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

[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 no value for forensic investigations. In the present study, hair shaft proteins were extracted from forty healthy subjects between the ages of twenty-one to forty years and profiled using gel-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 lysis method. The extracts were profiled by 2-dimensional electrophoresis (2DE) and resolved protein spots were identified by mass spectrometry and query of the human hair database.

Results. Separation of the human hair shaft proteins by 2DE generated improved and highly resolved profiles. Comparing the 2DE hair shaft protein profiles of ten female with ten 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). 2DE profiling of thirty 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 taken together, our data 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.

Introduction

The hair shaft is formed from epidermal keratinocytes undergoing a unique form of keratinization programme cell death termed cornification. This process involves degradation and destruction of the cell nucleus as well as the genetic material enclosed within it (*Eckhart et al., 2013*). As a result, nuclear DNA (nDNA) which may be used for genetic fingerprinting is usually not detected in the hair shaft (*Bender & Schneider, 2006*). Although mitochondrial DNA (mtDNA) remains intact during the keratinization process (*Pfeiffer et al., 1999*), it cannot be used for genotyping purposes as mtDNA lacks a region that contains short tandem repeat (STR) and people who share a common maternal blood line also share the same exact mtDNA profile (*McNevin et al., 2005*).

Keratin intermediate filament (KIF), makes up approximately 80% of the human hair shaft (*Yamauchi et al., 2008*). Human KIF can be assorted into two main families, the acidic type-I keratins and neutral basic type-II keratins, which consists of 28 and 26 members, respectively (*Jacob et al., 2018; Szeverenyi et al., 2008*). Aside from KIF, the human hair shaft also contains small amounts of keratin associated proteins (KAPs), which consist of 27 families, with each family comprising at least 100 KAP genes with high sequence identity (*Gong et al., 2012; Rogers et al., 2008*). Many KAPs in the human hair have been identified over the years and these proteins generally cross-link with KIFs but the precise role of each protein and the

mechanisms that give hairs their different shapes, colours and rigidities are not well characterized (*Rogers et al., 2006; Wolfram, 2003*).

The complex interaction of these hair shaft proteins provides robust rigidity to the hair structure and makes it resistant to many environmental factors such as pollutants, weathers, ultraviolet light or chemical treatments (*Wolfram,* 2003). Despite being well conserved, previous studies have shown that characterization of the hair shaft proteins can be challenging and difficult. This is mainly due to difficulties in solubilizing and extracting the proteins in solvents that are compatible with liquid chromatography gel or electrophoresis approaches (*Sun et al., 2014; Kollipara & Zahedi, 2013; Shin et al, 2010; Han et al., 2007; Smith & Parry, 2007; Langbein & Schweizer, 2005*). Whilst many attempts to improve the yield of proteins extracted from the human hair shaft have been reported, the quality of published 2-dimensional gel electrophoresis (2DE) profiles is still far from becoming a useful method for forensic investigations and the process is also time-consuming and not practical to be applied in a large scale study (*Takayama & Ito, 2013; Barthélemy et al., 2012; Thibaut et al., 2009; Plowman, 2007; Fuji, Nakamura et al., 2002*). However, an alkaline lysis method which is capable of extracting substantially higher percentage of hair shaft proteins within only two hours was recently developed (*Wong et al., 2016*). With this improved method and higher yield of hair shaft proteins, we have reanalysed the material using 2DE gel-based proteomics and compared the profiles to determine if they can be used to distinguish gender and/or ethnicity.

Methodology

Subject recruitment

A total of forty healthy individuals of different genders and ethnicities but with the same age range (21 to 40 years) were recruited in the study in accordance to the ethical clearance granted by the Medical Ethics Committee of University of Malaya Medical Center (Institutional Review Board), which adheres to the ICH-GCP guideline and the Declaration of Helsinki (MEC ID.NO: 20158-1577). *Table 1* demonstrates the distribution of subjects according to demographic factors including age, gender and ethnicity. Informed written consent was acquired from all subjects prior to collection of their head hair samples. At least three strands of hair were collected from each subject and stored at –20°C. Relevant phenotypic characteristics of the hair shaft including discoloration, whitening, bleaching, weathering and perming, if any, were recorded. Subjects with previous history of bacterial or fungal-borne skin diseases, inflammation or cancer and/or under treatment for such ailments and those with chemically-treated hair were excluded from the study.

Table 1:

Distribution of subjects according to demographic factors.

Hair samples were collected from healthy individuals based on their genders, ethnicities and age range. *Subethnic groups of the Malaysian population.

Isolation of human hair shaft proteins

Human hair shaft proteins were isolated using the alkaline lysis method as previously described (*Wong et al., 2016*). Briefly, the hair samples were sterilised with 90% ethanol, cut (1-4 mm) and incubated in lysis buffer at 90°C for 30 minutes. The resulting supernatant was isolated using a QIAquick spin column (Qiagen, Hilden, Germany) and kept at 4°C. The undissolved

hair shaft fractions were pulverised for 30 minutes, centrifuged and the supernatant was separated by centrifugation. Solution containing hair shaft protein extract was precipitated using cold acetone. Protein pellets was then resolubilised in sample buffer and their concentrations were assayed using the Bio-Rad protein assay (Bio-Rad, California, USA) in accordance to the manufacturer's guidelines.

