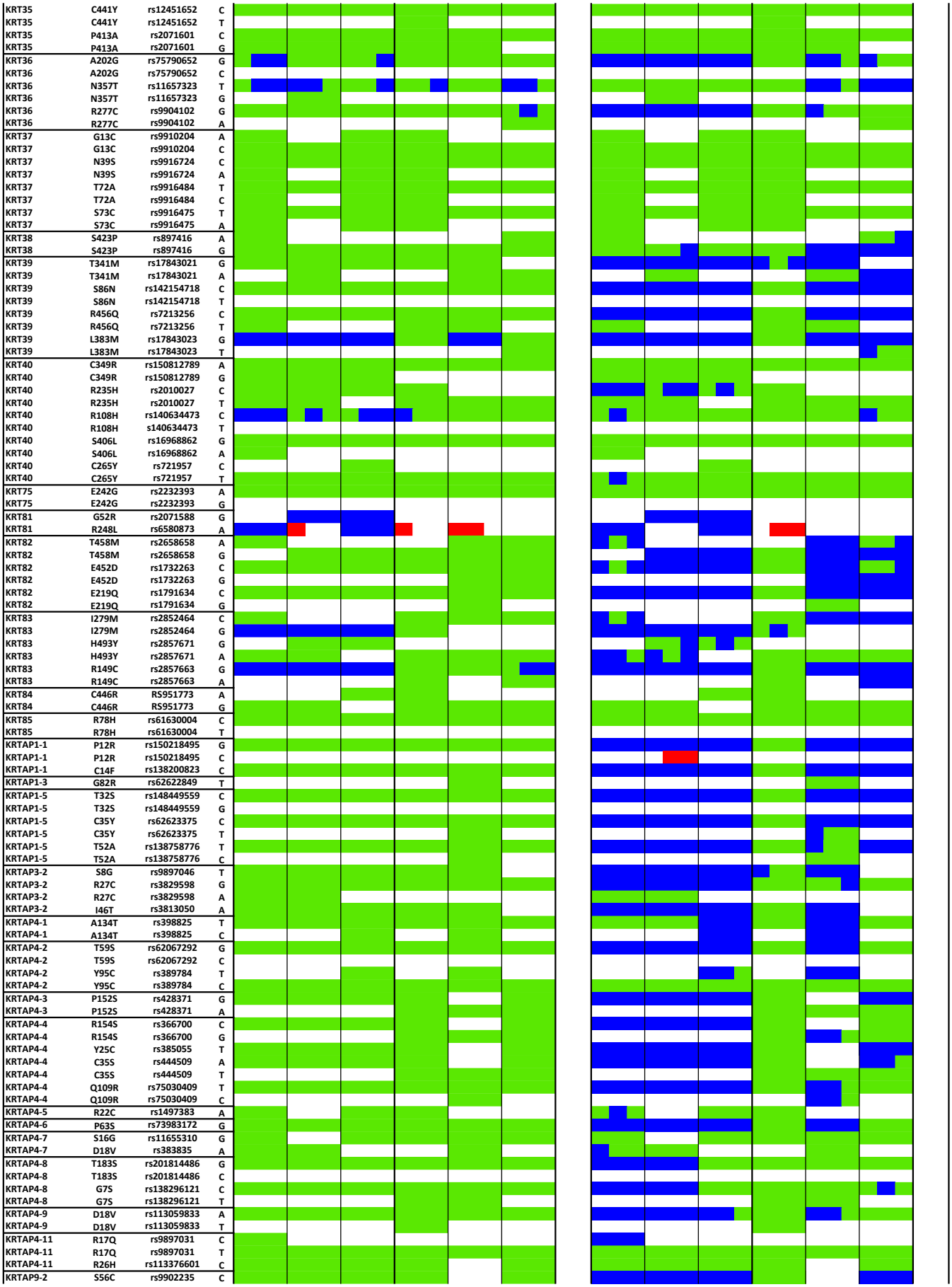
KRT84	C446R	RS951773	A
KRT84	C446R	RS951773	G
KRT85	R78H	rs61630004	C
KRT85	R78H	rs61630004	T
KRTAP1-1	P12R	rs150218495	G
KRTAP1-1	P12R	rs150218495	C
KRTAP1-1	C14F	rs138200823	C
KRTAP1-3	G82R	rs62622849	T
KRTAP1-5	T32S	rs148449559	C
KRTAP1-5	T32S	rs148449559	G
KRTAP1-5	C35Y	rs62623375	C
KRTAP1-5	C35Y	rs62623375	T
KRTAP1-5	T52A	rs138758776	T
KRTAP1-5	T52A	rs138758776	C
KRTAP3-2	S8G	rs9897046	T
KRTAP3-2	R27C	rs3829598	G
KRTAP3-2	R27C	rs3829598	A
KRTAP3-2	I46T	rs3813050	A
KRTAP4-1	A134T	rs398825	T
KRTAP4-1	A134T	rs398825	C
KRTAP4-2	T59S	rs62067292	G
KRTAP4-2	T59S	rs62067292	C
KRTAP4-2	Y95C	rs389784	T
KRTAP4-2	Y95C	rs389784	C
KRTAP4-3	P152S	rs428371	G
KRTAP4-3	P152S	rs428371	A
KRTAP4-4	R154S	rs366700	C
KRTAP4-4	R154S	rs366700	G
KRTAP4-4	Y25C	rs385055	T
KRTAP4-4	C35S	rs444509	A
KRTAP4-4	C35S	rs444509	T
KRTAP4-4	Q109R	rs75030409	T
KRTAP4-4	Q109R	rs75030409	C
KRTAP4-5	R22C	rs1497383	A
KRTAP4-6	P63S	rs73983172	G
KRTAP4-7	S16G	rs11655310	G
KRTAP4-7	D18V	rs383835	A
KRTAP4-8	T183S	rs201814486	G
KRTAP4-8	T183S	rs201814486	C
KRTAP4-8	G7S	rs138296121	C
KRTAP4-8	G7S	rs138296121	T
KRTAP4-9	D18V	rs113059833	A
KRTAP4-9	D18V	rs113059833	T
KRTAP4-11	R17Q	rs9897031	C
KRTAP4-11	R17Q	rs9897031	T
KRTAP4-11	R26H	rs113376601	C
KRTAP9-2	S56C	rs9902235	C
KRTAP9-2	S56C	rs9902235	G
KRTAP9-2	C149R	rs146532415	T
KRTAP9-2	C149R	rs146532415	C
KRTAP9-4	S146Y	rs2191379	A
KRTAP9-4	S146Y	rs2191379	C
KRTAP9-6	Y86C	rs12938692	A
KRTAP9-6	Y86C	rs12938692	G
KRTAP9-6	Y145C	rs576405629	A
KRTAP9-6	Y145C	rs576405629	G
KRTAP9-6	C146R	rs537301040	T
KRTAP9-6	C146R	rs537301040	C
KRTAP9-6	C51Y	rs12938374	G
