



基于头发蛋白质组学区分个人特质的方法学探究*

吴晓淋^{1,2)} 张涛²⁾ 徐平^{1,2,3)} 张亚莉^{1)**} 张振鹏^{2)**}

(¹⁾ 贵州大学医学院, 贵阳 550025; (²⁾ 军事科学院军事医学研究院生命组学研究所, 蛋白质组学国家重点实验室, 国家蛋白质科学中心 (北京), 北京蛋白质组研究中心, 北京 102206; (³⁾ 中国医学科学院蛋白质组学与新药研发创新单元, 北京 102206)

摘要 **目的** 头发是一类重要的皮肤附属物, 主要由角蛋白和角蛋白相关蛋白等组成。不同种族及性别样本的头发蛋白质组成和占比存在差异, 且目前缺乏高效率提取头发蛋白的方法。本文探究基于定量头发蛋白质组学方法, 旨在探索该方法区分不同个体的可能性。**方法** 以3例头发样本, 对样品处理方法和裂解缓冲液进行探究, 发展一种名为PLEE (PTM lab for protein extraction from hair with high efficiency) 的稳定、高效的头发蛋白质提取方法, 对7例人发样本, 以PLEE法进行提取, 结合胶内消化方法进行蛋白质组学实验, 产出蛋白质组学数据, 分析个体间的头发蛋白质组成及占比。**结果** 共鉴定274种蛋白质, 共有的蛋白质107种, 非共有蛋白质种类在57~119, 部分样本存在独特鉴定蛋白。使用共鉴定107种蛋白质进行定量蛋白质组分析, 通过聚类分析和主成分分析, 可将各样本进行区分, 且技术重复样本可聚在一起, 表明流程的稳定性。另外, 筛选出10个关键蛋白 (KRT33A、KRTAP9-6、KRT83、KRTAP7-1、KRT32、BLMH、KRT38、KRTAP11-1、NPAS1、KRTAP4-3) 在不同个体之间差异比较大且在个体中蛋白质鉴定稳定。**结论** 不同个体的头发蛋白质组成存在差异, 筛选出的10个蛋白质有望作为区分个体特质的关键蛋白, 在个体识别、刑侦方面具有重大的应用潜力。

关键词 头发, 蛋白质组学, 个人特质, PLEE

中图分类号 R33

DOI: 10.16476/j.pibb.2023.0165

头发是人体的一类结构复杂的表皮附属物^[1], 主要成分为蛋白质、水、脂类、色素和微量金属元素^[2]。蛋白质占总量的60%~95%, 其中角蛋白 (keratin, KRT) 与角蛋白相关蛋白 (keratin-associated proteins, KAP) 是最主要的蛋白质种类。它们相互作用形成交联结构, 保证头发的化学和机械稳定性^[3], 使头发具有较强的耐久性^[4]。人体头发的生长周期为2~7年, 据估计, 人体每天自然掉落的头发数量约50~100根^[5]。各种场景下发现的头发痕迹, 在考古学研究中的人类起源和遗传演化及犯罪学研究中的身份鉴定和犯罪嫌疑人追踪等具有重要意义^[6]。

据文献报道, 可利用蛋白质组学方法研究头发蛋白质组。Zhang等^[7]提出基于凝胶法的蛋白质组学方法, 分析头发样本蛋白质, 每个样本分为10个组分, 各组分鉴定到132~214种蛋白质。Laatsch等^[8]使用鸟枪法蛋白质组学技术探索非裔美国人、

高加索人、肯尼亚人和韩国人的头发蛋白质组, 发现不同种族的人群具有各自独特的头发蛋白表达水平, 主要体现在KRT和KAP两类蛋白。Parker等^[9]对欧美66名受试者头发蛋白质组表征, 共鉴定到401种蛋白质, 丰度最高的两类是KRT和KAP等。此外, 还有研究表明^[10], 人头发中I型和II型角蛋白如KRT33b、KRT81、KRT83和KRT86等可区分人类的性别和种族。目前尚未有文献报道使用头发蛋白质组学进行个人特质区分。

头发蛋白质高度交联的结构使得其结构稳定^[11]、难以被外界因素破坏, 蛋白质提取量低。已发展出多种头发蛋白质的处理方法, 包括高温、

* 国家重点研发计划 (2020YFE0202200) 资助项目。

** 通讯联系人。

张振鹏 Tel: 15898965955, E-mail: zhangzp1002@163.com

张亚莉 Tel: 13639069576, E-mail: 286802781@qq.com

收稿日期: 2023-04-23, 接受日期: 2023-06-20

高压、酸、碱或酶等, 可以将角蛋白分解成短肽或游离氨基酸^[12]。传统的头发处理方法包括物理法和化学法两大类。物理法是采用物理外力或条件, 如超声波、振荡、滚珠研磨等, 将头发的结构破坏, 从而将蛋白质释放出来, 其原理是超声压力波使头发的蛋白质分子振荡并分解、裂解, 或用滚珠研磨机或硅胶珠将头发样品压碎和摩擦, 使蛋白质分子溶解。化学法是采用化学试剂的作用将头发蛋白质从组织中溶解出来。主要方法包括直接溶解法、酸性裂解法等。直接溶解法使用含有变性剂和蛋白酶抑制剂的缓冲液将头发样品放入高温高压的环境中, 使蛋白质分子从头发释放。酸性裂解法将头发样品放入酸性溶液并加热, 通过酸性环境的作用, 使蛋白质分子从头发释放^[13]。

但是头发特殊的角蛋白高度交联结构, 相对其他组织而言更难以提取, 所需时间较长。使用单纯的加热或者震荡方法对头发蛋白质的提取存在一定程度的缺失, 导致鉴定量较少、覆盖率较低等问题。且在文献中需要高起始量的头发(至少150根1 cm长度的头发)用于实验, 在实际应用场景中难以满足。因此, 本研究采用混合法结合物理和化学法开发一种基于低起始头发量的高效率头发蛋白质提取方法, 提高提取效率。

本研究利用低头发起始量(10根1 cm长度的头发), 系统性探索样本处理方式对头发蛋白质的提取效率; 在此基础上探讨头发蛋白质裂解缓冲液成分对蛋白质提取的影响, 改进建立名为PLEE (PTM lab for protein extraction from hair with high efficiency) 的头发蛋白质稳定高效提取技术体系。通过PLEE法对7例样本的头发蛋白质组学数据进行分析, 研究头发在识别和区分不同个体特质方面的可行性。

