Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Keratin

Journal:	*Journal of Forensic Sciences*
Manuscript ID	JOFS-19-387.R2
Manuscript Type:	Paper
Keywords:	Forensic Science, Genetically Variant Peptide, Hair Protein Extraction, Cuticular Keratins, Peptide Mass Spectral Library, Trace Detection

ABSTRACT

Recent reports have demonstrated that genetically variant peptides derived from human hair shaft proteins can be used to differentiate individuals of different biogeographic origin. We report a method involving direct extraction of hair shaft proteins more sensitive than previously published methods regarding GVP detection. It involves one-step for protein extraction and was found to provide reproducible results. A detailed proteomic analysis of this data is presented that led to the following four results: 1) A peptide spectral library was created and made available for download. It contains all identified peptides from this work, including GVPs that, when appropriately expanded with diverse hair-derived peptides, can provide a routine, reliable and sensitive means of analyzing hair digests; 2) An analysis of artifact peptides arising from side reactions is also made using a new method for finding unexpected modifications; 3) Detailed analysis of the gel-based method employed clearly shows the high degree of crosslinking or protein association involved in hair digestion, with major GVPs eluting over a wide range of high molecular weights while others apparently arise from distinct non-crosslinked proteins; 4) Finally, we show that some of the specific GVP identifications depend on the sample preparation method.

KEYWORDS

Forensic Science, Genetically Variant Peptide, hair protein extraction, cuticular keratins, peptide mass spectral library, and trace detection

In recent publications from Lawrence Livermore National Laboratory (LLNL), genetically variant peptides (GVPs) derived from human hair have been shown to have forensic value (1,2). The publication (1) by Parker et al. showed that these peptides might serve as a source of evidence in addition to DNA for human identification due to several advantages that a hair sample carries: 1) commonly found – on average, humans shed 50 – 150 hairs per day; 2) stable – proteins in a hair sample usually last longer and are more resistant to degradation than DNA; 3) when good quality DNA is not available, hair proteins may serve as alternative evidence by detecting those GVPs in hair cuticular keratins and other hair proteins. A recent publication (2) by Mason et al. described protein-based or GVP-based human identification from a single hair as short as 1 inch-long. Another recent publication (3) by Carlson et al. described a sensitive method to extract proteins from 1-millimeter or less in total length of human anagen head hairs, and compared the proteins identified from hair shaft and hair root. The effectiveness of this method for detecting GVPs has not yet been determined.

The human hair shaft is made up of three main components (4). Starting from the center, the first component is the medulla which is rich in cross-links and highly insoluble. Next is the cortex which comprises most of the hair shaft and is made up of hair cuticular keratin fibrils as well as keratin-associated proteins. The thin outer layer is the cuticle which is also composed of keratin-associated proteins and is the component that would be visually inspected through microscopic examination. Hair cuticular keratins have been classified as type I (31-38) and type II (81-86) based on the finding that type I keratins are acidic and type II keratins are neutral or basic proteins (5,6). Two recent publications (1, 2) from LLNL have collectively identified a total of 88 GVP sites from multiple donors with bulk of hair samples: 32 sites from hair cuticular keratins, 7 sites from cytoskeletal keratins, 22 sites from keratin associated proteins, and 27 sites from non-keratins.

Based on these findings, a human hair sample has the potential to serve as alternative evidence for human identification if GVPs in hair keratins (mainly cuticular keratins), keratin associated proteins and other non-keratin hair proteins can be sensitively and reliably identified. To detect them, we first need an efficient method to extract proteins from human hair shafts. However, hair protein extraction is especially difficult due to extensive cross-linking and poor solubility of hair keratins (7,8,9). In this manuscript, we describe a direct protein extraction method (referred as the Direct method) that can efficiently extract hair proteins from a single hair shaft less than 1 cm in length. We performed GVP panel analyses and examined experimentally-introduced artifactual modifications among three methods: our newly developed Direct method and two of previously published methods – NaOH-based SDS repeated extraction method (we modified it to make it fit in small sample analysis, referred as modified NaOH+SDS method) (8) and ProteaseMax-based method (referred as Cleavable Surfactant method) (1,2). Considering the Direct method and modified NaOH+SDS method both utilize protein gel electrophoresis to separate extracted proteins, we made further comparisons between these two in-gel methods for sensitivity and reproducibility. We find that the Direct method is both sensitive and relatively convenient to carry out while generating reproducible results regarding to GVP detection from a single hair shaft from one individual donor. In the analysis of this data, we applied a number of proteomic data analysis methods including: 1) The development of a library of peptide ion spectra containing all identified peptides that, when extended, can contain all identifiable peptides from hair proteins. Spectral libraries provide a sensitive and reliable means of peptide identification and ultimately can contain spectra of all known GVPs. 2) Proteomic analysis that enable the detailed analysis of artifact peptides, generated by undesirable chemical analysis which can, in principle, lead to false positive analysis. 3) A gel-based method of analysis that reveals a wide distribution of molecular weights of proteins yielding keratin-based GVPs. 4) The finding that different digestion methods can identify different GVPs, suggesting the inadequacy of any current method of finding all potentially identifiable GVPs in a hair sample.

Materials and Methods

Human Hair Sample Preparation

Human hair samples were obtained commercially from BioreclamationIVT (LOT# BRH1363732, 5g of hair shaft per package from the same individual donor). Most of the results presented in this manuscript are derived from hair shafts from this single randomly selected donor: Asian male, 30 years old. Hair samples were briefly washed with 20% methanol and water, then dried and stored at -20°C. The related protocols have been reviewed and approved by National Institute of Standards and Technology (NIST) Human Subjects Review Board.

Direct Extraction Method

Hair shaft samples (5cm, 2.5cm, or 1cm) were cut using sterile laboratory scissors and then combined with 50 µl of the commercially obtained NuPAGE Lithium dodecyl sulfate (LDS) Sample Buffer (Catalog # NP0007, ThermoFisher Scientific) and 50 mmol/L reducing agent dithiothreitol (DTT). After heating the hair shaft in sample buffer at 90 °C for various lengths of time, extracted hair proteins (we call this the Direct method) were loaded onto NuPAGE 4-12% Bis-Tris Protein Gels (Catalog # NP0321, ThermoFisher Scientific) and then separated by size together with a Molecular Weight (MW) Standard (MW std) using sodium dodecyl sulfate - Polyacrylamide Gel Electrophoresis (SDS-PAGE) at 200 V for 30 minutes. The protein gel was stained with SimplyBlue SafeStain (Catalog # LC6060, ThermoFisher Scientific) for one hour. After overnight immersion in water, the destained-protein-containing gel was scanned, and intensities of the main bands were determined. From top to bottom, the gel was evenly cut in 10 fractions (about 4 mm-long per fraction) and in-gel-digestion was performed for each fraction by following a well-established in-gel-digestion protocol (10). Peptide concentrations were measured by a kit provided by Pierce (Quantitative Colorimetric Peptide Assay Kit, Catalog # 23275) after desalting by ZipTip (Catalog # ZTC18S960, EMD Millipore Corporation). Desalted peptides were injected to a Thermo Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. A simplified Direct method workflow is shown in Supplementary Document S1.

We performed a time course study to determine the optimal heating time for extracting hair proteins by this Direct method using six individual 5 cm-long hair shafts with each one processed at a different incubation time in the same amount of sample buffer (Fig. 1). The six different incubation times were: 5, 10, 15, 30, 60 and 90 min with net peptide yields measured by combining all ten fractions. The largest yield of peptides was found to occur at 30 minutes and was selected as the optimal incubation time. Note that the LDS sample buffer was unchanged at a pH of 8.5 through all incubation times. As Fig. 1A shows, we observed two distinct bands: the first was found to be enriched in type II (basic) hair cuticular keratins (Gene Name: KRT81 to 86, # Amino Acids: 486 to 600, MW 53.5 to 64.8), and the second enriched in type I (acidic) hair cuticular keratins (Gene Name: KRT31 to 38, # Amino Acids: 404 to 467, MW 45.9 to 52.2) (8). The orange thin lines in Fig. 1A also indicate an even fractionation of the gel in 10 slices per lane from top to bottom as F1 to F10. Fraction 6 (F6) contains the first main band which enriches type II cuticular keratins and fraction 7 (F7) contains the second main band which enriches type I cuticular keratins (discussion of this observation can be found in the Results and Discussion section). Fig. 1B shows the density reports of type I and type II bands at each time interval, reaching a maximum at 30 min (Fig. 1B), consistent with the time for maximum peptide yield described above. Fig. 1C shows the density ratios of all ten fractions obtained at 30 min, using F1 as the reference. The maximum is at F6, which is used as a keratin-enriched representative fraction. Fig. 1C indicates that the gel-based method both concentrates known GVP-rich keratin proteins and shows the hitherto unknown distribution of apparently crosslinked proteins.

We note that additional studies are needed to understand both the effect of heating and the influence of cysteine alkylation and other chemical processing details on peptide yields.

Modified NaOH-based SDS Repeated Extraction Method

To examine our newly developed Direct method, we compared it to a previously published NaOH-based SDS repeated extraction method (8). We modified the published protocol to fit the purpose of protein extraction from a single hair shaft. The modified work flow was performed as follows (also illustrated in Supplementary Document S1): 1) first, we used bead milling for sample preparation instead of incubation with lysis buffer: 5 cm-long hair shafts are ground by a bead mill (OMNI Bead Ruptor 24 Elite, OMNI-International Inc.) repeatedly (3 cycles, 30 second grinding at the speed of 5 m/s and 30 second dwell); 2) next, ground hair samples are incubated with a NaOH-based lysis buffer that contains SDS and beta-mercaptoethanol (ßME) for three cycles according to published (8) protocol and in each cycle, the hair residue is recycled through the process with bead milling; 3) pooled supernatant containing hair proteins are precipitated with acetone; 4) pellets from protein precipitation and leftover hair debris are combined for downstream SDS-PAGE; 5) in-gel-digestion was used to generate peptides.

Hair Peptide Mass Spectral Library Construction Including Published GVPs

Using the mass spectral library construction pipeline described in the literature (11), the raw mass spectral data files generated in the present studies were used to construct a hair-specific peptide mass spectral library. This relatively small library contains 6280 spectra (6280 peptide ions of 4343 distinct peptides, higher-energy collisional dissociation (HCD) =30eV), and among these – a total of 3754 spectra (3754 peptide ions of 2240 distinct peptides, HCD=30eV) arose from hair keratins or keratin associated proteins - using the National Center for Biotechnology Information (NCBI, downloaded March 2017) human protein FASTA file with 20,183 sequences plus additional 51 published GVP sequences (1). This provides a sequence coverage of hair cuticular keratins of about 70%. Of these spectra, 40 mass spectra are identified as GVP ions which cover 14 published GVP sites (a subset of total 88 published GVPs): 10 sites from hair cuticular keratins, 1 site from a keratin-associated protein, and 3 sites from non-keratin proteins. Detailed information can be found in the Results and Discussion section where we discuss GVP panel analysis.

Spectrum Library Searching

Freely available MSPepSearch software (peptide.nist.gov) (11) was used to perform mass spectral library searching using a precursor ion tolerance of 20 ppm (ppm was defined as parts per million) and fragment ion tolerance of 50 ppm. Label-free HCD human tryptic peptide spectral libraries (version September 23, 2016 contains 1,127,970 spectra, indicated as 'main' library) are available online (peptide.nist.gov) (12). A hair specific peptide spectral library (indicated as 'hair' library) (13) was created from 90 raw mass spectral data files generated during method development of processing 16 five cm-long hair shafts of this same individual Asian donor. Surprisingly, 40% of peptides contained in this 'hair' library were not present in the 'main' library even though it was constructed from a wide range of publicly available data files. Clearly hair was not a common protein-containing material in these studies. This 'hair' library was used in combination with the 'main' library for mass spectrum library searching. The 1% false discovery rate (FDR) level was determined by using the target-decoy method described in the literature (14,15). The NIST formatted mass spectral libraries were built using the program Lib2NIST freely available online at chemdata.nist.gov. This library and associated software are freely available online (13).

Sequence Database Searching

We used the Sequest (16) HT search node implemented in Proteome Discoverer (PD) 2.1 for initial peptide identification prior to entry into a library and comparison the results of spectral library searching. Mass tolerance settings were the same as in the library searches. The top scoring peptide identification was selected, and FDR level was set at 1% using the same FASTA file described above.

Proteomics Methods

GVP and its non-variant form designation: In this work, GVPs are tryptic peptides that are represented first by their Gene Name followed by the site of the amino acid substitution. For example, "DSP R1783Q_Q" indicates the tryptic peptide derived from Desmoplakin (GN=GSP) containing "Q" at position 1783. The corresponding non-variant form is "DSP R1738Q_R" where "R" is in place of "Q". The term "GVP ion" refers to not only tryptic peptide sequence, but also charge state and possible modifications. Peptides observed in different charge states or modifications are treated as different peptide ions. The most abundant form of a peptide ion is used to measure its intensity.

LC-MS/MS parameters: Digests were analyzed on an Eksigent Classic 2D Nano LC with an Acclaim PepMap RSLC column (75 µm x 15 cm, C18, 2 µm, 100 Å) with a nanospray source connected to a Thermo Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer in the positive ion mode. Mobile phase A consisted of 0.1% formic acid in water and mobile phase B consisted of 0.1% formic acid in Acetonitrile. The peptides were eluted by increasing mobile phase B from 1% to 90% over 200 minutes. Data was collected using a data dependent mode with a dynamic exclusion of 20 seconds. The top 10 most abundant precursor ions were selected from a 350-1600 m/z full scan for fragmentation. The resolution of full MS scan was set at 120,000 and the resolution of MS/MS scan was set at 30,000. In future work, we plan to perform a 2D-LC study to find more trace ions.

Modifications included in hair library are: (1) fixed carbamidomethyl (CAM) at Cysteine (C); (2) oxidation at Methionine (M); (3) acetylation (Acetyl) at peptide N-terminus; (4) acetaldehyde at peptide N-terminus; (5) Gln->pyro-Glu at Glutamine (Q) at peptide N-terminus; (6) Glu->pyro-Glu at Glutamic Acid (E) at peptide N-terminus. Other less abundant modifications may be added to future versions of the library, although these may be depended on the specific chemical processing involved in the digestion.

Incomplete digestion in proteomics: The inability to digest substantial portions of the proteome is common for the proteomics of biological material. Here are some examples: 1) In reference 8, the reference for the original NaOH+SDS method, hair pellets were simply discarded after incubation with lysis buffer containing NaOH+SDS; 2) In reference 9, scanning electron microscope images as Fig. 2 to show remaining undigested hair after extraction with SDS or with urea. In case 1 and 2, substantial portions of the hair undigested although it is method dependent; 3) In reference 17, heavy-isotope-labeled proteins were used to compare peptide recovery between laboratories and results showed that the digestion step was the greatest source of inconsistent recovery (median loss of 70%). These examples demonstrate that significant levels of incomplete digestion are expected in the proteomics of biological materials.

Results and Discussion

Identification of Hair Proteome including Cuticular Keratins by Direct Extraction Method

We examined overall protein and peptide identifications from all ten gel fractions and compared our library search results to the results from sequence (Sequest) searches. When searching spectral libraries, we added the 'hair' specific mass spectral library to our 'main' library (12,13) to obtain better search performance. The next A and B sub-sections discuss these results and demonstrate the effectiveness of spectral library searching for peptide identification. In sub-section C, we examine GVP detection with library searching in all ten fractions and compare the GVP panel analysis by the Direct method to the other two published methods (1,8).

A. Overall Gel Identification

Results for hair proteins extracted from a single 5 cm-long hair by the Direct method are presented in Table 1. They were derived from one raw MS data file for each of the ten gel fractions. All were independently analyzed to determine details of the gel separation and digestion process.

Using both spectral library and Sequest searching methods, results derived from F1 to F10 are compared in Table 1. As shown in Table 1, when the 'main' library was combined with the 'hair' library for spectral library searching, the overall library identification for proteins - for both hair proteome (7,9) and hair cuticular keratins (a major subset of the hair proteome) (1,8) was similar to that from Sequest, however for all peptides identified, the spectral library method was somewhat more sensitive at a given FDR level, consistent with previous observations (14).

Hair cuticular keratins are major components of hair proteome. Table 2 examined the sequence coverage of listed total 15 hair cuticular keratins of type I and type II by library and Sequest searches from all ten fractions. Peptides present in multiple proteins were used in calculating the sequence coverage of each protein. Since we are interested in GVPs, of course the better coverage, the greater the chance of detecting potential GVP sites. In general, library searching provides a fuller coverage than database searching, although except for the most abundant KRT31, some of these coverages are far less than 100%. There are several possible reasons for this: 1) cross-linking makes certain sites hard to reach by trypsin during the digestion; 2) extremely long (> 50) or short (< 6) peptides were not considered under the current search parameters; 3) loss of extremely hydrophilic or hydrophobic peptides occurs during sample preparation and LC analysis. 4) Incomplete conversion of proteins to peptides is common throughout proteomics, and according to reference 18, an approximately 70–80% of recovery is expected after extraction from the gel. Putting all ten fractions together, 8 out of 15 hair cuticular keratins reach more than 90% coverage, 5 out of the rest 7 reach more than 50%, and only 2 less than 50% (KRT37 and KRT84). Supplementary Document S2 shows sequence coverage in amino acids of 15 type I and type II hair cuticular keratins found by library and Sequest searches.

B. Major and Minor Gel Band Identification

We observed two distinct gel bands in fractions 6 and 7 (Fig. 1). The other fractions had several minor bands but most of the intensity was evenly distributed (Fig. 1C). Results are discussed below.

Fig. 2 shows the intensities over the fractions for selected peptides from type I (A) or type II (B) hair cuticular keratin. In both cases, both the GVP and non-variant form are shown along with another major peptide from each protein. The abundance of each peptide derived from its MS1 ion chromatogram peak area. These results indicate: 1) the major gel bands correspond to type I (fraction 7) and type II (fraction 6) hair cuticular keratins, consistent with literature (8) reports. Fractions 6 (type II) and 7 (type I) are enriched in individual hair cuticular keratins; 2) it is noteworthy that most peptides identified outside the main regions were the same as those inside that region. This behavior persisted in all analyses. This is presumably due to presence of significant quantities of cross-linked proteins or unseparated complexes with higher molecular weight with lower mobilities as well as fragments of these proteins at lower molecular weights with higher mobilities. We find that keratin GVPs are found in virtually all gel fractions suggesting that they distributed among a wide range of crosslinked proteins, suggests that the insoluble, crosslinked portion of the hair protein may not contain additional keratin-GVP identifications. According to reference 7, the insoluble, crosslinked portion has a higher content of non-keratin proteins and may contain additional non-keratin-GVP identifications. Further, we know of no way to enhance the method's digestion effectiveness, though such an improvement would be very welcome.

Note that in Table 1, fractions 6 and 7 show the highest peptide signal strengths but lowest numbers of peptide identifications (IDs). This is confirmed in Fig. 3, where the total ion currents (TICs) are inversely correlated with peptide IDs with a correlation coefficient of -0.75. This is a consequence of the higher concentrations of relatively a few proteins dominating fractions 6 (type II) and 7 (type I), which leads to higher concentrations of their tryptic peptides with consequent signal suppression of peptides from other, less abundant proteins. In other fractions, no individual proteins dominate, so tryptic peptides are more equally spread across a larger number of proteins, though many of them are crosslinked, fragmented or otherwise modified. Supplementary Table S1 shows when moving along the gel fractions from F1 to F10, the example big protein (Desmoplakin) decreases and the example small protein (a Keratin-associated protein) increases.

The major advantage of gel fractioning is that it separates the proteins by molecular weight, thereby showing more clearly the origin in individual GVPs. It can also minimize ion suppression leading to the identification of additional GVPs. Unfortunately, this approach is time-consuming. Our attempts to combine fractions led to loss of potential GVPs (see section C). Identifications of all GVPs in a single digest analysis is apparently not possible at present (discussed below). Finding optimal methods will be the topic of future research.

C. GVP Panel Analyses in All Ten Fractions and Among Three Methods

As described in the Method section, we identified a total of 14 published tryptic GVP sites from this Asian donor's hair samples. These sequences along with corresponding non-variant sequences, are listed in Supplementary Document S3. Table 3 shows the specific GVP identification for the three methods with three replicate runs for each method, namely: our Direct method, the modified NaOH+SDS method (8), and the Cleavable Surfactant method (1,2). For both the Direct method and modified NaOH+SDS method, GVP panel results from different fractions are combined in Table 3. Supplementary Document S3 uses the results from F1 to F10 as an example to illustrate how we performed this analysis for a complete data set by the Direct method. Analysis led to a number of general findings:

1) For high-abundance GVPs from major keratins, as shown in Fig. 2A or 2B, identifications are easily made. Scores are high [MF: 792 - 942], leading to highly confident identifications (14), retention times are reproducible (Supplementary Document S3), and identifications are made in all gel fractions for both the GVP and its non-variant form.

2) For low-abundance GVPs, mostly arising from less abundant proteins, identifications can be harder to assign, possibly involving lower and variable scores. Confidence can be increased by elution in the expected gel fraction as well as the determination of its non-variant form (sometimes this is made more difficult if GVP site involves a tryptic cleave site at R or K). This is illustrated with two examples:

(a) The GVP site 'DSP_R1738Q_Q: G[Q]SEADSDKNATILELR' (mutated site highlighted in brackets), was identified in the top gel fractions (F1 and F2). This is consistent with its very large precursor protein having 2871 residues, Desmoplakin (DSP). This is an example that R becomes Q and we identified both GVP and its non-variant form in the expected gel fractions with comparable intensity (Supplementary Document S3).

(b) Another GVP site 'KRTAP10-8_H26R_R: TYVIAASTMSVCSSDVG[R]' originates from a much smaller keratin-associated protein (KRTAP, 259 Amino Acids), and was recovered from bottom gel fractions (F9 and F10). This is an example that H becomes R and we only identified GVP but not its non-variant form. Such discrepancy happens because these are two different peptides when GVP site involves R/K. To solve this problem, we would need to choose a different digestion enzyme. Actual release rates for peptides in a protein are not easily predicted and depend on multiple factors (19). So, it is hard to estimate the relative intensities of a GVP and its non-variant if their lengths and possibly charge states are different.

3) The specific GVP identification depends on the experiments, with a number of different GVPs identified by the in-gel and in-solution digestion methods. Hence, false negative results appear to be a significant concern with the present methods, especially for the in-solution method.

4) We note that the identification of both a GVP and its non-variant will significantly increase the confidence of GVP identification. Of course, this is not possible if the source is homozygous or when the non-variant form is not an easily detectable peptide (as may be the case where tryptic cleavage sites are different in the GVP and non-variant form). In this work, the fact that several potential GVPs were observed (Supplementary Document S3), but not at high confidence (low abundance or matching score) reinforces the likelihood that they are not true GVPs.

Fractionating in the gel methods is part of a 2D study – the first dimension is separating hair proteins based on the MW during SDS-PAGE, the second dimension is separating extracted peptides by the LC gradient during LC-MS/MS. Analyzing each fraction enables very low abundance GVPs to be identified. It is why we detect more GVPs from the two in-gel methods than the in-solution method. However, we detect fewer GVPs if we combine these fractions and process as a mixture (Table 3). We also tried a brief 'short-gel' run by applying SDS-PAGE at 200 V for only 10 min (long-gel: 30 min at 200 V). We compare the GVPs between long-gel and short-gel runs and find that short-gel-mixture loses even more GVPs (Table 3). This can be explained by hair proteins not being effectively separated in a shorter run or possibly that SDS not being fully separated from proteins. In any case, this finding highlights the importance of both separation and sensitivity in finding all identifiable GVPs in a sample. While running 10 fractions is very time-consuming, possible GVPs were lost (Table 3) upon combining fractions indicates that more rapid analysis using a single LC-MS/MS run can lose less abundant GVPs. Moreover, the finding that different GVPs are found with different digestion protocols implies that no existing method can be relied on to identify all possible GVPs. Together, this clearly shows the need of future work for finding the most efficient way to maximize GVP identification.

Comparison Between the Direct Method and modified NaOH+SDS Method

Since the Direct method and modified NaOH+SDS method both use protein gel to separate hair proteins, for a direct comparison, we compared the Direct method with modified NaOH+SDS method for a further sensitivity and reproducibility check in this section.

A. Sensitivity

We examine the sensitivity of the Direct method to modified NaOH+SDS method by comparing multiple metrics across a dilution series. In Figure 4, we show the relative sensitivity of the two methods by comparing the degree of dilution needed for each method to yield the similar number of IDs. After comparing total number of ions (Fig. 4A), total number of peptides (Fig. 4B), total number of proteins (Fig. 4C), and total number of GVP ions (Fig. 4D), we found that the Direct method was about eight times more sensitive than modified NaOH+SDS method. The non-monotonic behavior of some of the irregular trends is a consequence of results from the general difficulty in obtaining highly reproducible proteomic results and, for GVPs, their small numbers and therefore greater statistical fluctuation. Note that since the GVPs are few in number and variable in intensity we could not reliably use GVPs alone to develop a reliable measure of method sensitivity based on their identifications alone. This was confirmed in a separate set of analyses: for example, GVP ions increased at 10D and then all the way decreased to minimum detection level at 1280D.

The present Direct method is both suitable for very small hair samples, and able to identify GVP ions across a broad range of ion intensity. Intensities of reliably identified GVP ions could differ by orders of magnitude in ion intensity. Fig. 5 illustrates this for two spectra of the same GVP ion ‘QVVSSSEQLQSYQ[V]EIIELR/3_0’. Even though intensities differ by four orders of magnitude, retention times were almost identical (161.7 min vs. 161.5 min) and spectral library match factors were quite high (over 800).

B. Reproducibility

In an examination of the reproducibility of the present method, the extraction was repeated eight times using eight individual 5 cm-long hair shafts (labeled as A to H in Fig. 6A) from the same donor, and particularly compared it to modified NaOH+SDS method (labeled as 1A to 1H in Fig. 6B, plus the last lane from 10 hairs included as a reference). We made the assumption that each individual 5 cm hair shaft contained the same protein mass. Fig. 6 clearly indicates that the Direct method is more reproducible than modified NaOH+SDS method. This presumably arises from lower sample loss for the Direct method since it only needs one-step/30 min for hair protein extraction, while the multiple-steps (also means much longer bench time) included in modified NaOH+SDS method are more prone to sample loss and generating variable results (workflows of the two methods are shown in S1) especially when the hair sample is very small.

We also compared the protein, peptide, and GVP identifications between the Direct method and modified NaOH+SDS method with analysis repeated three times for each method. Results of comparisons from a representative fraction (F6) are listed in Table 4 with three experimental repeats: 1) higher average peptide yield (µg) was obtained in the Direct method than in the modified NaOH+SDS method (11.5 vs. 2.9 µg); 2) more average peptides were identified by the Direct method than by the modified NaOH+SDS method (610 vs. 509); 3) although similar average number of GVP ions was observed in the Direct and modified NaOH+SDS methods, it is more reproducible with much smaller coefficient of variation (CV) in three experimental repeats in the Direct method (0.02 vs. 0.27 , respectively); 4) gel blank - only a few peptide IDs from gel blank and no GVP identification at all. Gel blank serves as a control to see if we introduce any contamination from handling the blank gel alone. Table 4 shows that the Direct method is not only a more sensitive, but also a more reproducible method when compared to the modified NaOH+SDS method.

Estimation of the digestion yield: The gel-based method we chose for analysis unfortunately did not allow us to use a conventional Bradford colorimetric (BCA) assay to measure protein concentration. Instead, yields of digested peptides using the Pierce method mentioned above served a similar, albeit less direct purpose. Based on a measured 5 cm hair mass of 100 µg (10 5-cm lengths were found to weigh 1.0 mg), we found that at the incubation time of 5, 10, 15, 30, 60 and 90 minutes, corresponding total yields of peptides to be 16%, 27%, 37%, 75%, 66% and 51%. The maximum of 75% at 30 min was selected as optimal (see above). For comparison, a yield of 47% was reported for an in-solution method (8) using BCA assay after precipitating extracted proteins.

Examination of Artifacts Among Three Methods

In most proteomics experiments, a large fraction of ions sampled are not identified. This not only reduces the efficiency of the experiment but also has potential to generate false positive results. Moreover, the identity of the unidentified ions may aid in understanding and optimizing the experiment and provide a measure of quality control.

In the present experiment almost 90% of ions are not directly identified as tryptic peptides using conventional library searching. Using our recently developed hybrid search (15), as shown in Supplementary Table S2, 11% can be identified as expected tryptic peptides, while about 75% can be identified via hybrid identification. These hybrid identifications find peptides that are chemically modified forms of conventional tryptic peptides. The reason we would like to examine experimentally introduced artifacts is because we must be aware of artifactual modifications that may masquerade as a GVP and therefore generate false positive identifications, the larger the number of spurious modifications the greater the chance that one will accidentally overlap a possible GVP. Proteomics cannot distinguish biological versus artifact origins of identified peptides. For example, a methylation at or near a serine might be interpreted as a serine to threonine GVP. IonPlot in Fig. 7 shows the classification of ions (GVP, Identified, and not-identified ions from F6 of the Direct method) by the hybrid search including a list of several interesting modifications that we would like to discuss more in this section. These analyses also show the nature and extent of certain spurious chemical processes that add to sample complexity and, in effect, diminish the sensitivity and overall quality of the experiment.

