Potential use of human hair shaft keratin peptide signatures to distinguish gender and ethnicity

Nurdiena Mohamed Nasir [1], **Jumriah Hiji** [2], **Jaime Jacqueline Jayapalan** [1, 2], **Onn Haji Hashim** [Corresp. 1, 2]

[1] Department of Molecular Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Wilayah Persekutuan, Malaysia

[2] University of Malaya Center for Proteomics Research, Faculty of Medicine, University of Malaya, Kuala Lumpur, Wilayah Persekutuan, Malaysia

Corresponding Author: Onn Haji Hashim
Email address: onnhashim@um.edu.my

Background. Most human hairs collected at old crime scenes do not contain nuclear DNA and are therefore of less value for forensic investigations. In the present study, hair shaft proteins were extracted from 40 healthy subjects between the ages of 21 to 40 years and profiled using gel electrophoresis-based proteomics to determine if they can be used to distinguish gender and ethnicity.

Methods. Extraction of the human hair shaft proteins was performed using a newly developed alkaline solubilisation method. The extracts were profiled by 2-dimensional electrophoresis and resolved protein spots were identified by mass spectrometry and queried against the human hair database. The study was then followed-up by immunoblotting of the identified hair shaft keratin of interest using commercially available antibodies.

Results. Separation of the human hair shaft proteins by 2-dimensional electrophoresis generated improved and highly resolved profiles. Comparing the hair shaft protein profiles of 10 female with 10 male subjects and their identification by mass spectrometry and query of the human hair database showed significant altered abundance of truncated/processed type-II keratin peptides K81 (2 spots), K83 (1 spot) and K86 (3 spots). The 2-dimensional electrophoresis profiling of 30 hair shaft samples taken from women of similar age range but from three distinctive ethnic subpopulations in Malaysia further showed significant altered abundance of one type-I and four type-II truncated/processed keratin peptides including K33b, K81, K83 and K86 (2 spots) between at least two of the ethnic groups. When a followed-up immunoblotting experiment was performed to detect the relative expression of the K86 peptides using commercialised antibodies, similar trends of expression were obtained. The present data, when taken together, demonstrated the potential use of keratin peptide signatures of the human hair shaft to distinguish gender and ethnicity although this needs to be further substantiated in a larger scale study.

Potential use of human hair shaft keratin peptide

signatures to distinguish gender and ethnicity

Nurdiena Mohamed Nasir[1], Jumriah Hiji[2], Jaime Jacqueline Jayapalan[1,2] and Onn Haji

Hashim[1,2,*]

[1] Department of Molecular Medicine, [2] University of Malaya Centre for Proteomics Research,

Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

Corresponding Author:

Onn Haji Hashim[*]

Department of Molecular Medicine, Faculty of Medicine, University of Malaya, 50603, Kuala

Lumpur, Malaysia

Email address: onnhashim@um.edu.my

Abstract

Background. Most human hairs collected at old crime scenes do not contain nuclear DNA and are therefore of less value for forensic investigations. In the present study, hair shaft proteins were extracted from 40 healthy subjects between the ages of 21 to 40 years and profiled using gel electrophoresis-based proteomics to determine if they can be used to distinguish gender and ethnicity.

Methods. Extraction of the human hair shaft proteins was performed using a newly developed alkaline solubilisation method. The extracts were profiled by 2-dimensional electrophoresis and resolved protein spots were identified by mass spectrometry and queried against the human hair database. The study was then followed-up by immunoblotting of the identified hair shaft keratin of interest using commercially available antibodies.

Results. Separation of the human hair shaft proteins by 2-dimensional electrophoresis generated improved and highly resolved profiles. Comparing the hair shaft protein profiles of 10 female with 10 male subjects and their identification by mass spectrometry and query of the human hair database showed significant altered abundance of truncated/processed type-II keratin peptides K81 (2 spots), K83 (1 spot) and K86 (3 spots). The 2-dimensional electrophoresis profiling of 30 hair shaft samples taken from women of similar age range but from three distinctive ethnic subpopulations in Malaysia further showed significant altered abundance of one type-I and four type-II truncated/processed keratin peptides including K33b, K81, K83 and K86 (2 spots) between at least two of the ethnic groups. When a followed-up immunoblotting experiment was performed to detect the relative expression of the K86 peptides using commercialised antibodies, similar trends of expression were obtained. The present data, when taken together, demonstrated

46   the potential use of keratin peptide signatures of the human hair shaft to distinguish gender and

47   ethnicity although this needs to be further substantiated in a larger scale study.

Introduction

49   The hair shaft is formed from epidermal keratinocytes undergoing a keratinization programme

50   cell death termed cornification. This process involves degradation and destruction of the cell

51   nucleus as well as the genetic material enclosed within it (*Eckhart et al., 2013*). As a result,

52   nuclear DNA which may be used for genetic fingerprinting is usually not detected in the hair

53   shaft (*Bender & Schneider, 2006*). However, the application of next generation sequencing for

54   the analysis of mitochondrial DNA, short tandem repeats, and single nucleotide polymorphisms

55   has also demonstrated a great promise for forensically challenging samples (*Shih et al., 2018*).

56   On the other hand, the potential of distinguishing individual hair donors based on their hair shaft

57   proteins has also been highlighted (*Wu et al., 2017*).

58   The complex interaction of these hair shaft proteins provides robust rigidity to the hair

59   structure and makes it resistant to many environmental factors such as pollutants, weather,

60   ultraviolet light or chemical treatments (*Wolfram,* 2003). Despite being well conserved, previous

61   studies have shown that characterization of the hair shaft proteins can be challenging and

62   difficult. This is mainly due to difficulties in solubilizing and extracting the proteins in solvents

63   that are compatible with liquid chromatography or gel electrophoresis approaches (*Sun et al.,*

64   *2014; Kollipara & Zahedi, 2013; Han et al., 2007; Smith & Parry, 2007; Langbein & Schweizer,*

65   *2005*). While many attempts to improve the yield of proteins extracted from the human hair shaft

66   have been reported, the quality of published 2-dimensional gel electrophoresis profiles is still far

67   from being a useful method for forensic investigations and the process is also time-consuming

68   and not practical to be applied in a large scale study (*Takayama & Ito, 2013; Barthelemy et al.,*

69   *2012; Thibaut et al., 2009; Plowman, 2007; Fuji, Nakamura et al., 2002*). However, an alkaline

70   solubilisation method which is capable of extracting substantially higher percentage of hair shaft

71   proteins within only 2 hrs was recently developed (*Wong et al., 2016*). With this improved

72   method and higher yield of hair shaft proteins, we analysed the hair material using 2-dimensional

73   electrophoresis gel-based proteomics as well as immunodetection methods and compared the

74   profiles to determine if they can be used to distinguish gender and/or ethnicity.