KRTAP9-6	C51Y	rs12938374	A
KRTAP9-9	T25S	rs150962386	C
KRTAP9-9	T25S	rs150962386	G
KRTAP10-5	R268P	rs464391	C
KRTAP10-5	R268P	rs464391	G
KRTAP10-6	S300P	rs465279	A
KRTAP10-8	H26R	rs411254	A
KRTAP10-8	H26R	rs411254	G
KRTAP10-9	C257R	rs9980129	T
KRTAP10-9	C257R	rs9980129	C
KRTAP10-10	V158M	rs4818950	G
KRTAP10-10	V158M	rs4818950	A
KRTAP10-10	V206I	rs7280841	G
KRTAP10-10	V206I	rs7280841	A
KRTAP10-12	C236Y	rs61745911	G
KRTAP10-12	C236Y	rs61745911	A
KRTAP10-12	G226S	rs34302939	G
KRTAP10-12	G226S	rs34302939	A

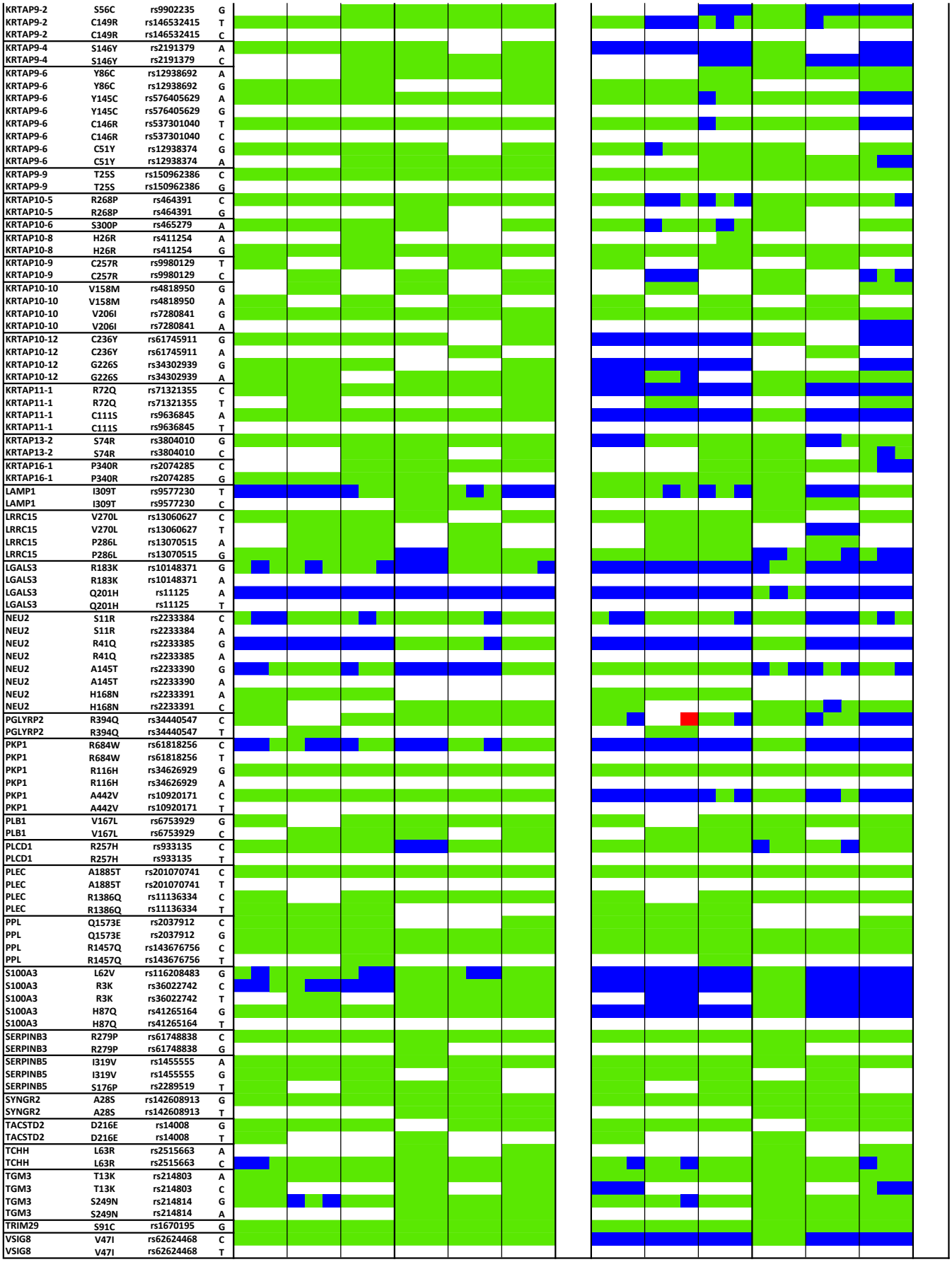


KRTAP11-1	R72Q	rs71321355	C
KRTAP11-1	R72Q	rs71321355	T
KRTAP11-1	C111S	rs9636845	A
KRTAP11-1	C111S	rs9636845	T
KRTAP13-2	S74R	rs3804010	G
KRTAP13-2	S74R	rs3804010	C
KRTAP16-1	P340R	rs2074285	C
KRTAP16-1	P340R	rs2074285	G
LAMP1	I309T	rs9577230	T
LAMP1	I309T	rs9577230	C
LRRC15	V270L	rs13060627	C
LRRC15	V270L	rs13060627	T
LRRC15	P286L	rs13070515	A
LRRC15	P286L	rs13070515	G
LGALS3	R183K	rs10148371	G
LGALS3	R183K	rs10148371	A
LGALS3	Q201H	rs11125	A
LGALS3	Q201H	rs11125	T
NEU2	S11R	rs2233384	C
NEU2	S11R	rs2233384	A
NEU2	R41Q	rs2233385	G
NEU2	R41Q	rs2233385	A
NEU2	A145T	rs2233390	G
NEU2	A145T	rs2233390	A
NEU2	H168N	rs2233391	A
NEU2	H168N	rs2233391	C
PGLYRP2	R394Q	rs34440547	C
PGLYRP2	R394Q	rs34440547	T
PKP1	R684W	rs61818256	C
PKP1	R684W	rs61818256	T
PKP1	R116H	rs34626929	G
PKP1	R116H	rs34626929	A
PKP1	A442V	rs10920171	C
PKP1	A442V	rs10920171	T
PLB1	V167L	rs6753929	G
PLB1	V167L	rs6753929	C
PLCD1	R257H	rs933135	C
PLCD1	R257H	rs933135	T
PLEC	A1885T	rs201070741	C
PLEC	A1885T	rs201070741	T
PLEC	R1386Q	rs11136334	C
PLEC	R1386Q	rs11136334	T
PPL	Q1573E	rs2037912	C
PPL	Q1573E	rs2037912	G