Since this issue is important for every sample preparation method regarding to GVP detection, below we examine the artifacts among the three methods: our Direct method, modified NaOH+SDS method, and Cleavable Surfactant method.

Table 5 compares the twenty most frequently identified DeltaMass values in three methods (15). For more information, Supplementary Document S4 shows the histograms of all DeltaMass values obtained from hybrid search identifications in each method to give a broad view of the distribution of all DeltaMass values. From the top 20 DeltaMass values listed in Table 5, we now further discuss four types of experimentally introduced artifactual modifications (Fig. 8).

Acetaldehyde adduction. We compared the occurrence of an acetaldehyde adduct across the three methods. Fig. 8 shows that this artifactual modification is more frequently identified in the Direct and modified NaOH+SDS methods due to the presence of ethanol in the SimplyBlue SafeStain that we used to stain the protein gels. We here included an example in Fig. 9 to show our main concern – a modification at peptide's N-terminus could be mistaken as a potential GVP: the DeltaMass value from the hybrid search for this hybrid identification is 26.0186 Da, within the mass tolerance range, which is likely due to acetaldehyde (26.01565 Da) but may be incorrectly identified as His (H) →Tyr (Y) (26.004417 Da) since His (H) is involved in the identification at the first amino acid in this peptide ion. Without the hybrid search, or without being aware of what type of artifactual modification exists, such a mis-identification will occur.

Acetylation. While acetylation at Lys (K) and the protein amino terminus are biological modifications, artifactual acetylation at the peptide N-terminus can be introduced during sample preparation. Although the source of acetic acid is not believed to have been introduced through sample preparation, this artifactual modification was identified more frequently in the Direct and modified NaOH+SDS methods.

Formylation. Formylation is less dissimilar across all three methods than that of the previous described two modifications. This is expected as formic acid is required in all three sample preparations.

Alkylation. Alkylation (CAM) is significantly greater in the Cleavable Surfactant method compared to the Direct and modified NaOH+SDS methods. This is consistent with the fact that iodoacetamide concentration we used in sample preparation of Cleavable Surfactant method is much higher than in the Direct and modified NaOH+SDS methods.

Table 5 and Supplementary Document S4 show that, overall, results of the three methods have similar degrees of experimentally introduced modifications. It seems likely that the artefactual modifications are a result of the inherent difficulty of digestion such an insoluble and crosslinked material as hair.

Regarding to GVP panel analysis, we find consistent results in regular and hybrid searches. Hybrid searching usually reports more GVP ions with many kinds of unexpected modifications but seems not gaining additional known GVP site detection. Verified GVP detection by the hybrid search (not only seeing the version that included in the library but also seeing the versions with some unexpected modifications) increases the confidence of GVP panel analysis.

Identification of Hair Proteome and Cuticular Keratins from as Little as 1 cm-long Human Hair Shaft by Direct Extraction Method

So far, the data we presented in this manuscript used 5 cm-long hair shafts as the starting material. While we learned about the sensitivity of the Direct method with the serial dilution study, we also wanted to check results using smaller lengths of hair. As the dilution series was a projection for low amounts based on similar extraction efficiencies for smaller lengths, one may expect further losses due to possible inefficiencies in digesting small lengths of hair. For this purpose, we undertook a series of studies where hair shaft varied from 5, 2.5, and 1 cm-long. Fig. 10A shows the separation of hair proteins by SDS-PAGE for three different hair lengths and Table 6 lists the total number of hair proteins and peptides identified as well as those that are specific for hair cuticular keratins and GVP ions. Fig. 10B shows the analysis of an example GVP ion whose abundance is almost linear in 5, 2.5, and 1 cm hair shaft samples to demonstrate the abundance is proportional to length. These results show that as little as 1 cm-long hair shaft sample can be analyzed by this Direct method. There is no reason to believe it would not work effectively for even smaller amounts of hair, suggesting that even forensic-relevant trace quantities of hair would be suitable for this analytical method.

Examination of the Direct Method in Another Donor

To ensure that these results were not unique to one donor, we applied the Direct method to another randomly selected donor's hair shaft samples obtained from BioreclamationIVT (LOT# BRH1363733, 5 g of hair shafts from a Caucasian male, 23 years old). Table 7 lists the total number of hair proteins and peptides identified as well as those from hair cuticular keratins and GVP ions. These results demonstrate that the Direct method works equally well for another donor's hair samples. The overall protein gel images, the peptide yields from in-gel-digestions, the hair keratins and their peptide identifications, and the number of found GVP ions are similar. Most of high abundance GVPs in this Caucasian donor overlap with previous described Asian donor in the GVP panel analysis. This manuscript is focused on the protein and peptide extraction from single hair shaft, that is the reason why we use hair samples from the same Asian donor for the development of protein extraction method. We believe our Direct method would work effectively for hair samples from any individual donor. These studies did not consider donors who heated or chemically treated their hair – this would be a useful topic for future research. The focus of this paper was only analytical methods and detailed proteomic analysis. Variations with hair origin will be the topic of future studies using the methods described here.

Summary and Conclusions

In summary, we have shown that the Direct extraction method is a sensitive, reliable, and relatively convenient method based on the depth of coverage of the human hair proteome and cuticular keratins: 1) It is a relatively sensitive method: it works for a hair shaft as short as 1 cm;

2) It is a relatively reliable method: it generates more consistent results in protein/peptide identification and GVP detection; 3) It is a relatively convenient method: it is simple to carry out since there is only one-step in protein extraction from hair, although to assure maximum GVP identification, it does require multiple LC-MS/MS runs.

Using our recently developed 'hybrid' spectral library search method, we have found that a very large fraction of the peptide spectra acquired were not simple tryptic peptides derived from known proteins. A conventional library search can identify only 11% of the peptides, who the hybrid search identifies 75%, including any previously unidentified GVPs (as our future work). We have also shown that the hybrid search, could be used to identify potential sources of false positives due to the presence of artifactual modifications that are experimentally introduced. Modifications that could be mistaken as a GVP should be the primary concern and a separated examination of artefactual modifications is needed. In difficult cases, a more careful manual checking of GVP spectra may also be needed.

Although we recommend the Direct method because of several advantages we described earlier, we also realize different methods may be most suitable for different GVP panel analysis. Each method will have its own strength and weakness. Unless we combine the results from all three tested methods, no single method covered all the identified published GVP sites in this study. This is largely because of the nature of the hair samples – heavy crosslinking makes hair mechanically strong and stable, but also very resistant to sample processing.

We have also shown that a GVP analysis can effectively done using a peptide spectral library containing all identifiable peptides derived from human hair samples. With this paper we provide a library containing all identified hair derived peptides (13). Future expansion of this library can include all known GVPs as well as all identifiable peptides derived from human hair. Further, it may be combined with the NIST-developed label-free HCD main peptide library (peptide.nist.gov) (12) to provide another layer of sensitivity and confidence for hair peptide identification and GVP detection.

Supporting Information

Supplementary Table S1. Example of a Big Protein and a Small Protein Amount Change in Ten Gel Fractions by the Direct Method

Supplementary Table S2. Percentages of Hybrid IDs in All Ten Gel Fractions by the Direct Method

Supplementary Document S1. Outline of Protein Extraction Work Flows for Direct Method and modified NaOH+SDS Method

Supplementary Document S2. Comparison of Sequences Coverage in Amino Acids of 15 type I and type II hair cuticular keratins by library and Sequest searching

Supplementary Document S3. GVP Panel Analyses in All Ten Fractions by the Direct Method

Supplementary Document S4. Histograms of the Distribution of All DeltaMass Values in Three Methods

References

1. Parker GJ, Leppert T, Anex DS, Hilmer JK, Matsunami N, Baird L et al. Demonstration of Protein-Based Human Identification Using the Hair Shaft Proteome. PLoS One 2016; 11(9): e0160653.

2. Mason KE, Paul PH, Chu F, Anex DS, Hart BR. Development of a Protein-based Human Identification Capability from a Single Hair. J Forensic Sci 2019 Jul; 64(4):1152-9.

3. Carlson TL, Moini M, Eckenrode BA, Allred BM, Donfack J. Protein extraction from human anagen head hairs 1-millimeter or less in total length. Biotechniques 2018; 64(4):170-6.

4. Bengtsson CF, Olsen ME, Brandt LØ, Bertelsen MF, Willerslev E, Tobin DJ et al. DNA from keratinous tissue. Part I: hair and nail. Ann Anat 2012; 194(1): 17-25.

5. Langbein L, Rogers MA, Winter H, Praetzel S, Beckhaus U, Rackwitz HR et al. The catalog of human hair keratins. I. Expression of the nine type I members in the hair follicle. J Biol Chem 1999; 274(28): 19874-84.

6. Langbein L, Rogers MA, Winter H, Praetzel S, Schweizer J. The catalog of human hair keratins. II. Expression of the six type II members in the hair follicle and the combined catalog of human type I and II keratins. J Biol Chem 2001; 276(37): 35123-32.

7. Lee YJ, Rice RH, Lee YM. Proteome analysis of human hair shaft: from protein identification to posttranslational modification. Mol Cell Proteomics 2006; 5(5): 789-800.

8. Wong SY, Lee CC, Ashrafzadeh A, Junit SM, Abrahim N, Hashim OH. A High-Yield Two-Hour Protocol for Extraction of Human Hair Shaft Proteins. PLoS One 2016; 11(10): e0164993.

9. Adav SS, Subbaiaih RS, Kerk SK, Lee AY, Lai HY, Ng KW et al. Studies on the Proteome of Human Hair - Identification of Histones and Deamidated Keratins. Sci Rep 2018 Jan; 8(1): 1599.

10. Jimenez CR, Huang L, Qiu Y, Burlingame AL. In-gel digestion of proteins for MALDI-MS fingerprint mapping. Current Protocols in Protein Science 1998; 14(1): 16.4.1-5.

11. Rudnick PA, Markey SP, Roth J, Mirokhin Y, Yan X, Tchekhovskoi DV et al. A Description of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) Common Data Analysis Pipeline. J Proteome Res 2016; 15(3): 1023-32.

12. The NIST Main Libraries of Peptide Tandem Mass Spectra

https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:lib:humanhcd20160503

13. The NIST Hair Libraries of Peptide Tandem Mass Spectra

https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:lib:human_hair_selected_with_gvps_passed

14. Zhang Z, Burke M, Mirokhin YA, Tchekhovskoi DV, Markey SP, Yu W et al. Reverse and Random Decoy Methods for False Discovery Rate Estimation in High Mass Accuracy Peptide Spectral Library Searches. J Proteome Res 2018; 17(2): 846-57.

15. Burke MC, Mirokhin YA, Tchekhovskoi DV, Markey SP, Heidbrink Thompson J, Larkin C et al. The Hybrid Search: A Mass Spectral Library Search Method for Discovery of Modifications in Proteomics. J Proteome Res 2017; 16(5): 1924-35.

16. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 1994; 5(11): 976-89.

17. Abbatiello SE, Schilling B, Mani DR, Zimmerman LJ, Hall SC, MacLean B et al. Large-Scale Interlaboratory Study to Develop, Analytically Validate and Apply Highly Multiplexed, Quantitative Peptide Assays to Measure Cancer-Relevant Proteins in Plasma. Mol Cell Proteomics 2015 Sep; 14(9): 2357-74.

18. Speicher K, Kolbas O, Harper S, Speicher D. Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. J Biomol Tech 2000 Jun; 11(2): 74–86.

19. Lowenthal MS, Liang Y, Phinney KW, Stein SE. Quantitative bottom-up proteomics depends on digestion conditions. Anal Chem 2014 Jan; 86(1):551-8.

Table 1. Comparison of Protein and Peptide Identifications from Spectral Library and Sequest Searching in All Ten Fractions at 1% FDR by the Direct Method from a 5 cm-long Hair Shaft*.

Direct	Yield (μg)	TIC	Main+Hair Spectral Library				Sequest			
			Hair Proteome		Cuticular Keratins		Hair Proteome		Cuticular Keratins	
			Proteins	Peptides	Proteins	Peptides	Proteins	Peptides	Proteins	Peptides
F1	1.76	3.91E+06	148	2040	14	583	98	1128	14	471
F2	3.81	6.54E+06	140	1888	15	614	84	1052	14	503
F3	5.46	1.03E+07	132	1744	14	614	73	1022	14	525
F4	8.95	1.44E+07	134	1789	14	628	83	1045	13	526
F5	5.86	8.27E+06	152	1781	14	594	93	1061	14	513
F6	13.25	2.06E+07	135	1617	15	620	68	906	15	503
F7	10.92	2.31E+07	146	1607	13	623	76	933	14	538
F8	7.06	8.17E+06	207	2167	15	631	129	1290	15	521
F9	5.98	4.72E+06	214	2268	14	589	138	1346	13	463
F10	12.24	8.59E+06	173	1744	14	470	120	1079	13	347

*Proteins were identified by ≥ 2 peptides throughout this manuscript. For peptide/protein identifications (IDs) under 'Hair Proteome', Fraction 8 (F8) and 9 (F9) gave more IDs in both spectral library and Sequest searches; for peptide/protein IDs under 'Cuticular Keratins', the distribution of IDs was more even across all 10 gel fractions in both spectral library and Sequest searches. TIC: an index of total ion current.

Table 2. Comparison of Sequence Coverage (%) of Hair Cuticular Keratins from Spectral Library and Sequest Searching in All Ten Fractions by the Direct Method.

Cuticular Keratins	From Library	From Sequest
KRT31	100.0	97.6
KRT32	54.2	49.6
KRT33A	97.0	93.3
KRT33B	97.0	93.6
KRT34	86.0	83.9
KRT35	91.0	86.4
KRT36	60.8	49.3
KRT37	43.0	34.7
KRT38	61.2	51.3
KRT81	96.2	91.9
KRT82	63.4	49.9
KRT83	97.0	87.2
KRT84	12.7	11.2
KRT85	96.8	89.4
KRT86	99.2	92.4
Average	**77.0**	**70.8**

Table 3. Genetically Variant Peptide (GVP) Panel Analyses in Three Methods*.

	DSP	GSDMA	KRT31	KRT32	KRT33A	KRT33B	KRT35	KRT35	KRT81	KRT82	KRT83	KRT83	KRTAP 10-8	TGM3
ONE 5 CM HAIR, ASIAN	R1738Q_Q	V128L_L	A82V_V	S222Y_Y	A270V_V	V279L_L	P443A_A	S36P_P	S13R_R	T458M_M	G362S_S	I279M_M	H26R_R	T13K_K
D_LG_F1_TO_F10_R1#	X		X		X			X	X		X	X	X	X
D_LG_F1_TO_F10_R2	X		X		X	X		X	X		X	X	X	X
D_LG_F1_TO_F10_R3	X		X	X	X	X		X	X		X	X	X	X
D_LG_COMBINED_R1			X		X	X		X	X			X		X
D_LG_COMBINED_R2			X		X	X		X	X			X		X
D_LG_COMBINED_R3			X		X			X	X			X		X
D_SG_COMBINED_R1			X		X			X	X			X		X
D_SG_COMBINED_R2			X		X			X	X			X		X
D_SG_COMBINED_R3			X		X		X	X	X			X		X
NS_LG_F1_TO_F10_R1	X	X	X		X		X	X	X		X	X	X	X
NS_LG_F1_TO_F10_R2	X	X	X	X	X			X	X		X	X		X
NS_LG_F1_TO_F10_R3	X		X	X	X		X	X	X		X	X	X	X
CS_R1			X				X			X		X		X
CS_R2			X				X	X		X		X		X
CS_R3			X				X			X				X

*all listed GVP analyses are derived from the same Asian donor's single 5 cm-long hair samples: GVP panel analyses by the Direct method with all 10 fractions from a long-gel (30 min run at 200 V) which have been individually processed by LC-MS/MS and then summarized the results in one row are labeled as 'D_LG_F1_TO_F10'; GVP panel analyses with combined fractions processed as a mixture from a long-gel run by the Direct method are labeled as 'D_LG_COMBINED'; with combined fractions from a short-gel run (10 min run at 200 V) are labeled as 'D_SG_COMBINED'; GVP panel analyses by the modified NaOH+SDS method with all 10 fractions from a long-gel run individually processed and then summarized are labeled as ‘NS_LG_F1_TO_F10’; GVP panel analyses by the Cleavable Surfactant method are labeled as ‘CS’. R1, R2, and R3 are three experiment repeats.

#results from F1 to F10 are listed in Supplementary Document S3, used as an example to demonstrate a GVP panel analysis from this ‘D_LG_F1_TO_F10_R1’ data set.

Table 4. Examination of Reproducibility for the Direct Method and modified NaOH+SDS method* from a Representative Gel Fraction (F6).

Methods (one 5 cm hair, Asian)	Yield (µg)	Main+Hair Spectral Library				
		Hair Proteome		Cuticular Keratins		GVP ions
		Proteins	Peptides	Proteins	Peptides	
Direct_R1	10.32	114	1427	14	593	43
Direct_R2	13.25	135	1617	15	620	44
Direct_R3	10.94	132	1725	14	618	45
NaOH+SDS_R1	3.36	101	1267	14	509	29
NaOH+SDS_R2	2.11	93	1178	14	497	51
NaOH+SDS_R3	3.32	83	1137	15	520	45
Blank Gel	0.04	6	17	2	7	0

*The result was obtained from fraction 6, a representative gel fraction. Three experimental repeats: R1, R2, and R3.

Table 5. The Twenty Most Frequently Identified DeltaMass Values Obtained from Hybrid Search Identifications in the Three Methods.

DeltaMass	Theorical Value of DeltaMass	Proposed Modification	Percent of Hybrid Identifications		
			Direct (Median)	NaOH+SDS (Median)	Cleavable Surfactant (Median)
1.001	1.00335483	1-C13	17.30	17.76	19.34
2.007	2.00670966	2-C13	6.73	8.82	6.71
42.013	42.010565	Acetyl	6.25	5.75	3.54
26.017	26.015650	Acetaldehyde	3.52	2.49	0.66
3.009	3.01006449	3-C13	3.59	4.96	3.55
27.999	27.994915	Formyl	1.87	3.03	1.57
14.018	14.015650	Methyl	3.08	2.60	1.12
-1.011	-1.00335483	-1-C13	2.31	3.05	
-17.023	-17.026549	-NH3	1.62	1.51	2.38
70.007	70.005480	Formyl + Acetyl	0.89	1.28	
4.009	4.01341932	4-C13	1.78	2.44	2.02
12.002	12.000000	Formaldehyde Adduct	1.45	1.20	
43.014	43.005814	Carbamyl/Acetyl + 1-C13	1.48	1.07	0.70
-18.008	-18.010565	Dehydration/Glu→pyro-Glu	1.34	1.35	2.01
-2.013	-2.00670966	-2-C13	1.36	1.58	1.43
23.986	23.98865266	Sodiated + 2C-13	1.17		
57.023	57.021464	CAM	1.78	1.87	4.21
15.997	15.994915	Oxidation	1.08	1.28	
120.028	120.024500	Desulferization + CAM + DTT	0.95		
58.010	58.005480	Deamidation + CAM	1.06	0.89	3.33
-91.009	-91.009185	Cys(CAM)→Dehydroalanine		0.82	
-16.019	-16.0231942	1C-13 + -NH3		0.76	0.93
-0.983	-0.984016	Amidation			3.44
5.014	5.01677415	5-C13			0.69
160.041	160.030654	Add-Cys+CAM			1.25
31.995	31.989829	Dioxidation			1.78
152.003	151.996571	+DTT			0.86

Table 6. Reduction of Starting Material to 1 cm-long Hair Shaft by the Direct Method*.

Hair Length (cm)	**Main+Hair Spectral Library**				
	Hair Proteome		**Cuticular Keratins**		**GVP ions**
	Proteins	**Peptides**	**Proteins**	**Peptides**	
5	135	1617	15	620	44
2.5	86	1203	14	563	40
1	78	1149	14	486	39

*The result was obtained from fraction 6, a representative gel fraction.

Table 7. Comparison of Protein and Peptide Identification from a 5 cm-long Hair Shaft from Asian and Caucasian Male Donor by the Direct Method*.

Donor	Yield (µg)	Main+Hair Spectral Library				
		Hair Proteome		Cuticular Keratins		GVP ions
		Proteins	Peptides	Proteins	Peptides	
Asian	13.25	135	1617	15	620	44
Caucasian	8.48	92	1177	14	581	45

*The result was obtained from fraction 6, a representative gel fraction.

Figure Legends

FIG. 1—*Time Course Study to Optimize the Best Heating Condition of the Direct Method. A time-course study was performed to find the optimal time that a 5 cm hair shaft sample need to be heated at 90°C. (A) The scanned gel image included a MW standard loaded in the first lane and six additional lanes where the samples were loaded on increasing length of time for which they have been heated at 90°C (5, 10, 15, 30, 60, and 90 min). The major bands that correspond to type I and type II hair cuticular keratins were labeled. The orange thin lines indicate fractionating the gel to 10 slices from top to bottom as "F1" to "F10". (B) The chart shows the density reports of type I and type II bands at each time interval. The density reports were obtained from gel scanning. The best time point (30 min) is labeled in red based on giving the maximum density reports for both type I and type II bands at 30 min. (C) The chart shows the density ratios of all 10 gel fractions obtained at 30 min, using fraction 1 as the reference.*

FIG. 2—*The Range of the Intensities of Example Peptide Ions Across All Ten Fractions from the Direct Method in Type I and Type II Cuticular Keratins. (A) Type I cuticular keratin KRT33A: The range of intensities of an example GVP peptide ion pair (KRT33A A270V_V: QVVSSSEQLQSYQ[V]EIIELR/3_0 (blue square linked by blue line) and KRT33A A270V_A: QVVSSSEQLQSYQ[A]EIIELR/3_0 (blue triangle linked by blue line)) as well as another peptide ion (SQQQEPLVCASYQSYFK/3_1/9, C, Carbamidomethyl (orange circle linked by orange line)) whose sequence is unique to KRT33A but not containing a known GVP site across all 10 fractions. 'KRT33A A270V_A' or 'KRT33A A270V_V' means the amino acid at position 270 of KRT33A can be a 'A' (regular version in human FASTA file) or a 'V' (published variable version). Dashed black line indicates these three peptide ions reach their maximum intensities at Fraction 7. (B) Type II cuticular keratin KRT83: The range of intensities of an example GVP*

peptide ion pair (KRT83 I279M_M DLNMDC[M]VAEIK/2_3/4,M,Oxidation/6,C,Carbamidomethyl/7,M,Oxidation (blue square linked by blue line) and KRT83 I279M_I DLNMDC[I]VAEIK/2_2/4,M,Oxidation/6,C,Carbamidomethyl (blue triangle linked by blue line)) as well as another peptide ion (LCEGVEAVNVCVSSSR/2_2/2,C,Carbamidomethyl/11,C,Carbamidomethyl (orange circle linked by orange line)) whose sequence is unique to KRT83 but not containing a known GVP site across all 10 fractions. 'KRT83 I279M_I' or 'KRT83 I279M_M' means the amino acid at position 279 of KRT83 can be an 'I' (regular version in human FASTA file) or a 'M' (published variable version). Dashed black line indicates these three peptide ions reach their maximum intensities at Fraction 6.

FIG. 3—*The range of total ion current (TIC, upper panel) and peptide identifications (lower panel) across all 10 fractions. Blue dashed lines indicate TIC values reach their maximum numbers at Fractions 6 & 7, where peptide IDs reach their minimum numbers at Fractions 6 & 7.*

FIG. 4—*Comparison of the Sensitivity in the Two Methods. The sensitivity of the two methods was measured by comparing multiple metrics across a dilution series from 5D to 1280D: (A) the total number of ions; (B) the total number of peptides; (C) the total number of proteins; (D) the total number of published GVP ions detected in mass spectral data from 5 cm-long hair shaft sample derived proteins that were extracted using the Direct method (blue) and modified NaOH+SDS method (green). Actual data has been labeled on the points of each dilution series.*

FIG. 5—*Identification of an Example GVP Ion with High and Low Abundance. The example GVP ions (KRT33A A270V_V: QVVSSSEQLQSYQ[V]EIIELR/3_0 higher-energy collisional*

dissociation (HCD) =30eV) were mapped to an IonPlot (x-axis: Retention Time (RT) in min, y-axis: Abundance in log 10 scale) to show the library identification with high abundance (upper blue dot) or with low abundance (lower blue dot). One blue dot indicates one peptide ion. For each blue dot, the RT and the abundance in log 10 scale were labeled underneath; blue arrows indicate their corresponding library identifications by searching the spectrum of this peptide ion as query spectrum against the hair specific peptide spectral library including known GVP ions. The match factor (MF) was labeled underneath its library identification.

FIG. 6—*Comparison of the Reproducibility of the Direct and modified NaOH+SDS Methods. The two gel images compare the reproducibility of method (A) the Direct method and (B) modified NaOH+SDS method using 5 cm-long hair shaft samples from the same individual donor across 8 replicates (A: A to H; B: 1A to 1H). A MW standard was loaded in the first lane. Note that the NaOH+SDS gel includes a 9th lane for which the extraction from ten 5cm-long hair shaft samples was included as a reference. The major bands that correspond to type I and type II hair cuticular keratins were labeled.*

FIG. 7—*Classification of Ions by the Hybrid Search. IonPlot shows the classification of GVP, identified, and not identified (NoID) ions, as well as several modifications: formylation (formyl), methylation (methyl), alkylation (CAM), acetaldehyde, and acetylation that present in fraction 6 (F6), a representative gel fraction from a protein gel separating proteins derived from a 5 cm-long hair shaft of this Asian donor by the Direct method. Solid: identified by regular library search; Hollowed: identified by hybrid library search. x-axis: Retention Time (RT) in minute (min), y-axis: Abundance in log 10 scale.*

FIG. 8—*Comparison of the Artifacts in the Three Methods. Comparison of experimentally introduced artifactual modifications among three methods using our recently developed hybrid*

search: Cleavable Surfactant method (red), modified NaOH+SDS method (green) and the Direct method (blue). The compared experimentally introduced artifactual modifications chosen as examples are: acetaldehyde (upper left), acetylation (upper right), formylation (lower left) and over alkylation (lower right).

FIG. 9—*An Example of a Modification at Peptide N-terminus Mistaken as a GVP. Spectral match of a hair-derived peptide to the peptide sequence HLQLAIR (Charge=2, Mods=0, Spectral Match Score=705) with a DeltaMass of 26.0186 Da, which is likely due to acetaldehyde (26.01565 Da) but may be incorrectly identified as His (H) →Tyr (Y) (26.004417 Da).*

FIG. 10—*Comparison of Hair Length Variation. Comparison of hair length variation. (A) This gel image shows the separation of hair proteins from 5, 2.5, and 1 cm-long hair shaft samples from the same individual donor. A MW standard was loaded in the first lane. Bands for type I and type II hair cuticular keratins were labeled. (B) spectral match (MF=921) of an example GVP ion (KRT31_A82V_V: DN[V]ELENLIR/2_0 HCD=30eV) is on the left. The spectrum shown in red is the query spectrum and the spectrum shown in blue is the reference library spectrum for this GVP ion. On the right is a plot that shows the abundance of this example GVP ion in the 1, 2.5, and 5 cm hair shaft samples is approximately linear. Note the y-axis is the log of the abundance value, plotted on a linear scale.*

ABSTRACT

Recent reports have demonstrated that genetically variant peptides derived from human hair shaft proteins can be used to differentiate individuals of different biogeographic origin. We report a method involving direct extraction of hair shaft proteins more sensitive than previously published methods regarding GVP detection. It involves one-step for protein extraction and was found to provide reproducible results. A detailed proteomic analysis of this data is presented that led to the following four results: 1) A peptide spectral library was created and made available for download. It contains all identified peptides from this work, including GVPs that, when appropriately expanded with diverse hair-derived peptides, can provide a routine, reliable and sensitive means of analyzing hair digests; 2) An analysis of artifact peptides arising from side reactions is also made using a new method for finding unexpected modifications; 3) Detailed analysis of the gel-based method employed clearly shows the high degree of crosslinking or protein association involved in hair digestion, with major GVPs eluting over a wide range of high molecular weights while others apparently arise from distinct non-crosslinked proteins; 4) Finally, we show that some of the specific GVP identifications depend on the sample preparation method.

KEYWORDS

Forensic Science, Genetically Variant Peptide, hair protein extraction, cuticular keratins, peptide mass spectral library, and trace detection

In recent publications from Lawrence Livermore National Laboratory (LLNL), genetically variant peptides (GVPs) derived from human hair have been shown to have forensic value (1,2). The publication (1) by Parker et al. showed that these peptides might serve as a source of evidence in addition to DNA for human identification due to several advantages that a hair sample carries: 1) commonly found – on average, humans shed 50 – 150 hairs per day; 2) stable – proteins in a hair sample usually last longer and are more resistant to degradation than DNA; 3) when good quality DNA is not available, hair proteins may serve as alternative evidence by detecting those GVPs in hair cuticular keratins and other hair proteins. A recent publication (2) by Mason et al. described protein-based or GVP-based human identification from a single hair as short as 1 inch-long. Another recent publication (3) by Carlson et al. described a sensitive method to extract proteins from 1-millimeter or less in total length of human anagen head hairs, and compared the proteins identified from hair shaft and hair root. The effectiveness of this method for detecting GVPs has not yet been determined.