75

76   **Methodology**

77   **Subject recruitment**

78   A total of 40 healthy individuals of different genders and ethnicities but with the same age range

79   (21 to 40 years) were recruited in the study in accordance to the ethical clearance granted by the

80   Medical Ethics Committee of University of Malaya Medical Centre (MEC ID. NO: 20158-1577)

81   (Institutional Review Board), which adheres to the Declaration of Helsinki. Informed written

82   consent was acquired from all subjects prior to collection of their head hair samples. At least

83   three strands of hair were collected from each subject and stored at –20 °C. In addition to visual

84   inspection, information regarding subject's hair treatment habits and relevant past medical

85   histories, if any, were collected using a questionnaire. Relevant phenotypic characteristics of the

86   hair shaft including discoloration, whitening, bleaching, weathering and perming, if any, were

87   recorded. Subjects with previous history of bacterial or fungal-borne skin diseases, inflammation

88   or cancer and/or under treatment for such ailments and those with chemically-treated hair were

89   excluded from the study.

90

91 **Isolation of human hair shaft proteins**

92 Human hair shaft proteins were isolated using the alkaline solubilisation method in two stages as

93 previously described (*Wong et al., 2016; Budowle & Acton, 1981*). Briefly, 5 mg of hair samples

94 were sterilised with 90% ethanol, cut (1-4 mm) and incubated in 300 µl of lysis buffer at 90 ºC

95 for 30 mins in the first stage. The resulting supernatant fluids were isolated using a QIAquick

96 spin column (Qiagen, Hilden, Germany) and kept at 4ºC. The undissolved hair shaft fractions on

97 the other hand, were transferred into a fresh tube containing 300 µl of lysis buffer with 0.1 M

98 NaOH and pulverised by magnetic stirring for 30 mins to increase the recovery of protein from

99 the second stage of extraction. The resulting solutions containing hair shaft protein extract were

100 centrifuged and the supernatant fluids were separated for subsequent precipitation of proteins

101 using 4ºC acetone. Protein pellets were then solubilised in sample buffer and their

102 concentrations were assayed using the Bio-Rad protein assay (Bio-Rad, California, USA) in

103 accordance to the manufacturer's guidelines.

104

105 **2-dimensional electrophoresis and silver staining**

106 Solubilised hair shaft proteins (100 µg) suspended in sample buffer were incubated in

107 rehydration buffer containing 7 M urea, 2 M thiourea, 4% 3-[(3-cholamidopropyl)

108 dimethylammonio]-1-propanesulfonate hydrate, 2% immobilized pH gradient buffer pH 4-7,

109 0.002% w/w orange G and 7 mg of dithiothreitol (DTT) for 30 mins at room temperature. The

110 sample mixtures were rehydrated with 13 cm DryStrip gel pH 4-7 (GE Healthcare, Uppsala,

111 Sweden) for 18 hrs at room temperature in a closed environment. The rehydrated strips were

112 subjected to first dimensional gel electrophoresis separation using the Ettan IPGphor 3

113 Isoelectric Focusing System (GE Healthcare, Uppsala, Sweden) according to the following

114   settings: (i) 500 V, 1 hr, step and hold; (ii) 1000 V, 1 hr, gradient; (iii) 8000 V, 2 hrs 30 mins,

115   gradient; (iv) 8000 V, 55 mins, step and hold. The focused strip was initially incubated with

116   SDS equilibration buffer consisting of 6 M urea, 75 mM Tris-HCl, pH 8.8, 30% (v/v) glycerol,

117   2% sodium dodecyl sulphate (SDS) and 1% (w/v) DTT for 20 mins, followed by a second

118   equilibration in the same buffer but containing 4.5% (w/v) iodoacetamide instead of DTT for

119   another 20 mins. Second dimension separation was performed on 8-15% gradient

120   polyacrylamide gel using the SE 600 Ruby electrophoresis system (GE Healthcare, Uppsala,

121   Sweden). A concentration of 2.6% crosslinker (Bis) was used for the preparation of gels (*% C*).

122   Cathode running buffer containing 0.025 M tris (pH 8.3), 0.192 M glycine and 0.1% SDS while,

123   the anode buffer containing 0.375 M tris (pH 8.8) were used for the second dimensional gel

124   electrophoresis separation. Four gels were run according to the following program: (i) 50 V, 150

125   mA, 100 W for 1 hr; (ii) 600 V, 150 mA, 100 W until tracker dye reached bottom of the gel. The

126   electrophoresed gels were stained using silver nitrate (*Yan et al., 2000*). In this process, gels

127   were firstly fixed with 40% (v/v) ethanol and 10% acetic acid for 30 mins, followed by

128   sensitization with 40% (v/v) ethanol, 0.5 M sodium acetate trihydrate and 8 mM sodium

129   thiosulphate for 30 mins. Gels were then washed thrice, each time for 5 mins, and stained with

130   14.7 mM silver nitrate solution. After washing two more times for 1 min, gels were developed

131   with 0.24 M sodium carbonate and 0.04% (v/v) formaldehyde. Development of spots was

132   stopped with 40 mM EDTA solution and gels were finally kept in distilled water before being

133   scanned (see next section). Subsequent to the present study, a report on modification of this 2-

134   dimensional electrophoresis method, which generated further improved image resolution, was

135   published (*Wong, Hayashi & Hashim, 2019*).

136

137 **Data analysis**

138 Silver-stained 2-dimensional electrophoresis gels were scanned using ImageScanner III (GE

139 Healthcare, Uppsala, Sweden) and analysis of protein spot volume was performed using

140 ImageMaster Platinum 7.0 software (GE Healthcare, Uppsala, Sweden). Image analysis was

141 restricted to protein spot clusters that appeared consistently within each group of hair shaft

142 proteins. The levels of proteins in each sample were calculated as a percentage of volume

143 contribution (% vol) in which the volume of contribution refers to the volume percentage of a

144 protein taken against the total spot volume of all the proteins, in order to eliminate the possible

145 variations due to staining and/or protein loading. Automatic spot detection was performed with

146 default parameters setting (cut-off parameters were Smooth-2; Saliency-

147 1; Min area-5), whereas spot editing and removal of artefacts were done manually.

148   Data were analysed using the Statistical Package for Social Sciences (SPSS) version 25.0

149 (IBM Corporation, New York, USA). Shapiro-Wilk test of normality was used to assess the

150 distribution of the datasets. Parametric Student's t-test and ANOVA was used to analyse and

151 identify significant changes in expression between two population means (gender) and several

152 population groups (ethnicities), respectively, when the normality assumption was met ($p > 0.05$).