PPL	R1457Q	rs143676756	C
PPL	R1457Q	rs143676756	T
S100A3	L62V	rs116208483	G
S100A3	R3K	rs36022742	C
S100A3	R3K	rs36022742	T
S100A3	H87Q	rs41265164	G
S100A3	H87Q	rs41265164	T
SERPINB3	R279P	rs61748838	C
SERPINB3	R279P	rs61748838	G
SERPINB5	I319V	rs1455555	A
SERPINB5	I319V	rs1455555	G
SERPINB5	S176P	rs2289519	T
SYNGR2	A28S	rs142608913	G
SYNGR2	A28S	rs142608913	T
TACSTD2	D216E	rs14008	G
TACSTD2	D216E	rs14008	T
TCHH	L63R	rs2515663	A
TCHH	L63R	rs2515663	C
TGM3	T13K	rs214803	A
TGM3	T13K	rs214803	C
TGM3	S249N	rs214814	G
TGM3	S249N	rs214814	A
TRIM29	S91C	rs1670195	G
VSIG8	V47I	rs62624468	C
VSIG8	V47I	rs62624468	T

Figure


Original Processing Method
Optimized Processing Method
Gene Name
SAP
RSID
Nuc
E1
E2
E3
A1
A2
A3
B
ACAA1 V294A rs2229528 T
ACAA1 V294A rs2229528 C
ALDH2 E504K rs671 G
ALDH2 E504K rs671 A
ATP5A1 A32S rs79011243 C
ATP5A1 A32S rs79011243 A
BLMH I443V rs1050565 C
BLMH I443V rs1050565 T
CSRP1 K108I rs3738283 A
CSRP1 K108I rs3738283 T
DSC3 K180Q rs35296997 T
DSC3 K180Q rs35296997 G
DSP N1526K rs28763966 C
DSP N1526K rs28763966 A
DSP R1537C rs28763967 C
DSP R1537C rs28763967 T
DSP R1738Q rs6929069 A
DSP R1738Q rs6929069 G
DYNC2H1 K1413R rs688906 G
DYNC2H1 K1413R rs688906 A
EFHD1 K90R rs11550699 A
EFHD1 K90R rs11550699 G
EIF2S2 E177D rs17856024 C
EIF2S2 E177D rs17856024 A
ENO3 V85A rs238239 C
ENO3 V85A rs238239 T
FLG T725I rs3120655 G
FLG A1167G rs58001094 G
FLG A1167G rs58001094 C
GIPC1 T238S rs45458894 G
GIPC1 T238S rs45458894 C
GPNMB P324L rs35363287 C
GPNMB P324L rs35363287 T
GSDMA R18Q rs3894194 A
GSDMA R18Q rs3894194 G
GSDMA T314N rs56030650 A
GSDMA T314N rs56030650 C
GSDMA V128L rs7212938 G