The human hair shaft is made up of three main components (4). Starting from the center, the first component is the medulla which is rich in cross-links and highly insoluble. Next is the cortex which comprises most of the hair shaft and is made up of hair cuticular keratin fibrils as well as keratin-associated proteins. The thin outer layer is the cuticle which is also composed of keratin-associated proteins and is the component that would be visually inspected through microscopic examination. Hair cuticular keratins have been classified as type I (31-38) and type II (81-86) based on the finding that type I keratins are acidic and type II keratins are neutral or basic proteins (5,6). Two recent publications (1, 2) from LLNL have collectively identified a total of 88 GVP sites from multiple donors with bulk of hair samples: 32 sites from hair cuticular keratins, 7 sites from cytoskeletal keratins, 22 sites from keratin associated proteins, and 27 sites from non-keratins.

Based on these findings, a human hair sample has the potential to serve as alternative evidence for human identification if GVPs in hair keratins (mainly cuticular keratins), keratin associated proteins and other non-keratin hair proteins can be sensitively and reliably identified. To detect them, we first need an efficient method to extract proteins from human hair shafts. However, hair protein extraction is especially difficult due to extensive cross-linking and poor solubility of hair keratins (7,8,9). In this manuscript, we describe a direct protein extraction method (referred as the Direct method) that can efficiently extract hair proteins from a single hair shaft less than 1 cm in length. We performed GVP panel analyses and examined experimentally-introduced artifactual modifications among three methods: our newly developed Direct method and two of previously published methods – NaOH-based SDS repeated extraction method (we modified it to make it fit in small sample analysis, referred as modified NaOH+SDS method) (8) and ProteaseMax-based method (referred as Cleavable Surfactant method) (1,2). Considering the Direct method and modified NaOH+SDS method both utilize protein gel electrophoresis to separate extracted proteins, we made further comparisons between these two in-gel methods for sensitivity and reproducibility. We find that the Direct method is both sensitive and relatively convenient to carry out while generating reproducible results regarding to GVP detection from a single hair shaft from one individual donor. In the analysis of this data, we applied a number of proteomic data analysis methods including: 1) The development of a library of peptide ion spectra containing all identified peptides that, when extended, can contain all identifiable peptides from hair proteins. Spectral libraries provide a sensitive and reliable means of peptide identification and ultimately can contain spectra of all known GVPs. 2) Proteomic analysis that enable the detailed analysis of artifact peptides, generated by undesirable chemical analysis which can, in principle, lead to false positive analysis. 3) A gel-based method of analysis that reveals a wide distribution of molecular weights of proteins yielding keratin-based GVPs. 4) The finding that different digestion methods can identify different GVPs, suggesting the inadequacy of any current method of finding all potentially identifiable GVPs in a hair sample.

Materials and Methods

Human Hair Sample Preparation

Human hair samples were obtained commercially from BioreclamationIVT (LOT# BRH1363732, 5g of hair shaft per package from the same individual donor). Most of the results presented in this manuscript are derived from hair shafts from this single randomly selected donor: Asian male, 30 years old. Hair samples were briefly washed with 20% methanol and water, then dried and stored at -20°C. The related protocols have been reviewed and approved by National Institute of Standards and Technology (NIST) Human Subjects Review Board.

Direct Extraction Method

Hair shaft samples (5cm, 2.5cm, or 1cm) were cut using sterile laboratory scissors and then combined with 50 µl of the commercially obtained NuPAGE Lithium dodecyl sulfate (LDS) Sample Buffer (Catalog # NP0007, ThermoFisher Scientific) and 50 mmol/L reducing agent dithiothreitol (DTT). After heating the hair shaft in sample buffer at 90 °C for various lengths of time, extracted hair proteins (we call this the Direct method) were loaded onto NuPAGE 4-12% Bis-Tris Protein Gels (Catalog # NP0321, ThermoFisher Scientific) and then separated by size together with a Molecular Weight (MW) Standard (MW std) using sodium dodecyl sulfate - Polyacrylamide Gel Electrophoresis (SDS-PAGE) at 200 V for 30 minutes. The protein gel was stained with SimplyBlue SafeStain (Catalog # LC6060, ThermoFisher Scientific) for one hour. After overnight immersion in water, the destained-protein-containing gel was scanned, and intensities of the main bands were determined. From top to bottom, the gel was evenly cut in 10 fractions (about 4 mm-long per fraction) and in-gel-digestion was performed for each fraction by following a well-established in-gel-digestion protocol (10). Peptide concentrations were measured by a kit provided by Pierce (Quantitative Colorimetric Peptide Assay Kit, Catalog # 23275) after desalting by ZipTip (Catalog # ZTC18S960, EMD Millipore Corporation). Desalted peptides were injected to a Thermo Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. A simplified Direct method workflow is shown in Supplementary Document S1.

We performed a time course study to determine the optimal heating time for extracting hair proteins by this Direct method using six individual 5 cm-long hair shafts with each one processed at a different incubation time in the same amount of sample buffer (Fig. 1). The six different incubation times were: 5, 10, 15, 30, 60 and 90 min with net peptide yields measured by combining all ten fractions. The largest yield of peptides was found to occur at 30 minutes and was selected as the optimal incubation time. Note that the LDS sample buffer was unchanged at a pH of 8.5 through all incubation times. As Fig. 1A shows, we observed two distinct bands: the first was found to be enriched in type II (basic) hair cuticular keratins (Gene Name: KRT81 to 86, # Amino Acids: 486 to 600, MW 53.5 to 64.8), and the second enriched in type I (acidic) hair cuticular keratins (Gene Name: KRT31 to 38, # Amino Acids: 404 to 467, MW 45.9 to 52.2) (8). The orange thin lines in Fig. 1A also indicate an even fractionation of the gel in 10 slices per lane from top to bottom as F1 to F10. Fraction 6 (F6) contains the first main band which enriches type II cuticular keratins and fraction 7 (F7) contains the second main band which enriches type I cuticular keratins (discussion of this observation can be found in the Results and Discussion section). Fig. 1B shows the density reports of type I and type II bands at each time interval, reaching a maximum at 30 min (Fig. 1B), consistent with the time for maximum peptide yield described above. Fig. 1C shows the density ratios of all ten fractions obtained at 30 min, using F1 as the reference. The maximum is at F6, which is used as a keratin-enriched representative fraction. Fig. 1C indicates that the gel-based method both concentrates known GVP-rich keratin proteins and shows the hitherto unknown distribution of apparently crosslinked proteins.

We note that additional studies are needed to understand both the effect of heating and the influence of cysteine alkylation and other chemical processing details on peptide yields.

Modified NaOH-based SDS Repeated Extraction Method

To examine our newly developed Direct method, we compared it to a previously published NaOH-based SDS repeated extraction method (8). We modified the published protocol to fit the purpose of protein extraction from a single hair shaft. The modified work flow was performed as follows (also illustrated in Supplementary Document S1): 1) first, we used bead milling for sample preparation instead of incubation with lysis buffer: 5 cm-long hair shafts are ground by a bead mill (OMNI Bead Ruptor 24 Elite, OMNI-International Inc.) repeatedly (3 cycles, 30 second grinding at the speed of 5 m/s and 30 second dwell); 2) next, ground hair samples are incubated with a NaOH-based lysis buffer that contains SDS and beta-mercaptoethanol (ßME) for three cycles according to published (8) protocol and in each cycle, the hair residue is recycled through the process with bead milling; 3) pooled supernatant containing hair proteins are precipitated with acetone; 4) pellets from protein precipitation and leftover hair debris are combined for downstream SDS-PAGE; 5) in-gel-digestion was used to generate peptides.

Hair Peptide Mass Spectral Library Construction Including Published GVPs

Using the mass spectral library construction pipeline described in the literature (11), the raw mass spectral data files generated in the present studies were used to construct a hair-specific peptide mass spectral library. This relatively small library contains 6280 spectra (6280 peptide ions of 4343 distinct peptides, higher-energy collisional dissociation (HCD) =30eV), and among these – a total of 3754 spectra (3754 peptide ions of 2240 distinct peptides, HCD=30eV) arose from hair keratins or keratin associated proteins - using the National Center for Biotechnology Information (NCBI, downloaded March 2017) human protein FASTA file with 20,183 sequences plus additional 51 published GVP sequences (1). This provides a sequence coverage of hair cuticular keratins of about 70%. Of these spectra, 40 mass spectra are identified as GVP ions which cover 14 published GVP sites (a subset of total 88 published GVPs): 10 sites from hair cuticular keratins, 1 site from a keratin-associated protein, and 3 sites from non-keratin proteins. Detailed information can be found in the Results and Discussion section where we discuss GVP panel analysis.

Spectrum Library Searching

Freely available MSPepSearch software (peptide.nist.gov) (11) was used to perform mass spectral library searching using a precursor ion tolerance of 20 ppm (ppm was defined as parts per million) and fragment ion tolerance of 50 ppm. Label-free HCD human tryptic peptide spectral libraries (version September 23, 2016 contains 1,127,970 spectra, indicated as 'main' library) are available online (peptide.nist.gov) (12). A hair specific peptide spectral library (indicated as 'hair' library) (13) was created from 90 raw mass spectral data files generated during method development of processing 16 five cm-long hair shafts of this same individual Asian donor. Surprisingly, 40% of peptides contained in this 'hair' library were not present in the ‘main’ library even though it was constructed from a wide range of publicly available data files. Clearly hair was not a common protein-containing material in these studies. This ‘hair’ library was used in combination with the ‘main’ library for mass spectrum library searching. The 1% false discovery rate (FDR) level was determined by using the target-decoy method described in the literature (14,15). The NIST formatted mass spectral libraries were built using the program Lib2NIST freely available online at chemdata.nist.gov. This library and associated software are freely available online (13).

Sequence Database Searching

We used the Sequest (16) HT search node implemented in Proteome Discoverer (PD) 2.1 for initial peptide identification prior to entry into a library and comparison the results of spectral library searching. Mass tolerance settings were the same as in the library searches. The top scoring peptide identification was selected, and FDR level was set at 1% using the same FASTA file described above.

Proteomics Methods

GVP and its non-variant form designation: In this work, GVPs are tryptic peptides that are represented first by their Gene Name followed by the site of the amino acid substitution. For example, “DSP R1783Q_Q” indicates the tryptic peptide derived from Desmoplakin (GN=GSP) containing “Q” at position 1783. The corresponding non-variant form is “DSP R1738Q_R” where “R” is in place of “Q”. The term “GVP ion” refers to not only tryptic peptide sequence, but also charge state and possible modifications. Peptides observed in different charge states or modifications are treated as different peptide ions. The most abundant form of a peptide ion is used to measure its intensity.

LC-MS/MS parameters: Digests were analyzed on an Eksigent Classic 2D Nano LC with an Acclaim PepMap RSLC column (75 µm x 15 cm, C18, 2 µm, 100 Å) with a nanospray source connected to a Thermo Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer in the positive ion mode. Mobile phase A consisted of 0.1% formic acid in water and mobile phase B consisted of 0.1% formic acid in Acetonitrile. The peptides were eluted by increasing mobile phase B from 1% to 90% over 200 minutes. Data was collected using a data dependent mode with a dynamic exclusion of 20 seconds. The top 10 most abundant precursor ions were selected from a 350-1600 m/z full scan for fragmentation. The resolution of full MS scan was set at 120,000 and the resolution of MS/MS scan was set at 30,000. In future work, we plan to perform a 2D-LC study to find more trace ions.

Modifications included in hair library are: (1) fixed carbamidomethyl (CAM) at Cysteine (C); (2) oxidation at Methionine (M); (3) acetylation (Acetyl) at peptide N-terminus; (4) acetaldehyde at peptide N-terminus; (5) Gln->pyro-Glu at Glutamine (Q) at peptide N-terminus; (6) Glu->pyro-Glu at Glutamic Acid (E) at peptide N-terminus. Other less abundant modifications may be added to future versions of the library, although these may be depended on the specific chemical processing involved in the digestion.

Incomplete digestion in proteomics: The inability to digest substantial portions of the proteome is common for the proteomics of biological material. Here are some examples: 1) In reference 8, the reference for the original NaOH+SDS method, hair pellets were simply discarded after incubation with lysis buffer containing NaOH+SDS; 2) In reference 9, scanning electron microscope images as Fig. 2 to show remaining undigested hair after extraction with SDS or with urea. In case 1 and 2, substantial portions of the hair undigested although it is method dependent; 3) In reference 17, heavy-isotope-labeled proteins were used to compare peptide recovery between laboratories and results showed that the digestion step was the greatest source of inconsistent recovery (median loss of 70%). These examples demonstrate that significant levels of incomplete digestion are expected in the proteomics of biological materials.

Results and Discussion

Identification of Hair Proteome including Cuticular Keratins by Direct Extraction Method

We examined overall protein and peptide identifications from all ten gel fractions and compared our library search results to the results from sequence (Sequest) searches. When searching spectral libraries, we added the 'hair' specific mass spectral library to our 'main' library (12,13) to obtain better search performance. The next A and B sub-sections discuss these results and demonstrate the effectiveness of spectral library searching for peptide identification. In sub-section C, we examine GVP detection with library searching in all ten fractions and compare the GVP panel analysis by the Direct method to the other two published methods (1,8).

A. Overall Gel Identification

Results for hair proteins extracted from a single 5 cm-long hair by the Direct method are presented in Table 1. They were derived from one raw MS data file for each of the ten gel fractions. All were independently analyzed to determine details of the gel separation and digestion process.

Using both spectral library and Sequest searching methods, results derived from F1 to F10 are compared in Table 1. As shown in Table 1, when the 'main' library was combined with the 'hair' library for spectral library searching, the overall library identification for proteins - for both hair proteome (7,9) and hair cuticular keratins (a major subset of the hair proteome) (1,8) was similar to that from Sequest, however for all peptides identified, the spectral library method was somewhat more sensitive at a given FDR level, consistent with previous observations (14).

Hair cuticular keratins are major components of hair proteome. Table 2 examined the sequence coverage of listed total 15 hair cuticular keratins of type I and type II by library and Sequest searches from all ten fractions. Peptides present in multiple proteins were used in calculating the sequence coverage of each protein. Since we are interested in GVPs, of course the better coverage, the greater the chance of detecting potential GVP sites. In general, library searching provides a fuller coverage than database searching, although except for the most abundant KRT31, some of these coverages are far less than 100%. There are several possible reasons for this: 1) cross-linking makes certain sites hard to reach by trypsin during the digestion; 2) extremely long (> 50) or short (< 6) peptides were not considered under the current search parameters; 3) loss of extremely hydrophilic or hydrophobic peptides occurs during sample preparation and LC analysis. 4) Incomplete conversion of proteins to peptides is common throughout proteomics, and according to reference 18, an approximately 70–80% of recovery is expected after extraction from the gel. Putting all ten fractions together, 8 out of 15 hair cuticular keratins reach more than 90% coverage, 5 out of the rest 7 reach more than 50%, and only 2 less than 50% (KRT37 and KRT84). Supplementary Document S2 shows sequence coverage in amino acids of 15 type I and type II hair cuticular keratins found by library and Sequest searches.

B. Major and Minor Gel Band Identification

We observed two distinct gel bands in fractions 6 and 7 (Fig. 1). The other fractions had several minor bands but most of the intensity was evenly distributed (Fig. 1C). Results are discussed below.

Fig. 2 shows the intensities over the fractions for selected peptides from type I (A) or type II (B) hair cuticular keratin. In both cases, both the GVP and non-variant form are shown along with another major peptide from each protein. The abundance of each peptide derived from its MS1 ion chromatogram peak area. These results indicate: 1) the major gel bands correspond to type I (fraction 7) and type II (fraction 6) hair cuticular keratins, consistent with literature (8) reports. Fractions 6 (type II) and 7 (type I) are enriched in individual hair cuticular keratins; 2) it is noteworthy that most peptides identified outside the main regions were the same as those inside that region. This behavior persisted in all analyses. This is presumably due to presence of significant quantities of cross-linked proteins or unseparated complexes with higher molecular weight with lower mobilities as well as fragments of these proteins at lower molecular weights with higher mobilities. We find that keratin GVPs are found in virtually all gel fractions suggesting that they distributed among a wide range of crosslinked proteins, suggests that the insoluble, crosslinked portion of the hair protein may not contain additional keratin-GVP identifications. According to reference 7, the insoluble, crosslinked portion has a higher content of non-keratin proteins and may contain additional non-keratin-GVP identifications. Further, we know of no way to enhance the method's digestion effectiveness, though such an improvement would be very welcome.

Note that in Table 1, fractions 6 and 7 show the highest peptide signal strengths but lowest numbers of peptide identifications (IDs). This is confirmed in Fig. 3, where the total ion currents (TICs) are inversely correlated with peptide IDs with a correlation coefficient of -0.75. This is a consequence of the higher concentrations of relatively a few proteins dominating fractions 6 (type II) and 7 (type I), which leads to higher concentrations of their tryptic peptides with consequent signal suppression of peptides from other, less abundant proteins. In other fractions, no individual proteins dominate, so tryptic peptides are more equally spread across a larger number of proteins, though many of them are crosslinked, fragmented or otherwise modified. Supplementary Table S1 shows when moving along the gel fractions from F1 to F10, the example big protein (Desmoplakin) decreases and the example small protein (a Keratin-associated protein) increases.

The major advantage of gel fractioning is that it separates the proteins by molecular weight, thereby showing more clearly the origin in individual GVPs. It can also minimize ion suppression leading to the identification of additional GVPs. Unfortunately, this approach is time-consuming. Our attempts to combine fractions led to loss of potential GVPs (see section C). Identifications of all GVPs in a single digest analysis is apparently not possible at present (discussed below). Finding optimal methods will be the topic of future research.

C. GVP Panel Analyses in All Ten Fractions and Among Three Methods

As described in the Method section, we identified a total of 14 published tryptic GVP sites from this Asian donor’s hair samples. These sequences along with corresponding non-variant sequences, are listed in Supplementary Document S3. Table 3 shows the specific GVP identification for the three methods with three replicate runs for each method, namely: our Direct method, the modified NaOH+SDS method (8), and the Cleavable Surfactant method (1,2). For both the Direct method and modified NaOH+SDS method, GVP panel results from different fractions are combined in Table 3. Supplementary Document S3 uses the results from F1 to F10 as an example to illustrate how we performed this analysis for a complete data set by the Direct method. Analysis led to a number of general findings:

1) For high-abundance GVPs from major keratins, as shown in Fig. 2A or 2B, identifications are easily made. Scores are high [MF: 792 - 942], leading to highly confident identifications (14), retention times are reproducible (Supplementary Document S3), and identifications are made in all gel fractions for both the GVP and its non-variant form.

2) For low-abundance GVPs, mostly arising from less abundant proteins, identifications can be harder to assign, possibly involving lower and variable scores. Confidence can be increased by elution in the expected gel fraction as well as the determination of its non-variant form (sometimes this is made more difficult if GVP site involves a tryptic cleave site at R or K). This is illustrated with two examples:

(a) The GVP site 'DSP_R1738Q_Q: G[Q]SEADSDKNATILELR' (mutated site highlighted in brackets), was identified in the top gel fractions (F1 and F2). This is consistent with its very large precursor protein having 2871 residues, Desmoplakin (DSP). This is an example that R becomes Q and we identified both GVP and its non-variant form in the expected gel fractions with comparable intensity (Supplementary Document S3).

(b) Another GVP site 'KRTAP10-8_H26R_R: TYVIAASTMSVCSSDVG[R]' originates from a much smaller keratin-associated protein (KRTAP, 259 Amino Acids), and was recovered from bottom gel fractions (F9 and F10). This is an example that H becomes R and we only identified GVP but not its non-variant form. Such discrepancy happens because these are two different peptides when GVP site involves R/K. To solve this problem, we would need to choose a different digestion enzyme. Actual release rates for peptides in a protein are not easily predicted and depend on multiple factors (19). So, it is hard to estimate the relative intensities of a GVP and its non-variant if their lengths and possibly charge states are different.

3) The specific GVP identification depends on the experiments, with a number of different GVPs identified by the in-gel and in-solution digestion methods. Hence, false negative results appear to be a significant concern with the present methods, especially for the in-solution method.

4) We note that the identification of both a GVP and its non-variant will significantly increase the confidence of GVP identification. Of course, this is not possible if the source is homozygous or when the non-variant form is not an easily detectable peptide (as may be the case where tryptic cleavage sites are different in the GVP and non-variant form). In this work, the fact that several potential GVPs were observed (Supplementary Document S3), but not at high confidence (low abundance or matching score) reinforces the likelihood that they are not true GVPs.

Fractionating in the gel methods is part of a 2D study – the first dimension is separating hair proteins based on the MW during SDS-PAGE, the second dimension is separating extracted peptides by the LC gradient during LC-MS/MS. Analyzing each fraction enables very low abundance GVPs to be identified. It is why we detect more GVPs from the two in-gel methods than the in-solution method. However, we detect fewer GVPs if we combine these fractions and process as a mixture (Table 3). We also tried a brief 'short-gel' run by applying SDS-PAGE at 200 V for only 10 min (long-gel: 30 min at 200 V). We compare the GVPs between long-gel and short-gel runs and find that short-gel-mixture loses even more GVPs (Table 3). This can be explained by hair proteins not being effectively separated in a shorter run or possibly that SDS not being fully separated from proteins. In any case, this finding highlights the importance of both separation and sensitivity in finding all identifiable GVPs in a sample. While running 10 fractions is very time-consuming, possible GVPs were lost (Table 3) upon combining fractions indicates that more rapid analysis using a single LC-MS/MS run can lose less abundant GVPs. Moreover, the finding that different GVPs are found with different digestion protocols implies that no existing method can be relied on to identify all possible GVPs. Together, this clearly shows the need of future work for finding the most efficient way to maximize GVP identification.

Comparison Between the Direct Method and modified NaOH+SDS Method

Since the Direct method and modified NaOH+SDS method both use protein gel to separate hair proteins, for a direct comparison, we compared the Direct method with modified NaOH+SDS method for a further sensitivity and reproducibility check in this section.

A. Sensitivity

We examine the sensitivity of the Direct method to modified NaOH+SDS method by comparing multiple metrics across a dilution series. In Figure 4, we show the relative sensitivity of the two methods by comparing the degree of dilution needed for each method to yield the similar number of IDs. After comparing total number of ions (Fig. 4A), total number of peptides (Fig. 4B), total number of proteins (Fig. 4C), and total number of GVP ions (Fig. 4D), we found that the Direct method was about eight times more sensitive than modified NaOH+SDS method. The non-monotonic behavior of some of the irregular trends is a consequence of results from the general difficulty in obtaining highly reproducible proteomic results and, for GVPs, their small numbers and therefore greater statistical fluctuation. Note that since the GVPs are few in number and variable in intensity we could not reliably use GVPs alone to develop a reliable measure of method sensitivity based on their identifications alone. This was confirmed in a separate set of analyses: for example, GVP ions increased at 10D and then all the way decreased to minimum detection level at 1280D.

The present Direct method is both suitable for very small hair samples, and able to identify GVP ions across a broad range of ion intensity. Intensities of reliably identified GVP ions could differ by orders of magnitude in ion intensity. Fig. 5 illustrates this for two spectra of the same GVP ion 'QVVSSSEQLQSYQ[V]EIIELR/3_0'. Even though intensities differ by four orders of magnitude, retention times were almost identical (161.7 min vs. 161.5 min) and spectral library match factors were quite high (over 800).

B. Reproducibility

In an examination of the reproducibility of the present method, the extraction was repeated eight times using eight individual 5 cm-long hair shafts (labeled as A to H in Fig. 6A) from the same donor, and particularly compared it to modified NaOH+SDS method (labeled as 1A to 1H in Fig. 6B, plus the last lane from 10 hairs included as a reference). We made the assumption that each individual 5 cm hair shaft contained the same protein mass. Fig. 6 clearly indicates that the Direct method is more reproducible than modified NaOH+SDS method. This presumably arises from lower sample loss for the Direct method since it only needs one-step/30 min for hair protein extraction, while the multiple-steps (also means much longer bench time) included in modified NaOH+SDS method are more prone to sample loss and generating variable results (workflows of the two methods are shown in S1) especially when the hair sample is very small.

We also compared the protein, peptide, and GVP identifications between the Direct method and modified NaOH+SDS method with analysis repeated three times for each method. Results of comparisons from a representative fraction (F6) are listed in Table 4 with three experimental repeats: 1) higher average peptide yield (µg) was obtained in the Direct method than in the modified NaOH+SDS method (11.5 vs. 2.9 µg); 2) more average peptides were identified by the Direct method than by the modified NaOH+SDS method (610 vs. 509); 3) although similar average number of GVP ions was observed in the Direct and modified NaOH+SDS methods, it is more reproducible with much smaller coefficient of variation (CV) in three experimental repeats in the Direct method (0.02 vs. 0.27 , respectively); 4) gel blank - only a few peptide IDs from gel blank and no GVP identification at all. Gel blank serves as a control to see if we introduce any contamination from handling the blank gel alone. Table 4 shows that the Direct method is not only a more sensitive, but also a more reproducible method when compared to the modified NaOH+SDS method.

Estimation of the digestion yield: The gel-based method we chose for analysis unfortunately did not allow us to use a conventional Bradford colorimetric (BCA) assay to measure protein concentration. Instead, yields of digested peptides using the Pierce method mentioned above served a similar, albeit less direct purpose. Based on a measured 5 cm hair mass of 100 µg (10 5-cm lengths were found to weigh 1.0 mg), we found that at the incubation time of 5, 10, 15, 30, 60 and 90 minutes, corresponding total yields of peptides to be 16%, 27%, 37%, 75%, 66% and 51%. The maximum of 75% at 30 min was selected as optimal (see above). For comparison, a yield of 47% was reported for an in-solution method (8) using BCA assay after precipitating extracted proteins.

Examination of Artifacts Among Three Methods

In most proteomics experiments, a large fraction of ions sampled are not identified. This not only reduces the efficiency of the experiment but also has potential to generate false positive results. Moreover, the identity of the unidentified ions may aid in understanding and optimizing the experiment and provide a measure of quality control.

In the present experiment almost 90% of ions are not directly identified as tryptic peptides using conventional library searching. Using our recently developed hybrid search (15), as shown in Supplementary Table S2, 11% can be identified as expected tryptic peptides, while about 75% can be identified via hybrid identification. These hybrid identifications find peptides that are chemically modified forms of conventional tryptic peptides. The reason we would like to examine experimentally introduced artifacts is because we must be aware of artifactual modifications that may masquerade as a GVP and therefore generate false positive identifications, the larger the number of spurious modifications the greater the chance that one will accidentally overlap a possible GVP. Proteomics cannot distinguish biological versus artifact origins of identified peptides. For example, a methylation at or near a serine might be interpreted as a serine to threonine GVP. IonPlot in Fig. 7 shows the classification of ions (GVP, Identified, and not-identified ions from F6 of the Direct method) by the hybrid search including a list of several interesting modifications that we would like to discuss more in this section. These analyses also show the nature and extent of certain spurious chemical processes that add to sample complexity and, in effect, diminish the sensitivity and overall quality of the experiment.

Since this issue is important for every sample preparation method regarding to GVP detection, below we examine the artifacts among the three methods: our Direct method, modified NaOH+SDS method, and Cleavable Surfactant method.

Table 5 compares the twenty most frequently identified DeltaMass values in three methods (15). For more information, Supplementary Document S4 shows the histograms of all DeltaMass values obtained from hybrid search identifications in each method to give a broad view of the distribution of all DeltaMass values. From the top 20 DeltaMass values listed in Table 5, we now further discuss four types of experimentally introduced artifactual modifications (Fig. 8).

Acetaldehyde adduction. We compared the occurrence of an acetaldehyde adduct across the three methods. Fig. 8 shows that this artifactual modification is more frequently identified in the Direct and modified NaOH+SDS methods due to the presence of ethanol in the SimplyBlue SafeStain that we used to stain the protein gels. We here included an example in Fig. 9 to show our main concern – a modification at peptide's N-terminus could be mistaken as a potential GVP: the DeltaMass value from the hybrid search for this hybrid identification is 26.0186 Da, within the mass tolerance range, which is likely due to acetaldehyde (26.01565 Da) but may be incorrectly identified as His (H) →Tyr (Y) (26.004417 Da) since His (H) is involved in the identification at the first amino acid in this peptide ion. Without the hybrid search, or without being aware of what type of artifactual modification exists, such a mis-identification will occur.

Acetylation. While acetylation at Lys (K) and the protein amino terminus are biological modifications, artifactual acetylation at the peptide N-terminus can be introduced during sample preparation. Although the source of acetic acid is not believed to have been introduced through sample preparation, this artifactual modification was identified more frequently in the Direct and modified NaOH+SDS methods.

Formylation. Formylation is less dissimilar across all three methods than that of the previous described two modifications. This is expected as formic acid is required in all three sample preparations.

Alkylation. Alkylation (CAM) is significantly greater in the Cleavable Surfactant method compared to the Direct and modified NaOH+SDS methods. This is consistent with the fact that iodoacetamide concentration we used in sample preparation of Cleavable Surfactant method is much higher than in the Direct and modified NaOH+SDS methods.

Table 5 and Supplementary Document S4 show that, overall, results of the three methods have similar degrees of experimentally introduced modifications. It seems likely that the artefactual modifications are a result of the inherent difficulty of digestion such an insoluble and crosslinked material as hair.