153 Following ANOVA, a pairwise comparison procedure was performed to test all possible

154 pairwise differences of the means using either Tukey HSD when equal variances was assumed

155 or, Dunnett's T3, if otherwise. On the other hand, non-parametric Mann-Whitney U and

156 Kruskal-Wallis tests with rank-based *post-hoc* test was used as their respective counterparts

157 when the assumption for normality was violated ($p < 0.05$). All values were presented as mean ±

158 95% confidence interval, unless otherwise stated. A p value of less than 0.05 and fold change of

159 more than 1.5-fold was considered significant.

160

161   **Mass spectrometry and database search**

162   Identification of proteins was performed as previously described with minor modifications

163   (*Seriramalu et al., 2010*). Briefly, protein spots of interests were carefully cut out from 2-

164   dimensional electrophoresis gels and kept in high-purity water at –20ºC. Gel plugs were first

165   destained using 15 mM potassium ferricyanide (III) and 50 mM sodium thiosulphate for 15 mins

166   at room temperature. The destaining procedure was repeated until the gel plugs became clear

167   and transparent. The proteins in gel plugs were then reduced and alkylated using 10 mM DTT

168   and 55 mM iodoacetamide both in 100 mM ammonium bicarbonate. They were then washed

169   thrice with 50% acetonitrile in 100 mM ammonium bicarbonate, dehydrated with 100%

170   acetonitrile and dried using vacuum centrifugation. The dried gels were treated with trypsin (6

171   µg/mL in 50 mM ammonium bicarbonate) for 18 hrs at 37ºC. The resulting peptides were then

172   dried, reconstituted in formic acid (0.1%) and desalted using ZipTip with C18 resin (Millipore,

173   Massachusetts, USA). The desalted and concentrated peptides were mixed with equal volume of

174   α-cyano-4-hydroxycinnamic acid (6 mg/ml), before being spotted onto the OptiToF 384-well

175   insert (0.7 µl) of the 5800 MALDI ToF/ToF analyser (SCIEX, Framingham, USA).

176   The proteins were identified using MASCOT search engine (*Perkins et al., 1999*) and the

177   resulting mass spectral data were thoroughly queried against the human hair entries in the

178   Uniprot database (last update: January 17, 2019, containing 1329 sequences). The following

179   parameters were set: enzyme: trypsin; maximum missed cleavages: 1; fixed modification:

180   carbamidomethylation of cysteine; variable modification oxidation of methionine; precursor ion

181   mass tolerance: 100 ppm; fragment ion mass tolerance: 0.2 Da. An individual ion score of more

182   than 17 indicates extensive homology or identity ($p < 0.05$).

183

184 Verification of altered protein abundance using immunoblotting

185 Immunoblotting was performed on pooled hair shaft protein samples of each groups of subjects,

186 to verify the altered abundance of proteins observed in the 2-dimensional electrophoresis

187 analysis. A total of 30 μg of proteins for each pooled sample, determined using Bio-Rad protein

188 assay (Bio-Rad, California, USA), was separated on 12.5% SDS polyacrylamide gel and

189 transferred onto PVDF membranes (0.2 μm, PALL Life Sciences, Port-Washington, USA) using

190 a Mini Trans-Blot® Electrophoretic Transfer Cell (Bio-Rad, Hercules, USA) for 1 h at 100 V.

191 Membranes were stained for total protein using MemCode™ Reversible Protein Stain Kit

192 (Thermo Fisher Scientific, Waltham, USA), dried and scanned to be used as a loading control.

193 After removal of Mem-Code™ staining, the membranes were blocked with 5% skim milk in 1×

194 TBST (25 mM Tris, 500 mM NaCl, 0.05% Tween, pH 7.5) for 30 mins, washed thrice, 5 mins

195 with 1× TBST, and incubated with anti-KRT86 antibody (ab192754) (Abcam, Cambridge, UK)

196 overnight at 4 °C. After washing with 1× TBST thrice for 5 mins each wash, the membranes

197 were incubated for 1 h with horseradish peroxidase conjugated goat anti-guinea pig IgG H&L

198 antibody (ab6908) (Abcam, Cambridge, UK). The membranes were washed again thrice, 5 mins

199 with 1×TBST and developed either using 3, 3′-diaminobenzidine chromogen solutions (Thermo

200 Fisher Scientific, Waltham, USA) containing 0.02% hydrogen peroxide or WesternBright Sirius

201 enhanced chemiluminescence system (Advansta, California, USA). MemCode™ signal and

202 protein band intensities were obtained using the ImageJ Software (NIH, Bethesda, USA).

203

204 **Results**

Hair shaft protein profiling - distinguishing genders and ethnicities

205

206 The extraction yields of protein (%) using the alkaline solubilisation method as previously

207 described by Wong et al. (*2016*) were generally higher compared to the methods of Lee et al.

208 (*2006*) and Shindai (*Fujii & Daguang, 2008*). When alkaline solubilisation protein extraction

209 method was applied to 5 mg of hair strands from 40 different individuals (30 females and 10

210 males), an average extraction yield of 63.56 ± 6.59 (mean $\pm$ SD) was obtained.

211 When the human hair shaft protein extracts from female and male subjects of the Malay

212 ethnic group were separated by 2-dimensional electrophoresis and subjected to silver staining,

213 similar profiles were obtained (*Fig.1*, panels A and B). The gel profiles were then subjected to

214 ImageMaster 2D Platinum Software and statistical analyses (see section: Data analysis).

215 Abundance of 6 hair shaft protein spots was found to be significantly higher in female compared

216 to male subjects (*Fig. 2*). Spot 3 showed the highest mean percentage of volume contribution for

217 females compared to males. The highest fold change difference was observed in spot 6 (*Fig. 2*).

218 When the same gel profiles of the female Malay subjects were compared to those

219 generated from subjects of Chinese and Indian ethnicities of the same age range, 5 protein spots

220 were significantly different in at least one ethnic group compared to the others (*Fig. 3*). Among

221 the protein spots of altered abundance, spot 7 appeared the most intense. However, spot 11

222 demonstrated the highest fold change differences between Indian (0.244 ± 0.028) and Chinese

223 (0.038 ± 0.079) subjects, as well as between Malay (0.147 ± 0.042) and Chinese (0.038 ± 0.027)

224 subjects. When taken together, the analysis generally showed that the Indian female subjects had

225 the highest mean percentage of volume contribution for spots 7, 9 and 11 compared to other

226 ethnicities (females) whilst their spots 8 and 10 were the least intense. The Chinese subjects

227   showed the highest mean percentage of volume contribution in spot 10 and lowest values for

228   spots 9 and 11. In the Malay female subjects, spot 8 had the highest mean percentage of volume

229   contribution, while spot 7 was the least intense.