2DE and silver staining

Solubilised hair shaft proteins (100 µg) suspended in sample buffer were incubated in rehydration buffer containing 7M urea, 2M thiourea, 4% CHAPS, 2% IPG buffer pH 4-7, 0.002% w/w orange G and 7 mg of dithiothreitol (DTT) for 30 minutes at room temperature. The sample mixture was rehydrated with 13 cm DryStrip gel pH 4-7 (GE Healthcare, Uppsala, Sweden) for 18 hours at room temperature in a closed environment. The rehydrated strips were subjected to first dimensional separation using the Ettan IPGphor 3 IEF System (GE Healthcare, Uppsala, Sweden) according to the following settings: (i) 500 V, 1 hour, step and hold; (ii) 1000 V, 1 hour, gradient; (iii) 8000 V, 2 hours 30 minutes, gradient; (iv) 8000 V, 55 minutes, step and hold. Focused strips were equilibrated using DTT and iodoacetamide for 20 minutes, respectively, before being subjected to second dimensional separation using gradient polyacrylamide gel (8-15%) using SE 600 Ruby electrophoresis system (GE Healthcare, Uppsala, Sweden). Four 2DE gels were run according to the following program: (i) 50 V, 150 mA, 100 W for 1 hour; (ii) 600 V, 150 mA, 100 W until tracker dye reached bottom of the gel. The electrophoresed gels were stained using silver nitrate. In this process, gels were firstly fixed

with 40% (v/v) ethanol and 10% acetic acid for 30 minutes, followed by sensitization with 30% (v/v) ethanol, 0.5 M sodium acetate trihydrate and 12.7 mM sodium phosphate for 30 minutes. Gels were then washed thrice, each time for 5 minutes, and stained with 14.7 mM silver nitrate solution. After washing two more times for one min, gels were developed with 0.24 M sodium carbonate and 0.04% (v/v) formaldehyde. Development of spots was stopped with 40 mM EDTA solution and gels were finally kept in distilled water before being scanned. Subsequent to the conduct of the present study, a report on modification of this 2DE method, which generated further improved image resolution, has been published (*Wong, Hashim & Hayashi, 2019*).

Data analysis

Silver-stained 2DE gels were scanned using ImageScanner III (GE Healthcare, Uppsala, Sweden) and analysis of protein spot volume was performed using ImageMaster Platinum 7.0 software (GE Healthcare, Uppsala, Sweden). Data were analysed using the Statistical Package for Social Sciences (SPSS) version 25.0 (IBM Corporation, New York, USA). All values were presented as mean ± SEM. Levene's test for equality of variances was used to assess the distribution of the t-test data sets between female and male subjects, whilst ANOVA was used to analyse the differences in the abundance of proteins between three groups of subjects of different ethnicities. A p value of less than 0.05 and fold change of more than 1.5-folds were considered significant.

Mass spectrometry and database search

Identification of proteins was performed as previously described with minor modifications (*Seriramalu et al., 2010*). Briefly, protein spots of interests were carefully cut out from 2DE gels

and kept in high-purity water at –20°C. Gel plugs were first destained using potassium ferricyanide (III) (15 mM) and sodium thiosulphate (50 mM) for 15 minutes at room temperature. The destain procedure was repeated until the gel plugs became clear and transparent. The proteins in gel plugs were reduced and alkylated using DTT (10 mM) and iodoacetamide (55 mM) both in 100 mM ammonium. They were then washed thrice with 50% acetonitrile (ACN) in 100 mM ammonium bicarbonate, parched with 100% ACN and dried using vacuum centrifugation. The dried gels were treated with trypsin (6 ng/μL in 50 mM ammonium bicarbonate) for 18 hours at 37°C. The peptides were then dried, reconstituted in formic acid (0.1%) and desalted using ZipTip with C18 resin (Millipore, Massachusetts, USA). The desalted and concentrated peptides were mixed with equal volume of α-cyano-4-hydroxycinnamic acid (6 mg/ml), before being spotted onto the OptiToF 384-well insert (0.7 μl) of the 5800 MALDI ToF/ToF analyser (SCIEX, Framingham, USA).

The proteins were identified using MASCOT search engine (*Perkins et al., 1999*) and the resulting mass spectral data were thoroughly queried against the human hair entries in the Uniprot database (last update: January 17, 2019, containing 1329 sequences). The following parameters were set: enzyme: trypsin; maximum missed cleavages: 1; fixed modification: carbamidomethylation of cysteine; variable modification oxidation of methionine; precursor ion mass tolerance: 100 ppm; fragment ion mass tolerance: 0.2 Da. An individual ion score of more than seventeen indicates extensive homology or identity ($p<0.05$).

Results

2DE hair shaft protein profiling - distinguishing genders

When the human hair shaft protein extracts from male and female subjects were separated by 2DE and subjected to silver staining, similar profiles were obtained. *Figure 1* demonstrates representative protein profiles of female and male subjects of the same ethnicity (Malaysian Malay) and range of age. The 2DE gel profiles were then subjected to ImageMaster 2D Platinum Software analysis. In this analysis, Levene's test for equality was used to assess the distribution of the data set and independent sample t-test was used to compare the mean difference of the volume distribution between the subjects. Abundance of six hair shaft protein spots was found to be significantly higher in female compared to male subjects (*Fig. 2*). Spot 3 showed the highest mean percentage of volume contribution for females (0.387 ± 0.037) compared to males (0.244 ± 0.031). The highest fold change difference was observed in spot 6, with 2.15-folds higher mean percentage of volume contribution in females (0.072 ± 0.011) compared to males (0.155 ± 0.018).

Figure 1: Representative 2DE hair shaft protein profiles.
Hair shaft protein profiles of (A) male and (B) female subjects. Circled are protein spots that were significantly different in abundance between male (n=10) and female (n=10) subjects.

Figure 2: Mean percentage of volume contribution of hair shaft protein spots that were significantly different between male and female subjects.
2DE gels were analysed by ImageMaster 2D Platinum Software (mean ± SEM). Panels demonstrate the six protein spots that were significantly different in abundance between male and female subjects. FC is fold change between the mean values for males and females.