Regarding to GVP panel analysis, we find consistent results in regular and hybrid searches. Hybrid searching usually reports more GVP ions with many kinds of unexpected modifications but seems not gaining additional known GVP site detection. Verified GVP detection by the hybrid search (not only seeing the version that included in the library but also seeing the versions with some unexpected modifications) increases the confidence of GVP panel analysis.

Identification of Hair Proteome and Cuticular Keratins from as Little as 1 cm-long Human Hair Shaft by Direct Extraction Method

So far, the data we presented in this manuscript used 5 cm-long hair shafts as the starting material. While we learned about the sensitivity of the Direct method with the serial dilution study, we also wanted to check results using smaller lengths of hair. As the dilution series was a projection for low amounts based on similar extraction efficiencies for smaller lengths, one may expect further losses due to possible inefficiencies in digesting small lengths of hair. For this purpose, we undertook a series of studies where hair shaft varied from 5, 2.5, and 1 cm-long. Fig. 10A shows the separation of hair proteins by SDS-PAGE for three different hair lengths and Table 6 lists the total number of hair proteins and peptides identified as well as those that are specific for hair cuticular keratins and GVP ions. Fig. 10B shows the analysis of an example GVP ion whose abundance is almost linear in 5, 2.5, and 1 cm hair shaft samples to demonstrate the abundance is proportional to length. These results show that as little as 1 cm-long hair shaft sample can be analyzed by this Direct method. There is no reason to believe it would not work effectively for even smaller amounts of hair, suggesting that even forensic-relevant trace quantities of hair would be suitable for this analytical method.

Examination of the Direct Method in Another Donor

To ensure that these results were not unique to one donor, we applied the Direct method to another randomly selected donor's hair shaft samples obtained from BioreclamationIVT (LOT# BRH1363733, 5 g of hair shafts from a Caucasian male, 23 years old). Table 7 lists the total number of hair proteins and peptides identified as well as those from hair cuticular keratins and GVP ions. These results demonstrate that the Direct method works equally well for another donor's hair samples. The overall protein gel images, the peptide yields from in-gel-digestions, the hair keratins and their peptide identifications, and the number of found GVP ions are similar. Most of high abundance GVPs in this Caucasian donor overlap with previous described Asian donor in the GVP panel analysis. This manuscript is focused on the protein and peptide extraction from single hair shaft, that is the reason why we use hair samples from the same Asian donor for the development of protein extraction method. We believe our Direct method would work effectively for hair samples from any individual donor. These studies did not consider donors who heated or chemically treated their hair – this would be a useful topic for future research. The focus of this paper was only analytical methods and detailed proteomic analysis. Variations with hair origin will be the topic of future studies using the methods described here.

Summary and Conclusions

In summary, we have shown that the Direct extraction method is a sensitive, reliable, and relatively convenient method based on the depth of coverage of the human hair proteome and cuticular keratins: 1) It is a relatively sensitive method: it works for a hair shaft as short as 1 cm;

2) It is a relatively reliable method: it generates more consistent results in protein/peptide identification and GVP detection; 3) It is a relatively convenient method: it is simple to carry out since there is only one-step in protein extraction from hair, although to assure maximum GVP identification, it does require multiple LC-MS/MS runs.

Using our recently developed ‘hybrid’ spectral library search method, we have found that a very large fraction of the peptide spectra acquired were not simple tryptic peptides derived from known proteins. A conventional library search can identify only 11% of the peptides, who the hybrid search identifies 75%, including any previously unidentified GVPs (as our future work). We have also shown that the hybrid search, could be used to identify potential sources of false positives due to the presence of artifactual modifications that are experimentally introduced. Modifications that could be mistaken as a GVP should be the primary concern and a separated examination of artefactual modifications is needed. In difficult cases, a more careful manual checking of GVP spectra may also be needed.

Although we recommend the Direct method because of several advantages we described earlier, we also realize different methods may be most suitable for different GVP panel analysis. Each method will have its own strength and weakness. Unless we combine the results from all three tested methods, no single method covered all the identified published GVP sites in this study. This is largely because of the nature of the hair samples – heavy crosslinking makes hair mechanically strong and stable, but also very resistant to sample processing.

We have also shown that a GVP analysis can effectively done using a peptide spectral library containing all identifiable peptides derived from human hair samples. With this paper we provide a library containing all identified hair derived peptides (13). Future expansion of this library can include all known GVPs as well as all identifiable peptides derived from human hair. Further, it may be combined with the NIST-developed label-free HCD main peptide library (peptide.nist.gov) (12) to provide another layer of sensitivity and confidence for hair peptide identification and GVP detection.

Supporting Information

Supplementary Table S1. Example of a Big Protein and a Small Protein Amount Change in Ten Gel Fractions by the Direct Method

Supplementary Table S2. Percentages of Hybrid IDs in All Ten Gel Fractions by the Direct Method

Supplementary Document S1. Outline of Protein Extraction Work Flows for Direct Method and modified NaOH+SDS Method

Supplementary Document S2. Comparison of Sequences Coverage in Amino Acids of 15 type I and type II hair cuticular keratins by library and Sequest searching

Supplementary Document S3. GVP Panel Analyses in All Ten Fractions by the Direct Method

Supplementary Document S4. Histograms of the Distribution of All DeltaMass Values in Three Methods

References

1. Parker GJ, Leppert T, Anex DS, Hilmer JK, Matsunami N, Baird L et al. Demonstration of Protein-Based Human Identification Using the Hair Shaft Proteome. PLoS One 2016; 11(9): e0160653.

2. Mason KE, Paul PH, Chu F, Anex DS, Hart BR. Development of a Protein-based Human Identification Capability from a Single Hair. J Forensic Sci 2019 Jul; 64(4):1152-9.

3. Carlson TL, Moini M, Eckenrode BA, Allred BM, Donfack J. Protein extraction from human anagen head hairs 1-millimeter or less in total length. Biotechniques 2018; 64(4):170-6.

4. Bengtsson CF, Olsen ME, Brandt LØ, Bertelsen MF, Willerslev E, Tobin DJ et al. DNA from keratinous tissue. Part I: hair and nail. Ann Anat 2012; 194(1): 17-25.

5. Langbein L, Rogers MA, Winter H, Praetzel S, Beckhaus U, Rackwitz HR et al. The catalog of human hair keratins. I. Expression of the nine type I members in the hair follicle. J Biol Chem 1999; 274(28): 19874-84.

6. Langbein L, Rogers MA, Winter H, Praetzel S, Schweizer J. The catalog of human hair keratins. II. Expression of the six type II members in the hair follicle and the combined catalog of human type I and II keratins. J Biol Chem 2001; 276(37): 35123-32.

7. Lee YJ, Rice RH, Lee YM. Proteome analysis of human hair shaft: from protein identification to posttranslational modification. Mol Cell Proteomics 2006; 5(5): 789-800.

8. Wong SY, Lee CC, Ashrafzadeh A, Junit SM, Abrahim N, Hashim OH. A High-Yield Two-Hour Protocol for Extraction of Human Hair Shaft Proteins. PLoS One 2016; 11(10): e0164993.

9. Adav SS, Subbaiaih RS, Kerk SK, Lee AY, Lai HY, Ng KW et al. Studies on the Proteome of Human Hair - Identification of Histones and Deamidated Keratins. Sci Rep 2018 Jan; 8(1): 1599.

10. Jimenez CR, Huang L, Qiu Y, Burlingame AL. In-gel digestion of proteins for MALDI-MS fingerprint mapping. Current Protocols in Protein Science 1998; 14(1): 16.4.1-5.

11. Rudnick PA, Markey SP, Roth J, Mirokhin Y, Yan X, Tchekhovskoi DV et al. A Description of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) Common Data Analysis Pipeline. J Proteome Res 2016; 15(3): 1023-32.

12. The NIST Main Libraries of Peptide Tandem Mass Spectra

https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:lib:humanhcd20160503

13. The NIST Hair Libraries of Peptide Tandem Mass Spectra

https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:lib:human_hair_selected_with_gvps_passed

14. Zhang Z, Burke M, Mirokhin YA, Tchekhovskoi DV, Markey SP, Yu W et al. Reverse and Random Decoy Methods for False Discovery Rate Estimation in High Mass Accuracy Peptide Spectral Library Searches. J Proteome Res 2018; 17(2): 846-57.

15. Burke MC, Mirokhin YA, Tchekhovskoi DV, Markey SP, Heidbrink Thompson J, Larkin C et al. The Hybrid Search: A Mass Spectral Library Search Method for Discovery of Modifications in Proteomics. J Proteome Res 2017; 16(5): 1924-35.

16. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 1994; 5(11): 976-89.

17. Abbatiello SE, Schilling B, Mani DR, Zimmerman LJ, Hall SC, MacLean B et al. Large-Scale Interlaboratory Study to Develop, Analytically Validate and Apply Highly Multiplexed, Quantitative Peptide Assays to Measure Cancer-Relevant Proteins in Plasma. Mol Cell Proteomics 2015 Sep; 14(9): 2357-74.

18. Speicher K, Kolbas O, Harper S, Speicher D. Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. J Biomol Tech 2000 Jun; 11(2): 74–86.

19. Lowenthal MS, Liang Y, Phinney KW, Stein SE. Quantitative bottom-up proteomics depends on digestion conditions. Anal Chem 2014 Jan; 86(1):551-8.

Table 1. Comparison of Protein and Peptide Identifications from Spectral Library and Sequest Searching in All Ten Fractions at 1% FDR by the Direct Method from a 5 cm-long Hair Shaft*.

Direct	Yield (µg)	TIC	Main+Hair Spectral Library				Sequest			
			Hair Proteome		Cuticular Keratins		Hair Proteome		Cuticular Keratins	
			Proteins	Peptides	Proteins	Peptides	Proteins	Peptides	Proteins	Peptides
F1	1.76	3.91E+06	148	2040	14	583	98	1128	14	471
F2	3.81	6.54E+06	140	1888	15	614	84	1052	14	503
F3	5.46	1.03E+07	132	1744	14	614	73	1022	14	525
F4	8.95	1.44E+07	134	1789	14	628	83	1045	13	526
F5	5.86	8.27E+06	152	1781	14	594	93	1061	14	513
F6	13.25	2.06E+07	135	1617	15	620	68	906	15	503
F7	10.92	2.31E+07	146	1607	13	623	76	933	14	538
F8	7.06	8.17E+06	207	2167	15	631	129	1290	15	521
F9	5.98	4.72E+06	214	2268	14	589	138	1346	13	463
F10	12.24	8.59E+06	173	1744	14	470	120	1079	13	347

*Proteins were identified by ≥ 2 peptides throughout this manuscript. For peptide/protein identifications (IDs) under ‘Hair Proteome’, Fraction 8 (F8) and 9 (F9) gave more IDs in both spectral library and Sequest searches; for peptide/protein IDs under ‘Cuticular Keratins’, the distribution of IDs was more even across all 10 gel fractions in both spectral library and Sequest searches. TIC: an index of total ion current.

Table 2. Comparison of Sequence Coverage (%) of Hair Cuticular Keratins from Spectral Library and Sequest Searching in All Ten Fractions by the Direct Method.

Cuticular Keratins	From Library	From Sequest
KRT31	100.0	97.6
KRT32	54.2	49.6
KRT33A	97.0	93.3
KRT33B	97.0	93.6
KRT34	86.0	83.9
KRT35	91.0	86.4
KRT36	60.8	49.3
KRT37	43.0	34.7
KRT38	61.2	51.3
KRT81	96.2	91.9
KRT82	63.4	49.9
KRT83	97.0	87.2
KRT84	12.7	11.2
KRT85	96.8	89.4
KRT86	99.2	92.4
Average	**77.0**	**70.8**

Table 3. Genetically Variant Peptide (GVP) Panel Analyses in Three Methods*.

	DSP	GSDMA	KRT31	KRT32	KRT33A	KRT33B	KRT35	KRT35	KRT81	KRT82	KRT83	KRT83	KRTAP 10-8	TGM3
ONE 5 CM HAIR, ASIAN	R1738Q_Q	V128L_L	A82V_V	S222Y_Y	A270V_V	V279L_L	P443A_A	S36P_P	S13R_R	T458M_M	G362S_S	I279M_M	H26R_R	T13K_K
D_LG_F1_TO_F10_R1#	X		X		X			X	X		X	X	X	X
D_LG_F1_TO_F10_R2	X		X		X	X		X	X		X	X	X	X
D_LG_F1_TO_F10_R3	X		X	X	X	X		X	X		X	X	X	X
D_LG_COMBINED_R1			X		X	X		X	X			X		X
D_LG_COMBINED_R2			X		X	X		X	X			X		X
D_LG_COMBINED_R3			X		X			X	X			X		X
D_SG_COMBINED_R1			X		X			X	X			X		X
D_SG_COMBINED_R2			X		X			X	X			X		X
D_SG_COMBINED_R3			X		X		X	X	X			X		X
NS_LG_F1_TO_F10_R1	X	X	X		X		X	X	X		X	X	X	X
NS_LG_F1_TO_F10_R2	X	X	X	X	X			X	X		X	X		X
NS_LG_F1_TO_F10_R3	X		X	X	X		X	X	X		X	X	X	X
CS_R1			X				X			X		X		X
CS_R2			X				X	X		X		X		X
CS_R3			X				X			X				X

*all listed GVP analyses are derived from the same Asian donor's single 5 cm-long hair samples: GVP panel analyses by the Direct method with all 10 fractions from a long-gel (30 min run at 200 V) which have been individually processed by LC-MS/MS and then summarized the results in one row are labeled as 'D_LG_F1_TO_F10'; GVP panel analyses with combined fractions processed as a mixture from a long-gel run by the Direct method are labeled as 'D_LG_COMBINED'; with combined fractions from a short-gel run (10 min run at 200 V) are labeled as 'D_SG_COMBINED'; GVP panel analyses by the modified NaOH+SDS method with all 10 fractions from a long-gel run individually processed and then summarized are labeled as ‘NS_LG_F1_TO_F10’; GVP panel analyses by the Cleavable Surfactant method are labeled as ‘CS’. R1, R2, and R3 are three experiment repeats.

#results from F1 to F10 are listed in Supplementary Document S3, used as an example to demonstrate a GVP panel analysis from this ‘D_LG_F1_TO_F10_R1’ data set.

Table 4. Examination of Reproducibility for the Direct Method and modified NaOH+SDS method* from a Representative Gel Fraction (F6).

Methods (one 5 cm hair, Asian)	**Yield (µg)**	**Main+Hair Spectral Library**				
		Hair Proteome		**Cuticular Keratins**		**GVP ions**
		Proteins	**Peptides**	**Proteins**	**Peptides**	
Direct_R1	10.32	114	1427	14	593	43
Direct_R2	13.25	135	1617	15	620	44
Direct_R3	10.94	132	1725	14	618	45
NaOH+SDS_R1	3.36	101	1267	14	509	29
NaOH+SDS_R2	2.11	93	1178	14	497	51
NaOH+SDS_R3	3.32	83	1137	15	520	45
Blank Gel	0.04	6	17	2	7	0

*The result was obtained from fraction 6, a representative gel fraction. Three experimental repeats: R1, R2, and R3.

Table 5. The Twenty Most Frequently Identified DeltaMass Values Obtained from Hybrid Search Identifications in the Three Methods.

DeltaMass	Theorical Value of DeltaMass	Proposed Modification	Percent of Hybrid Identifications		
			Direct (Median)	NaOH+SDS (Median)	Cleavable Surfactant (Median)
1.001	1.00335483	1-C13	17.30	17.76	19.34
2.007	2.00670966	2-C13	6.73	8.82	6.71
42.013	42.010565	Acetyl	6.25	5.75	3.54
26.017	26.015650	Acetaldehyde	3.52	2.49	0.66
3.009	3.01006449	3-C13	3.59	4.96	3.55
27.999	27.994915	Formyl	1.87	3.03	1.57
14.018	14.015650	Methyl	3.08	2.60	1.12
-1.011	-1.00335483	-1-C13	2.31	3.05	
-17.023	-17.026549	-NH3	1.62	1.51	2.38
70.007	70.005480	Formyl + Acetyl	0.89	1.28	
4.009	4.01341932	4-C13	1.78	2.44	2.02
12.002	12.000000	Formaldehyde Adduct	1.45	1.20	
43.014	43.005814	Carbamyl/Acetyl + 1-C13	1.48	1.07	0.70
-18.008	-18.010565	Dehydration/Glu→pyro-Glu	1.34	1.35	2.01
-2.013	-2.00670966	-2-C13	1.36	1.58	1.43
23.986	23.98865266	Sodiated + 2C-13	1.17		
57.023	57.021464	CAM	1.78	1.87	4.21
15.997	15.994915	Oxidation	1.08	1.28	
120.028	120.024500	Desulferization + CAM + DTT	0.95		
58.010	58.005480	Deamidation + CAM	1.06	0.89	3.33
-91.009	-91.009185	Cys(CAM)→Dehydroalanine		0.82	
-16.019	-16.0231942	1C-13 + -NH3		0.76	0.93
-0.983	-0.984016	Amidation			3.44
5.014	5.01677415	5-C13			0.69
160.041	160.030654	Add-Cys+CAM			1.25
31.995	31.989829	Dioxidation			1.78
152.003	151.996571	+DTT			0.86

Table 6. Reduction of Starting Material to 1 cm-long Hair Shaft by the Direct Method*.

Hair Length (cm)	Main+Hair Spectral Library				
	Hair Proteome		Cuticular Keratins		GVP ions
	Proteins	Peptides	Proteins	Peptides	
5	135	1617	15	620	44
2.5	86	1203	14	563	40
1	78	1149	14	486	39

*The result was obtained from fraction 6, a representative gel fraction.

Table 7. Comparison of Protein and Peptide Identification from a 5 cm-long Hair Shaft from Asian and Caucasian Male Donor by the Direct Method*.

Donor	Yield (µg)	Main+Hair Spectral Library				
		Hair Proteome		Cuticular Keratins		GVP ions
		Proteins	Peptides	Proteins	Peptides	
Asian	13.25	135	1617	15	620	44
Caucasian	8.48	92	1177	14	581	45

*The result was obtained from fraction 6, a representative gel fraction.

Figure Legends

FIG. 1—*Time Course Study to Optimize the Best Heating Condition of the Direct Method. A time-course study was performed to find the optimal time that a 5 cm hair shaft sample need to be heated at 90°C. (A) The scanned gel image included a MW standard loaded in the first lane and six additional lanes where the samples were loaded on increasing length of time for which they have been heated at 90°C (5, 10, 15, 30, 60, and 90 min). The major bands that correspond to type I and type II hair cuticular keratins were labeled. The orange thin lines indicate fractionating the gel to 10 slices from top to bottom as "F1" to "F10". (B) The chart shows the density reports of type I and type II bands at each time interval. The density reports were obtained from gel scanning. The best time point (30 min) is labeled in red based on giving the maximum density reports for both type I and type II bands at 30 min. (C) The chart shows the density ratios of all 10 gel fractions obtained at 30 min, using fraction 1 as the reference.*

FIG. 2—*The Range of the Intensities of Example Peptide Ions Across All Ten Fractions from the Direct Method in Type I and Type II Cuticular Keratins. (A) Type I cuticular keratin KRT33A: The range of intensities of an example GVP peptide ion pair (KRT33A A270V_V: QVVSSSEQLQSYQ[V]EIIELR/3_0 (blue square linked by blue line) and KRT33A A270V_A: QVVSSSEQLQSYQ[A]EIIELR/3_0 (blue triangle linked by blue line)) as well as another peptide ion (SQQQEPLVCASYQSYFK/3_1/9, C, Carbamidomethyl (orange circle linked by orange line)) whose sequence is unique to KRT33A but not containing a known GVP site across all 10 fractions. 'KRT33A A270V_A' or 'KRT33A A270V_V' means the amino acid at position 270 of KRT33A can be a 'A' (regular version in human FASTA file) or a 'V' (published variable version). Dashed black line indicates these three peptide ions reach their maximum intensities at Fraction 7. (B) Type II cuticular keratin KRT83: The range of intensities of an example GVP*

peptide ion pair (KRT83 I279M_M DLNMDC[M]VAEIK/2_3/4,M,Oxidation/6,C,Carbamidomethyl/7,M,Oxidation (blue square linked by blue line) and KRT83 I279M_I DLNMDC[I]VAEIK/2_2/4,M,Oxidation/6,C,Carbamidomethyl (blue triangle linked by blue line)) as well as another peptide ion (LCEGVEAVNVCVSSSR/2_2/2,C,Carbamidomethyl/11,C,Carbamidomethyl (orange circle linked by orange line)) whose sequence is unique to KRT83 but not containing a known GVP site across all 10 fractions. 'KRT83 I279M_I' or 'KRT83 I279M_M' means the amino acid at position 279 of KRT83 can be an 'I' (regular version in human FASTA file) or a 'M' (published variable version). Dashed black line indicates these three peptide ions reach their maximum intensities at Fraction 6.

FIG. 3—*The range of total ion current (TIC, upper panel) and peptide identifications (lower panel) across all 10 fractions. Blue dashed lines indicate TIC values reach their maximum numbers at Fractions 6 & 7, where peptide IDs reach their minimum numbers at Fractions 6 & 7.*

FIG. 4—*Comparison of the Sensitivity in the Two Methods. The sensitivity of the two methods was measured by comparing multiple metrics across a dilution series from 5D to 1280D: (A) the total number of ions; (B) the total number of peptides; (C) the total number of proteins; (D) the total number of published GVP ions detected in mass spectral data from 5 cm-long hair shaft sample derived proteins that were extracted using the Direct method (blue) and modified NaOH+SDS method (green). Actual data has been labeled on the points of each dilution series.*

FIG. 5—*Identification of an Example GVP Ion with High and Low Abundance. The example GVP ions (KRT33A A270V_V: QVVSSSEQLQSYQ[V]EIIELR/3_0 higher-energy collisional*

dissociation (HCD) =30eV) were mapped to an IonPlot (x-axis: Retention Time (RT) in min, y-axis: Abundance in log 10 scale) to show the library identification with high abundance (upper blue dot) or with low abundance (lower blue dot). One blue dot indicates one peptide ion. For each blue dot, the RT and the abundance in log 10 scale were labeled underneath; blue arrows indicate their corresponding library identifications by searching the spectrum of this peptide ion as query spectrum against the hair specific peptide spectral library including known GVP ions. The match factor (MF) was labeled underneath its library identification.

FIG. 6—*Comparison of the Reproducibility of the Direct and modified NaOH+SDS Methods. The two gel images compare the reproducibility of method (A) the Direct method and (B) modified NaOH+SDS method using 5 cm-long hair shaft samples from the same individual donor across 8 replicates (A: A to H; B: 1A to 1H). A MW standard was loaded in the first lane. Note that the NaOH+SDS gel includes a 9th lane for which the extraction from ten 5cm-long hair shaft samples was included as a reference. The major bands that correspond to type I and type II hair cuticular keratins were labeled.*

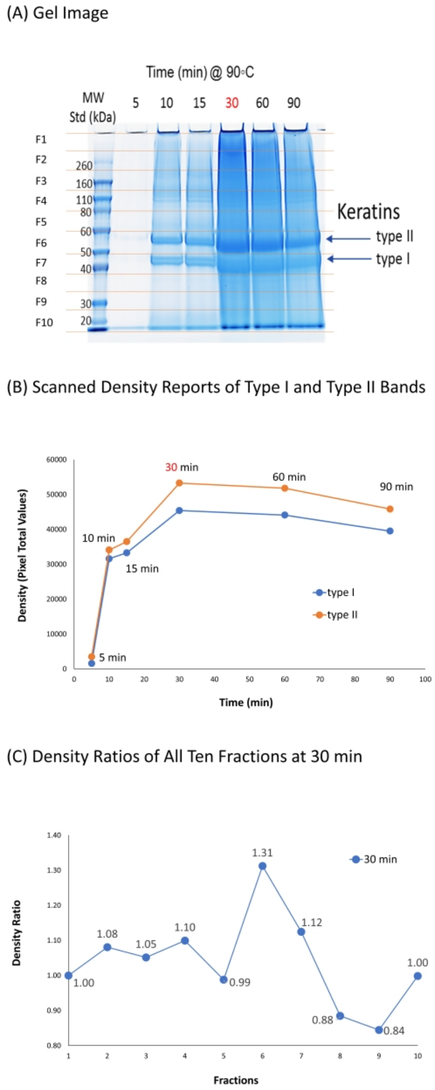
FIG. 7—*Classification of Ions by the Hybrid Search. IonPlot shows the classification of GVP, identified, and not identified (NoID) ions, as well as several modifications: formylation (formyl), methylation (methyl), alkylation (CAM), acetaldehyde, and acetylation that present in fraction 6 (F6), a representative gel fraction from a protein gel separating proteins derived from a 5 cm-long hair shaft of this Asian donor by the Direct method. Solid: identified by regular library search; Hollowed: identified by hybrid library search. x-axis: Retention Time (RT) in minute (min), y-axis: Abundance in log 10 scale.*

FIG. 8—*Comparison of the Artifacts in the Three Methods. Comparison of experimentally introduced artifactual modifications among three methods using our recently developed hybrid*

search: Cleavable Surfactant method (red), modified NaOH+SDS method (green) and the Direct method (blue). The compared experimentally introduced artifactual modifications chosen as examples are: acetaldehyde (upper left), acetylation (upper right), formylation (lower left) and over alkylation (lower right).