230

231   Identification of hair shaft proteins by MALDI-ToF/ToF

232   The hair shaft protein spots of significant altered abundance between subjects of different

233   genders and ethnicities were identified using MALDI-ToF/ToF analysis and search of the human

234   hair database (Table 1). Analysis of the 11 hair shaft proteins of interest (*Fig. 1*) identified five

235   different types of keratins, including (1) type II cuticular Hb6 (K86) for spots 3, 4, 5, 7 and 8, (2)

236   type II cuticular Hb3 (K83) for spot 9, (3) type II cuticular Hb1 (K81) for spots 2, 6 and 10, (4)

237   type II cuticular Hb5 (K85) for spot 1 and (5) type I cuticular Ha3-II (K33B) for spot 11. While

238   the hair shaft of subjects of different genders demonstrated significant altered levels of K81, K85

239   and K86, those from different ethnicities showed significant different abundance of K33B, K81,

240   K83 and K86.

241

242   Verification of K86 abundance

243   In this study, K86 was selected for verification of its abundances in the hair shaft of the subjects

244   using immunoblotting method in view of its highest altered frequency as well as availability of

245   commercialised antibodies against the protein. However, since the anti-K86 used is known to

246   specifically recognise amino acids 446-469 of the keratin peptide, one of its peptide spots (spot

247   8), with amino acid sequences not in the region recognised by the antisera (Supplemental Table

248   1), was not included in the analysis. Figure 4 demonstrates the representative images generated

249   from the immunoblotting experiments. The results obtained for K86 spots (3, 4, 5 and 7)

250   appeared to corroborate the findings previously obtained via two-dimensional gel electrophoresis

251   separation. K86 bands equivalent to the molecular weights of spots 3, 4 and 5 (spots 4 and 5

252   share similar molecular weight) appeared intensely stained in the female Malay subjects, relative

253   to the male counterparts, while the female Indian subjects demonstrated more intense K86 bands

254   compared to the female Malay and Chinese subjects. When analysed by densitometry, the

255   intensity ratios for female Malay:male Malay bands for spots 3-, 4- and 5-equivalent K86 bands

256   were < 1.5-fold different, but significant ($p < 0.05$). On the other hand, the intensity ratios for

257   the K86 spot 7-equivalent band for female Malay versus Indian subjects showed > 1.5-fold

258   difference (FC = 1.62, $p = 0.008$). The levels of spot 7-equivalent bands seemed comparable

259   between the female Malay and Chinese ($p > 0.05$) subjects.

260

Discussion

262   Protein profiling of the human hair shaft have shown promising avenue for providing definitive

263   identification in aid of forensic investigations (*Wu et al., 2017*). Formed via cornification of

264   keratinocytes, the hair shaft mainly contains 17 different types of keratins, (i.e. K31-40 (including

265   K33a and K33b) and K81-86) (*Moll, Divo & Langbein, 2008*), which are poorly analysed mainly

266   because of their limited solubility and extraction yield. In the present study, we have adopted a

267   newly developed protocol that enhanced the extraction yield of proteins from the human hair shaft

268   (*Wong et al., 2016*) and reanalysed the material. The gel profiles that were generated showed

269   reasonably improved resolution of the separated hair shaft proteins compared to those that were

270   earlier reported (*Barthelemy et al., 2012; Thibaut et al., 2009; Nakamura et al., 2002*).

271   Identification of these proteins by mass spectrometry and search of the human hair protein database

272   showed that they comprised the different types of keratins, the cysteine-rich helicoidal proteins

273   that protect the hair because of their insolubility and impermeability.

274   In view of its reproducibility, the profiling of the human hair shaft proteins was further

275   utilised in a pilot study to determine if the developed profiles could be used to distinguish gender

276   of individuals. Our analysis of 20 hair shaft samples from healthy individuals between the age of

277   21 to 40 years demonstrated significant higher abundance of six different type-II keratin spots,

278   including two K81, one K85 and three K86, in the women subjects compared to men. These

279   different types of keratins are known to be restricted to the hair shaft and not present in the follicle

280   (*Moll, Divo & Langbein, 2008*). However, all these spots of altered abundance appeared to be

281   truncated or processed keratin products as they were resolved within the molecular weight regions

282   lower than their putative primary translated precursor polypeptides (Table 1). The influence or

283   effects of processing of proteins in this material remains to be elucidated in future investigations.

284   Similar marked shifts in molecular weights and isoelectric point (p*I*) values of the hair shaft

285   proteins that were separated by 2-dimensional electrophoresis, which may be due to deamidation,

286   post-translational modifications, or processing at the RNA level as sequence variants (*Person et*

287   *al., 2006*), have been previously reported (*Barthelemy et al., 2012*). Whether this was an effect of

288   the strong alkali used at 90°C in the hair shaft protein extraction, or that the proteins were further

289   biologically processed during their various stages of genetic expression, subsequent to their

290   translation or during cornification, remains to be investigated.

291   The gel electrophoresis profiling of 30 hair shaft samples taken from women of similar age

292   range but from three distinctive Malaysian ethnic subpopulations further showed significant

293   altered abundance of one type-I (K33b) and four type-II (K81, K83 and two K86) keratins between

294   the ethnic groups that were analysed. Like the earlier detected keratins, the type-I K33b and all

295   the type-II keratins detected are also known to be localised exclusively in the hair shaft (*Moll, Divo*

296   *& Langbein, 2008*). Based on the resolved experimental molecular weights, all the five spots of

297   altered abundance also appeared to be truncated or processed keratins. In addition, the K81 and

298   K86 spots that were also detected in this analysis were different from their counterparts that were

299   detected in the earlier gender analysis as they showed distinctive experimental molecular weight

300   and p*I* values.

301   The results of the latter study also demonstrated that the Indian ethnic group to be most

302   distinctive as they showed four abundantly different keratins (K33B, K81, K83 and K86 (spot 7))

303   compared to the Chinese and three keratins that were differently altered (K33B and two K86 (spots

304   7 and 8)) compared to the Malays. On the other hand, the Chinese and Malay ethnic groups only

305   appeared to be distinctive in the abundance of K33B (3.87-fold of difference) and their levels of

306   the type-I keratin were both significantly different compared to the Indians. These results are

307   generally comparable with the genetic data that were earlier reported. In a study using multi-

308   dimensional scale analysis on the population genetic structure of the different ethnic groups in

309   Peninsular Malaysia, Hatin et al. (*2011*) had previously reported that the Malay and Chinese

310   populations were clustered together while the Indians were further apart.

311   Data generated from the verification study using immunoblotting method generally

312   substantiated the abundances of the K86 truncated peptides in the hair shaft of subjects previously

313   obtained via two dimensional gel separation method, with little variation in the levels of

314   significance and/or intensity ratios (FC). The intensity of spots 3-, 4-, 5- and 7-equivalent K86

315   peptide bands appeared to corroborate the findings previously obtained via two-dimensional gel

316   electrophoresis separation. Together, these data further emphasise the potential of the keratin

317   peptide markers for distinguishing gender and ethnicity, and show promises for future forensic

318   applications.