204

205 **2DE hair shaft protein profiling - distinguishing ethnicities**

206 *Figure 3* demonstrates the representative protein profiles of female subjects from the Malaysian
207 Malay, Chinese and Indian ethnic groups of the same age range. The 2DE gel profiles were
208 similarly analysed using ImageMaster 2D Platinum Software. However, statistical analysis was
209 performed using ANOVA with use of appropriate post-hoc test to analyse the differences in
210 abundance of the proteins between the distinct groups of subjects. In this analysis, five protein
211 spots were shown to be significantly different in at least one ethnic group compared to the others
212 (*Fig. 4*). Among the protein spots of altered abundance, spot 7 appeared the most intense.
213 However, spot 11 demonstrated the highest fold change differences between Indian (0.244 ±
214 0.028) and Chinese (0.038 ± 0.010) subjects, as well as between Malay (0.147 ± 0.018) and
215 Chinese (0.038 ± 0.010) subjects. When taken together, the analysis generally showed that the
216 Indian subjects had the highest mean percentage of volume contribution for spots 7, 9 and 11
217 compared to other ethnicities whilst their spots 8 and 10 were the least intense. The Chinese
218 subjects showed the highest mean percentage of volume contribution in spot 10 and lowest
219 values for spots 9 and 11. In the Malay subjects, spot 8 had the highest mean percentage of
220 volume contribution, whilst spot 7 was the least intense.

221

222 **Figure 3: Representative 2DE hair shaft protein profiles.**
223 Hair shaft protein profiles of (A) Malay, (B) Chinese and (C) Indian ethnic groups. Protein spots
224 that were significantly different in abundance between the three ethnicities are shown in circles.

225

Figure 4: Mean percentage of volume contribution of hair shaft protein spots that were significantly different between the three different ethnicities.

Panels demonstrate the five protein spots of significant altered abundance between Malay (n=10), Chinese (n=10) and Indian (n=10) subjects as analysed by ImageMaster 2D Platinum Software of 2DE gel profiles (mean ± SEM). FC is fold change between mean values.

Identification of hair shaft proteins by MALDI-ToF/ToF

The hair shaft protein spots of significant altered abundance between subjects of different genders and ethnicities were identified using MALDI-ToF/ToF analysis and search of the human hair database. *Table 2* demonstrates the results of the database query. Analysis of the 11 hair shaft proteins of interest identified five different types of keratins, including (1) type II cuticular Hb6 (K86) for spots 3, 4, 5, 7 and 8, (2) type II cuticular Hb3 (K83) for spot 9, (3) type II cuticular Hb1 (K81) for spots 2, 6 and 10, (4) type II cuticular Hb5 (K85) for spot 1 and (5) type I cuticular Ha3-II (K33B) for spot 11. Whilst the hair shaft of subjects of different genders demonstrated significant altered levels of K81, K85 and K86, those from different ethnicities showed significant different abundance of K33B, K81, K83 and K86.

Table 2: MS identification of 2DE hair shaft protein spots of altered abundance.

cov – coverage; p*I* – isoelectric point. Spots that were significantly different in abundance between subjects of distinct genders (spots 1-6 of *Fig. 1*) and ethnicities (spots 7-11 of *Fig. 2*)

were excised from 2DE gels and subjected to in-gel trypsin digestion, MALDI-ToF/ToF analysis and human hair database query. Experimental mass was calculated based on relative mobilities (R_f) of the spots.

Discussion

In the absence of nDNA, the human hair shaft is a useless biological material for forensic investigations. Formed via cornification of keratinocytes, the hair shaft mainly contains seventeen different types of keratins, (i.e. K31-40 (including K33a and K33b) and K81-86) (*Moll, Divo & Langbein, 2008*), which are poorly analysed mainly because of their limited solubility and extraction yield. In the present study, we have adopted a newly developed protocol that enhanced the extraction yield of proteins from the human hair shaft (*Wong et al., 2016*) and reanalysed the material by 2DE. The 2DE profiles that were generated showed improved resolution of the separated hair shaft proteins compared to those that were earlier reported (*Barthélemy et al., 2012; Thibaut et al., 2009; Nakamura et al., 2002*). Identification of these well-resolved proteins by mass spectrometry and search of the human hair protein database showed that they comprised the different types of keratins, the cysteine-rich helicoidal proteins that protect the hair because of their insolubility and impermeability.

In view of its high resolution and reproducibility, the 2DE profiling of the human hair shaft proteins was further utilised in a pilot study to determine if the developed profiles could be used to distinguish gender of individuals. Our analysis of twenty hair shaft samples from healthy individuals between the age of twenty-one to forty years by 2DE and mass spectrometry demonstrated significant higher abundance of six different type-II keratin spots, including two

20 + 20 = 40

269 K81, one K85 and three K86, in the women subjects compared to men. These different types of
270 keratins are known to be restricted to the hair shaft and not present in the follicle (*Moll, Divo &*
271 *Langbein, 2008*). However, all these spots of altered abundance appeared to be truncated or
272 processed keratin products as they were resolved within the molecular weight regions lower than
273 their putative primary translated precursor polypeptides (*Table 2*).

no data
not reasonable

274 Similar marked shifts in molecular weights and isoelectric point (p*I*) values of the hair shaft
275 proteins that were separated by 2DE, which may be due to deamidation, post-translational
276 modifications, or processing at the RNA level as sequence variants (*Person et al., 2006*), have
277 been previously reported (*Barthélemy et al., 2012*). In the present study, SDS and urea were used
278 during the hair shaft protein extraction. Although these chemicals are known to cause deamidation
279 and disruption of disulphide linkages (*Adav et al., 2018*), they are unlikely to have generated
280 substantial difference in the molecular weights of the hair shaft proteins as observed in the present
281 study. Whether this was an effect of the strong alkali used at 90°C in the hair shaft protein
282 extraction, or that the proteins were further biologically processed during their various stages of
283 genetic expression, subsequent to their translation or during cornification, remains to be
284 investigated.

all hypothesis

Pure Hypothesis

need more work

gel electrophoresis

30 ♀ Table 1

285 The ~~2DE~~ profiling of 30 hair shaft samples taken from women of similar age range but
286 from three distinctive Malaysian ethnic subpopulations further showed significant altered
287 abundance of one type-I (K33b) and four type-II (K81, K83 and two K86) keratins between the
288 ethnic groups that were analysed. Like the earlier detected keratins, the type-I K33b and all the
289 type-II keratins detected are also known to be localised exclusively in the hair shaft (*Moll, Divo &*
290 *Langbein, 2008*). Based on the resolved experimental molecular weights, all the five spots of
291 altered abundance also appeared to be truncated or processed keratins. In addition, the K81 and

Table 2 13 KRT Figs 2+4 which?