FIG. 9—*An Example of a Modification at Peptide N-terminus Mistaken as a GVP. Spectral match of a hair-derived peptide to the peptide sequence HLQLAIR (Charge=2, Mods=0, Spectral Match Score=705) with a DeltaMass of 26.0186 Da, which is likely due to acetaldehyde (26.01565 Da) but may be incorrectly identified as His (H) →Tyr (Y) (26.004417 Da).*

FIG. 10—*Comparison of Hair Length Variation. Comparison of hair length variation. (A) This gel image shows the separation of hair proteins from 5, 2.5, and 1 cm-long hair shaft samples from the same individual donor. A MW standard was loaded in the first lane. Bands for type I and type II hair cuticular keratins were labeled. (B) spectral match (MF=921) of an example GVP ion (KRT31_A82V_V: DN[V]ELENLIR/2_0 HCD=30eV) is on the left. The spectrum shown in red is the query spectrum and the spectrum shown in blue is the reference library spectrum for this GVP ion. On the right is a plot that shows the abundance of this example GVP ion in the 1, 2.5, and 5 cm hair shaft samples is approximately linear. Note the y-axis is the log of the abundance value, plotted on a linear scale.*

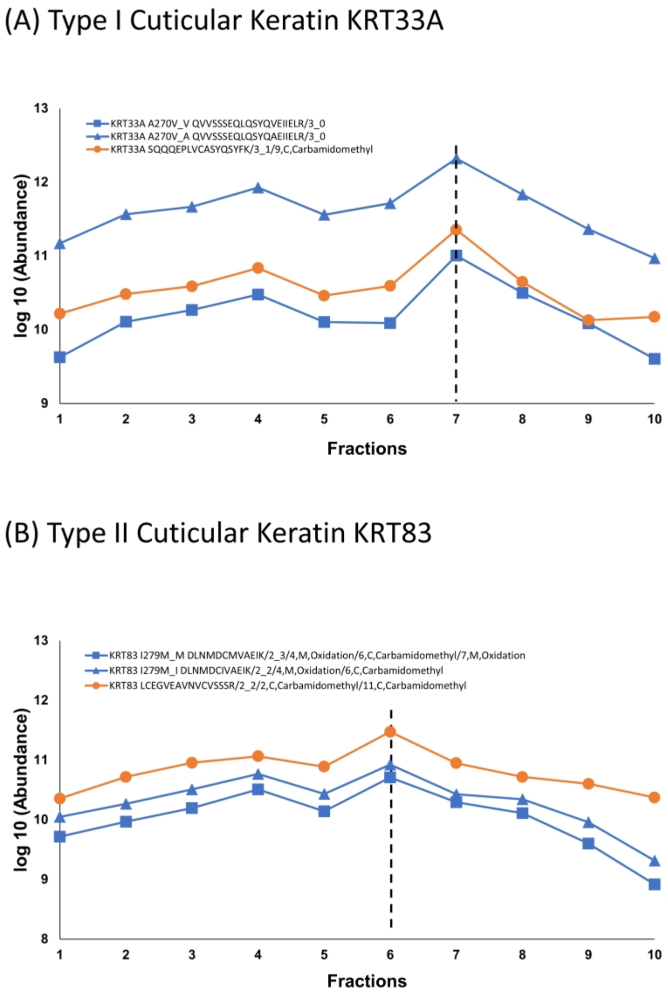


(A) Gel Image

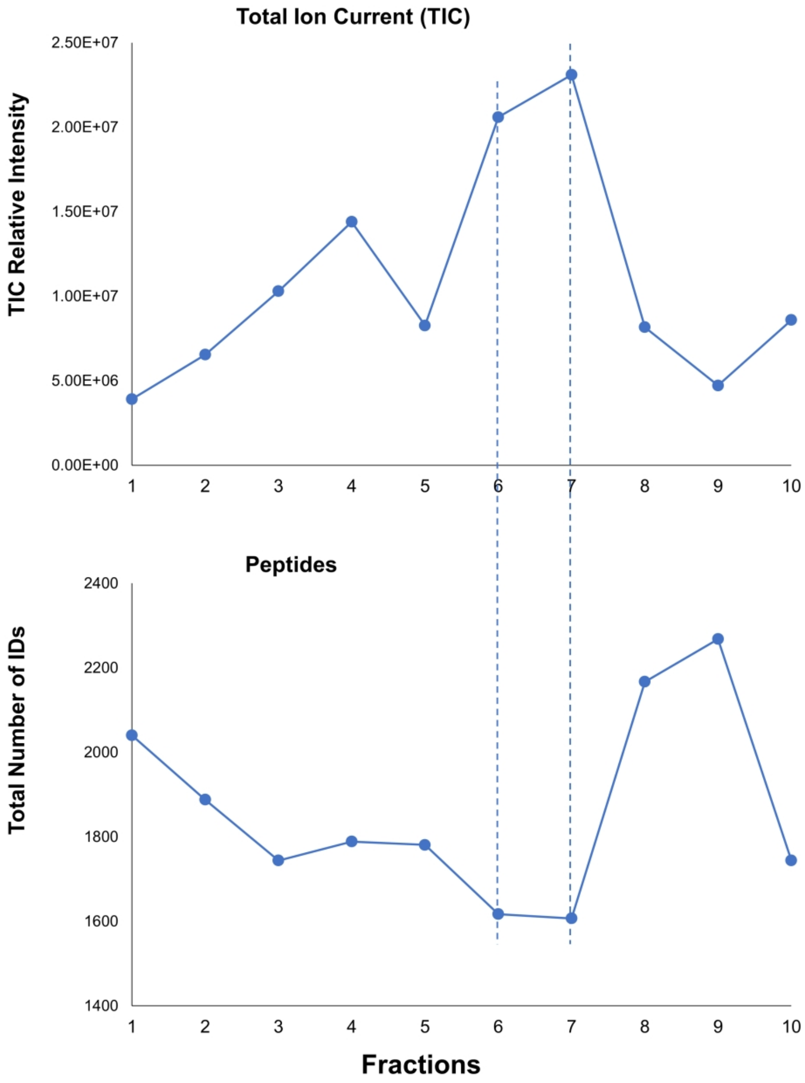
(B) Scanned Density Reports of Type I and Type II Bands

(C) Density Ratios of All Ten Fractions at 30 min

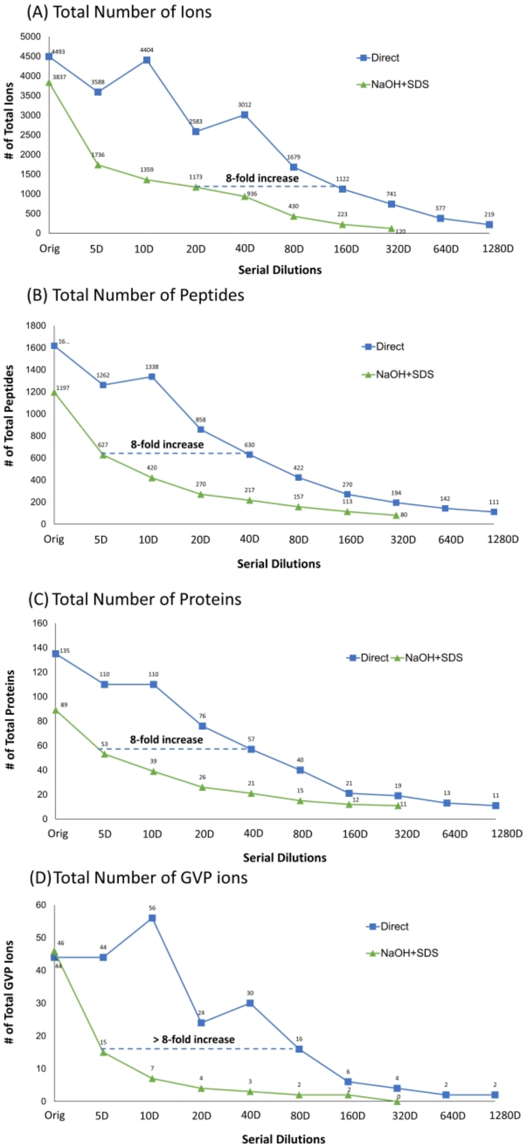
338x635mm (300 x 300 DPI)



304x381mm (300 x 300 DPI)



304x381mm (300 x 300 DPI)

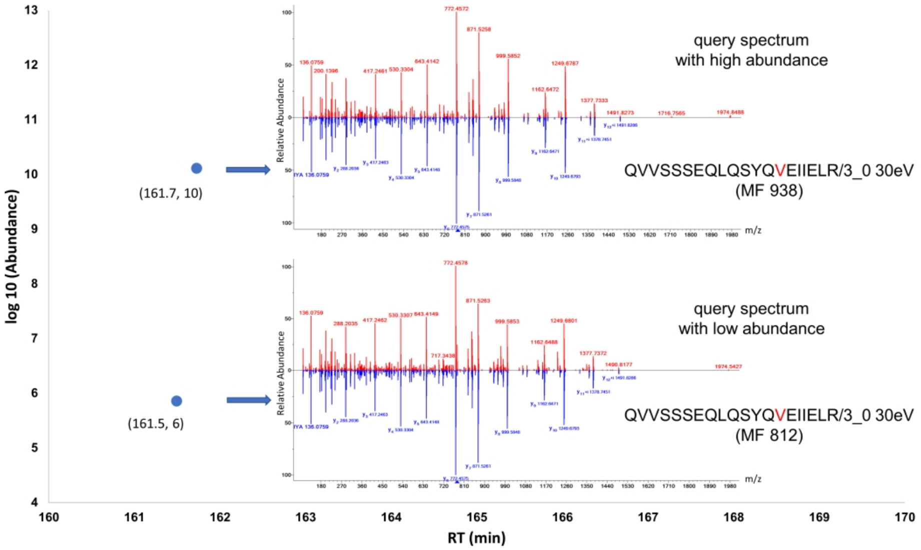

(A) Total Number of Ions

(B) Total Number of Peptides

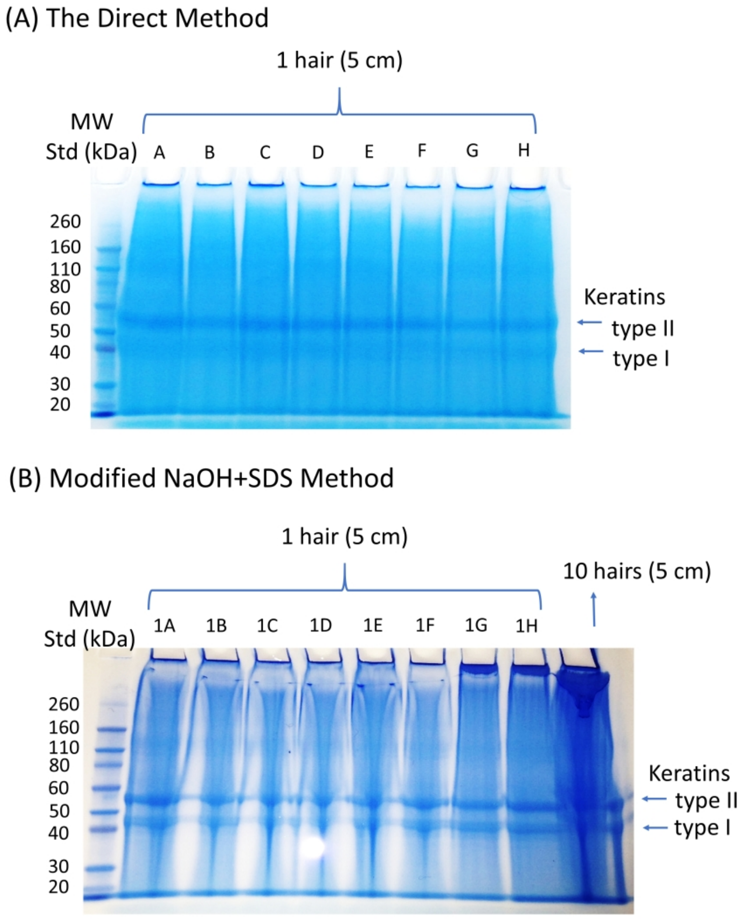
(C) Total Number of Proteins

(D) Total Number of GVP ions

338x635mm (300 x 300 DPI)



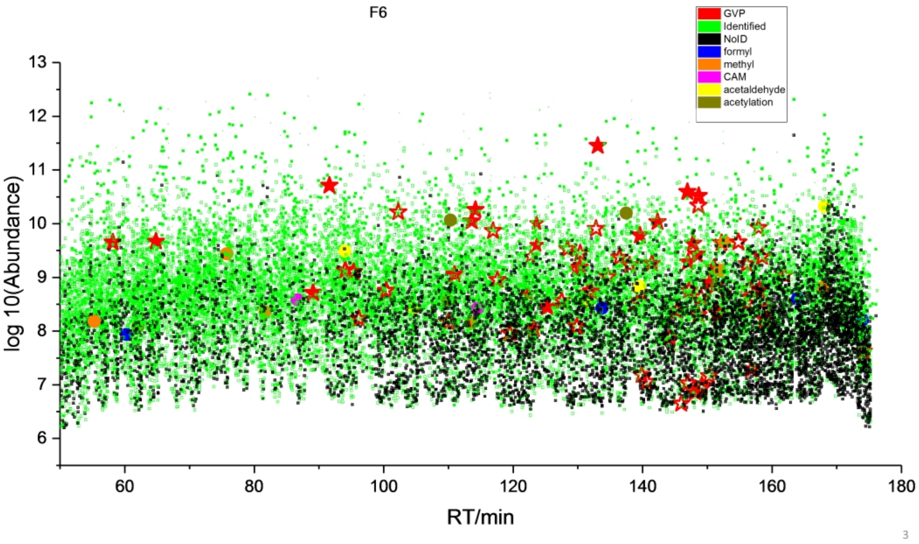
338x190mm (300 x 300 DPI)



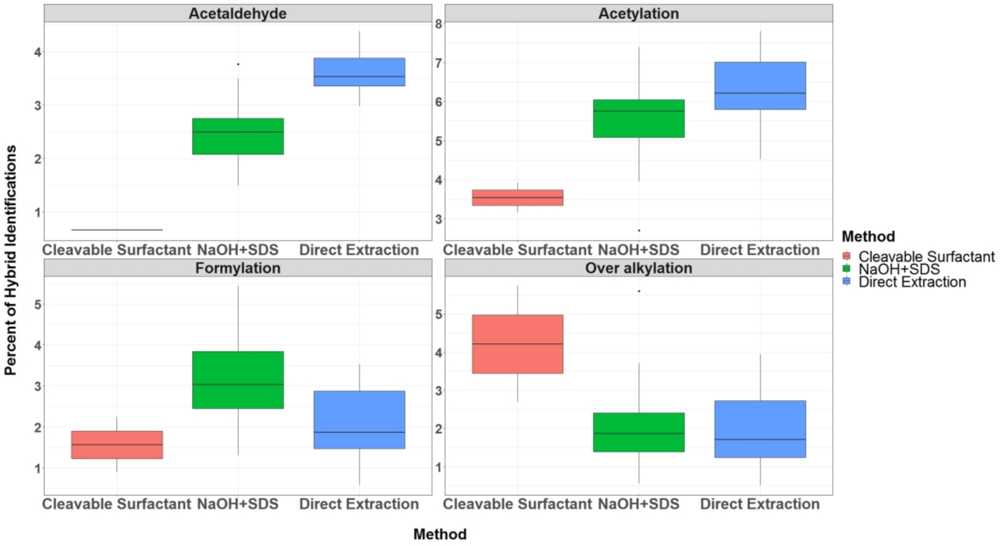
(A) The Direct Method

(B) Modified NaOH+SDS Method

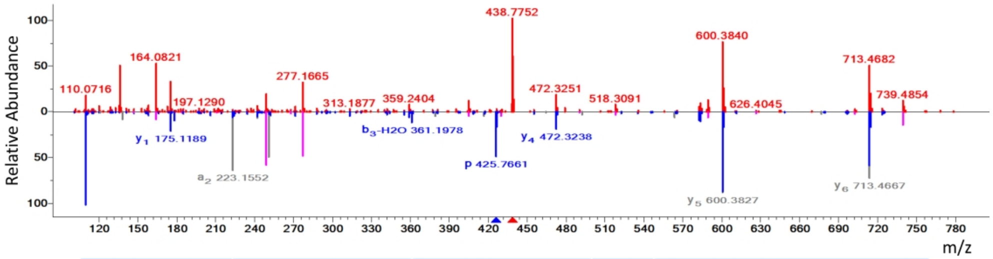
338x330mm (300 x 300 DPI)



338x190mm (300 x 300 DPI)

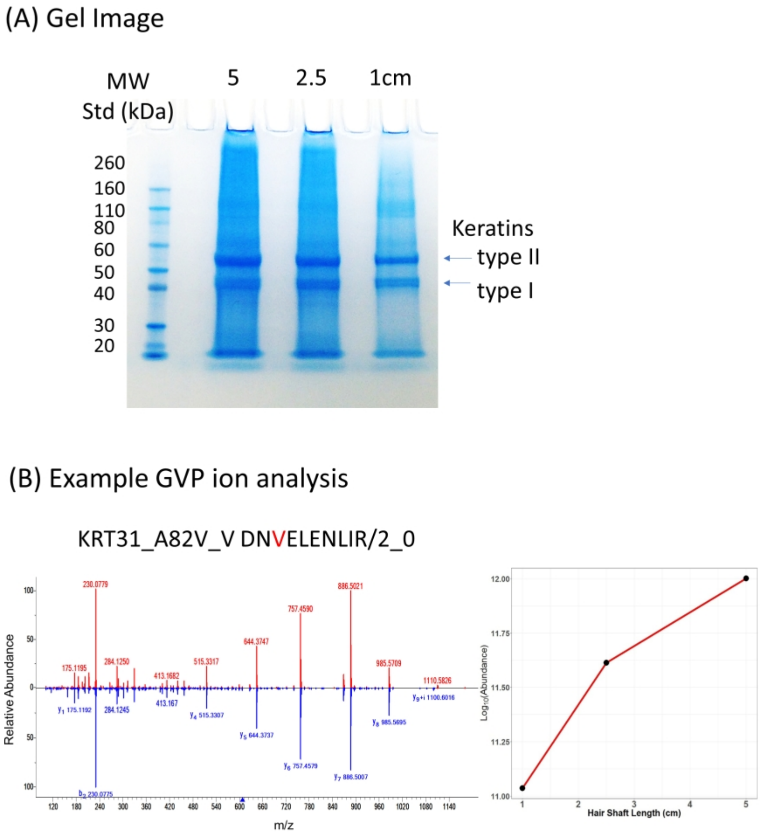


338x190mm (300 x 300 DPI)



Peptide Sequence	Protein	Localization	Score	Possible Amino Acid Substitution
HLQLAIR	Histone H2A type 1-D	H(+26.0186)LQLAIR	705	His→Tyr

338x190mm (300 x 300 DPI)


(A) Gel Image

(B) Example GVP ion analysis

338x330mm (300 x 300 DPI)

SUPPORTING INFORMATION:

Title: Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Ke

Authors: Zheng Zhang, Meghan C. Burke, William E. Wallace, Yuxue Liang, Sergey L. Sheetlin, Yuri A. Mir

Affiliations: Mass Spectrometry Data Center, National Institute of Standards and Technology, 100 Burea

Table of Contents:

eratin

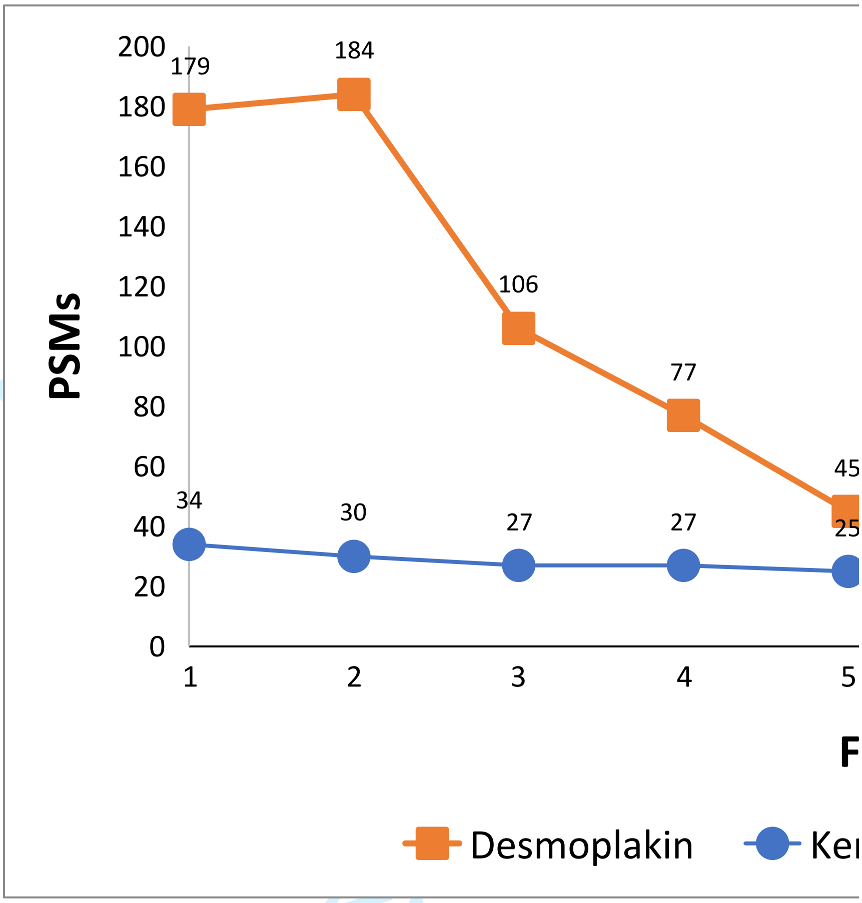
rokhin, Dmitrii V. Tchekhovskoi, Stephen E. Stein

au Drive, Gaithersburg, Maryland 20899 USA

ct Method

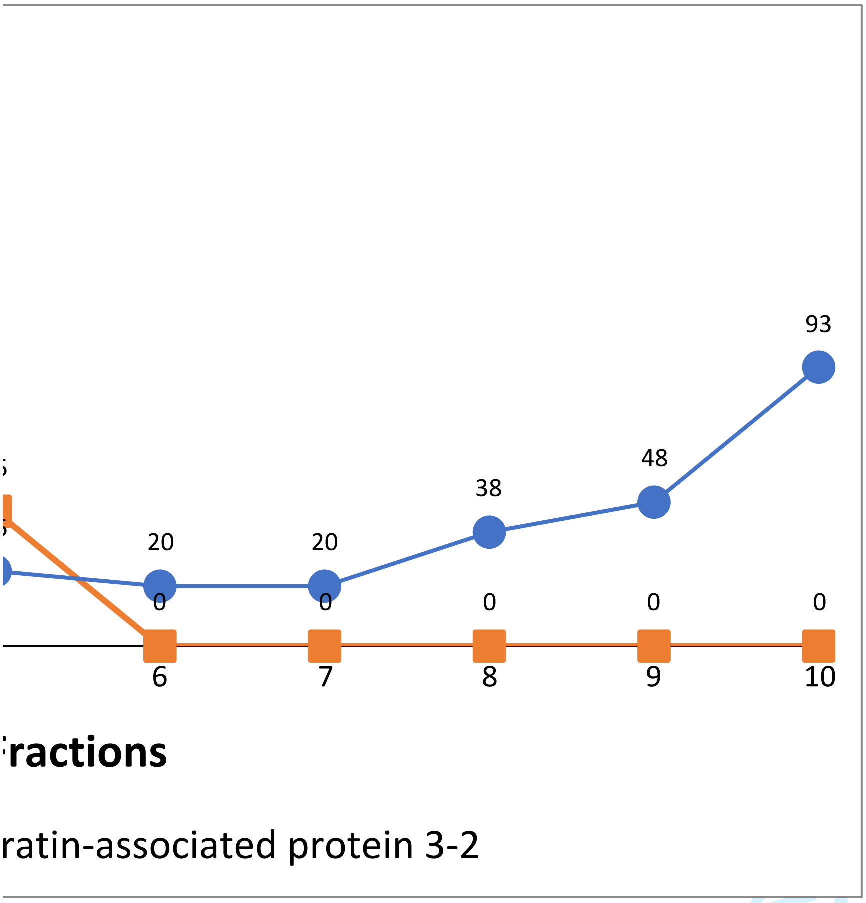
Table S1. Example of a Big Protein and a Small Protein Amount Change in Ten Gel Fractions by the Direct M

	Accession number	AA	Coverage_%	Fr_1	Fr_2
Desmoplakin	P15924	2871	35%	179	184
Keratin-associated protein 3-2	Q9BYR7	98	59%	34	30



ethod

PSMs							
Fr_3	**Fr_4**	**Fr_5**	**Fr_6**	**Fr_7**	**Fr_8**	**Fr_9**	**Fr_10**
106	77	45	0	0	0	0	0
27	27	25	20	20	38	48	93



SUPPORTING INFORMATION:

Title: Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Ke

Authors: Zheng Zhang, Meghan C. Burke, William E. Wallace, Yuxue Liang, Sergey L. Sheetlin, Yuri A. Mir

Affiliations: Mass Spectrometry Data Center, National Institute of Standards and Technology, 100 Burea

Table of Contents:

eratin

rokhin, Dmitrii V. Tchekhovskoi, Stephen E. Stein

au Drive, Gaithersburg, Maryland 20899 USA

Table S2. Percentages of Hybrid IDs in All Ten Gel Fractions by the Direct Method from a 5 cm-long Hair Sha

Samples	ID	ID_%	Hybrid_ID	Hybrid_ID_	No_ID	No_ID_%
F1	3871	0.11	25596	0.727	5729	0.163
F2	3522	0.101	26358	0.755	5044	0.144
F3	3027	0.119	20144	0.791	2288	0.09
F4	4094	0.106	28466	0.736	6131	0.158
F5	3649	0.097	28783	0.762	5352	0.142
F6	2831	0.112	20527	0.811	1954	0.077
F7	2869	0.114	19432	0.77	2929	0.116
F8	4301	0.106	29048	0.715	7299	0.180
F9	3804	0.134	20624	0.727	3942	0.139
F10	2694	0.097	19456	0.697	5748	0.206
Average(%)		**0.11**		**0.75**		**0.14**

aft

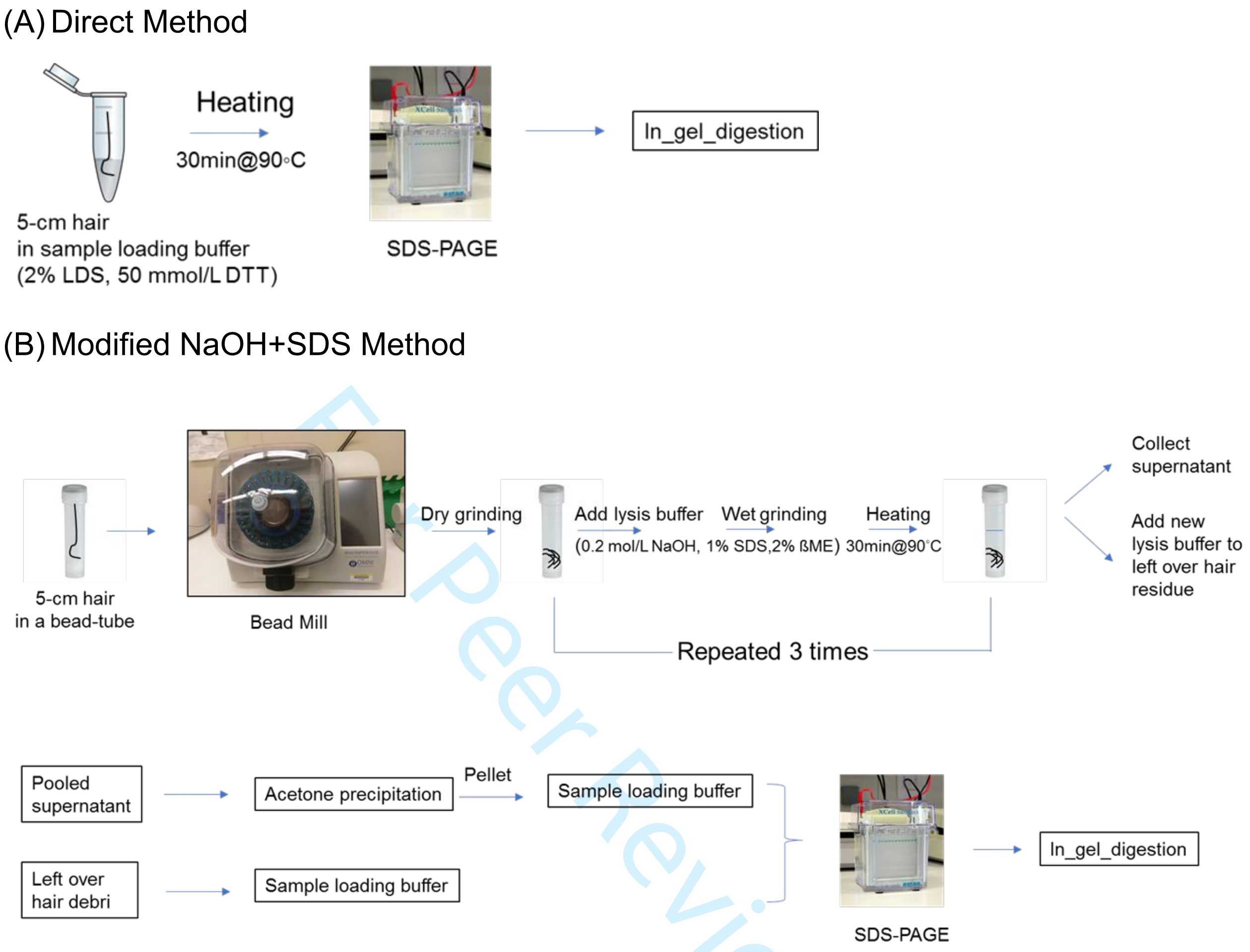
Supporting Information:

Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Keratin

Authors: Zheng Zhang, Meghan C. Burke, William E. Wallace, Yuxue Liang, Sergey L. Sheetlin, Yuri A. Mirokhin, Dmitrii V. Tchekhovskoi, Stephen E. Stein

Mass Spectrometry Data Center, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, United States

Supplemental Document S1: Outline of Protein Extraction Work Flows for the Direct Method and Modified NaOH+SDS Method.



Supplementary Document S1. Work flows of the Direct method and modified NaOH+SDS method are illustrated.