319

Conclusion

321   When taken together, the human hair keratin profiling that was conducted in this pilot study

322   provided a potential method that can be used to distinguish gender and ethnicity of individuals

323   based on their hair shaft samples. However, a larger scale analysis of the hair shaft proteins using

324   antibodies that are specific to the different types of keratins that were highlighted in the present

325   study would be needed to increase the robustness of the results. This large scale study could

326   eventually lead to the development of a searchable database as well as signature keratin biomarkers

327   that could facilitate determination of one's gender and ethnicity based on their hair shaft keratin

328   profiles. Nonetheless, several other factors such as the effects of chemical exposure as well as

329   dietary and environmental influences on the hair shaft keratin profiles are also required for

330   confirmation of the accuracy of the results.

331

References

333   Barthélemy NR, Bednarczyk A, Schaeffer-Reiss C, Jullien D, Van Dorsselaer A, and Cavusoglu

334   N. 2012. Proteomic tools for the investigation of human hair structural proteins and evidence

335   of weakness sites on hair keratin coil segments. *Analytical Biochemistry* 421:43-55. DOI:

336   10.1016/j.ab.2011.10.011

337   Bender K, and Schneider PM. 2006. Validation and casework testing of the BioPlex-11 for STR

338   typing of telogen hair roots. *Forensic Science International* 161:52-59. DOI:

339   10.1016/j.forsciint.2005.10.024

340   Budowle B, and Acton RT. 1981. A technique for the detection of variable electrophoretic patterns

341   of hair proteins. *Electrophoresis* 2:333-334. DOI:10.1002/elps.1150020516

342   Eckhart L, Lippens S, Tschachler E, and Declercq W. 2013. Cell death by cornification.

343   *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1833:3471-3480. DOI:

344   10.1016/j.bbamcr.2013.06.010

345   Fujii T, Takayama S, and Ito Y. 2013. A novel purification procedure for keratin-associated

346   proteins and keratin from human hair. *Journal of Biological Macromolecules* 13.

347   Fujii T, and Daguang, L. 2008. Preparation and properties of protein films and particles from

348   chicken feather. *Journal of Biological Macromolecules* 8:48-55.

349   Han MO, Chun JA, Lee WH, Lee JW, and Chung CH. 2007. A simple improved method for protein

350   extraction from human head hairs. *Journal of Cosmetic Science* 58:527-534.

351   Hatin WI, Zahri M-K, Xu S, Jin L, Tan S-G, Rizman-Idid M, Zilfalil BA, and Consortium HP-AS.

352   2011. Population genetic structure of Peninsular Malaysia Malay sub-ethnic groups. *PloS*

353   *One* 6:e18312. DOI: 10.1371/journal.pone.0018312

354   Kollipara L, and Zahedi RP. 2013. Protein carbamylation: In vivo modification or in vitro artefact?

355   *Proteomics* 13:941-944. DOI: 10.1002/pmic.201200452

356   Langbein L, and Schweizer J. 2005. Keratins of the human hair follicle. *International Review of*

357   *Cytology* 243:1-78. DOI: 10.1016/S0074-7696(05)43001-6

358   Lee YJ, Rice RH, and Lee YM. 2006. Proteome analysis of human hair shaft: From protein
359   identification to posttranslational modification. *Molecular & Cellular Proteomics* 5:789-
360   800. DOI: 10.1074/mcp.M500278-MCP200

361   McNevin D, Wilson-Wilde L, Robertson J, Kyd J, and Lennard C. 2005. Short tandem repeat
362   (STR) genotyping of keratinised hair - Part 1. Review of current status and knowledge gaps.
363   *Forensic Science International* 153:237-246. DOI: 10.1016/j.forsciint.2005.05. 006

364   Moll R, Divo M, and Langbein L. 2008. The human keratins: biology and pathology.
365   *Histochemistry and Cell Biology* 129:705. DOI: 10.1007/s00418-008-0435-6

366   Nakamura A, Arimoto M, Takeuchi K, and Fujii T. 2002. A rapid extraction procedure of human
367   hair proteins and identification of phosphorylated species. *Biological and Pharmaceutical*
368   *Bulletin* 25:569-572. DOI: 10.1248/Bpb.25.569

369   Perkins DN, Pappin DJ, Creasy DM, and Cottrell JS. 1999. Probability-based protein identification
370   by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551-
371   3567. DOI: 10.1002/(SICI)1522-2683(19991201)20:18 <3551::AID-ELPS3551>3.0.CO;2-
372   2

373   Person MD, Shen J, Traner A, Hensley SC, Lo H-H, Abbruzzese JL, and Li D. 2006. Protein
374   fragment domains identified using 2D gel electrophoresis/MALDI-TOF. *Journal of*
375   *Biomolecular Techniques* 17:145.

376   Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, and Forster P.
377   1999. Expanding the forensic German mitochondrial DNA control region database: genetic
378   diversity as a function of sample size and microgeography. *International Journal of Legal*
379   *Medicine* 112:291-298. DOI 10.1007/s004140050252

380 Plowman JE. 2007. The proteomics of keratin proteins. *Journal of Chromatography B* 849:181-
381   189. DOI: 10.1016/j.jchromb.2006.07.055

382 Seriramalu R, Pang WW, Jayapalan JJ, Mohamed E, Abdul-Rahman PS, Bustam AZ, Khoo AS,
383   and Hashim OH. 2010. Application of champedak mannose-binding lectin in the
384   glycoproteomic profiling of serum samples unmasks reduced expression of alpha-2
385   macroglobulin and complement factor B in patients with nasopharyngeal carcinoma.
386   *Electrophoresis* 31:2388-2395. DOI: 10.1002/elps.201000164

387 Shih S, Bose N, Gonçalves A, Erlich H, and Calloway C. 2018. Applications of probe capture
388   enrichment next generation sequencing for whole mitochondrial genome and 426 nuclear
389   SNPs for forensically challenging samples. *Genes* 9:49. DOI:10.3390/genes9010049

390 Smith TA, and Parry DAD. 2007. Sequence analyses of Type I and Type II chains in human hair
391   and epithelial keratin intermediate filaments: Promiscuous obligate heterodimers, Type II
392   template for molecule formation and a rationale for heterodirner formation. *Journal of*
393   *Structural Biology* 158:344-357. DOI: 10.1016/j.jsb.2006.12.002

394 Sun SS, Zhou JY, Yang WM, and Zhang H. 2014. Inhibition of protein carbamylation in urea
395   solution using ammonium-containing buffers. *Analytical Biochemistry* 446:76-81. DOI:
396   10.1016/j.ab.2013.10.024

397 Thibaut S, Cavusoglu N, de Becker E, Zerbib F, Bednarczyk A, Schaeffer C, van Dorsselaer A,
398   and Bernard BA. 2009. Transglutaminase-3 Enzyme: A Putative Actor in Human Hair Shaft
399   Scaffolding? *Journal of Investigative Dermatology* 129:449-459. DOI:
400   10.1038/jid.2008.231

401 Wong SY, Hashim OH, and Hayashi N. 2019. Development of high-performance two-dimensional
402   gel electrophoresis for human hair shaft proteome. *PloS One* 14:e0213947.