1 材料与方法

1.1 样本来源

7例人头发样本(编号为P1至P7), 在后顶头皮1 cm处进行采样。其中1例进行4次技术重复(P7为技术重复组, 编号为P7-R1至P7-R4)。

1.2 头发蛋白质提取方法

1.2.1 头发样本预处理

将10根1 cm长度的头发剪成约1 mm长度的片段, 加入1 ml 20%甲醇进行清洗, 涡旋30 s, 离心1 min, 更换20%甲醇再次清洗, 重复3次, 最后

一次吸出上清, 头发在室温下自然风干后备用。

1.2.2 头发样本处理方法优化

a. 玻璃匀浆研磨器法 (GH): 头发样品中加入240 μ l SDS缓冲液 (50 mmol/L Tris (pH=7.5), 2% SDS和20 mmol/L DTT), 使用玻璃匀浆器 (Kimble, Manzanillo, Mexico) 研磨, 直到没有可见的纤维颗粒, 将裂解液体系转移到EP管中。

b. 液氮加硅胶珠研磨法 (LN-BH): 头发样品用研钵在液氮中研磨头发碎段, 将头发碎段放入EP管中, 加入240 μ l SDS缓冲液, 1/3底部的二氧化硅珠 (BioSpec, Bartlesville, USA), 涡旋提取蛋白质30 s, 并在冰上静置30 s, 重复5次。

c. 室温震荡提取方法 (RTO): 头发样品中加入240 μ l 尿素缓冲液 (50 mmol/L Tris (pH=7.5), 7 mol/L 尿素, 2 mol/L 硫脲和40 mmol/L DTT), 室温下1 200 r/min的条件下振荡18 h。

d. 液氮研磨加加热法 (LN-H95): 头发样品用研钵在液氮中研磨, 将头发碎段放入EP管中, 加入240 μ l 尿素缓冲液, 在95℃下加热5 min。

e. 加热震荡法 (SH65): 头发样品中加入240 μ l SDS缓冲液, 65℃下, 1 200 r/min的条件下振荡12 h。

f. 高温加热震荡法 (H-90-SH-65): 头发样品中加入120 μ l SDS缓冲液, 90℃加热5 h。4℃, 17 000 g离心, 收集上清液作为馏分1 (F1)。随后, 将120 μ l的SDS缓冲液加入到剩余的头发沉淀中, 65℃, 1 200 r/min下振荡12 h。4℃, 17 000 g离心, 收集上清液作为馏分2 (F2)。F1和F2的混合物被保存为H-90-SH-65法蛋白质组样品的完整蛋白质。

1.3 头发样本蛋白裂解缓冲液比较

基于高温加热震荡法 (H-90+SH-65) 法, 分别使用SDS缓冲液和尿素缓冲液进行提取, 对裂解液进行探索。

SDS缓冲液: 50 mmol/L Tris (pH=7.5), 2% SDS和20 mmol/L DTT。

尿素缓冲液: 50 mmol/L Tris (pH=7.5), 7 mol/L 尿素, 2 mol/L 硫脲和40 mmol/L DTT。

1.4 蛋白质质量检测及蛋白质定量

使用SDS-PAGE电泳检测头发蛋白质提取质量, 考马斯亮蓝染色液染色2 h, 脱色至背景无色。扫描胶片, 使用ScanImage分析, 计算灰度值, 确定蛋白质的浓度及裂解的蛋白质质量。

1.5 质谱样本制备

各取 50 μg 蛋白质样品进行蛋白质酶切。每份样品加终浓度 10 mmol/L DTT, 80°C 金属浴还原反应 10 min 后, 加 40 mmol/L 碘乙酰胺 (IAM), 室温避光烷基化反应 30 min。使用 SDS-PAGE 胶分离, 考马斯亮蓝染色及脱色。将各条带切为大约 1 mm³ 的胶粒。加入适量脱色液 (30% 乙腈 (acetonitrile, ACN), 50 mmol/L NH₄HCO₃) 振荡 30 min, 换脱色液继续振荡, 直至胶粒无色。加入适量 ACN 脱水, 更换 ACN, 重复以上步骤直至胶粒为白色硬颗粒, 最后使用真空干燥机蒸干胶粒。取出干燥的胶粒置于冰上, 加酶解缓冲液 (50 mmol/L NH₄HCO₃, 13.3 mg/L 胰蛋白酶 (trypsin), 6.7 mg/L Lys-C), 冰上溶胀 40 min, 补加酶解缓冲液至没过胶粒。将酶解体系置于 37°C 恒温箱反应 12~16 h。

酶解后, 每管加 100 μL 抽提液 (5% 甲酸 (formic acid, FA), 50% ACN), 13 300 r/min 常温离心 1 min, 静置 4 min, 重复 4 次。将抽提液上清转移到新的 2 ml 离心管中。重复上述步骤 1 次。

加入 ACN 弹匀, 13 300 r/min 常温离心 1 min, 转移上清, 重复上述步骤至胶粒变为白色硬颗粒, 将所有抽提液及 ACN 混匀。使用真空干燥机蒸干样本, -80°C 保存备用。

1.6 液相色谱及质谱参数

溶样缓冲液 (1% ACN, 0.1% FA) 溶解肽段样品, 浓度调整至 0.2 g/L。使用自动进样器自动进样。选择 LTQ Orbitrap Velos 质谱仪 (Thermo Fisher Scientific, Palo Alto, CA) 与 UPLC EASY-nLC 1000 系统 (Thermo Fisher Scientific, Palo Alto, CA) 配备一个自封的毛细管柱 (75 μm × 15 cm, 1.9 μm C₁₈ reverse-phase fused-silica) 联用进行质谱检测。

洗脱条件如下。流动相 A: 98% H₂O+2% ACN+0.1% FA; 流动相 B: 100% ACN+0.1% FA。60 min 的梯度洗脱样品: 0~6 min, B 相由 4% 升至 8%; 6~40 min, B 相从 8% 升至 26%; 40~55 min, B 相从 26% 升至 43%; 55~56 min, B 相从 43% 升至 80%; 58~60 min, B 相维持 80%, 流速 0.4 $\mu\text{L}/\text{min}$ 。洗脱的组分直接通过纳喷离子源 (Nano Spray Ionization, NSI) 进入质谱检测。

参数设置: 一级质谱扫描离子质荷比范围为 300~1 600, 分辨率为 30 000。自动增益控制 (automatic gain control, AGC) 设置为 1×10^6 , 最