GSDMA V128L rs7212938 T
GSTP1 A114V rs1138272 C
GSTP1 A114V rs1138272 T
GSTP1 I105V rs1695 A
GSTP1 I105V rs1695 G
HEXB I207V rs10805890 A
HEXB I207V rs10805890 G
HEXB I420V rs77499935 A
HEXB I420V rs77499935 G
IL1F10 T44I rs6761276 C
IL1F10 T44I rs6761276 T
JUP R142H rs41283425 C
JUP R142H rs41283425 T
JUP V648I rs143043662 C
JUP V648I rs143043662 T
Keratinocyte pr C413S rs4329520 T
Keratinocyte pr C413S rs4329520 A
KRT1 A454S rs17678945 A
KRT1 A454S rs17678945 C
KRT2 G219D rs638043 C
KRT2 G219D rs638043 T
KRT2 S101G rs2634041 T
KRT2 S101G rs2634041 C
KRT6A T552S rs11540301 C
KRT9 C406R rs77688767 A
KRT9 C406R rs77688767 G
KRT13 T298A rs4796697 C
KRT13 T298A rs4796697 T
KRT16 T306A rs112891689 T
KRT16 T306A rs112891689 C
KRT18 I159V rs59979366 A
KRT18 I159V rs59979366 G
KRT31 A82V rs6503627 A
KRT32 S222Y rs2071561 T
KRT32 R280H rs72830046 C
KRT32 R280H rs72830046 T
KRT32 T339M rs16966929 A
KRT32 T395M rs2071563 G
KRT32 T395M rs2071563 A
KRT32 N402S rs2604955 T
KRT32 N402S rs2604955 C
KRT32 P427T rs2604953 G
KRT32 P427T rs2604953 T
KRT32 D371E rs2604956 C
KRT32 Q72R rs3744786 T
KRT32 Q72R rs3744786 C
KRT32 I171T rs2071560 T
KRT32 I171T rs2071560 C
KRT32 A255T rs146792525 C
KRT32 A255T rs146792525 T
KRT33A A270V rs12937519 A
KRT34 I238T rs2239710 T
KRT34 H348R rs2071599 A
KRT34 R334W rs61740668 G
KRT34 R334W rs61740668 A
KRT35 S36P rs743686 A
KRT35 S36P rs743686 G
KRT35 R163W rs138303882 A
KRT35 R163W rs138303882 G


KRT35 C441Y rs12451652 C
KRT35 C441Y rs12451652 T
KRT35 P413A rs2071601 C
KRT35 P413A rs2071601 G
KRT36 A202G rs75790652 G
KRT36 A202G rs75790652 C
KRT36 N357T rs11657323 T
KRT36 N357T rs11657323 G