403   Wong SY, Lee CC, Ashrafzadeh A, Junit SM, Abrahim N, and Hashim OH. 2016. A high-yield

404   two-hour protocol for extraction of human hair shaft proteins. *PloS One* 11:e0164993. DOI:

405   10.1371/journal.pone.0164993

406   Wu P-W, Mason KE, Durbin-Johnson BP, Salemi M, Phinney BS, Rocke DM, Parker GJ and Rice

407   RH. 2017. Proteomic analysis of hair shafts from monozygotic twins: Expression profiles

408   and genetically variant peptides. *Proteomics* 17:1600462. DOI: 10.1002/pmic.201600462

409   Yan JX, Wait R, Berkelman T, Harry RA, Westbrook JA, Wheeler CH, and Dunn MJ. 2000. A

410   modified silver staining protocol for visualization of proteins compatible with

411   matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry.

412   *Electrophoresis* 21:3666-3672.

Figure 1

Figure 1. Representative hair shaft protein profiles.

In panel A, circled are 11 protein spots that were significantly different in abundance between gender [male (*n*=10) and female (*n*=10) subjects of the same ethnicity (Malaysian Malay)] (spots 1-6) and three different ethnicities [Malaysian Malay (*n*=10), Chinese (*n*=10) and Indian (*n*=10) female subjects] (spots 7-11). Panels B, C and D demonstrate representative hair shaft protein profiles of male Malay, female Chinese and Indian subjects, respectively.
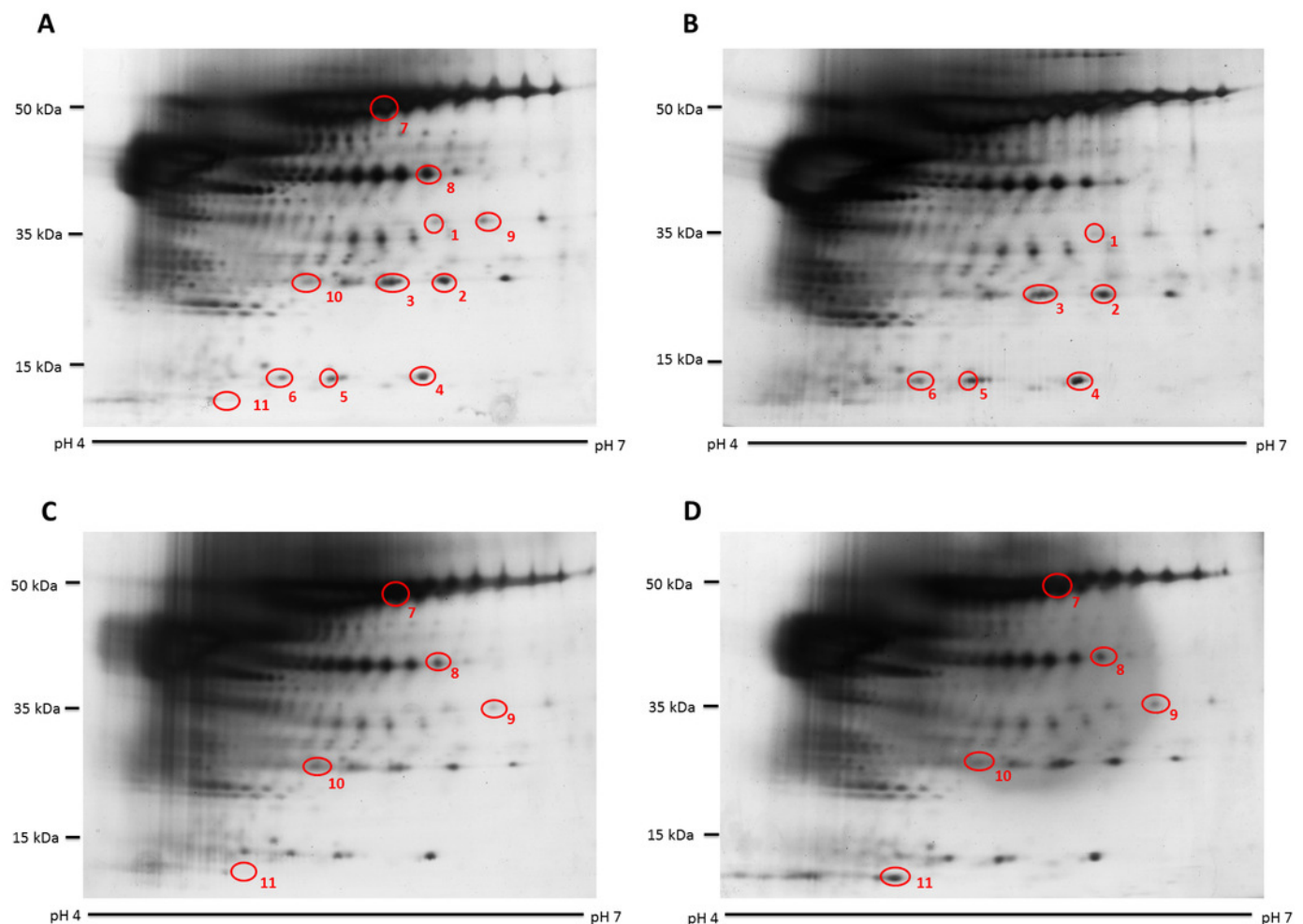
Figure 2

Figure 2. Volume contribution of 6 hair shaft protein spots that were significantly different between male and female subjects (*Fig. 1*).