大离子注入时间 (maximum ion injection time, MIT) 设置为 50 ms。二级质谱采用碰撞诱导解离 (collision-induced dissociation, CID) 模式碎裂, 选取丰度最高的前 20 个离子进行二级碎裂分析, 碰撞能量归一化为 35%, AGC 的设置为 5×10^4 。动态排除设定为 50 s 避免冗余检测。

1.7 蛋白质组学数据处理

所有的数据原始文件都由 MaxQuant (1.6.17 版) 对照 UniProt 数据库 (2017 年 3 月版, Uniprot, <https://www.uniprot.org/>) 中的蛋白质 FASTA 文件进行搜索。搜库参数: 非标 (label free) 定量方法, 选择 iBAQ 值; 胰蛋白酶的特异性消化; 最多允许 2 次漏切; 最小肽的长度设定为 7 个氨基酸; 前体质量偏差小于 20×10^{-6} ; MS/MS 质量偏差小于 0.5 u; 固定修饰为 Carbamidomethyl (C), 可变修饰为 Oxidation (M), Acetyl (Protein N-term); 肽和蛋白质水平上的错误发现率 (FDR) 小于 1%。使用 Perseus (1.6.6.0 版)^[14] 对所有样品进行基于质谱的定量蛋白质组学实验的统计分析。

2 结果与分析

2.1 PLEE法可高效、稳定提取头发蛋白质

采用低的头发起始量 (10 根 1 cm 长度的头发), 从样本处理方法和裂解液成分两方面优化头发蛋白质提取效率和质量, 流程如图 1a 所示。采用相同起始量的蛋白质裂解液进行 SDS-PAGE 凝胶电泳, 使用考马斯亮蓝染色, 将各方法的条带进行灰度定量量化蛋白质提取量, 以最高的方法为 100%, 根据灰度比例计算各方法的蛋白质提取效率 (图 1b), GH 法和 LN-BH 法的头发蛋白质提取效率较低, 分别为 1% 和 0.48%, 而 RTO 法和 LN-H95 法的头发提取效率则分别为 43.25% 和 13.20%, 高于 GH 法和 LN-BH 法。然而, RTO 法和 LN-H95 法在 3 个重复中的提取稳定性较低。SH-65 法能够以高效率 and 稳定性在 3 次重复中提取蛋白质, 提取率为 66.52%~86.48%。H-90-SH-6 法从头发中高效提取最多含量的蛋白质, 且在 3 个生物重复中具有良好重现性。H-90-SH-65 法提取的平均效率比 SH-65 法高出 22.71%, 与其他提取方法相比, 效率至少增加 1.3 倍。结果显示, 强烈的物理震荡和长时间的煮沸处理明显提高蛋白质的提取效率。比较使用 SDS 缓冲液和尿素缓冲液提取头发蛋白质的效率 (图 1c)。结果表明, SDS 缓冲液提取的蛋白质是尿素缓冲液提取蛋白质的 1.67 倍。

通过比较头发样本处理方法和裂解液成分, 在 SDS 缓冲液的作用下, 结合强冲击和长时间加热, 可高效提取头发中的蛋白质。本研究中将

H-90-SH-65 法和 SDS 缓冲液相结合的技术流程命名为 PLEE (PTM lab for protein extraction from hair with high efficiency) 法。