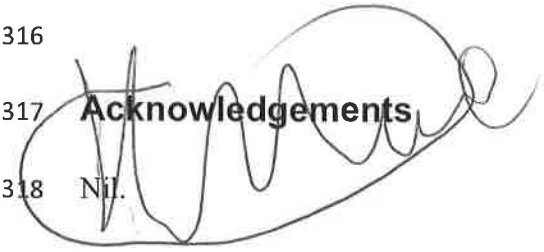
K86 spots that were also detected in this analysis were different from their counterparts that were detected in the earlier gender analysis as they showed distinctive experimental molecular weight and p*I* values.

The results of the latter study also demonstrated that the Indian ethnic group to be most distinctive as they showed four abundantly different keratins (K33B, K81, K83 and K86 (spot 7)) compared to the Chinese and three keratins that were differently altered (K33B and two K86 (spots 7 and 8)) compared to the Malays. On the other hand, the Chinese and Malay ethnic groups only appeared to be distinctive in the abundance of K33B (3.87-fold of difference) and their levels of the type-I keratin were both significantly different compared to the Indians. These results are generally comparable with the genetic data that were earlier reported. In a study using multi-dimensional scale analysis on the population genetic structure of the different ethnic groups in Peninsular Malaysia, Hatin et al. had previously reported that the Malay and Chinese populations were clustered together while the Indians were further apart (*Hatin et al. 2011*).

Conclusion

When taken together, the human hair 2DE keratin profiling that was conducted in this pilot study provided a potential method that can be used to distinguish gender and ethnicity of individuals based on their hair shaft samples. However, a larger scale analysis of the hair shaft proteins using similar proteomics methods or immuno-analysis utilizing antibodies that are specific to the different types of keratins that were highlighted in the present study would certainly increase the robustness of the results. The large scale study could eventually lead to the development of a searchable database as well as signature keratin biomarkers that could facilitate determination of one's gender and ethnicity based on his/her 2DE hair shaft keratin profiles. However, several other

factors such as the effects of chemical exposure as well as dietary and environmental influences on the hair shaft keratin profiles are also required for confirmation of the accuracy of the results.



Acknowledgements

Nil.

References

Adav SS, Subbaiaih RS, Kerk SK, Lee AY, Lai HY, Ng KW, Sze SK, Schmidtchen A. 2018. Studies on the proteome of human hair-identification of histones and deamidated keratins. *Scientific Reports* 8:1599 DOI: 10.1038/s41598-018-20041-9.

Barthélemy NR, 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* 421:43-55 DOI: 10.1016/j.ab.2011.10.011.

Bender K, Schneider PM. 2006. Validation and casework testing of the BioPlex-11 for STR typing of telogen hair roots. *Forensic Science International* 161:52-59 DOI: 10.1016/j.forsciint.2005.10.024.

Eckhart L, Lippens S, Tschachler E, Declercq W. 2013. Cell death by cornification. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1833:3471-3480 DOI: 10.1016/j.bbamcr.2013.06.010.

Fujii T, Takayama S, Ito Y. 2013. A novel purification procedure for keratin-associated proteins and keratin from human hair. *Journal of Biological Macromolecules* 13:92-106.

359 Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. 1999. Probability-based protein identification
360 by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551-3567
361 DOI: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.
362 Person MD, Shen J, Traner A, Hensley SC, Lo H-H, Abbruzzese JL, Li D. 2006. Protein
363 fragment domains identified using 2D gel electrophoresis/MALDI-ToF. *Journal of Biomolecular*
364 *Techniques* 17:145.
365 Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P.
366 1999. Expanding the forensic German mitochondrial DNA control region database: genetic
367 diversity as a function of sample size and microgeography. *International Journal of Legal*
368 *Medicine* 112:291-298 DOI 10.1007/s004140050252.
369 Plowman JE. 2007. The proteomics of keratin proteins. *Journal of Chromatography B* 849:181-
370 189 DOI: 10.1016/j.jchromb.2006.07.055.
371 Rogers M, Langbein L, Praetzel-Wunder S, Giehl K. 2008. Characterization and expression
372 analysis of the hair keratin associated protein KAP26. 1. *British Journal of Dermatology*
373 159:725-729 DOI: 10.1111/j.1365-2133.2008.08743.x.
374 Rogers MA, Langbein L, Praetzel-Wunder S, Winter H, Schweizer J. 2006. Human hair
375 keratin-associated proteins (KAPs). *International Review of Cytology* 251:209-263 DOI:
376 10.1016/S0074-7696(06)51006-X.
377 Shin S, Lee A, Lee S, Lee K, Kwon J, Yoon MY, Hong J, Lee D, Lee GH, Kim J. 2010.
378 Microwave-assisted extraction of human hair proteins. *Analytical Biochemistry* 407:281-283
379 DOI: 10.1016/j.ab.2010.08.021.
380 Smith TA, Parry DAD. 2007. Sequence analyses of Type I and Type II chains in human hair and
381 epithelial keratin intermediate filaments: Promiscuous obligate heterodimers, Type II template

Gong H, Zhou H, McKenzie GW, Yu Z, Clerens S, Dyer JM, Plowman JE, Wright MW, Arora R, Bawden CS. 2012. An updated nomenclature for keratin-associated proteins (KAPs). *International Journal of Biological Sciences* 8:258 DOI: 10.7150/ijbs.3278.