Gel images were analysed by ImageMaster 2D Platinum Software (mean ± 95% confidence interval; *n* = 20). Panels A - F demonstrate the 6 protein spots (K85, three protein species of K86 and two protein species of K81) that were significantly different in abundance between male and female subjects. FC is fold change between the mean values for males and females. Error bars represent 95% confidence intervals.
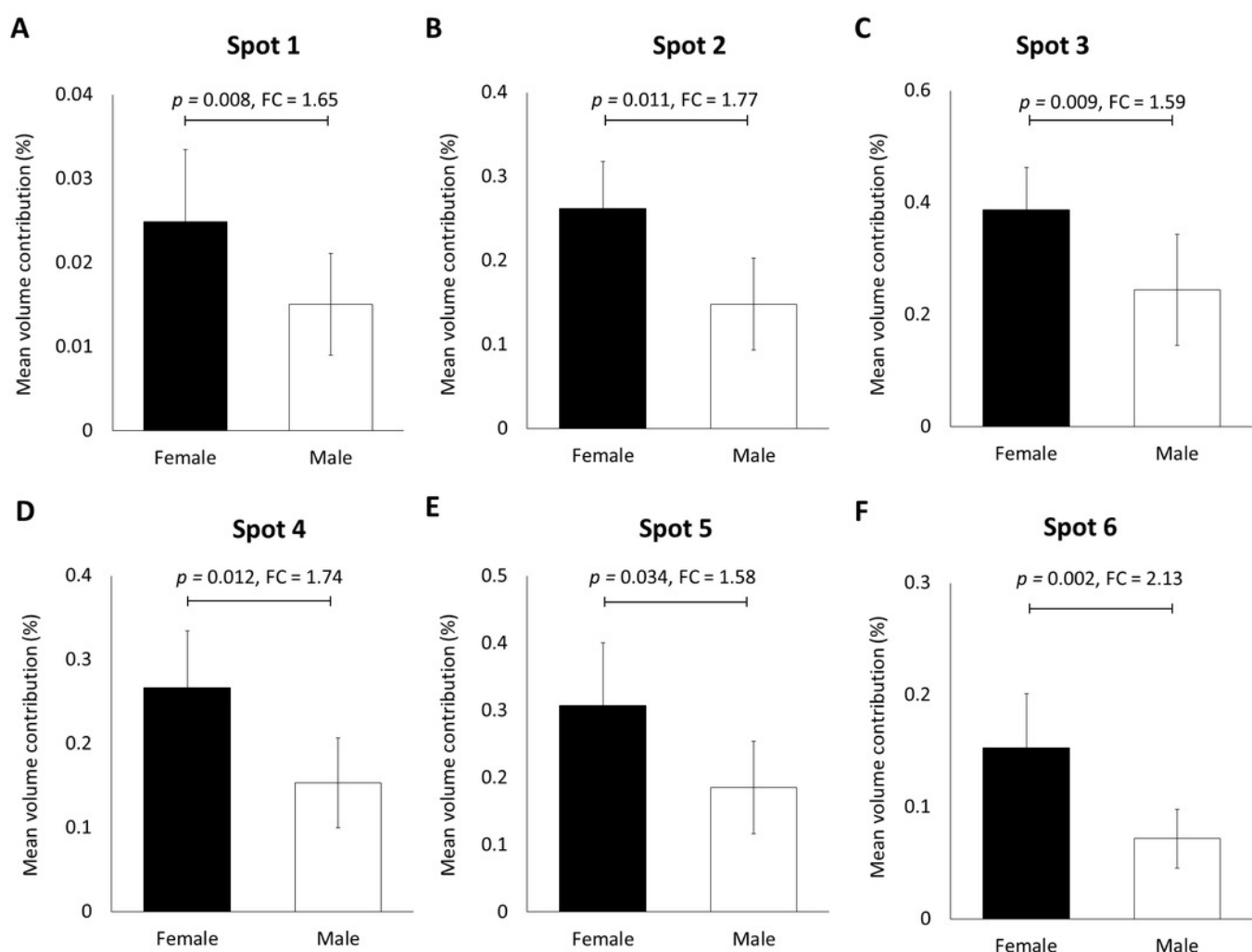
Figure 3

Figure 3. Volume contribution of hair shaft protein spots that were significantly different between the three different ethnicities.

The 5 protein spots (two protein species of K86, K83, K81 and K33B corresponding to the spot numbers 7-11 shown in *Fig. 1* – panels B - D, respectively) of significant altered abundance between Malay (*n* = 10), Chinese (*n* = 10) and Indian (*n* = 10) subjects as analysed by ImageMaster 2D Platinum Software of gel profiles (mean ± 95% confidence interval) were demonstrated (Panels A - E). FC is fold change between the mean values. Error bars represent 95% confidence intervals.
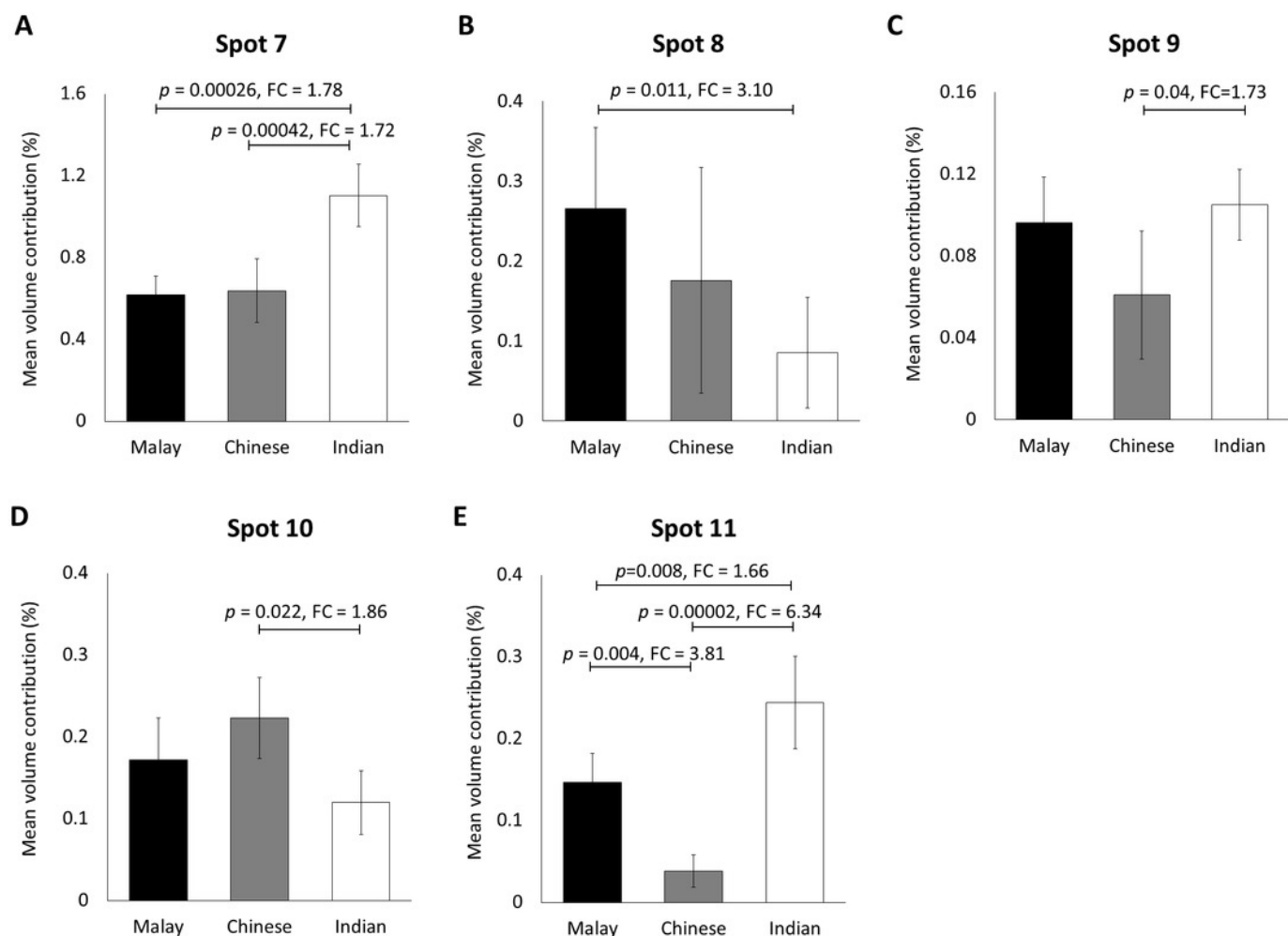
Figure 4