KRT36 R277C rs9904102 G
KRT36 R277C rs9904102 A
KRT37 G13C rs9910204 A
KRT37 G13C rs9910204 C
KRT37 N39S rs9916724 C
KRT37 N39S rs9916724 A
KRT37 T72A rs9916484 T
KRT37 T72A rs9916484 C
KRT37 S73C rs9916475 T
KRT37 S73C rs9916475 A
KRT38 S423P rs897416 A
KRT38 S423P rs897416 G
KRT39 T341M rs17843021 G
KRT39 T341M rs17843021 A
KRT39 S86N rs142154718 C
KRT39 S86N rs142154718 T
KRT39 R456Q rs7213256 C
KRT39 R456Q rs7213256 T
KRT39 L383M rs17843023 G
KRT39 L383M rs17843023 T
KRT40 C349R rs150812789 A
KRT40 C349R rs150812789 G
KRT40 R235H rs2010027 C
KRT40 R235H rs2010027 T
KRT40 R108H rs140634473 C
KRT40 R108H s140634473 T
KRT40 S406L rs16968862 G
KRT40 S406L rs16968862 A
KRT40 C265Y rs721957 C
KRT40 C265Y rs721957 T
KRT75 E242G rs2232393 A
KRT75 E242G rs2232393 G
KRT81 G52R rs2071588 G
KRT81 R248L rs6580873 A
KRT82 T458M rs2658658 A
KRT82 T458M rs2658658 G
KRT82 E452D rs1732263 C
KRT82 E452D rs1732263 G
KRT82 E219Q rs1791634 C
KRT82 E219Q rs1791634 G
KRT83 I279M rs2852464 C
KRT83 I279M rs2852464 G
KRT83 H493Y rs2857671 G
KRT83 H493Y rs2857671 A
KRT83 R149C rs2857663 G
KRT83 R149C rs2857663 A
KRT84 C446R RS951773 A
KRT84 C446R RS951773 G
KRT85 R78H rs61630004 C
KRT85 R78H rs61630004 T
KRTAP1-1 P12R rs150218495 G
KRTAP1-1 P12R rs150218495 C
KRTAP1-1 C14F rs138200823 C
KRTAP1-3 G82R rs62622849 T
KRTAP1-5 T32S rs148449559 C
KRTAP1-5 T32S rs148449559 G
KRTAP1-5 C35Y rs62623375 C
KRTAP1-5 C35Y rs62623375 T
KRTAP1-5 T52A rs138758776 T
KRTAP1-5 T52A rs138758776 C
KRTAP3-2 S8G rs9897046 T
KRTAP3-2 R27C rs3829598 G
KRTAP3-2 R27C rs3829598 A
KRTAP3-2 I46T rs3813050 A
KRTAP4-1 A134T rs398825 T
KRTAP4-1 A134T rs398825 C