Han MO, Chun JA, Lee WH, Lee JW, Chung CH. 2007. A simple improved method for protein extraction from human head hairs. *Journal of Cosmetic Science* 58:527-534.

Hatin WI, Zahri M-K, Xu S, Jin L, Tan S-G, Rizman-Idid M, Zilfalil BA, Consortium HP-AS. 2011. Population genetic structure of peninsular Malaysia Malay sub-ethnic groups. *PloS One* 6:e18312 DOI: 10.1371/journal.pone.0018312.

Jacob JT, Coulombe PA, Kwan R, Omary MB. 2018. Types I and II keratin intermediate filaments. *Cold Spring Harbor Perspectives in Biology* 10:a018275 DOI: 10.1101/cshperspect.a018275.

Kollipara L, Zahedi RP. 2013. Protein carbamylation: *In vivo* modification or *in vitro* artefact? *Proteomics* 13:941-944 DOI: 10.1002/pmic.201200452.

Langbein L, Schweizer J. 2005. Keratins of the human hair follicle. *International Review of Cytology* 243:1-78 DOI: 10.1016/S0074-7696(05)43001-6.

McNevin D, Wilson-Wilde L, Robertson J, Kyd J, Lennard C. 2005. Short tandem repeat (STR) genotyping of keratinised hair - Part 1. Review of current status and knowledge gaps. *Forensic Science International* 153:237-246 DOI: 10.1016/j.forsciint.2005.05.006.

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

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* 25:569-572 DOI: 10.1248/Bpb.25.569.

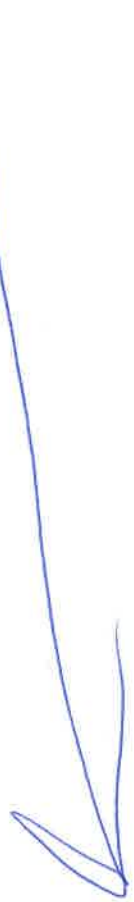
382 for molecule formation and a rationale for heterodimer formation. *Journal of Structural Biology*
383 158:344-357 DOI: 10.1016/j.jsb.2006.12.002.
384 Sun SS, Zhou JY, Yang WM, Zhang H. 2014. Inhibition of protein carbamylation in urea
385 solution using ammonium-containing buffers. *Analytical Biochemistry* 446:76-81 DOI:
386 10.1016/j.ab.2013.10.024.
387 Szeverenyi I, Cassidy AJ, Chung CW, Lee BT, Common JE, Ogg SC, Chen H, Sim SY, Goh
388 WL, Ng KW. 2008. The human intermediate filament database: Comprehensive information on a
389 gene family involved in many human diseases. *Human Mutation* 29:351-360 DOI:
390 10.1002/humu.20652.
391 Thibaut S, Cavusoglu N, de Becker E, Zerbib F, Bednarczyk A, Schaeffer C, van Dorsselaer A,
392 Bernard BA. 2009. Transglutaminase-3 enzyme: A putative actor in human hair shaft
393 scaffolding? *Journal of Investigative Dermatology* 129:449-459 DOI: 10.1038/jid.2008.231.
394 Wolfram LJ. 2003. Human hair: A unique physicochemical composite. *Journal of the American*
395 *Academy of Dermatology* 48:S106-S114 DOI: 10.1067/mjd.2003.276.
396 Wong SY, Hashim OH, Hayashi N. 2019. Development of high-performance two-dimensional
397 gel electrophoresis for human hair shaft proteome. *PloS One* 14:e0213947 DOI:
398 10.1371/journal.pone.0213947.
399 Wong SY, Lee CC, Ashrafzadeh A, Junit SM, Abrahim N, Hashim OH. 2016. A high-yield two-
400 hour protocol for extraction of human hair shaft proteins. *PloS One* 11:e0164993 DOI:
401 10.1371/journal.pone.0164993.
402 Yamauchi C, Okazaki W, Yoshida T, Karasawa A. 2008. Enzymatic degradation of keratin films
403 and keratin fibers prepared from human hair. *Biological and Pharmaceutical Bulletin* 31:994-
404 997 DOI: 10.1248/Bpb.31.994.

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

Table 1(on next page)

Distribution of subjects according to demographic factors.

Hair samples were collected from healthy individuals based on their genders, ethnicities and age range.



add Table / Table 1 to be included needed data but not in Report

Sample	Age (years)	Gender	Ethnicity*	Skin/scalp diseases	Aberrant hair phenotypic characteristics
1	23	Female	Malay	Nil	Nil
2	27	Female	Malay	Nil	Nil
3	36	Female	Malay	Nil	Nil
4	30	Female	Malay	Nil	Nil
5	27	Female	Malay	Nil	Nil
6	29	Female	Malay	Nil	Nil
7	25	Female	Malay	Nil	Nil
8	29	Female	Malay	Nil	Nil
9	40	Female	Malay	Nil	Nil
10	23	Female	Malay	Nil	Nil
11	24	Male	Malay	Nil	Nil
12	36	Male	Malay	Nil	Nil
13	23	Male	Malay	Nil	Nil
14	24	Male	Malay	Nil	Nil
15	24	Male	Malay	Nil	Nil
16	24	Male	Malay	Nil	Nil
17	27	Male	Malay	Nil	Nil
18	22	Male	Malay	Nil	Nil
19	28	Male	Malay	Nil	Nil
20	23	Male	Malay	Nil	Nil
21	33	Female	Chinese	Nil	Nil
22	24	Female	Chinese	Nil	Nil
23	25	Female	Chinese	Nil	Nil
24	24	Female	Chinese	Nil	Nil
25	33	Female	Chinese	Nil	Nil
26	36	Female	Chinese	Nil	Nil
27	25	Female	Chinese	Nil	Nil
28	24	Female	Chinese	Nil	Nil
29	32	Female	Chinese	Nil	Nil
30	33	Female	Chinese	Nil	Nil
31	22	Female	Indian	Nil	Nil
32	33	Female	Indian	Nil	Nil
33	39	Female	Indian	Nil	Nil

all Nil

all Nil

all Nil

34	28	Female	Indian	Nil	Nil
35	33	Female	Indian	Nil	Nil
36	21	Female	Indian	Nil	Nil
37	32	Female	Indian	Nil	Nil
38	28	Female	Indian	Nil	Nil
39	21	Female	Indian	Nil	Nil
40	23	Female	Indian	Nil	Nil