Figure 4. Immunoblotting of hair shaft K86.

Pooled hair shaft proteins were initially subjected to SDS-polyacrylamide gel electrophoresis and transferred electrophoretically onto PVDF membranes. Immunodetection was performed using polyclonal guinea pig antisera against K86, whilst peroxidase-conjugated goat anti-guinea pig antibody was used as secondary antisera. The PVDF membranes was developed using WesternBright Sirius enhanced chemiluminescence system. Normalised intensity values are ratios of the relative intensities of the bands to their respective loading controls (intensities of total protein using MemCode™ Reversible Protein Stain Kit). Panels A, B and C demonstrate cropped images of bands while, panels D, E & F shows normalised intensities of the four K86 peptides (spots 3, 4, 5 and 7, respectively). Error bars represent 95% confidence intervals.
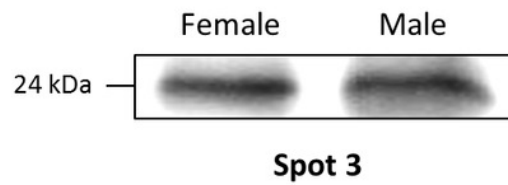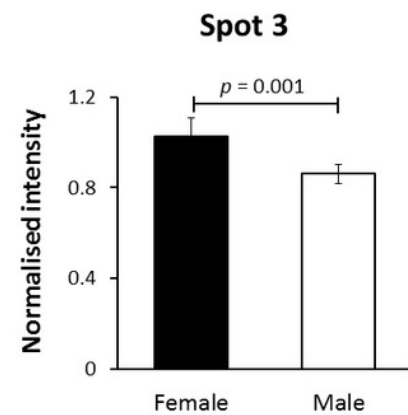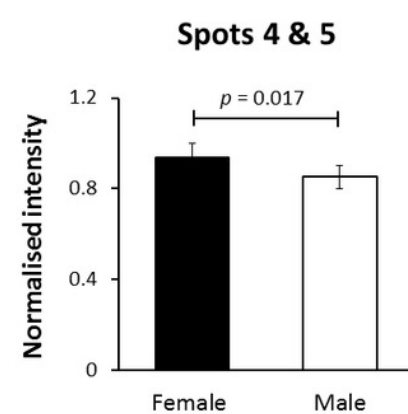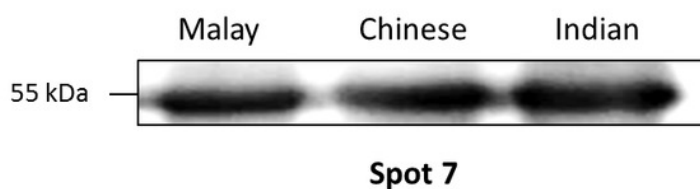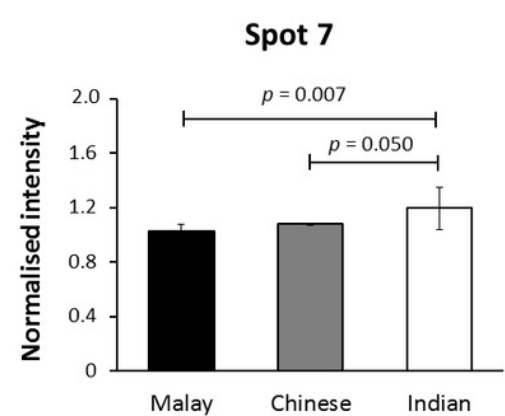
Table 1. MS identification of 2-dimensional electrophoresis hair shaft protein gel spots of altered abundance.

Spots that were significantly different in abundance between subjects of distinct genders (spots 1-6) and ethnicities (spots 7-11) were excised from gels and subjected to in-gel trypsin digestion, MALDI-ToF/ToF analysis and human hair database query (*Fig 1.*). Experimental mass was calculated based on relative mobilities (R_f) of the spots. cov – coverage; p*I* – isoelectric point.

Spot number	Accession number	Protein name	Abbreviation	Mascot score	Sequence cov (%)	Distinct peptides	Theoretical mass (Da)	Experimental mass (Da)	Theoretical pI	Experimental pI
1	P78386	Keratin, type-II cuticular Hb5	K85	74	6	4	57,306	32,359	6.28	5.90
2	Q14533	Keratin, type-II cuticular Hb1	K81	51	3	2	56,832	24,547	5.40	6.00
3	O43790	Keratin, type-II cuticular Hb6	K86	106	10	4	55,120	24,547	5.56	5.70
4	O43790	Keratin, type-II cuticular Hb6	K86	105	15	5	55,120	13,803	5.56	5.70
5	O43790	Keratin, type-II cuticular Hb6	K86	49	7	3	55,120	13,803	5.56	5.30
6	Q14533	Keratin, type-II cuticular Hb1	K81	42	1	1	56,832	13,803	5.40	5.00
7	O43790	Keratin, type-II cuticular Hb6	K86	509	35	15	55,120	54,954	5.56	5.50
8	O43790	Keratin, type-II cuticular Hb6	K86	164	11	7	55,120	39,810	5.56	5.90
9	P78385	Keratin, type-II cuticular Hb3	K83	55	2	2	55,928	32,359	5.54	6.30
10	Q14533	Keratin, type-II cuticular Hb1	K81	51	3	2	56,832	24,547	5.40	5.20
11	Q14525	Keratin, type-I cuticular Ha3-II	K33B	18	2	1	47,338	11,749	4.81	4.80