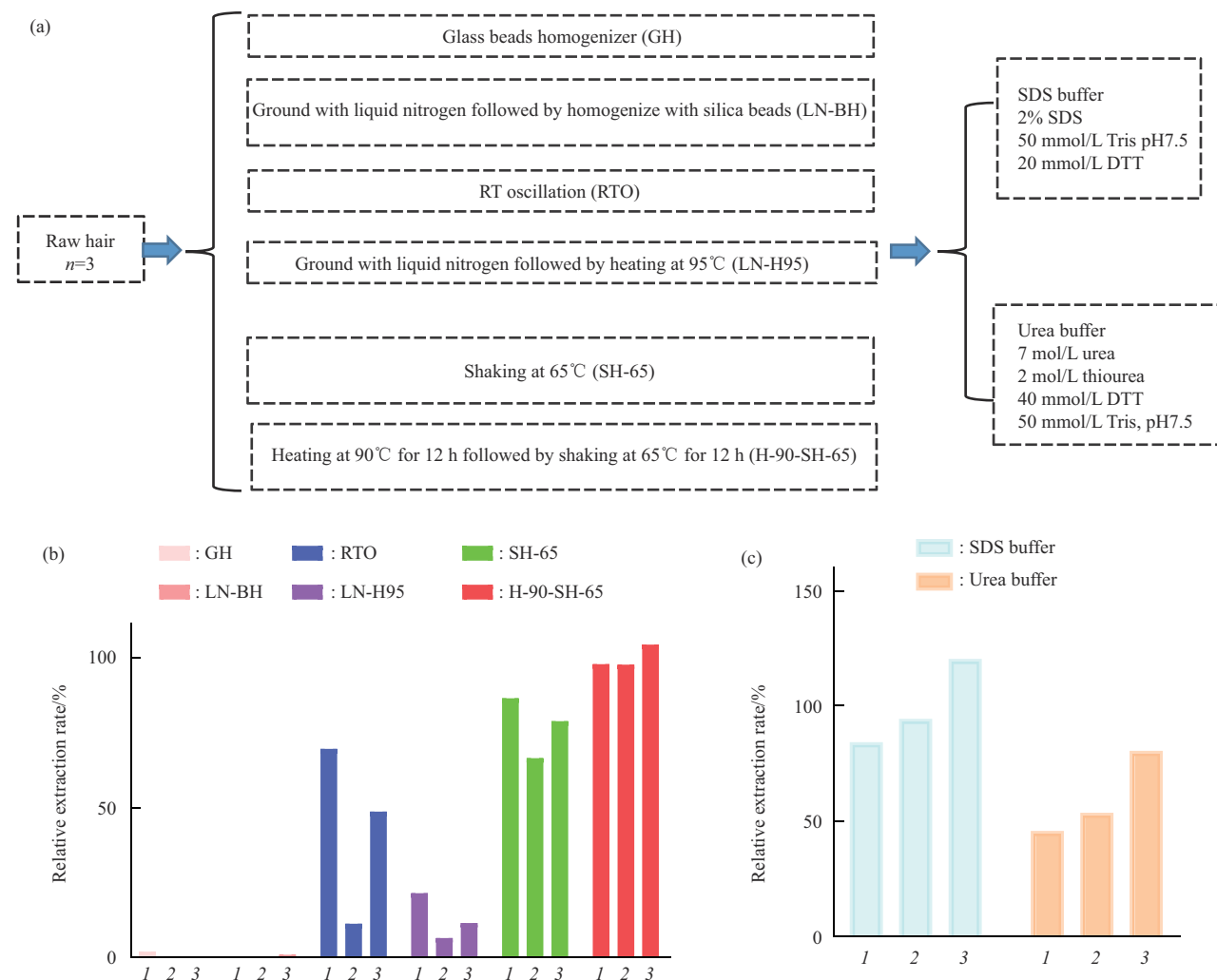


Fig. 1 PLEE method for effective hair protein extraction

(a) Workflow for the development of a highly efficient protein extraction method from hair. (b) Quantitative analysis of the SDS-PAGE to compare the protein extraction efficiency among 6 methods. GH: glass homogenizer; LN-BH: ground with liquid nitrogen (LN) followed by homogenize with Silica beads (BH); RTO: RT oscillation; LN-H95: ground with liquid nitrogen followed by heating at 95°C; SH-65: shaking at 65°C; H-90-SH-65: heating at 90°C for 12 h followed by shaking at 65°C for 12 h; 1, 2, 3 are 3 biological replicates. (c) Quantitative analysis of the SDS-PAGE to compare the protein extraction efficiency among two lysis buffers. 1, 2, 3 are 3 biological replicates.

2.2 基于质谱的蛋白质组学技术可稳定高效鉴定头发蛋白质

基于质谱可稳定、高效地分析 PLEE 法提取头发蛋白质。采集 7 例头发样本, 使用 PLEE 法蛋白质提取, 通过胶内消化和质谱分析获取质谱数据, 其中 P7 样本重复提取检测 4 次, 评估流程稳定性。对产出质谱数据进行严格的质量控制分析, 85% 的肽段二级谱谱图数在 1~2 之间 (图 2a), 82% 的肽

段长度分布在 8~20 个氨基酸残基之间 (图 2b), 99% 的肽段评分分布在 40~380 之间 (图 2c), 99.9% 的肽段质量偏差在 5×10^{-6} 之内 (图 2d)。结果提示, 质谱数据质量较好, 鉴定结果可靠。

蛋白质组学分析头发样本蛋白质鉴定情况 (表 1), 样本鉴定到总蛋白质数量在 169~229 之间, 其中角蛋白和角蛋白相关蛋白的数量在 46~49 之间。对 P7 样本的 4 次技术重复分析, 共同鉴定蛋白质

131种(图3a),重复间的定量值进行分析结果显示各重复间的定量值分布(图3b)及累计曲线(图3c)相似,具有强的相关性系数,均为0.99及

以上(图3d)。表明基于PLEE法和质谱的头发蛋白质组学流程稳定、重复性高。

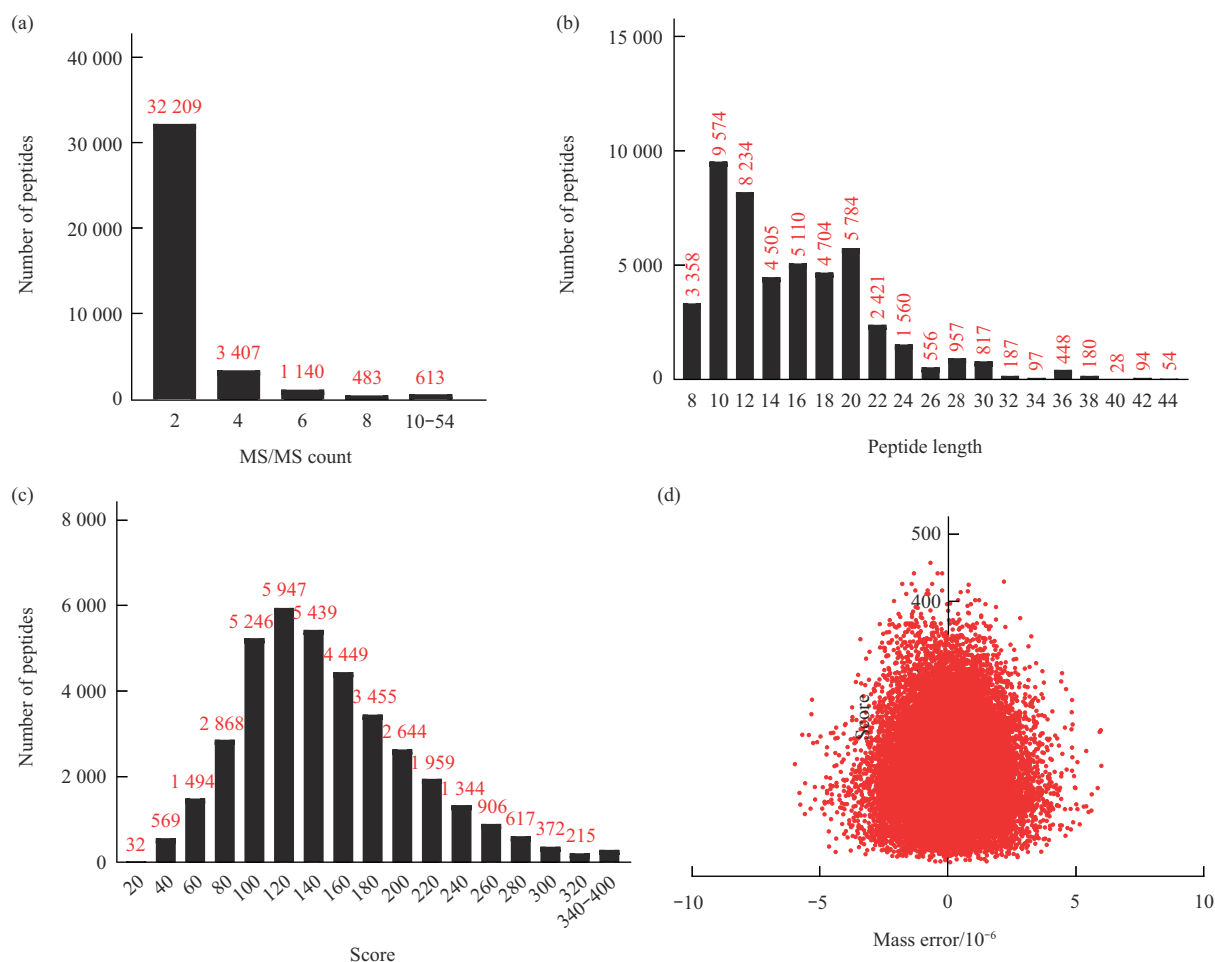


Fig. 2 Mass spectrometry quality control analysis

(a) Number of MS/MS count of peptides. (b) Distribution of peptide lengths. (c) Distribution of score values of peptides. (d) Distribution of mass errors.