KRTAP4-2 T59S rs62067292 G
KRTAP4-2 T59S rs62067292 C
KRTAP4-2 Y95C rs389784 T
KRTAP4-2 Y95C rs389784 C
KRTAP4-3 P152S rs428371 G
KRTAP4-3 P152S rs428371 A
KRTAP4-4 R154S rs366700 C
KRTAP4-4 R154S rs366700 G
KRTAP4-4 Y25C rs385055 T
KRTAP4-4 C35S rs444509 A
KRTAP4-4 C35S rs444509 T
KRTAP4-4 Q109R rs75030409 T
KRTAP4-4 Q109R rs75030409 C
KRTAP4-5 R22C rs1497383 A
KRTAP4-6 P63S rs73983172 G
KRTAP4-7 S16G rs11655310 G
KRTAP4-7 D18V rs383835 A
KRTAP4-8 T183S rs201814486 G
KRTAP4-8 T183S rs201814486 C
KRTAP4-8 G7S rs138296121 C
KRTAP4-8 G7S rs138296121 T
KRTAP4-9 D18V rs113059833 A
KRTAP4-9 D18V rs113059833 T
KRTAP4-11 R17Q rs9897031 C
KRTAP4-11 R17Q rs9897031 T
KRTAP4-11 R26H rs113376601 C
KRTAP9-2 S56C rs9902235 C



KRTAP9-2	S56C	rs9902235	G
KRTAP9-2	C149R	rs146532415	T
KRTAP9-2	C149R	rs146532415	C
KRTAP9-4	S146Y	rs2191379	A
KRTAP9-4	S146Y	rs2191379	C
KRTAP9-6	Y86C	rs12938692	A
KRTAP9-6	Y86C	rs12938692	G
KRTAP9-6	Y145C	rs576405629	A
KRTAP9-6	Y145C	rs576405629	G
KRTAP9-6	C146R	rs537301040	T
KRTAP9-6	C146R	rs537301040	C
KRTAP9-6	C51Y	rs12938374	G
KRTAP9-6	C51Y	rs12938374	A
KRTAP9-9	T25S	rs150962386	C
KRTAP9-9	T25S	rs150962386	G
KRTAP10-5	R268P	rs464391	C
KRTAP10-5	R268P	rs464391	G
KRTAP10-6	S300P	rs465279	A
KRTAP10-8	H26R	rs411254	A
KRTAP10-8	H26R	rs411254	G
KRTAP10-9	C257R	rs9980129	T
KRTAP10-9	C257R	rs9980129	C
KRTAP10-10	V158M	rs4818950	G
KRTAP10-10	V158M	rs4818950	A
KRTAP10-10	V206I	rs7280841	G
KRTAP10-10	V206I	rs7280841	A
KRTAP10-12	C236Y	rs61745911	G
KRTAP10-12	C236Y	rs61745911	A
KRTAP10-12	G226S	rs34302939	G
KRTAP10-12	G226S	rs34302939	A
KRTAP11-1	R72Q	rs71321355	C
KRTAP11-1	R72Q	rs71321355	T