Supporting Information:

Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Keratin

Authors: Zheng Zhang, Meghan C. Burke, William E. Wallace, Yuxue Liang, Sergey L. Sheetlin, Yuri A. Mirokhin, Dmitrii V. Tchekhovskoi, Stephen E. Stein

Mass Spectrometry Data Center, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, United States

Supplementary Document S2:

GN=KRT31: Keratin, type I cuticular Ha1 OS=Homo sapiens

From Library (100%) and Sequest (97.6%):

M**PYNFCLPSL SCRTSCSSRP CVPPSCHSCT LPGACNIPAN VSNCNWFCEG SFNGSEKETM QFLNDR**LASY

LEKVR**QLERD NAELENLIRE RSQQQEPLLC PSYQSYFKTI EELQQKILCT KSENARLVVQ IDNAKLAADD**

FRTKYQTELS LRQLVESDIN GLRRILDELT LCKSDLEAQV ESLKEELLCL KSNHEQEVNT LRCQLGDRLN

VEVDAAPTVD LNRVLNETRS QYEALVETNR REVEQWFTTQ TEELNKQVVS SSEQLQSYQA EIIELRRTVN

ALEIELQAQH NLRDSLENTL TESEARYSSQ LSQVQSLITN VESQLAEIRS DLERQNQEYQ VLLDVRARLE

CEINTYRSLL ESEDCNLPSN PCATTNACSK PIGPCLSNPC TSCVPPAPCT PCAPRPRCGP CNSFVR

GN=KRT32: Keratin, type I cuticular Ha2 OS=Homo sapiens

From Library (54.2%) and Sequest (49.6%):

MTSSCCVTNN LQASLKSCPR PASVCSSGVN CRPELCLGYV CQPMACLPSV CLPTTFRPAS CLSKTYLSSS

CQAASGISGS MGPGSWYSEG AFNGNEK**ETM QFLNDR**LASY LTRVR**QLEQE NAELESRIQE ASHSQVLTMT**

PDYQSHFRTI EELQQKILCT KAENAR**MVVN IDNAKLAADD FRAKYEAELA MRQLVEADIN GLRRILDDLT**

LCKADLEAQV ESLKEELMCL KKNHEEEVGS LRCQLGDR**LN IEVDAAPPVD LTR**VLEEMRC QYEAMVEANR

RDVEEWFNMQ MEELNQQVAT SSEQLQNYQS DIIDLRR**TVN TLEIELQAQH SLRDSLENTL TESEARYSSQ**

LAQMQCMITN VEAQLAEIRA DLE**RQNQEYQ VLLDVRARLE GEINTYR**SLL ENEDCKLPCN PCSTPSCTTC

VPSPCVPRTV CVPRTVGMPC SPCPQGRY

GN=KRT33A: Keratin, type I cuticular Ha3-I OS=Homo sapiens

From Library (97.0%) and Sequest (93.3%):

M**SYSCGLPSL SCRTSCSSRP CVPPSCHGCT LPGACNIPAN VSNCNWFCEG SFNGSEKETM QFLNDR**LASY

LEKVR**QLERD NAELENLIRE RSQQQEPLVC ASYQSYFKTI EELQQKILCS KSENARLVVQ IDNAKLASDD**

FRTKYETELS LRQLVESDIN GLRRILDELT LCRSDLEAQV ESLKEELLCL KQNHEQEVNT LRCQLGDR**LN**

VEVDAAPTVD LNQVLNETRS QYEALVETNR REVEQWFATQ TEELNKQVVS SSEQLQSYQA EIIELRRTVN

ALEIELQAQH NLRDSLENTL TESEARYSSQ LSQVQRLITN VESQLAEIRS DLERQNQEYQ VLLDVRARLE

CEINTYRSLL ESEDCKLPSN PCATTNACDK STGPCISNPC GLRARCGPCN TFGY

GN=KRT33B: Keratin, type I cuticular Ha3-II OS=Homo sapiens

From Library (97.0%) and Sequest (93.6%):

M**PYNFCLPSL SCRTSCSSRP CVPPSCHGYT LPGACNIPAN VSNCNWFCEG SFNGSEKETM QFLNDR**LASY

LEKVR**QLERD NAELENLIRE RSQQQEPLLC PSYQSYFKTI EELQQKILCS KSENARLVVQ IDNAKLAADD**

FRTKYQTEQS LRQLVESDIN SLRRILDELT LCRSDLEAQM ESLKEELLSL KQNHEQEVNT LRCQLGD**RLN**

VEVDAAPAVD LNQVLNETRN QYEALVETNR REVEQWFATQ TEELNKQVVS SSEQLQSYQA EIIELRRTVN

ALEIELQAQH NLRYSLENTL TESEARYSSQ LSQVQSLITN VESQLAEIRS DLERQNQEYQ VLLDVRARLE

CEINTYRSLL ESEDCKLPSN PCATTNACEK PIGSCVTNPC GPRSRCGPCN TFGY

GN=KRT34: Keratin, type I cuticular Ha4 OS=Homo sapiens

From Library (86.0%) and Sequest (83.9%):

MLYAKPPPTI NGIKGLQRKE RLKPAHIHLQ QLTCFSITCS STM**SYSCCLP SLGCRTSCSS RPCVPPSCHG**

YTLPGACNIP ANVSNCNWFC EGSFNGSEKE TMQFLNDRLA SYLEKVR**QLE RDNAELEKLI QERSQQQEPL**

LCPSYQSYFK TIEELQQKIL CAKAENARLV VNIDNAKLAS DDFRSKYQTE QSLRLLVESD INSIRRILDE

LTLCKSDLES QVESLREELI CLKKNHEEEV NTLRSQLGDR LNVEVDTAPT VDLNQVLNET RSQYEALVEI

N**RREVEQWFA TQTEELNKQV VSSSEQLQSC QAEIIELRRT VNALEIELQA QHNLRDSLEN TLTESEAHYS**

SQLSQVQSLI TNVESQLAEI RCDLE**RQNQE YQVLLDVRAR LECEINTYRS LLESEDCKLP CNPCATTNAS**

GNSCGPCGTS QKGCCN

GN=KRT35: Keratin, type I cuticular Ha5 OS=Homo sapiens

From Library (91.0%) and Sequest (86.4%):

MASKCLK**AGF SSGSLKSPGG ASGGSTRVSA MYSSSSCKLP SLSPVARSFS ACSVGLGR**SS YR**ATSCLPAL**

CLPAGGFATS YSGGGGWFGE GILTGNEKET MQSLNDRLAG YLEKVR**QLEQ ENASLESRIR EWCEQQVPYM**

CPDYQSYFRT IEELQKKTLC SKAENAR**LVV EIDNAKLAAD DFRTKYETEV SLRQLVESDI NGLRRILDDL**

TLCKSDLEAQ VESLKEELLC LKKNHEEEVN SLRCQLGDR**L NVEVDAAPPV DLNR**VLEEMR **CQYETLVENN**

RR**DAEDWLDT QSEELNQQVV SSSEQLQSCQ AEIIELRRTV NALEIELQAQ HSMRDALEST LAETEARYSS**

QLAQMQCMIT NVEAQLAEIR ADLE**RQNQEY QVLLDVRARL ECEINTYRGL LESEDSKLPC NPCAPDYSPS**

K**SCLPCLPAA SCGPSAAR**TN CSPR**PICVPC PGGR**F

GN=KRT36: Keratin, type I cuticular Ha6 OS=Homo sapiens

From Library (60.8%) and Sequest (49.3%):

MATQTCTPTF STGSIKGLCG TAGGISRVSS IRSVGSCR**VP SLAGAAGYIS SAR**SGLSGLG SCLPGSYLSS

ECHTSGFVGS GGWF**CEGSFN GSEKETMQFL NDR**LANYLEK VRQLERENAE LESRIQEWYE FQIPYICPDY

QSYFKTIEDF QQKILLTKSE NARLVL**QIDN AKLAADDFRT KYETELSLRQ LVEADINGLR RILDELTLCK**

ADLEAQVESL KEELMCLKK**N HEEEVSVLR**C QLGDR**LNVEV DAAPPVDLNK** ILEDMRCQYE ALVENNRRDV

EAWFNTQTEE LNQQVVSSSE QLQCCQTEII ELR**RTVNALE IELQAQHSMR NSLESTLAET EARYSSQLAQ**

MQCLISNVEA QLSEIRCDLE R**QNQEYQVLL DVK**AR**LEGEI ATYRHLLEGE DCK**LPPQPCA TACKPVIR**VP**

SVPPVPCVPS VPCTPAPQVG TQIRTITEEI RDGKVISSRE HVQSRPL

GN=KRT37: Keratin, type I cuticular Ha7 OS=Homo sapiens

From Library (43.0%) and Sequest (34.7%):

MTSFYSTSSC PLGCTMAPGA RNVFVSPIDV GCQPVAEANA ASM**CLLANVA HANR**VRVGST PLGRPSLCLP

PTSHTACPLP GTCHIPGNIG ICGAYGKNTL NGHEKETMKF LNDRLANYLE KVR**QLEQENA ELETTLLER**S

KCHESTVCPD YQSY**FRTIEE LQQKILCSK**A ENAR**LIVQID NAKLAADDFR** IKLESERSLH QLVEADKCGT

QKLLDDATLA K**ADLEAQQES LKEEQLSLK**S NHEQEVKILR SQLGEKFR**IE LDIEPTIDLN R**VLGEMRAQY

EAMVETNHQD VEQWFQAQSE GISLQAMSCS EELQCCQSEI LELR**CTVNAL EVER**QAQHTL K**DCLQNSLCE**

AEDRYGTELA QMQSLISNLE EQLSEIRADL ER**QNQEYQVL LDVK**AR**LENE IATYR**NLLES EDCKLPCNPC

STPASCTSCP SCGPVTGGSP SGHGASMGR

GN=KRT38: Keratin, type I cuticular Ha8 OS=Homo sapiens

From Library (61.2%) and Sequest (51.3%):

MTSSYSSSSC PLGCTMAPGA R**NVSVSPIDI GCQPGAEANI APMCLLANVA HANR**VRVGST PLGRPSLCLP

PTCHTACPLP GTCHIPGNIG ICGAYGENTL NGHEK**ETMQF LNDR**LANYLE KVR**QLEQENA ELEATLLER**S

K**CHESTVCPD YQSYFHTIEE LQQKILCSK**A ENAR**LIVQID NAKLAADDFR** IKLESERSLR QLVEADKCGT

QKLLDDATLA K**ADLEAQQES LKEEQLSLK**S NHEQEVKILR SQLGEKLR**IE LDIEPTIDLN R**VLGEMR**AQY**

EAMLETNRQD VEQWFQAQSE GISLQDMSCS EELQCCQSEI LELR**CTVNAL EVER**QAQHTL K**DCLQNSLCE**

AEDRFGTELA QMQSLISNVE EQLSEIRADL ER**QNQEYQVL LDVKTRLENE IATYR**NLLES EDCKLPCNPC

STSPSCVTAP CAPR**PSCGPC TTCGPTCGAS TTGSR**F

GN=KRT81: Keratin, type II cuticular Hb1 OS=Homo sapiens

From Library (96.2%) and Sequest (91.9%):

M**TCGSGFGGR AFSCISACGP RPGRCCITAA PYR**GISCY**RG LTGGFGSHSV CGGFR**AGSCG RSFGYR**SGGV**

CGPSPPCITT VSVNESLLTP LNLEIDPNAQ CVKQEEKEQI KSLNSRFAAF IDKVRFLEQQ NKLLETKLQF

YQNRECCQSN LEPLFEGYIE TLRREAECVE ADSGRLASEL NHVQEVLEGY KKKYEEEVSL RATAENEFVA

LKKDVDCAYL RKSDLEANVE ALIQEIDFLR RLYEEEILIL QSHISDTSVV VKLDNSRDLN MDCIIAEIKA

QYDDIVTRSR AEAESWYRSK CEEMKATVIR HGETLR**RTKE EINELNR**MIQ **RLTAEVENAK CQNSKLEAAV**

AQSEQQGEAA LSDARCKLAE LEGALQKAKQ DMACLIREYQ EVMNSKLGLD IEIATYRRLL EGEEQRLCEG

IGAVNVCVSS SRGGVVCGDL CVSGSRPVTG SVCSAPCNGN VAVSTGLCAP C**GQLNTTCGG GSCGVGSCGI**

SSLGVGSCGS SCRKC

GN=KRT82: Keratin, type II cuticular Hb2 OS=Homo sapiens

From Library (63.4%) and Sequest (49.9%):

MSYHSFQPGS RCGSQSFSSY SAVMPRMVTH YAVSKGPCRP GGGRGLRALG CLGSRSLCNV GFGRPRVASR

CGGTLPGFGY RLGATCGPSA CITPVTINES LLVPLALEID PTVQRVKRDE KEQIKCLNNR FASFINKVRF

LEQKNKLLET KWNFMQQQR**C CQTNIEPIFE GYISALR**R**QL DCVSGDRVRL ESELCSLQAA LEGYKK**K**YEE**

ELSLRPCVEN EFVALKKDVD TAFLMK**ADLE TNAEALVQEI DFLKSLYEEE ICLLQSQISE TSVIVK**MDNS

R**ELDVDGIIA EIK**AQYDDIA SR**SKAEAEAW YQCR**YEELR**V TAGNHCDNLR** NRKNEILEMN KLIQRLQQET

ENVKAQRCK**L EGAIAEAEQQ GEAALNDAK**C K**LAGLEEALQ KAKQDMACLL KEYQEVMNSK LGLDIEIATY**

RRLLEGEEHR LCEGIGPVNI SVSSSKGAFL YEPCGVSTPV LSTGVLRSNG GCSIVGTGEL YVPCEPQGLL

SCGSGRKSSM TLGAGGSSPS HKH

GN=KRT83: Keratin, type II cuticular Hb3 OS=Homo sapiens

From Library (97.0%) and Sequest (87.2%):

MTCGFNSIGC GFRPGNFSCV SACGPRPS**RC CITAAPYR**GI SCY**RGLTGGF GSHSVCGGFR** AGSCGRSFGY

R**SGGVCGPSP PCITTVSVNE SLLTPLNLEI DPNAQCVKQE EKEQIKSLNS RFAAFIDKVR FLEQQNKLLE**

TKLQFYQNRE CCQSNLEPLF AGYIETLRRE AECVEADSGR LASELNHVQE VLEGYKKKYE EEVALRATAE

NEFVALKKDV DCAYLRKSDL EANVEALIQE IDFLRRLYEE EIRILQSHIS DTSVVVKLDN SRDLNMDCIV

AEIKAQYDDI ATRSRAEAES WYRSKCEEMK ATVIRHGETL R**RTKEEINEL NR**MIQ**RLTAE VENAKCQNSK**

LEAAVAQSEQ QGEAALSDAR CKLAELEGAL QKAKQDMACL IREYQEVMNS KLGLDIEIAT YRRLLEGEEQ

RLCEGVEAVN VCVSSSRGGV VCGDLCVSGS RPVTGSVCSA PCNGNLVVST GL**CKPCGQLN TTCGGGSCGQ**

GRH

GN=KRT84: Keratin, type II cuticular Hb4 OS=Homo sapiens

From Library (12.7%) and Sequest (11.2%):

MSCRSYRVSS GHRVGNFSSC SAMTPQNLNR FRANSVSCWS GPGFRGLGSF GSRSVITFGS YSPRIAAVGS

RPIHCGVRFG AGCGMGFGDG RGVGLGPRAD SCVGLGFGAG SGIGYGFGGP GFGYRVGGVG VPAAPSITAV

TVNKSLLTPL NLEIDPNAQR VKKDEKEQIK TLNNK**FASFI DKVRFLEQQN KLLETK**WSFL QEQKCIRSNL

EPLFESYITN LRRQLEVLVS DQARLQAERN HLQDVLEGFK KKYEEEVVCR **ANAENEFVAL KK**DVDAAFMN

KSDLEANVDT LTQEIDFLKT LYMEEIQLLQ SHISETSVIV KMDNSRDLNL DGIIAEVKAQ YEEVARRSRA

DAEAWYQTKY EEMQVTAGQH CDNLRNIRNE INELTRLIQR LKAEIEHAKA QR**AKLEAAVA EAEQQGEATL**

SDAKCKLADL ECALQQAKQD MARQLCEYQE LMNAK**LGLDI EIATYRR**LLE GEESRLCEGV GPVNISVSSS

RGGLVCGPEP LVAGSTLSRG GVTFSGSSSV CATSGVLASC GPSLGGARVA PATGDLLSTG TRSGSMLISE

ACVPSVPCPL PTQGGFSSCS GGRSSSVRFV STTTSCRTKY

GN=KRT85: Keratin, type II cuticular Hb5 OS=Homo sapiens

From Library (96.8%) and Sequest (89.4%):

MSCRSYR**ISS GCGVTRNFSS CSAVAPKTGN RCCISAAPYR** GVSCYRGLTG FGSR**SLCNLG SCGPR**IAVGG

FRAGSCGRSF GYR**SGGVCGP SPPCITTVSV NESLLTPLNL EIDPNAQCVK QEEKEQIKSL NSRFAAFIDK**

VRFLEQQNKL LETKWQFYQN QRCCESNLEP LFSGYIETLR REAECVEADS GRLASELNHV QEVLEGYKKK

YEEEVALRAT AENEFVVLKK DVDCAYLRKS DLEANVEALV EESSFLRRLY EEEIRVLQAH ISDTSVIVKM

DNSRDLNMDC IIAEIKAQYD DVASRSRAEA ESWYRSKCEE MKATVIRHGE TLR**RTKEEIN ELNR**MIQR**LT**

AEIENAKCQR AKLEAAVAEA EQQGEAALSD ARCKLAELEG ALQKAKQDMA CLLKEYQEVM NSKLGLDIEI

ATYRRLLEGE EHRLCEGVGS VNVCVSSSRG GVSCGGLSYS TTPGRQITSG PSAIGGSITV VAPDSCAPCQ

PRSSSFSCGS SRSVRFA

GN=KRT86: Keratin, type II cuticular Hb6 OS=Homo sapiens

From Library (99.2%) and Sequest (92.4%):

MTCGSYCGGR **AFSCISACGP RPGRCCITAA PYR**GISCY**RG LTGGFGSHSV CGGFR**AGSCG RSFGYR**SGGV**

CGPSPPCITT VSVNESLLTP LNLEIDPNAQ CVKQEEKEQI KSLNSRFAAF IDKVRFLEQQ NKLLETKLQF

YQNRECCQSN LEPLFEGYIE TLRREAECVE ADSGRLASEL NHVQEVLEGY KKKYEEEVSL RATAENEFVA

LKKDVDCAYL RKSDLEANVE ALIQEIDFLR RLYEEEIRVL QSHISDTSVV VKLDNSRDLN MDCIIAEIKA

QYDDIVTRSR AEAESWYRSK CEEMKATVIR HGETLR**RTKE EINELNR**MIQ **RLTAEVENAK CQNSKLEAAV**

AQSEQQGEAA LSDARCKLAE LEGALQKAKQ DMACLIREYQ EVMNSKLGLD IEIATYRRLL EGEEQRLCEG

VGSVNVCVSS SRGGVVCGDL CASTTAPVVS TRVSSVPSNS NVVVGTTNAC APSARVGVCG GSCKRC

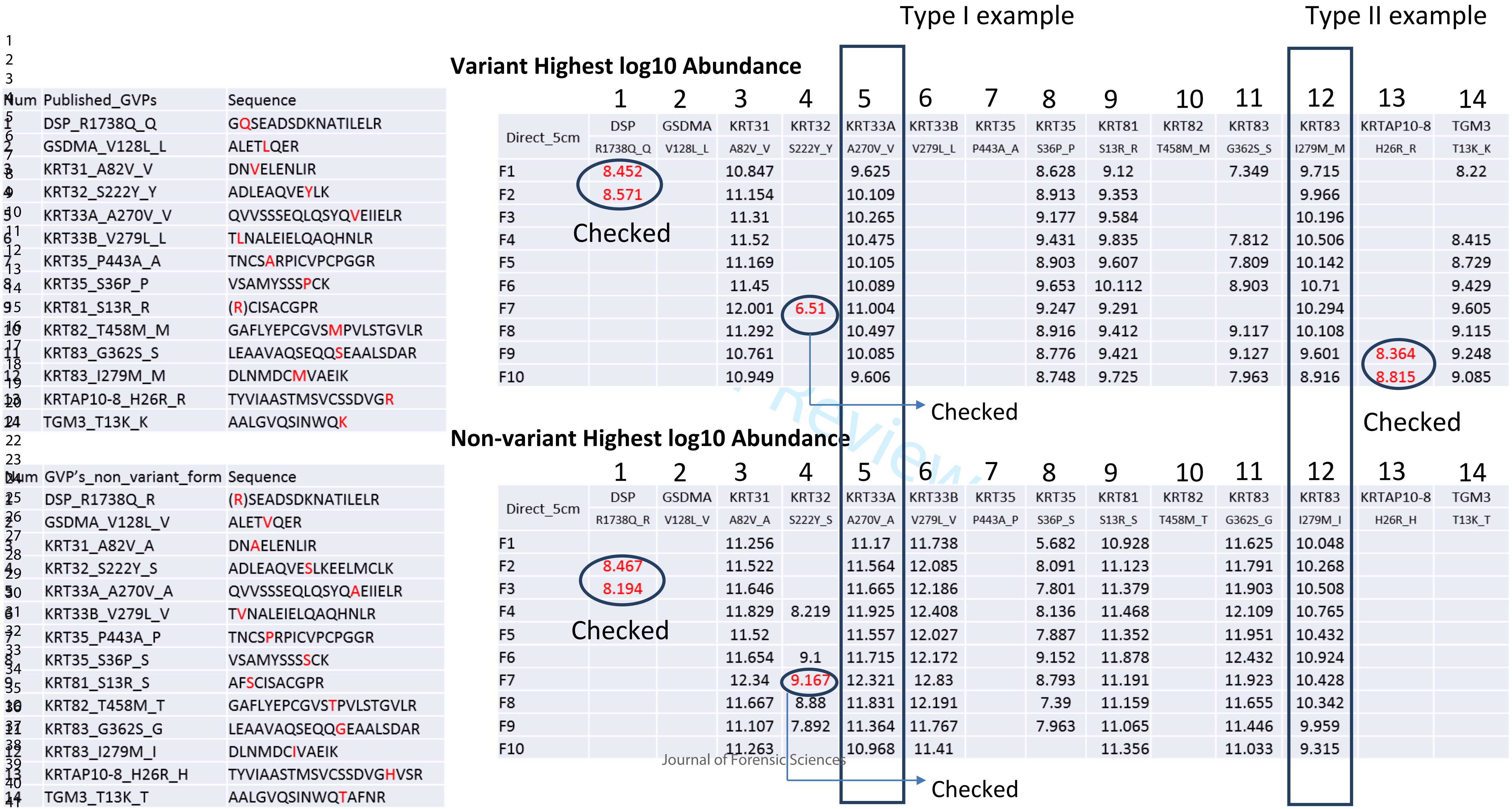
Supplementary Document S2: Amino acid sequence highlighted in green indicates peptide identified with high confidence (FDR at 1% level) by Sequest and library searching; in yellow indicates peptide identified with high confidence **by library searching only**. This sheet is sorted by type I cuticular keratins (from KRT31 to KRT38) and type II cuticular keratins (from KRT81 to KRT86). The coverage analyses were combined from all ten gel fractions.

Example GVP Panel Analysis

(D_LG_F1_TO_F10_R1)

Outlines:

- GVP Panel Analysis Overall
    - Left: Names and sequences of total 14 GVPs and non-variants
    - Right: Highest abundance in each fraction of “D_LG_F1_TO_F10_R1”
- High Abundance GVP pairs
    - Type I example
    - Type II example
- Check Low Abundance GVPs
    - Check spectral match
        - MS Search: search inquiry spectrum against library spectra
    - Check MS1 peak
        - Xcalibur Qual Brower
- Check Low Abundance Regular Forms (if applicable)
    - Check spectral match
        - MS Search: search inquiry spectrum against library spectra
    - Check MS1 peak
        - Xcalibur Qual Brower
- Summary Sheet



Type I example

Type II example

Num	Published_GVPs	Sequence
1	DSP_R1738Q_Q	GQSEADSDKNATILELR
2	GSDMA_V128L_L	ALETLQER
3	KRT31_A82V_V	DNVELENLIR
4	KRT32_S222Y_Y	ADLEAQVEYLK
5	KRT33A_A270V_V	QVVSSSEQLQSYQVEIIELR
6	KRT33B_V279L_L	TLNALEIELQAQHNLR
7	KRT35_P443A_A	TNCSARPICVPCPGGR
8	KRT35_S36P_P	VSAMYSSSPCK
9	KRT81_S13R_R	(R)CISACGPR
10	KRT82_T458M_M	GAFLYEPCGVSMPVLSTGVLR
11	KRT83_G362S_S	LEAAVAQSEQQSEAALSDAR
12	KRT83_I279M_M	DLNMDCMVAEIK
13	KRTAP10-8_H26R_R	TYVIAASTMSVCSSDVGR
14	TGM3_T13K_K	AALGVQSINWQK

Variant Highest log10 Abundance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Direct_5cm	DSP	GSDMA	KRT31	KRT32	KRT33A	KRT33B	KRT35	KRT35	KRT81	KRT82	KRT83	KRT83	KRTAP10-8	TGM3
	R1738Q_Q	V128L_L	A82V_V	S222Y_Y	A270V_V	V279L_L	P443A_A	S36P_P	S13R_R	T458M_M	G362S_S	I279M_M	H26R_R	T13K_K
F1	8.452		10.847		9.625			8.628	9.12		7.349	9.715		8.22
F2	8.571		11.154		10.109			8.913	9.353			9.966		
F3			11.31		10.265			9.177	9.584			10.196		
F4			11.52		10.475			9.431	9.835		7.812	10.506		8.415
F5			11.169		10.105			8.903	9.607		7.809	10.142		8.729
F6			11.45		10.089			9.653	10.112		8.903	10.71		9.429
F7			12.001	6.51	11.004			9.247	9.291			10.294		9.605
F8			11.292		10.497			8.916	9.412		9.117	10.108		9.115
F9			10.761		10.085			8.776	9.421		9.127	9.601	8.364	9.248
F10			10.949		9.606			8.748	9.725		7.963	8.916	8.815	9.085

Checked

Checked

Checked

Num	GVP's_non_variant_form	Sequence
1	DSP_R1738Q_R	(R)SEADSDKNATILELR
2	GSDMA_V128L_V	ALETVQER
3	KRT31_A82V_A	DNAELENLIR
4	KRT32_S222Y_S	ADLEAQVESLKEELMCLK
5	KRT33A_A270V_A	QVVSSSEQLQSYQAEIIELR
6	KRT33B_V279L_V	TVNALEIELQAQHNLR
7	KRT35_P443A_P	TNCSPRPICVPCPGGR
8	KRT35_S36P_S	VSAMYSSSSCK
9	KRT81_S13R_S	AFSCISACGPR
10	KRT82_T458M_T	GAFLYEPCGVSTPVLSTGVLR
11	KRT83_G362S_G	LEAAVAQSEQQGEAALSDAR
12	KRT83_I279M_I	DLNMDCIVAEIK
13	KRTAP10-8_H26R_H	TYVIAASTMSVCSSDVGHVSR
14	TGM3_T13K_T	AALGVQSINWQTAFNR

Non-variant Highest log10 Abundance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Direct_5cm	DSP	GSDMA	KRT31	KRT32	KRT33A	KRT33B	KRT35	KRT35	KRT81	KRT82	KRT83	KRT83	KRTAP10-8	TGM3
	R1738Q_R	V128L_V	A82V_A	S222Y_S	A270V_A	V279L_V	P443A_P	S36P_S	S13R_S	T458M_T	G362S_G	I279M_I	H26R_H	T13K_T
F1			11.256		11.17	11.738		5.682	10.928		11.625	10.048		
F2	8.467		11.522		11.564	12.085		8.091	11.123		11.791	10.268		
F3	8.194		11.646		11.665	12.186		7.801	11.379		11.903	10.508		
F4			11.829	8.219	11.925	12.408		8.136	11.468		12.109	10.765		
F5			11.52		11.557	12.027		7.887	11.352		11.951	10.432		
F6			11.654	9.1	11.715	12.172		9.152	11.878		12.432	10.924		
F7			12.34	9.167	12.321	12.83		8.793	11.191		11.923	10.428		
F8			11.667	8.88	11.831	12.191		7.39	11.159		11.655	10.342		
F9			11.107	7.892	11.364	11.767		7.963	11.065		11.446	9.959		
F10			11.263		10.968	11.41			11.356		11.033	9.315		

Checked

Checked

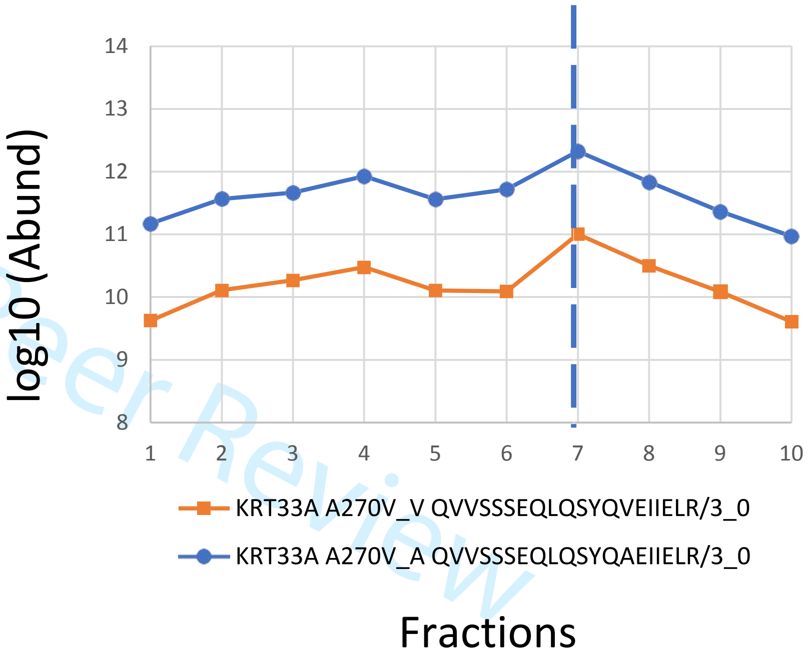
High Abundance GVP pairs

Type I example

Direct_5cm	KRT33A	RT	MF
	A270V_V QVVSSSEQLQSYQVEIIELR/3_0		
F1	9.625	162.2	935
F2	10.109	161.7	938
F3	10.265	161.1	942
F4	10.475	160.5	938
F5	10.105	160.4	869
F6	10.089	155.1	607
F7	11.004	157.7	803
F8	10.497	156.8	813
F9	10.085	156.7	906
F10	9.606	155.3	792

Direct_5cm	KRT33A	RT	MF
	A270V_A QVVSSSEQLQSYQAEIIELR/3_0		
F1	11.17	160.5	902
F2	11.564	160.0	897
F3	11.665	159.2	796
F4	11.925	158.7	892
F5	11.557	158.6	904
F6	11.715	152.5	900
F7	12.321	155.1	900
F8	11.831	154.2	903
F9	11.364	154.3	908
F10	10.968	152.8	910