Table 1 Search results in MaxQuant based on the PLEE method

Sample	Total spectra	PSM	MS2 identified/%	Protein groups	Peptides	Full-cleavage/%	KRT&KAP	Pep/Prot
P1	47 104	7 134	15.15	207	1 256	76.59	44	6.07
P2	46 178	5 293	11.46	217	1 284	78.89	44	5.92
P3	48 539	5 527	11.39	200	1 199	77.06	47	6.00
P4	47 273	7 307	15.46	229	1 453	77.70	47	6.34
P5	46 067	6 233	13.53	222	1 366	78.48	49	6.15
P6	48 711	6 829	14.02	222	1 458	78.74	49	6.57
P7-R1	57 442	6 623	11.53	169	956	76.67	47	5.66
P7-R2	56 978	6 278	11.02	172	935	76.79	46	5.44
P7-R3	58 281	7 359	12.63	179	1 006	76.94	48	5.62
P7-R4	57 418	7 445	12.97	184	1 004	77.29	46	5.46

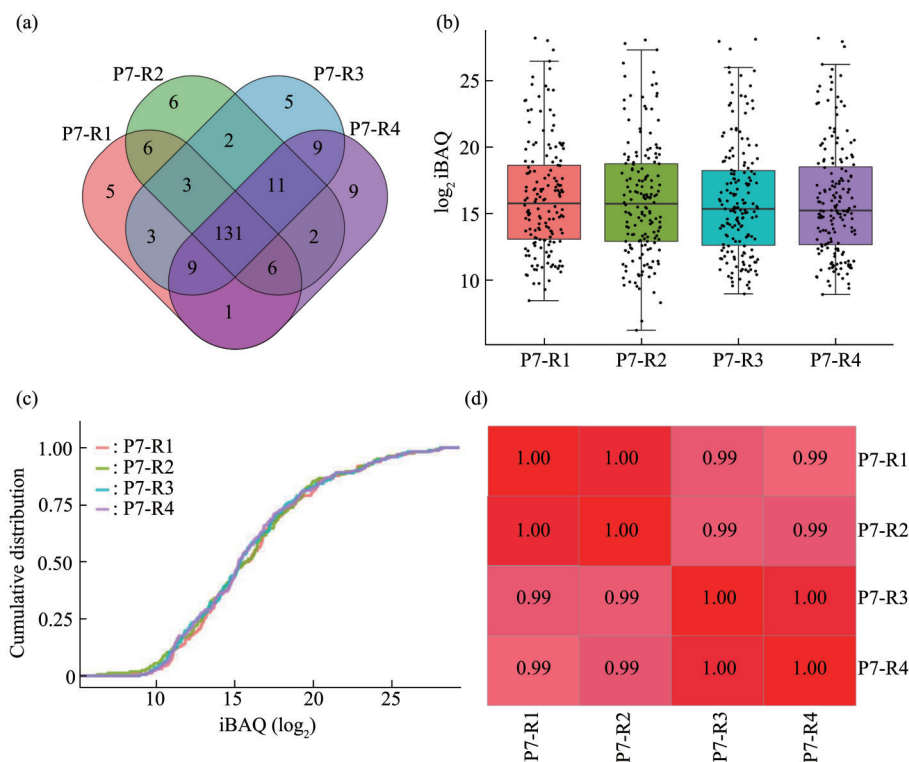


Fig. 3 Proteomics analysis of technical replicates

(a) Analysis of identified protein in 4 technical replicates. (b) Distribution of quantitative values in 4 technical replicates. (c) Cumulative curve of quantitative values in 4 technical replicates. (d) Analysis of correlation between 4 technical replicates.

2.3 个体的头发蛋白质组学差异分析

分析各样本的鉴定蛋白质种类及共鉴定蛋白质分布的情况,发现不同个体的头发蛋白质组成和占比存在差异。10份样本共有蛋白质为107种(图4a),各样本非共有蛋白质种类在57~119之间(图4b),其中5例样本(P1、P3、P4、P5、P7-R4)各鉴定到1种独特蛋白(PARP4、RRP15、SERPINB4、ANK3、MACF1),1例(P6)鉴定到2种独特蛋白(TUBA1C、TPP1)(图4c,表2)。选择共有蛋白质对所有样本进行定量蛋白质组学分析。主成分分析(图4d)及聚类分析(图4e)结果显示,技术重复样本聚为一类,不同样本被区分。以上结果提示,不同个体间存在独特鉴定的蛋白质,且共有蛋白质在不同个体间定量值存在

差异。

筛选在不同个体来源的头发中表达量差异大、在同个体中稳定鉴定的头发蛋白质簇。P1~P6样本中的定量值方差最大的30个蛋白质进行定量值分布情况可视化,其中有17个为KRT或KAP(KRT33A、KRTAP9-6、KRT6A、KRTAP2-2、KRT83、KRTAP3-2、KRTAP7-1、KRTAP19-5等)(图5a)。筛选4个技术重复中定量值方差最小的30个蛋白质(图5b)。同时符合两个条件的10个蛋白为:KRT33A、KRTAP9-6、KRT83、KRTAP7-1、KRT32、BLMH、KRT38、KRTAP11-1、NPAS1、KRTAP4-3(图5c,表3)。这些蛋白质在不同样本间的分布存在显著差异,在同个体的不同技术重复内稳定鉴定。