KRTAP11-1	C111S	rs9636845	A
KRTAP11-1	C111S	rs9636845	T
KRTAP13-2	S74R	rs3804010	G
KRTAP13-2	S74R	rs3804010	C
KRTAP16-1	P340R	rs2074285	C
KRTAP16-1	P340R	rs2074285	G
LAMP1	I309T	rs9577230	T
LAMP1	I309T	rs9577230	C
LRRC15	V270L	rs13060627	C
LRRC15	V270L	rs13060627	T
LRRC15	P286L	rs13070515	A
LRRC15	P286L	rs13070515	G
LGALS3	R183K	rs10148371	G
LGALS3	R183K	rs10148371	A
LGALS3	Q201H	rs11125	A
LGALS3	Q201H	rs11125	T
NEU2	S11R	rs2233384	C
NEU2	S11R	rs2233384	A
NEU2	R41Q	rs2233385	G
NEU2	R41Q	rs2233385	A
NEU2	A145T	rs2233390	G
NEU2	A145T	rs2233390	A
NEU2	H168N	rs2233391	A
NEU2	H168N	rs2233391	C
PGLYRP2	R394Q	rs34440547	C
PGLYRP2	R394Q	rs34440547	T
PKP1	R684W	rs61818256	C
PKP1	R684W	rs61818256	T
PKP1	R116H	rs34626929	G
PKP1	R116H	rs34626929	A
PKP1	A442V	rs10920171	C
PKP1	A442V	rs10920171	T
PLB1	V167L	rs6753929	G
PLB1	V167L	rs6753929	C
PLCD1	R257H	rs933135	C
PLCD1	R257H	rs933135	T
PLEC	A1885T	rs201070741	C
PLEC	A1885T	rs201070741	T
PLEC	R1386Q	rs11136334	C
PLEC	R1386Q	rs11136334	T
PPL	Q1573E	rs2037912	C
PPL	Q1573E	rs2037912	G
PPL	R1457Q	rs143676756	C
PPL	R1457Q	rs143676756	T
S100A3	L62V	rs116208483	G
S100A3	R3K	rs36022742	C
S100A3	R3K	rs36022742	T
S100A3	H87Q	rs41265164	G
S100A3	H87Q	rs41265164	T
SERPINB3	R279P	rs61748838	C
SERPINB3	R279P	rs61748838	G
SERPINB5	I319V	rs1455555	A
SERPINB5	I319V	rs1455555	G
SERPINB5	S176P	rs2289519	T
SYNGR2	A28S	rs142608913	G
SYNGR2	A28S	rs142608913	T
TACSTD2	D216E	rs14008	G
TACSTD2	D216E	rs14008	T
TCHH	L63R	rs2515663	A
TCHH	L63R	rs2515663	C
TGM3	T13K	rs214803	A
TGM3	T13K	rs214803	C
TGM3	S249N	rs214814	G
TGM3	S249N	rs214814	A
TRIM29	S91C	rs1670195	G
VSIG8	V47I	rs62624468	C
VSIG8	V47I	rs62624468	T