1

Table 2 (on next page)

MS identification of 2DE hair shaft protein spots of altered abundance.

Spots that were significantly different in abundance between subjects of distinct genders (spots 1-6 of *Fig. 1*) and ethnicities (spots 7-11 of *Fig. 2*) were excised from 2DE gels and subjected to in-gel trypsin digestion, MALDI-ToF/ToF analysis and human hair database query. Experimental mass was calculated based on relative mobilities (R_f) of the spots.

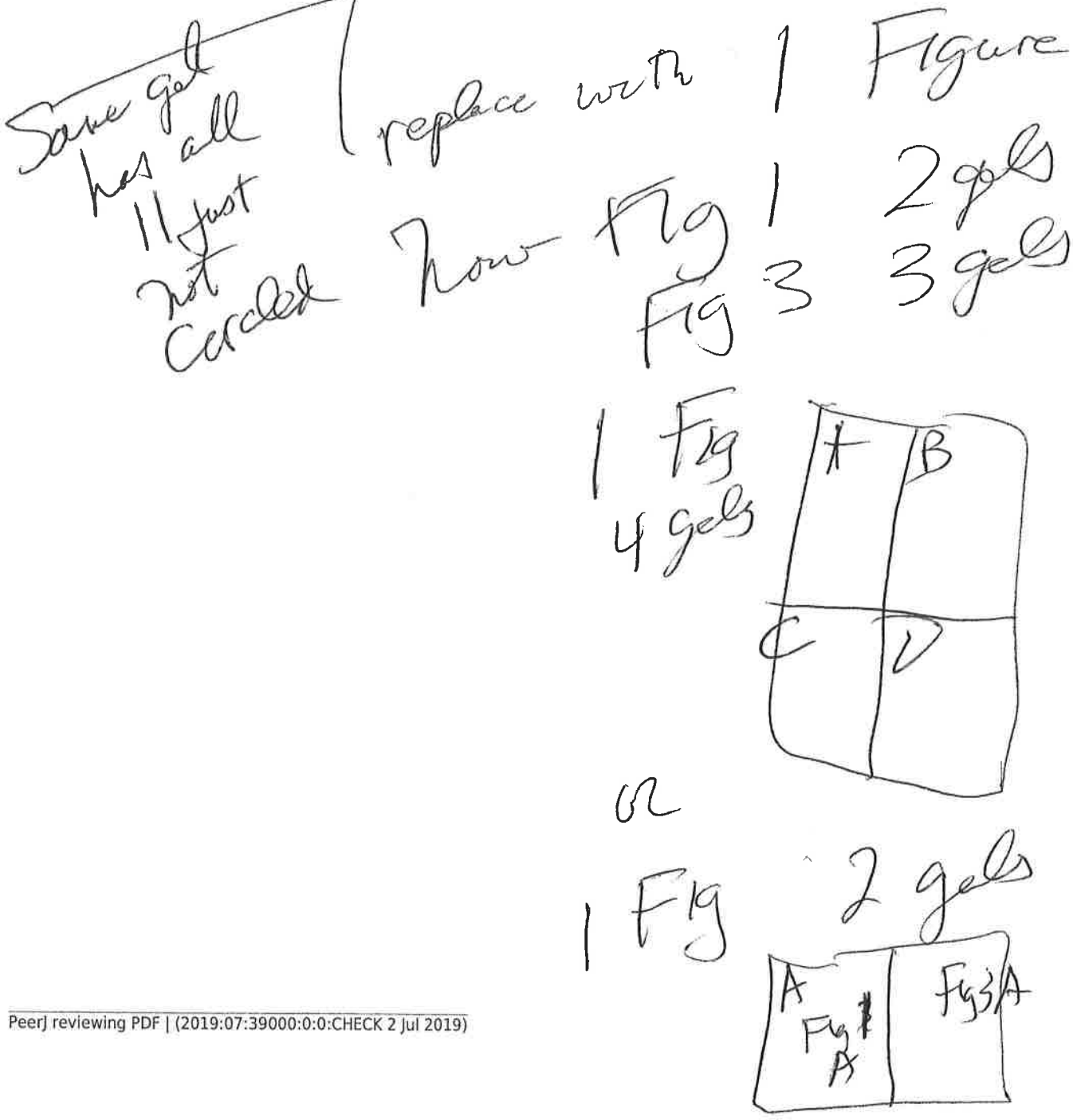


Table 2

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	28,379	6.28	5.90
2	Q14533	Keratin, type-II cuticular Hb1	K81	51	3	2	56,832	24,381	5.40	6.00
3	O43790	Keratin, type-II cuticular Hb6	K86	106	10	4	55,120	24,381	5.56	5.70
4	O43790	Keratin, type-II cuticular Hb6	K86	105	15	5	55,120	14,023	5.56	5.70
5	O43790	Keratin, type-II cuticular Hb6	K86	49	7	3	55,120	14,023	5.56	5.30
6	Q14533	Keratin, type-II cuticular Hb1	K81	42	1	1	56,832	14,023	5.40	5.00
7	O43790	Keratin, type-II cuticular Hb6	K86	509	35	15	55,120	45,078	5.56	5.50
8	O43790	Keratin, type-II cuticular Hb6	K86	164	11	7	55,120	36,353	5.56	5.90
9	P78385	Keratin, type-II cuticular Hb3	K83	55	2	2	55,928	24,381	5.54	6.30
10	Q14533	Keratin, type-II cuticular Hb1	K81	51	3	2	56,832	14,023	5.40	5.20
11	Q14525	Keratin, type-I cuticular Ha3-II	K33B	18	2	1	47,338	11,309	4.81	4.80