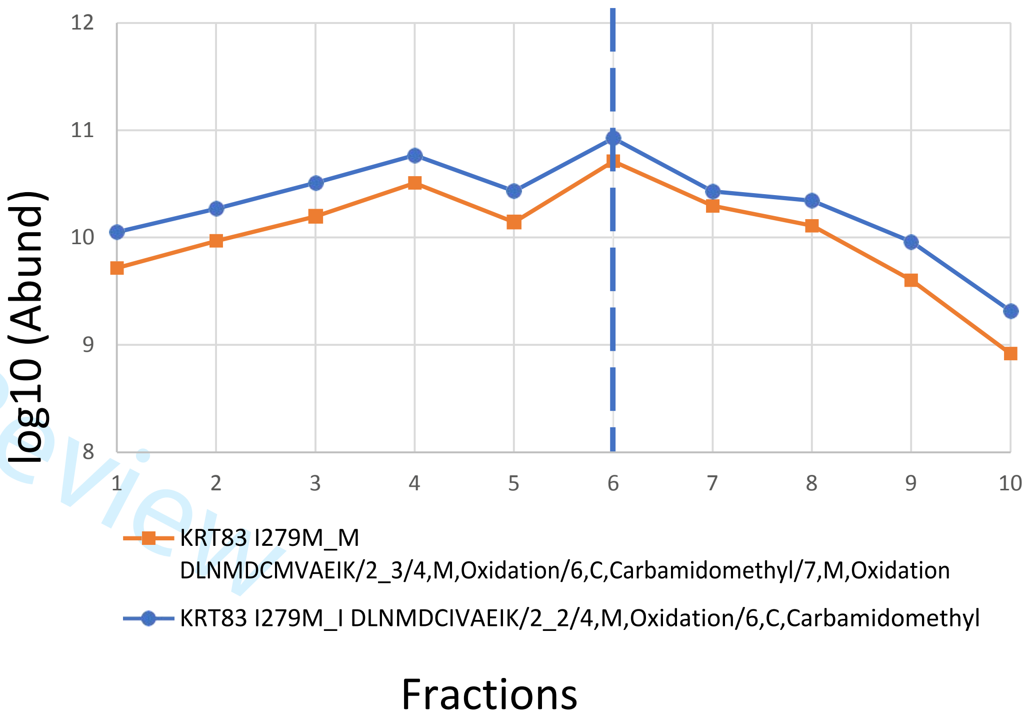
Type I example, also shown in Figure 2A



Type II example

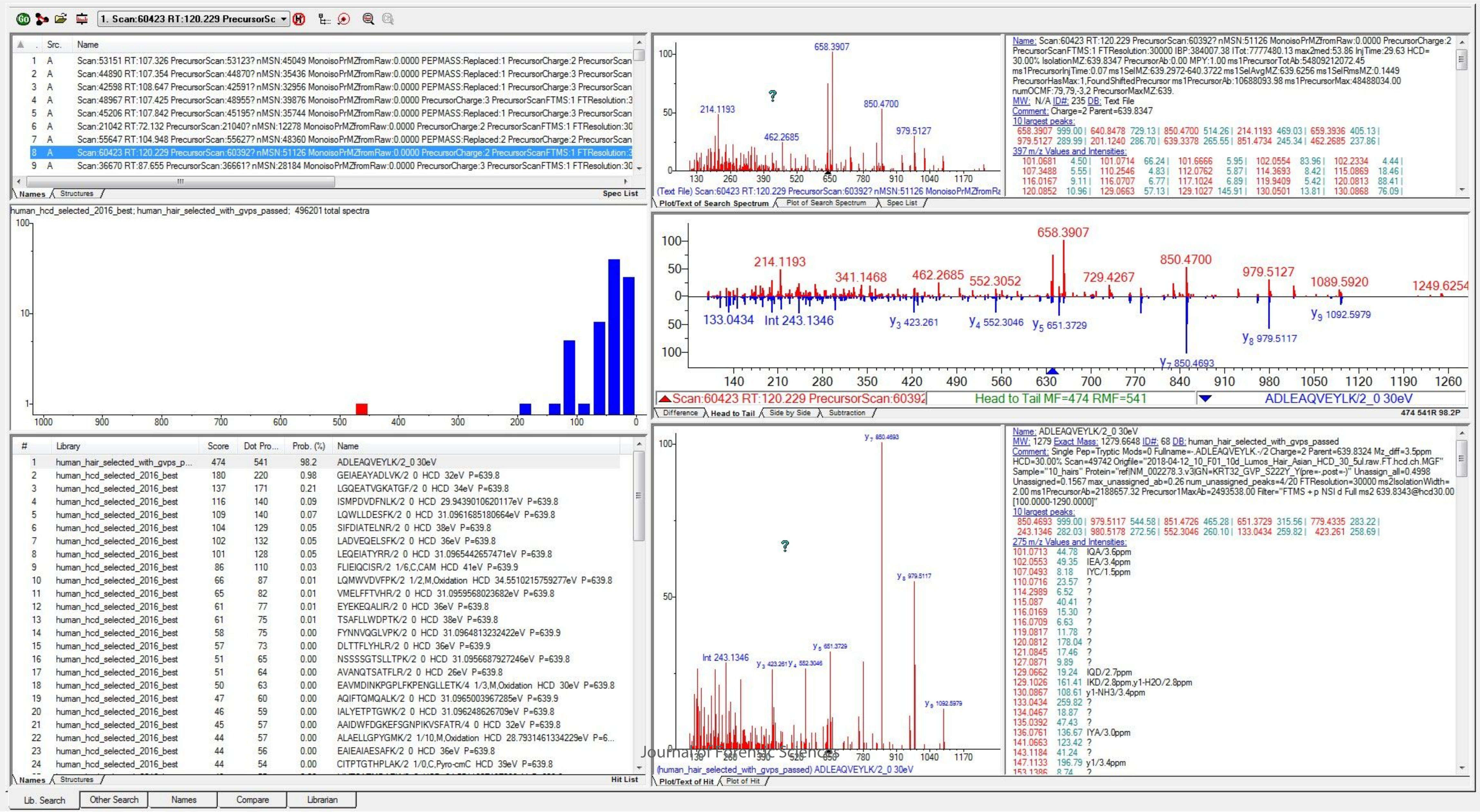
Direct_5cm	KRT83	RT	MF
	I279M_M DLNMDCMVAEIK/2_3/4,M,Oxidation/6,C,Carbamidomethyl/7,M,Oxidation		
F1	9.715	104.8	813
F2	9.966	103.5	802
F3	10.196	103.3	829
F4	10.506	102.1	847
F5	10.142	101.2	819
F6	10.71	91.6	792
F7	10.294	95.5	853
F8	10.108	93.5	807
F9	9.601	94.6	873
F10	8.916	93.1	349
Direct_5cm	KRT83	RT	MF
	I279M_I DLNMDCIVAEIK/2_2/4,M,Oxidation/6,C,Carbamidomethyl		
F1	10.048	137.2	887
F2	10.268	135.7	909
F3	10.508	134.1	865
F4	10.765	134.2	891
F5	10.432	134.2	818
F6	10.924	124.8	921
F7	10.428	128.6	895
F8	10.342	127.2	885
F9	9.959	128.4	931
F10	9.315	125.8	887

Type II example, also shown in Figure 2B



Check Low Abundance GVPs

KRT32_S222Y_Y: ADLEAQVEYLK in Fraction 7 'hold' to be confirmed



▲	Src.	Name
1	A	Scan:53151 RT:107.326 PrecursorScan:53123? nMSN:45049 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:2 PrecursorScan
2	A	Scan:44890 RT:107.354 PrecursorScan:44870? nMSN:35436 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
3	A	Scan:42598 RT:108.647 PrecursorScan:42591? nMSN:32956 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
4	A	Scan:48967 RT:107.425 PrecursorScan:48955? nMSN:39876 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:3
5	A	Scan:45206 RT:107.842 PrecursorScan:45195? nMSN:35744 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
6	A	Scan:21042 RT:72.132 PrecursorScan:21040? nMSN:12278 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30
7	A	Scan:55647 RT:104.948 PrecursorScan:55627? nMSN:48360 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:2 PrecursorCharge:2 PrecursorScan
8	A	Scan:60423 RT:120.229 PrecursorScan:60392? nMSN:51126 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:3
9	A	Scan:36670 RT:87.655 PrecursorScan:36661? nMSN:28184 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30

Name: Scan:60423 RT:120.229 PrecursorScan:60392? nMSN:51126 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30000 IBP:384007.38 ITot:7777480.13 max2med:53.86 InjTime:29.63 HCD=30.00% IsolationMZ:639.8347 PrecursorAb:0.00 MPY:1.00 ms1PrecursorTotAb:54809212072.45 ms1PrecursorInjTime:0.07 ms1SelMZ:639.2972-640.3722 ms1SelAvgMZ:639.6256 ms1SelRmsMZ:0.1449 PrecursorHasMax:1,FoundShiftedPrecursor ms1PrecursorAb:10688093.98 ms1PrecursorMax:48488034.00 numOCMF:79,79,-3,2 PrecursorMaxMZ:639.
MW: N/A ID#: 235 DB: Text File
Comment: Charge=2 Parent=639.8347
10 largest peaks:

```
 658.3907  999.00 | 640.8478  729.13 | 850.4700  514.26 | 214.1193  469.03 | 659.3936  405.13 |
 979.5127  289.99 | 201.1240  286.70 | 639.3378  265.55 | 851.4734  245.34 | 462.2685  237.86 |
```

397 m/z Values and Intensities:

```
  101.0681     4.50 |  101.0714   66.24 |  101.6666     5.95 |  102.0554   83.96 |  102.2334     4.44 |
  107.3488     5.55 |  110.2546    4.83 |  112.0762     5.87 |  114.3693    8.42 |  115.0869    18.46 |
  116.0167     9.11 |  116.0707    6.77 |  117.1024     6.89 |  119.9409    5.42 |  120.0813    88.41 |
  120.0852    10.96 |  129.0663   57.13 |  129.1027   145.91 |  130.0501   13.81 |  130.0868    76.09 |
```

human_hcd_selected_2016_best; human_hair_selected_with_gvps_passed; 496201 total spectra

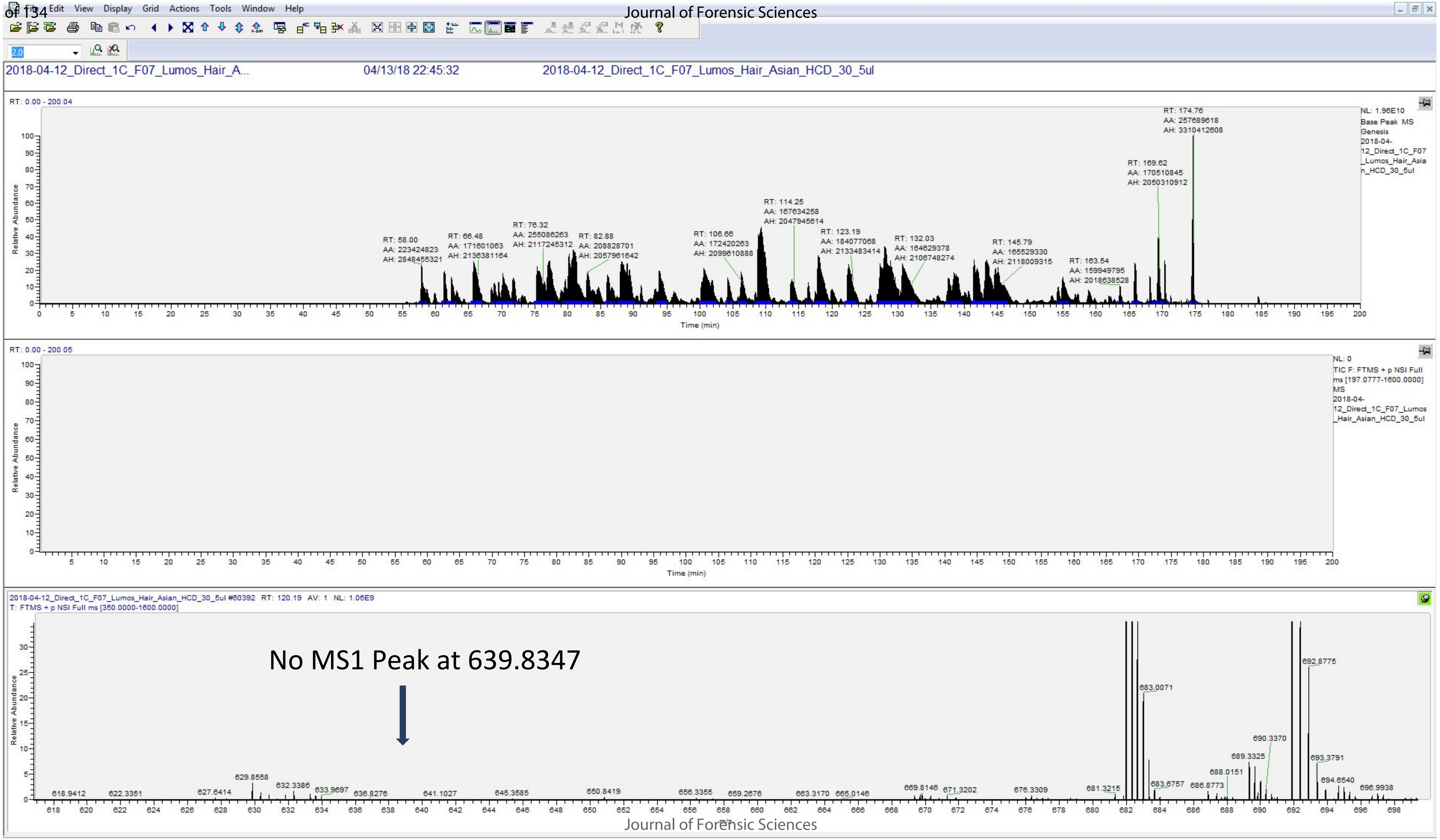
#	Library	Score	Dot Pro...	Prob. (%)	Name
1	human_hair_selected_with_gvps_p...	474	541	98.2	ADLEAQVEYLK/2_0 30eV
2	human_hcd_selected_2016_best	180	220	0.98	GEIAEAYADLVK/2 0 HCD 32eV P=639.8
3	human_hcd_selected_2016_best	137	171	0.21	LGQEATVGKATGF/2 0 HCD 34eV P=639.8
4	human_hcd_selected_2016_best	116	140	0.09	ISMPDVDFNLK/2 0 HCD 29.9439010620117eV P=639.8
5	human_hcd_selected_2016_best	109	140	0.07	LQWLLDESFK/2 0 HCD 31.0961685180664eV P=639.8
6	human_hcd_selected_2016_best	104	129	0.05	SIFDIATELNR/2 0 HCD 38eV P=639.8
7	human_hcd_selected_2016_best	102	132	0.05	LADVEQELSFK/2 0 HCD 36eV P=639.8
8	human_hcd_selected_2016_best	101	128	0.05	LEQEIATYRR/2 0 HCD 31.0965442657471eV P=639.8
9	human_hcd_selected_2016_best	86	110	0.03	FLIEIQCISR/2 1/6,C,CAM HCD 41eV P=639.9
10	human_hcd_selected_2016_best	66	87	0.01	LQMWVDVFPK/2 1/2,M,Oxidation HCD 34.5510215759277eV P=639.8
11	human_hcd_selected_2016_best	65	82	0.01	VMELFFTVHR/2 0 HCD 31.0959568023682eV P=639.8
12	human_hcd_selected_2016_best	61	77	0.01	EYEKEQALIR/2 0 HCD 36eV P=639.8
13	human_hcd_selected_2016_best	61	75	0.01	TSAFLLWDPTK/2 0 HCD 38eV P=639.8
14	human_hcd_selected_2016_best	58	75	0.00	FYNNVQGLVPK/2 0 HCD 31.0964813232422eV P=639.9
15	human_hcd_selected_2016_best	57	73	0.00	DLTTFLYHLR/2 0 HCD 36eV P=639.9
16	human_hcd_selected_2016_best	51	65	0.00	NSSSSGTSLLTPK/2 0 HCD 31.0956687927246eV P=639.8
17	human_hcd_selected_2016_best	51	64	0.00	AVANQTSATFLR/2 0 HCD 26eV P=639.8
18	human_hcd_selected_2016_best	50	63	0.00	EAVMDINKPGPLFKPENGLLETK/4 1/3,M,Oxidation HCD 30eV P=639.8
19	human_hcd_selected_2016_best	47	60	0.00	AQIFTQMQALK/2 0 HCD 31.0965003967285eV P=639.9
20	human_hcd_selected_2016_best	46	59	0.00	IALYETPTGWK/2 0 HCD 31.096248626709eV P=639.8
21	human_hcd_selected_2016_best	45	57	0.00	AAIDWFDGKEFSGNPIKVSFATR/4 0 HCD 32eV P=639.8
22	human_hcd_selected_2016_best	44	57	0.00	ALAELLGPYGMK/2 1/10,M,Oxidation HCD 28.7931461334229eV P=6...
23	human_hcd_selected_2016_best	44	56	0.00	EAIEAIAESAFK/2 0 HCD 36eV P=639.8
24	human_hcd_selected_2016_best	44	54	0.00	CITPTGTHPLAK/2 1/0,C,Pyro-cmC HCD 39eV P=639.8

Name: ADLEAQVEYLK/2_0 30eV
MW: 1279 Exact Mass: 1279.6648 ID#: 68 DB: human_hair_selected_with_gvps_passed
Comment: Single Pep=Tryptic Mods=0 Fullname=-.ADLEAQVEYLK.-/2 Charge=2 Parent=639.8324 Mz_diff=3.5ppm HCD=30.00% Scan=49742 Origfile="2018-04-12_10_F01_10d_Lumos_Hair_Asian_HCD_30_5ul.raw.FT.hcd.ch.MGF" Sample="10_hairs" Protein="ref|NM_002278.3.v3|GN=KRT32_GVP_S222Y_Y(pre=-,post=-)" Unassign_all=0.4998 Unassigned=0.1567 max_unassigned_ab=0.26 num_unassigned_peaks=4/20 FTResolution=30000 ms2IsolationWidth=2.00 ms1PrecursorAb=2188657.32 Precursor1MaxAb=2493538.00 Filter="FTMS + p NSI d Full ms2 639.8343@hcd30.00 [100.0000-1290.0000]"
10 largest peaks:

```
 850.4693  999.00 | 979.5117  544.58 | 851.4726  465.28 | 651.3729  315.56 | 779.4335  283.22 |
 243.1346  282.03 | 980.5178  272.56 | 552.3046  260.10 | 133.0434  259.82 |  423.261  258.69 |
```

275 m/z Values and Intensities:

```
101.0713   44.78   IQA/3.6ppm
102.0553   49.35   IEA/3.4ppm
107.0493   8.18    IYC/1.5ppm
110.0716   23.57   ?
114.2989   6.52    ?
115.087    40.41   ?
116.0169   15.30   ?
116.0709   6.63    ?
119.0817   11.78   ?
120.0812   178.04  ?
121.0845   17.46   ?
127.0871   9.89    ?
129.0662   19.24   IQD/2.7ppm
129.1026   161.41  IKD/2.8ppm,y1-H2O/2.8ppm
130.0867   108.61  y1-NH3/3.4ppm
133.0434   259.82  ?
134.0467   18.87   ?
135.0392   47.43   ?
136.0761   136.67  IYA/3.0ppm
141.0663   123.42  ?
143.1184   41.24   ?
147.1133   196.79  y1/3.4ppm
153.1386   8.74    ?
```



2018-04-12_Direct_1C_F07_Lumos_Hair_A... 04/13/18 22:45:32 2018-04-12_Direct_1C_F07_Lumos_Hair_Asian_HCD_30_5ul

RT: 0.00 - 200.04

NL: 1.96E10
Base Peak MS
Genesis
2018-04-12_Direct_1C_F07_Lumos_Hair_Asian_HCD_30_5ul

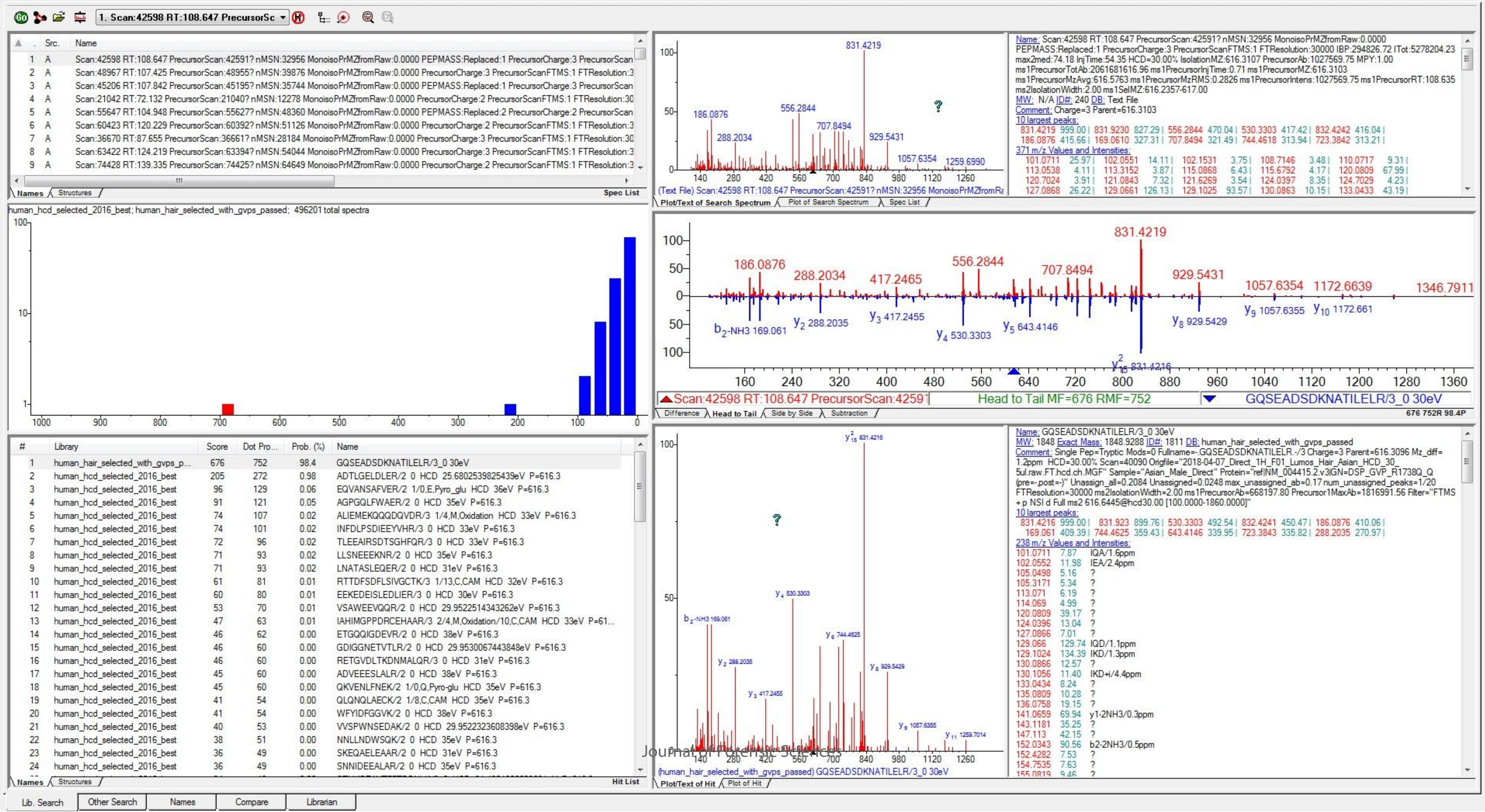
RT: 0.00 - 200.05

NL: 0
TIC F: FTMS + p NSI Full ms [197.0777-1600.0000]
MS
2018-04-12_Direct_1C_F07_Lumos_Hair_Asian_HCD_30_5ul

2018-04-12_Direct_1C_F07_Lumos_Hair_Asian_HCD_30_5ul #60392 RT: 120.19 AV: 1 NL: 1.06E9
T: FTMS + p NSI Full ms [350.0000-1600.0000]

No MS1 Peak at 639.8347

DSP_R1738Q_Q: GQSEADSDKNATILELR in Fraction 1 confirmed



▲	Src.	Name
1	A	Scan:42598 RT:108.647 PrecursorScan:42591? nMSN:32956 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
2	A	Scan:48967 RT:107.425 PrecursorScan:48955? nMSN:39876 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:3
3	A	Scan:45206 RT:107.842 PrecursorScan:45195? nMSN:35744 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
4	A	Scan:21042 RT:72.132 PrecursorScan:21040? nMSN:12278 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30
5	A	Scan:55647 RT:104.948 PrecursorScan:55627? nMSN:48360 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:2 PrecursorCharge:2 PrecursorScan
6	A	Scan:60423 RT:120.229 PrecursorScan:60392? nMSN:51126 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:3
7	A	Scan:36670 RT:87.655 PrecursorScan:36661? nMSN:28184 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30
8	A	Scan:63422 RT:124.219 PrecursorScan:63394? nMSN:54044 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:3
9	A	Scan:74428 RT:139.335 PrecursorScan:74425? nMSN:64649 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:3

Name: Scan:42598 RT:108.647 PrecursorScan:42591? nMSN:32956 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30000 IBP:294826.72 ITot:5278204.23 max2med:74.18 InjTime:54.35 HCD=30.00% IsolationMZ:616.3107 PrecursorAb:1027569.75 MPY:1.00 ms1PrecursorTotAb:2061681616.96 ms1PrecursorInjTime:0.71 ms1PrecursorMZ:616.3103 ms1PrecursorMzAvg:616.5763 ms1PrecursorMzRMS:0.2826 ms1PrecursorIntens:1027569.75 ms1PrecursorRT:108.635 ms2IsolationWidth:2.00 ms1SelMZ:616.2357-617.00
MW: N/A ID#: 240 DB: Text File
Comment: Charge=3 Parent=616.3103
10 largest peaks:

```
 831.4219  999.00 |  831.9230  827.29 |  556.2844  470.04 |  530.3303  417.42 |  832.4242  416.04 |
 186.0876  415.66 |  169.0610  327.31 |  707.8494  321.49 |  744.4618  313.94 |  723.3842  313.21 |
```

371 m/z Values and Intensities:

```
 101.0711   25.97 |  102.0551   14.11 |  102.1531    3.75 |  108.7146    3.48 |  110.0717    9.31 |
 113.0538    4.11 |  113.3152    3.87 |  115.0868    6.43 |  115.6792    4.17 |  120.0809   67.99 |
 120.7024    3.91 |  121.0843    7.32 |  121.6269    3.54 |  124.0397    8.35 |  124.7029    4.23 |
 127.0868   26.22 |  129.0661  126.13 |  129.1025   93.57 |  130.0863   10.15 |  133.0433   43.19 |
```

human_hcd_selected_2016_best; human_hair_selected_with_gvps_passed; 496201 total spectra