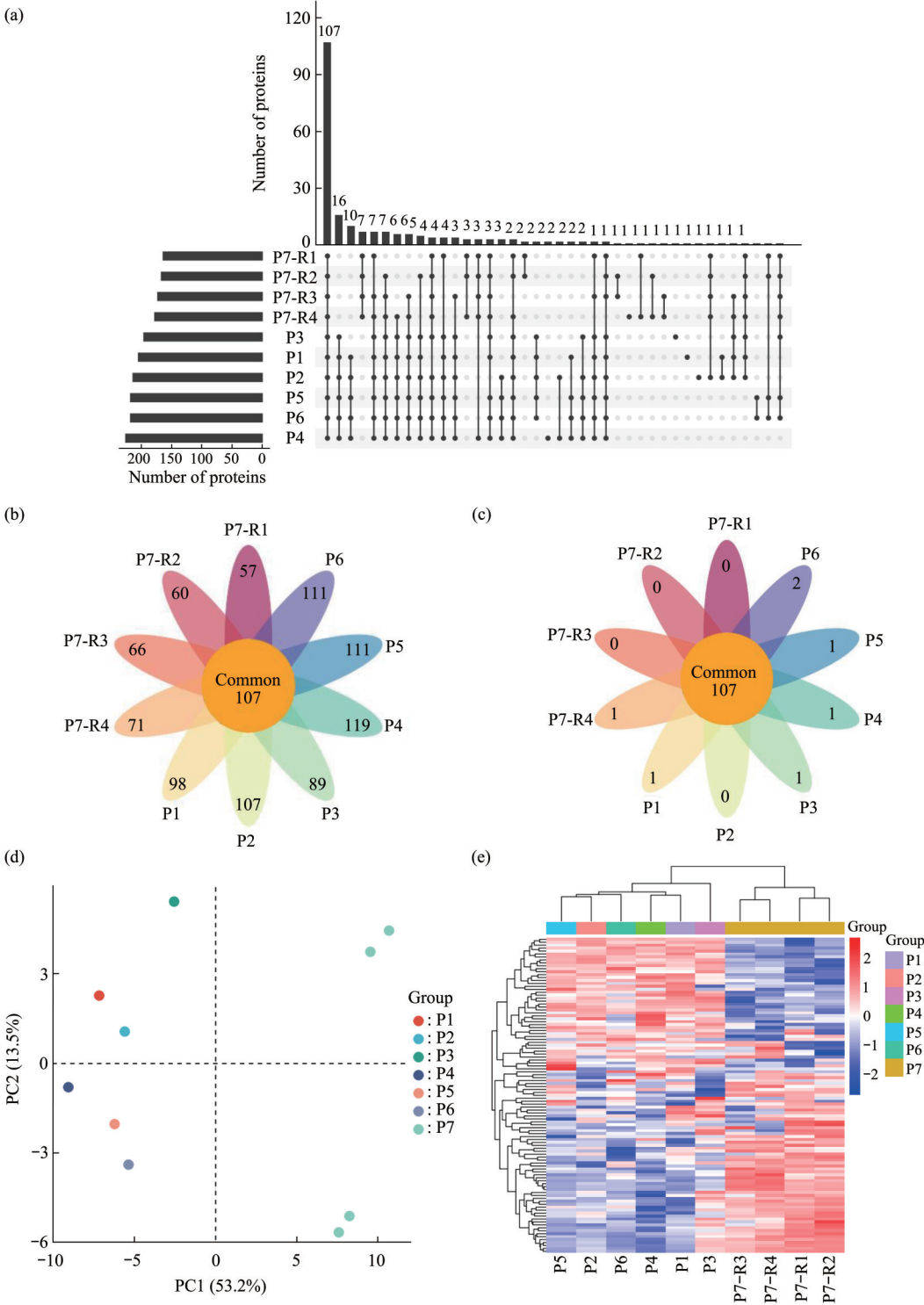


Fig. 4 Proteomics analysis among different samples

(a) Analysis of identified protein in 10 samples. (b) Petal plot showing the number of shared proteins and the number of non-shared proteins after removal of shared proteins in 10 samples. (c) Petal plot showing the number of shared proteins and the number of independently existing proteins in each of the 10 samples. (d) PCA analysis based on 107 shared proteins. (e) Heat map clustering based on 107 shared proteins.

Table 2 Protein statistics uniquely present in a sample			
No.	Group	Protein name	Gene name
1	P1	Poly [ADP-ribose] polymerase 4	<i>PARP4</i>
2	P3	RRP15-like protein	<i>RRP15</i>
3	P4	Serpin B4	<i>SERPINB4</i>
4	P5	Ankyrin-3	<i>ANK3</i>
5	P6	Tubulin alpha-1C chain	<i>TUBA1C</i>
6	P6	Tripeptidyl-peptidase 1	<i>TPP1</i>
7	P7-R4	Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5	<i>MACF1</i>