from sequence?
from gel
from gel
from gel
what is this?
why very poor agreement

Figure 1

Representative 2DE hair shaft protein profiles

Hair shaft protein profiles of (A) male and (B) female subjects. Circled are protein spots that were significantly different in abundance between male (n=10) and female (n=10) subjects

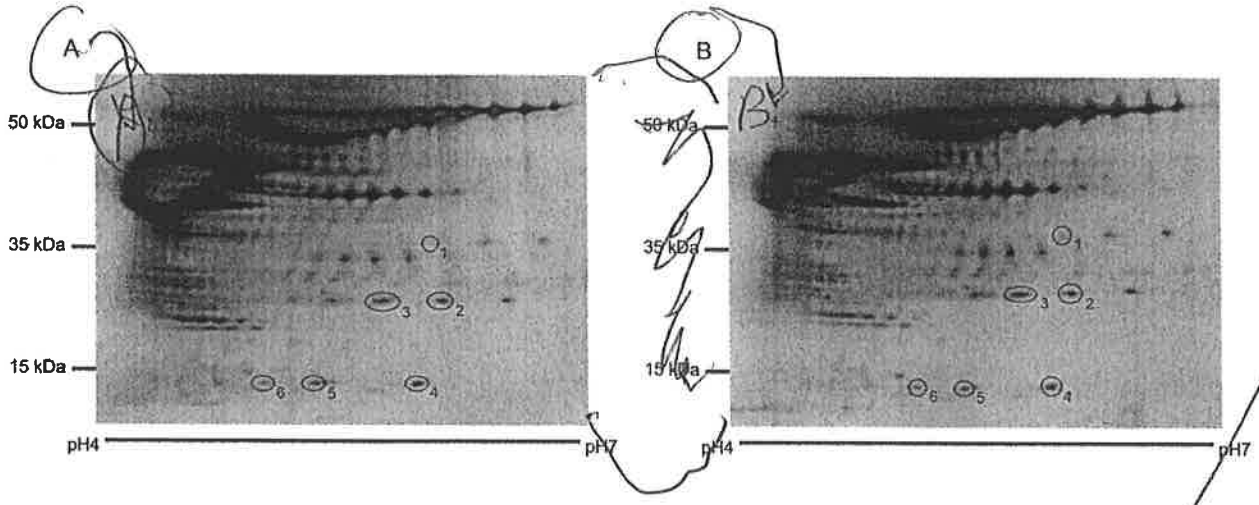
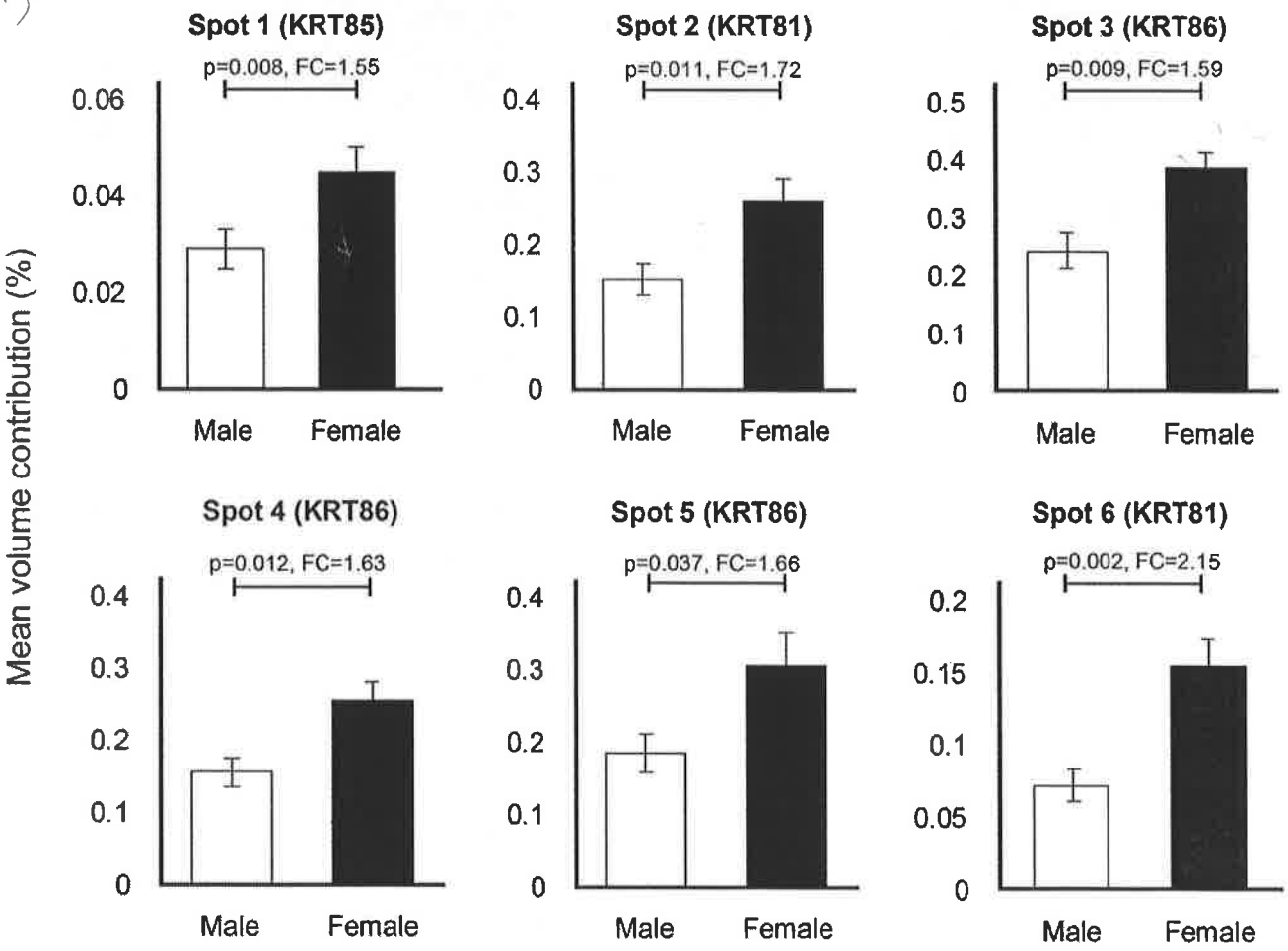