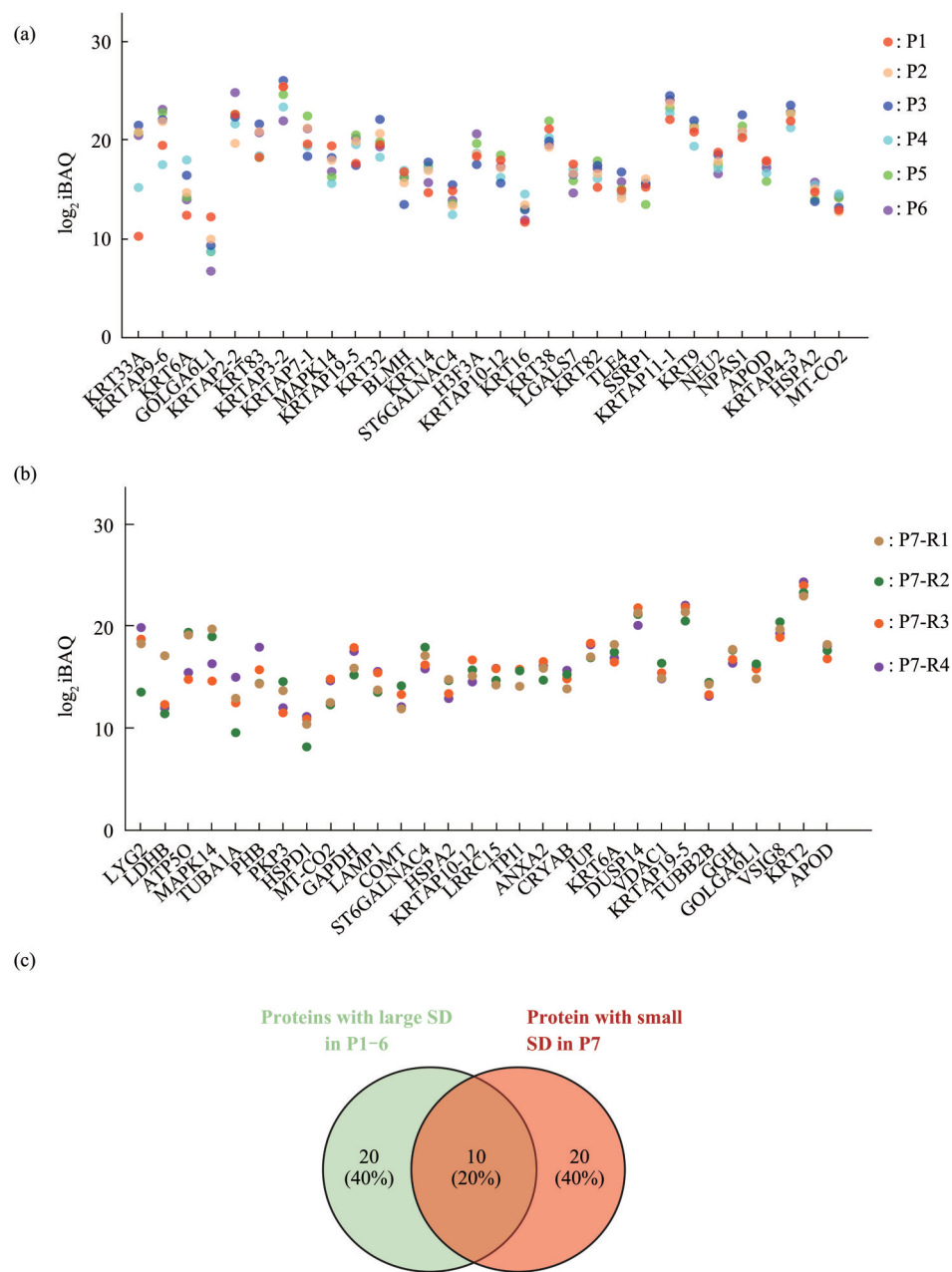


Fig. 5 Screening of differentially expressed proteins among different samples

(a) Distribution of quantification values for the top 30 proteins with the highest SD value in different individual samples of shared proteins. (b) Distribution of quantification values for the top 30 proteins with the lowest SD value in technical replicates of shared proteins. (c) Venn diagram statistics of the proteins screened in (a) and (b).

Table 3 Statistics of proteins expressed stably between the same individuals and expressed differently between different individuals

Gene name	log ₂ iBAQ									
	P1	P2	P3	P4	P5	P6	P7R1	P7R2	P7R3	P7R4
KRT33A	27.69	27.91	28.74	27.33	27.78	27.45	29.18	29.11	28.75	28.93
KRTAP9-6	19.51	21.94	22.12	17.54	22.83	23.16	23.99	24.39	23.70	24.09
KRT83	18.31	20.89	21.67	18.46	18.26	20.78	20.93	21.04	20.73	21.18
KRTAP7-1	19.64	21.25	18.40	19.42	22.50	21.19	22.42	22.37	22.20	22.43
KRT32	19.58	20.70	22.14	18.31	19.86	19.36	21.34	21.23	21.42	21.68
BLMH	16.82	15.70	13.51	16.98	16.23	16.28	14.35	14.42	15.04	14.84
KRT38	21.17	19.32	19.94	20.31	22.00	19.55	20.06	20.19	19.97	20.06
KRTAP11-1	22.11	23.79	24.55	22.77	23.28	24.10	25.76	26.08	25.49	25.58
NPAS1	20.29	20.92	22.59	20.58	21.48	20.94	22.41	22.69	22.15	22.23
KRTAP4-3	21.96	22.70	23.59	21.28	22.75	22.88	23.97	24.21	23.49	23.97

3 结 论

追踪人员身份的方法包括DNA检测^[15]、指纹识别^[16]、面部识别、虹膜识别^[17]等。然而,这些方法都有一定的限制和缺陷。例如,在DNA检测中,样本污染或者样本量不足导致结果不准确,指纹识别可能会出现误判或者无法识别,面部识别和虹膜识别会受到光线、角度等因素的影响,导致识别失败^[18-19]。并且这些技术都需要预先录入人员的信息,才能通过比对确定身份。在刑侦和考古工作中,经常出现追踪人员身份的困难。侦查人员搜集到的犯罪人员信息有限,考古人员在工作中存在人员身份、年代确认的问题。在现场,往往能追寻头发样品的踪迹。头发可长时间稳定储存,可记录身体一段时间内的状态^[20-22]。头发的特殊蛋白质和结构成分使蛋白质在室温下保持稳定,对头发蛋白质组进行更多的生物学信息挖掘可作为区分个体特质的有效手段,从而进行身份追踪^[23]。

为提高头发蛋白质的提取率和提取质量,对样本处理方法和蛋白质裂解缓冲液进行优化。结果显示,与SH-65方法相比,H-90-SH-65方法的提取率增加22.71%,比另外4种方法至少增加1.3倍。强烈的物理震荡和长时间煮沸都有助于有效地从头发中提取蛋白质。SDS缓冲液的蛋白质提取效率为尿素缓冲液的1.67倍。且PLEE法在3个生物学重复的提取效率具有高度的重现性,表明PLEE法对头发蛋白质的提取效率高且稳定。

通过对质谱的质量检测及技术重复分析,表明该流程稳定且高效的。对不同个人的特质头发蛋白

质组学数据进行分析。在10份样本共有的蛋白质为107种。每个样本中还含有57~119种非共有蛋白质,其中部分样本鉴定出1~2种独特蛋白质。这提示,对个体特异性蛋白质的鉴定有望作为个人特质的区分。另外,发现在样本间表达存在较大差异的蛋白质,且在同个体的技术重复间稳定表达的10个蛋白质:KRT33A、KRTAP9-6、KRT83、KRTAP7-1、KRT32、BLMH、KRT38、KRTAP11-1、NPAS1、KRTAP4-3。这些蛋白质有望作为区分个体特质的关键蛋白。

综上所述,基于PLEE方法的头发蛋白质组学提供不同个体之间更详细的生物学信息,表明角蛋白及角蛋白相关蛋白是至关重要的一类蛋白质,可以用于区分个人特质。基于PLEE法的头发蛋白质组学研究将是下一代强大而有用的工具。还可用于洗涤剂有效性评估和头发健康评估。基于不同表达的蛋白质簇用于区分个体特质的灵敏性、准确性还有待进一步探索。

致谢 感谢北京国家蛋白质科学中心的孙浩帆博士在实验中提供的帮助。

参 考 文 献

- [1] Müller-Röver S, Foitzik K, Paus R, *et al.* A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. *J Invest Dermatol*, 2001, **117**(1): 3-15
- [2] Popescu C, Hocker H. Hair-the most sophisticated biological composite material. *Chem Soc Rev*, 2007, **36**(8): 1282-1291
- [3] Barthelemy N R, Bednarczyk A, Schaeffer-Reiss C, *et al.* Proteomic tools for the investigation of human hair structural proteins and evidence of weakness sites on hair keratin coil

- segments. *Anal Biochem*, 2012, **421**(1): 43-55
- [4] Lubec G, Nauefl G, Seifertb K, *et al.* Structural stability of hair over three thousand years. *J Archaeol Sci*, 1987, **14**: 113-120
- [5] Semwal B, Agrawal K, Singh K, *et al.* Alopecia: switch to herbal medicine. *J Phar Rev Opinion*, 2011, **1**(4): 101-104
- [6] Siegel J A, Mirakovits K. *Crime scene investigation*//Mirakovits K, Siegel J A. *Forensic Science*. Boca Raton: CRC Press, 2021: 41-65
- [7] Zhang Z, Burke M C, Wallace W E, *et al.* Sensitive method for the confident identification of genetically variant peptides in human hair keratin. *J Forensic Sci*, 2020, **65**(2): 406-420
- [8] Laatsch C N, Durbin-Johnson B P, Rocke D M, *et al.* Human hair shaft proteomic profiling: individual differences, site specificity and cuticle analysis. *PeerJ*, 2014, **2**: e506
- [9] Parker G J, Leppert T, Anex D S, *et al.* Demonstration of protein-based human identification using the hair shaft proteome. *PLoS One*, 2016, **11**(9): e0160653
- [10] Nasir N M, Hiji J, Jayapalan J J, *et al.* Potential use of human hair shaft keratin peptide signatures to distinguish gender and ethnicity. *PeerJ*, 2020, **8**: e8248
- [11] Kempson I M, Lombi E. Hair analysis as a biomonitor for toxicology, disease and health status. *Chem Soc Rev*, 2011, **40**(7): 3915-3940
- [12] Feroz S, Muhammad N, Ratnayake J, *et al.* Keratin-based materials for biomedical applications. *Bioact Mater*, 2020, **5**(3): 496-509
- [13] Agarwal V, Panicker A G, Indrakumar S, *et al.* Comparative study of keratin extraction from human hair. *Int J Biol Macromol*, 2019, **133**: 382-390
- [14] Tyanova S, Temu T, Sinitcyn P, *et al.* The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nat Methods*, 2016, **13**(9): 731-740
- [15] Chakraborty R, Jin L. Determination of relatedness between individuals using DNA fingerprinting. *Hum Biol*, 1993, **65**(6): 875-895
- [16] Bose P K, Kabir M J. Fingerprint: a unique and reliable method for identification. *J Enam Med Col*, 2017, **7**(1): 29-34
- [17] Wildes R P. Iris recognition: an emerging biometric technology. *IEEE Proceedings*, 1997, **85**(9): 1348-1363
- [18] Kaur G, Singh G, Kumar V. A review on biometric recognition. *Int J Biotech Biosci*, 2014, **6**(4): 69-76
- [19] Jain A K. Biometric recognition. *Nature*, 2007, **449**(7158): 38-40
- [20] Zhang H, Chai Z, Sun H. Human hair as a potential biomonitor for assessing persistent organic pollutants. *Environ Int*, 2007, **33**(5): 685-693
- [21] Usman M, Naseer A, Baig Y, *et al.* Forensic toxicological analysis of hair: a review. *Egypt J Forensic Sci*, 2019, **9**: 17
- [22] Angerer J, Ewers U, Wilhelm M. Human biomonitoring: state of the art. *Int J Hyg Environ Health*, 2007, **210**(3-4): 201-228
- [23] Trüeb R M. The impact of oxidative stress on hair. *Int J Cosmet Sci*, 2015, **37**: 25-30

A Methodological Investigation of Hair Proteomics-based Differentiation of Individual Traits*

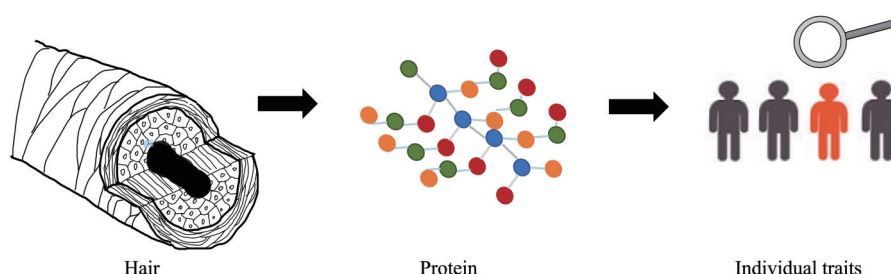
WU Xiao-Lin^{1,2)}, ZHANG Tao²⁾, XU Ping^{1,2,3)}, ZHANG Ya-Li¹⁾**, ZHANG Zhen-Peng²⁾**

⁽¹⁾School of Medicine, Guizhou University, Guiyang 550025, China;

²⁾State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Institute of Military Medical Sciences, Academy of Military Sciences, Beijing 102206, China;

³⁾Research Unit of Proteomics & Research and Development of New Drug, Chinese Academy of Medical Sciences, Beijing 102206, China)

Graphical abstract



Abstract Objective Hair is an essential skin appendage, primarily composed of keratins and keratin-associated proteins. The protein composition and proportion of hair samples vary among different races and sexes. Currently, there is a lack of efficient methods to extract hair proteins. This study aims to explore the application of quantitative hair proteomics in distinguishing individual hair characteristics. **Methods** Based on the exploration of sample processing and lysis buffer using three hair samples, we developed a stable and efficient hair protein extraction method, named PLEE (PTM lab for protein extraction from hair with high efficiency). We used the PLEE method to extract seven human hair samples and performed proteomic experiments on them using in-gel digestion method to produce data for analyzing hair protein composition and proportion among individuals. **Results** A total of 274 proteins were identified, among which 107 proteins were commonly present, and the number of non-common proteins ranged from 57–119, with some samples having unique identification proteins. Using the 107 commonly identified proteins for quantitative protein fractionation analysis, various samples were distinguished by clustering and principal component analysis, and technical repeated samples were merged, indicating the stability of the process. In addition, 10 key proteins (KRT33A, KRTAP9-6, KRT83, KRTAP7-1, KRT32, BLMH, KRT38, KRTAP11-1, NPAS1, KRTAP4-3) with large differences between individuals and stable protein identification within the same individual were screened. **Conclusion** The protein composition of hair varies among different individuals, and the 10 selected proteins are expected to be key proteins for distinguishing individual hair characteristics and have significant potential applications in individual identification and criminal investigation.

Key words hair, proteomics, individual traits, PLEE

DOI: 10.16476/j.pibb.2023.0165

* This work was supported by a grant from National Key Research and Development Plan (2020YFE0202200).

** Corresponding author.

ZHANG Zhen-Peng. Tel: 86-15898965955, E-mail: zhangzp1002@163.com

ZHANG Ya-Li. Tel: 86-13639069576, E-mail: 286802781@qq.com

Received: April 23, 2023 Accepted: June 20, 2023