Figure 2

Mean percentage of volume contribution of hair shaft protein spots that were significantly different between male and female subjects.

2DE gels were analysed by ImageMaster 2D Platinum Software (mean ± SEM). Panels demonstrate the six protein spots that were significantly different in abundance between male and female subjects. FC is fold change between the mean values for males and females.



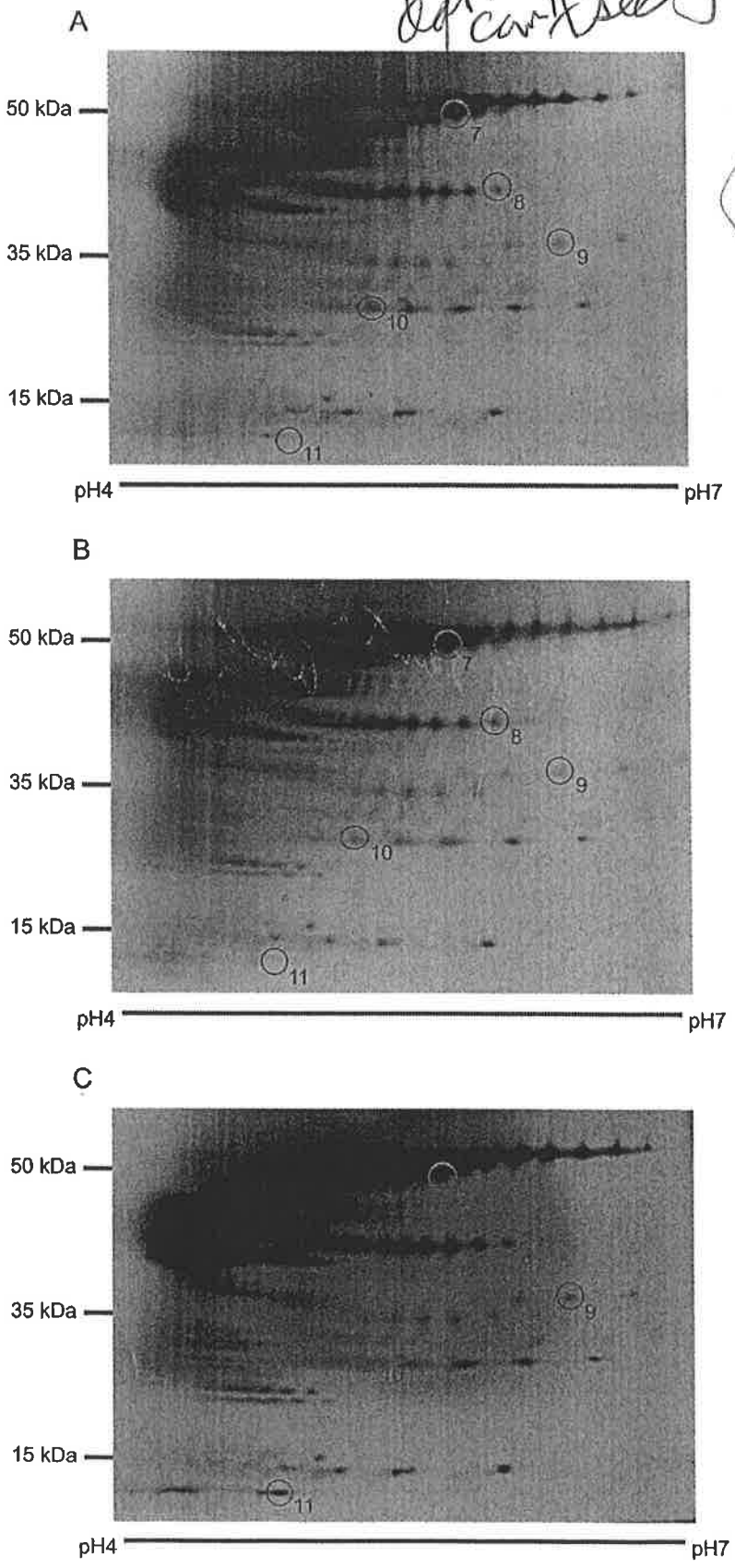


Figure 3

Representative 2DE hair shaft protein profiles.

Hair shaft protein profiles of (A) Malay, (B) Chinese and (C) Indian ethnic groups. Protein spots that were significantly different in abundance between the three ethnicities are shown in circles.

Figure 4

Mean percentage of volume contribution of hair shaft protein spots that were significantly different between the three different ethnicities.

Panels demonstrate the five protein spots of significant altered abundance between Malay (n=10), Chinese (n=10) and Indian (n=10) subjects as analysed by ImageMaster 2D Platinum Software of 2DE gel profiles (mean ± SEM). FC is fold change between mean values.