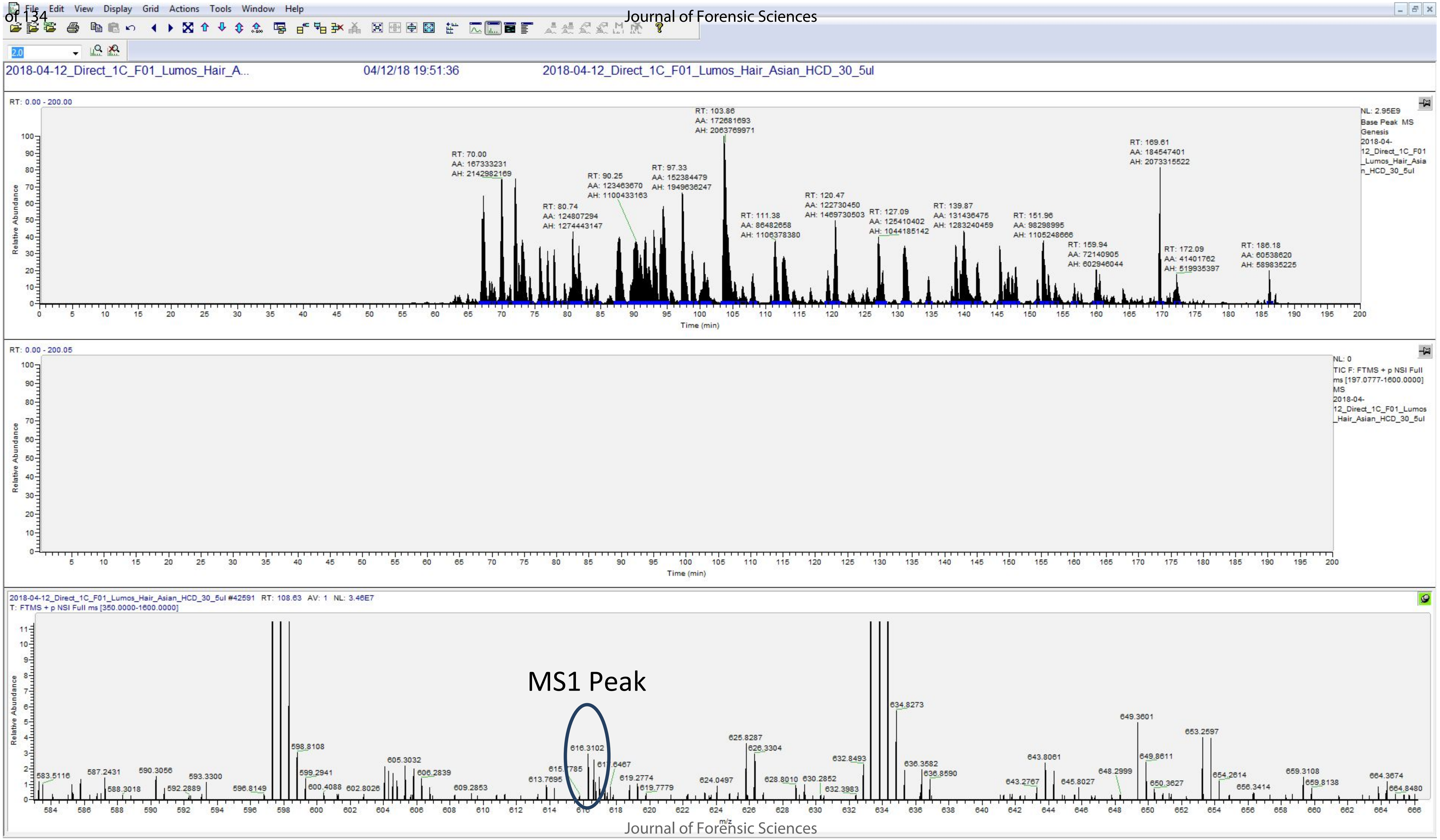
#	Library	Score	Dot Pro...	Prob. (%)	Name
1	human_hair_selected_with_gvps_p...	676	752	98.4	GQSEADSDKNATILELR/3_0 30eV
2	human_hcd_selected_2016_best	205	272	0.98	ADTLGELDLER/2 0 HCD 25.6802539825439eV P=616.3
3	human_hcd_selected_2016_best	96	129	0.06	EQVANSAFVER/2 1/0,E,Pyro_glu HCD 36eV P=616.3
4	human_hcd_selected_2016_best	91	121	0.05	AGPGQLFWAER/2 0 HCD 35eV P=616.3
5	human_hcd_selected_2016_best	74	107	0.02	ALIEMEKQQQDQVDR/3 1/4,M,Oxidation HCD 33eV P=616.3
6	human_hcd_selected_2016_best	74	101	0.02	INFDLPSDIEEYVHR/3 0 HCD 33eV P=616.3
7	human_hcd_selected_2016_best	72	96	0.02	TLEEAIRSDTSGHFQR/3 0 HCD 33eV P=616.3
8	human_hcd_selected_2016_best	71	93	0.02	LLSNEEEKNR/2 0 HCD 35eV P=616.3
9	human_hcd_selected_2016_best	71	93	0.02	LNATASLEQER/2 0 HCD 31eV P=616.3
10	human_hcd_selected_2016_best	61	81	0.01	RTTDFSDFLSIVGCTK/3 1/13,C,CAM HCD 32eV P=616.3
11	human_hcd_selected_2016_best	60	80	0.01	EEKEDEISLEDLIER/3 0 HCD 30eV P=616.3
12	human_hcd_selected_2016_best	53	70	0.01	VSAWEEVQQR/2 0 HCD 29.9522514343262eV P=616.3
13	human_hcd_selected_2016_best	47	63	0.01	IAHIMGPPDRCEHAAR/3 2/4,M,Oxidation/10,C,CAM HCD 33eV P=61...
14	human_hcd_selected_2016_best	46	62	0.00	ETGQQIGDEVR/2 0 HCD 38eV P=616.3
15	human_hcd_selected_2016_best	46	60	0.00	GDIGGNETVTLR/2 0 HCD 29.9530067443848eV P=616.3
16	human_hcd_selected_2016_best	46	60	0.00	RETGVDLTKDNMALQR/3 0 HCD 31eV P=616.3
17	human_hcd_selected_2016_best	45	60	0.00	ADVEEESLALR/2 0 HCD 38eV P=616.3
18	human_hcd_selected_2016_best	45	60	0.00	QKVENLFNEK/2 1/0,Q,Pyro-glu HCD 35eV P=616.3
19	human_hcd_selected_2016_best	41	54	0.00	QLQNQLAECK/2 1/8,C,CAM HCD 35eV P=616.3
20	human_hcd_selected_2016_best	41	54	0.00	WFYIDFGGVK/2 0 HCD 38eV P=616.3
21	human_hcd_selected_2016_best	40	53	0.00	VVSPWNSEDAK/2 0 HCD 29.9522323608398eV P=616.3
22	human_hcd_selected_2016_best	38	51	0.00	NNLLNDWSQK/2 0 HCD 35eV P=616.3
23	human_hcd_selected_2016_best	36	49	0.00	SKEQAELEAAR/2 0 HCD 31eV P=616.3
24	human_hcd_selected_2016_best	36	49	0.00	SNNIDEEALAR/2 0 HCD 35eV P=616.3

Name: GQSEADSDKNATILELR/3_0 30eV
MW: 1848 Exact Mass: 1848.9288 ID#: 1811 DB: human_hair_selected_with_gvps_passed
Comment: Single Pep=Tryptic Mods=0 Fullname=-.GQSEADSDKNATILELR.-/3 Charge=3 Parent=616.3096 Mz_diff=1.2ppm HCD=30.00% Scan=40090 Origfile="2018-04-07_Direct_1H_F01_Lumos_Hair_Asian_HCD_30_5ul.raw.FT.hcd.ch.MGF" Sample="Asian_Male_Direct" Protein="ref|NM_004415.2.v3|GN=DSP_GVP_R1738Q_Q (pre=-,post=-)" Unassign_all=0.2084 Unassigned=0.0248 max_unassigned_ab=0.17 num_unassigned_peaks=1/20 FTResolution=30000 ms2IsolationWidth=2.00 ms1PrecursorAb=668197.80 Precursor1MaxAb=1816991.56 Filter="FTMS + p NSI d Full ms2 616.6445@hcd30.00 [100.0000-1860.0000]"
10 largest peaks:

```
 831.4216  999.00 |   831.923  899.76 |  530.3303  492.54 |  832.4241  450.47 |  186.0876  410.06 |
  169.061  409.39 |  744.4625  359.43 |  643.4146  339.95 |  723.3843  335.82 |  288.2035  270.97 |
```

238 m/z Values and Intensities:

```
101.0711   7.87    IQA/1.6ppm
102.0552   11.98   IEA/2.4ppm
105.0498   5.16    ?
105.3171   5.34    ?
113.071    6.19    ?
114.069    4.99    ?
120.0809   39.17   ?
124.0396   13.04   ?
127.0866   7.01    ?
129.066    129.74  IQD/1.1ppm
129.1024   134.39  IKD/1.3ppm
130.0866   12.57   ?
130.1056   11.40   IKD+i/4.4ppm
133.0434   8.24    ?
135.0809   10.28   ?
136.0758   19.15   ?
141.0659   69.94   y1-2NH3/0.3ppm
143.1181   35.25   ?
147.113    42.15   ?
152.0343   90.56   b2-2NH3/0.5ppm
152.4282   7.53    ?
154.7535   7.63    ?
155.0819   9.46    ?
```



2018-04-12_Direct_1C_F01_Lumos_Hair_A... 04/12/18 19:51:36 2018-04-12_Direct_1C_F01_Lumos_Hair_Asian_HCD_30_5ul

RT: 0.00 - 200.00

NL: 2.95E9
Base Peak MS
Genesis
2018-04-12_Direct_1C_F01_Lumos_Hair_Asian_HCD_30_5ul

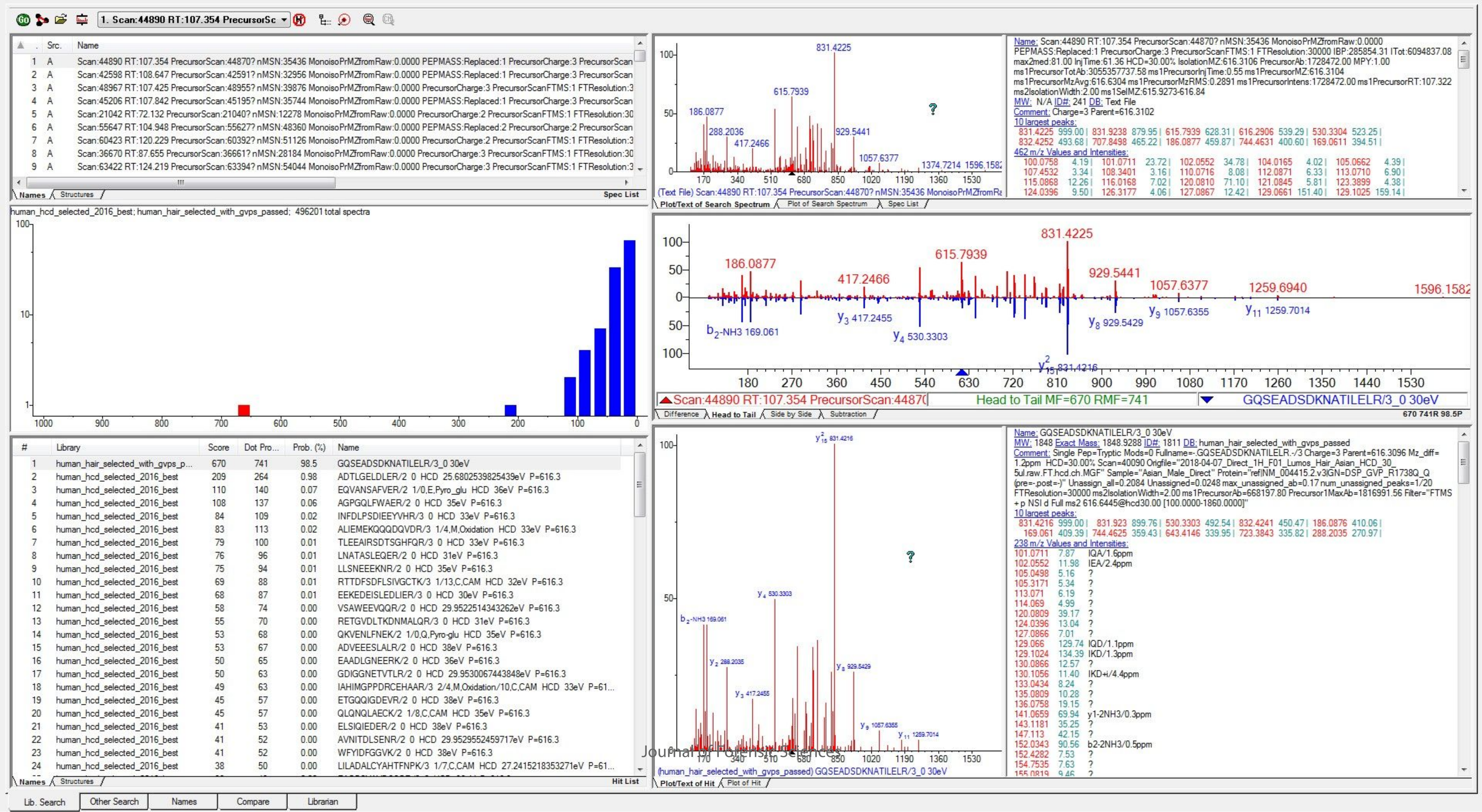
RT: 0.00 - 200.05

NL: 0
TIC F: FTMS + p NSI Full ms [197.0777-1600.0000] MS
2018-04-12_Direct_1C_F01_Lumos_Hair_Asian_HCD_30_5ul

2018-04-12_Direct_1C_F01_Lumos_Hair_Asian_HCD_30_5ul #42591 RT: 108.63 AV: 1 NL: 3.46E7
T: FTMS + p NSI Full ms [350.0000-1600.0000]

MS1 Peak

DSP_R1738Q_Q: GQSEADSDKNATILELR in Fraction 2 confirmed



▲	Src.	Name
1	A	Scan:44890 RT:107.354 PrecursorScan:44870? nMSN:35436 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
2	A	Scan:42598 RT:108.647 PrecursorScan:42591? nMSN:32956 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
3	A	Scan:48967 RT:107.425 PrecursorScan:48955? nMSN:39876 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:3
4	A	Scan:45206 RT:107.842 PrecursorScan:45195? nMSN:35744 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
5	A	Scan:21042 RT:72.132 PrecursorScan:21040? nMSN:12278 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30
6	A	Scan:55647 RT:104.948 PrecursorScan:55627? nMSN:48360 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:2 PrecursorCharge:2 PrecursorScan
7	A	Scan:60423 RT:120.229 PrecursorScan:60392? nMSN:51126 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:3
8	A	Scan:36670 RT:87.655 PrecursorScan:36661? nMSN:28184 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30
9	A	Scan:63422 RT:124.219 PrecursorScan:63394? nMSN:54044 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:3

Spec List

(Text File) Scan:44890 RT:107.354 PrecursorScan:44870? nMSN:35436 MonoisoPrMZfromRa

Name: Scan:44890 RT:107.354 PrecursorScan:44870? nMSN:35436 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30000 IBP:285854.31 ITot:6094837.08 max2med:81.00 InjTime:61.36 HCD=30.00% IsolationMZ:616.3106 PrecursorAb:1728472.00 MPY:1.00 ms1PrecursorTotAb:3055357737.58 ms1PrecursorInjTime:0.55 ms1PrecursorMZ:616.3104 ms1PrecursorMzAvg:616.6304 ms1PrecursorMzRMS:0.2891 ms1PrecursorIntens:1728472.00 ms1PrecursorRT:107.322 ms2IsolationWidth:2.00 ms1SelMZ:615.9273-616.84
MW: N/A ID#: 241 DB: Text File
Comment: Charge=3 Parent=616.3102
10 largest peaks:

```
 831.4225  999.00 |  831.9238  879.95 |  615.7939  628.31 |  616.2906  539.29 |  530.3304  523.25 |
 832.4252  493.68 |  707.8498  465.22 |  186.0877  459.87 |  744.4631  400.60 |  169.0611  394.51 |
```

462 m/z Values and Intensities:

```
  100.0758    4.19 |  101.0711   23.72 |  102.0552   34.78 |  104.0165    4.02 |  105.0662    4.39 |
  107.4532    3.34 |  108.3401    3.16 |  110.0716    8.08 |  112.0871    6.33 |  113.0710    6.90 |
  115.0868   12.26 |  116.0168    7.02 |  120.0810   71.10 |  121.0845    5.81 |  123.3899    4.38 |
  124.0396    9.50 |  126.3177    4.06 |  127.0867   12.42 |  129.0661  151.40 |  129.1025  159.14 |
```

human_hcd_selected_2016_best; human_hair_selected_with_gvps_passed; 496201 total spectra

Scan:44890 RT:107.354 PrecursorScan:4487(Head to Tail MF=670 RMF=741 GQSEADSDKNATILELR/3_0 30eV

670 741R 98.5P

#	Library	Score	Dot Pro...	Prob. (%)	Name
1	human_hair_selected_with_gvps_p...	670	741	98.5	GQSEADSDKNATILELR/3_0 30eV
2	human_hcd_selected_2016_best	209	264	0.98	ADTLGELDLER/2 0 HCD 25.6802539825439eV P=616.3
3	human_hcd_selected_2016_best	110	140	0.07	EQVANSAFVER/2 1/0,E,Pyro_glu HCD 36eV P=616.3
4	human_hcd_selected_2016_best	108	137	0.06	AGPGQLFWAER/2 0 HCD 35eV P=616.3
5	human_hcd_selected_2016_best	84	109	0.02	INFDLPSDIEEYVHR/3 0 HCD 33eV P=616.3
6	human_hcd_selected_2016_best	83	113	0.02	ALIEMEKQQQDQVDR/3 1/4,M,Oxidation HCD 33eV P=616.3
7	human_hcd_selected_2016_best	79	100	0.01	TLEEAIRSDTSGHFQR/3 0 HCD 33eV P=616.3
8	human_hcd_selected_2016_best	76	96	0.01	LNATASLEQER/2 0 HCD 31eV P=616.3
9	human_hcd_selected_2016_best	75	94	0.01	LLSNEEEKNR/2 0 HCD 35eV P=616.3
10	human_hcd_selected_2016_best	69	88	0.01	RTTDFSDFLSIVGCTK/3 1/13,C,CAM HCD 32eV P=616.3
11	human_hcd_selected_2016_best	68	87	0.01	EEKEDEISLEDLIER/3 0 HCD 30eV P=616.3
12	human_hcd_selected_2016_best	58	74	0.00	VSAWEEVQQR/2 0 HCD 29.9522514343262eV P=616.3
13	human_hcd_selected_2016_best	55	70	0.00	RETGVDLTKDNMALQR/3 0 HCD 31eV P=616.3
14	human_hcd_selected_2016_best	53	68	0.00	QKVENLFNEK/2 1/0,Q,Pyro-glu HCD 35eV P=616.3
15	human_hcd_selected_2016_best	53	67	0.00	ADVEEESLALR/2 0 HCD 38eV P=616.3
16	human_hcd_selected_2016_best	50	65	0.00	EAADLGNEERK/2 0 HCD 36eV P=616.3
17	human_hcd_selected_2016_best	50	63	0.00	GDIGGNETVTLR/2 0 HCD 29.9530067443848eV P=616.3
18	human_hcd_selected_2016_best	49	63	0.00	IAHIMGPPDRCEHAAR/3 2/4,M,Oxidation/10,C,CAM HCD 33eV P=61...
19	human_hcd_selected_2016_best	45	57	0.00	ETGQQIGDEVR/2 0 HCD 38eV P=616.3
20	human_hcd_selected_2016_best	45	57	0.00	QLQNQLAECK/2 1/8,C,CAM HCD 35eV P=616.3
21	human_hcd_selected_2016_best	41	53	0.00	ELSIQIEDER/2 0 HCD 38eV P=616.3
22	human_hcd_selected_2016_best	41	52	0.00	AVNITDLSENR/2 0 HCD 29.9529552459717eV P=616.3
23	human_hcd_selected_2016_best	41	52	0.00	WFYIDFGGVK/2 0 HCD 38eV P=616.3
24	human_hcd_selected_2016_best	38	50	0.00	LILADALCYAHTFNPK/3 1/7,C,CAM HCD 27.2415218353271eV P=61...

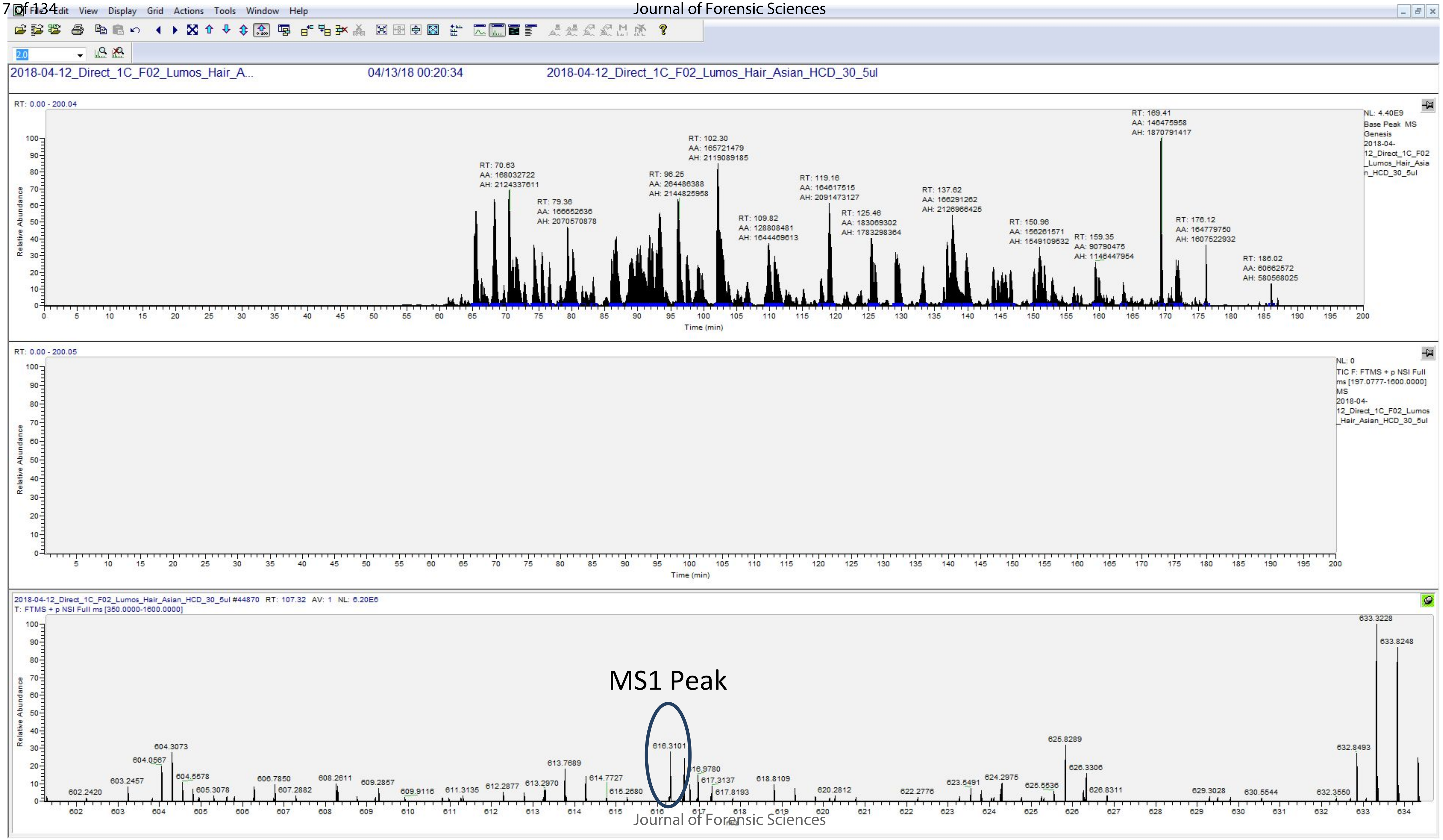
(human_hair_selected_with_gvps_passed) GQSEADSDKNATILELR/3_0 30eV

Name: GQSEADSDKNATILELR/3_0 30eV
MW: 1848 Exact Mass: 1848.9288 ID#: 1811 DB: human_hair_selected_with_gvps_passed
Comment: Single Pep=Tryptic Mods=0 Fullname=-.GQSEADSDKNATILELR.-/3 Charge=3 Parent=616.3096 Mz_diff=1.2ppm HCD=30.00% Scan=40090 Origfile="2018-04-07_Direct_1H_F01_Lumos_Hair_Asian_HCD_30_5ul.raw.FT.hcd.ch.MGF" Sample="Asian_Male_Direct" Protein="ref|NM_004415.2.v3|GN=DSP_GVP_R1738Q_Q (pre=-,post=-)" Unassign_all=0.2084 Unassigned=0.0248 max_unassigned_ab=0.17 num_unassigned_peaks=1/20 FTResolution=30000 ms2IsolationWidth=2.00 ms1PrecursorAb=668197.80 Precursor1MaxAb=1816991.56 Filter="FTMS + p NSI d Full ms2 616.6445@hcd30.00 [100.0000-1860.0000]"
10 largest peaks:

```
 831.4216  999.00 |   831.923  899.76 |  530.3303  492.54 |  832.4241  450.47 |  186.0876  410.06 |
  169.061  409.39 |  744.4625  359.43 |  643.4146  339.95 |  723.3843  335.82 |  288.2035  270.97 |
```

238 m/z Values and Intensities:

```
101.0711   7.87    IQA/1.6ppm
102.0552   11.98   IEA/2.4ppm
105.0498   5.16    ?
105.3171   5.34    ?
113.071    6.19    ?
114.069    4.99    ?
120.0809   39.17   ?
124.0396   13.04   ?
127.0866   7.01    ?
129.066    129.74  IQD/1.1ppm
129.1024   134.39  IKD/1.3ppm
130.0866   12.57   ?
130.1056   11.40   IKD+i/4.4ppm
133.0434   8.24    ?
135.0809   10.28   ?
136.0758   19.15   ?
141.0659   69.94   y1-2NH3/0.3ppm
143.1181   35.25   ?
147.113    42.15   ?
152.0343   90.56   b2-2NH3/0.5ppm
152.4282   7.53    ?
154.7535   7.63    ?
155.0819   9.46    ?
```

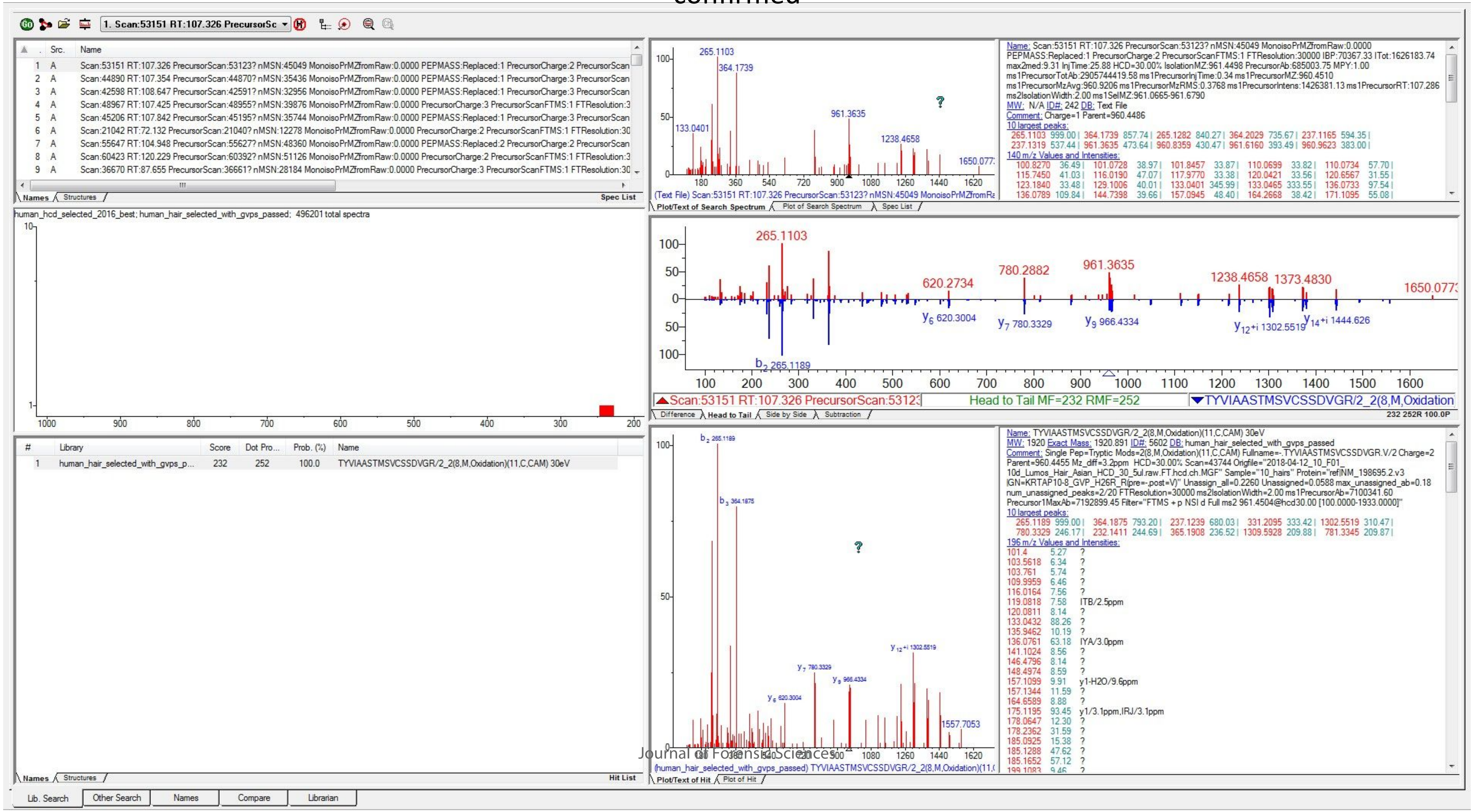


2018-04-12_Direct_1C_F02_Lumos_Hair_A... 04/13/18 00:20:34 2018-04-12_Direct_1C_F02_Lumos_Hair_Asian_HCD_30_5ul

2018-04-12_Direct_1C_F02_Lumos_Hair_Asian_HCD_30_5ul #44870 RT: 107.32 AV: 1 NL: 6.20E6
T: FTMS + p NSI Full ms [350.0000-1600.0000]

MS1 Peak

KRTAP10-8_H26R_R: TYVIAASTMSVCSSDVGR in Fraction 9 confirmed



▲	.	Src.	Name
	1	A	Scan:53151 RT:107.326 PrecursorScan:53123? nMSN:45049 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:2 PrecursorScan
	2	A	Scan:44890 RT:107.354 PrecursorScan:44870? nMSN:35436 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
	3	A	Scan:42598 RT:108.647 PrecursorScan:42591? nMSN:32956 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
	4	A	Scan:48967 RT:107.425 PrecursorScan:48955? nMSN:39876 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:3
	5	A	Scan:45206 RT:107.842 PrecursorScan:45195? nMSN:35744 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
	6	A	Scan:21042 RT:72.132 PrecursorScan:21040? nMSN:12278 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30
	7	A	Scan:55647 RT:104.948 PrecursorScan:55627? nMSN:48360 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:2 PrecursorCharge:2 PrecursorScan
	8	A	Scan:60423 RT:120.229 PrecursorScan:60392? nMSN:51126 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:3
	9	A	Scan:36670 RT:87.655 PrecursorScan:36661? nMSN:28184 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30

Name: Scan:53151 RT:107.326 PrecursorScan:53123? nMSN:45049 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30000 IBP:70367.33 ITot:1626183.74 max2med:9.31 InjTime:25.88 HCD=30.00% IsolationMZ:961.4498 PrecursorAb:685003.75 MPY:1.00 ms1PrecursorTotAb:2905744419.58 ms1PrecursorInjTime:0.34 ms1PrecursorMZ:960.4510 ms1PrecursorMzAvg:960.9206 ms1PrecursorMzRMS:0.3768 ms1PrecursorIntens:1426381.13 ms1PrecursorRT:107.286 ms2IsolationWidth:2.00 ms1SelMZ:961.0665-961.6790
MW: N/A ID#: 242 DB: Text File
Comment: Charge=1 Parent=960.4486
10 largest peaks:

```
 265.1103  999.00 |  364.1739  857.74 |  265.1282  840.27 |  364.2029  735.67 |  237.1165  594.35 |
 237.1319  537.44 |  961.3635  473.64 |  960.8359  430.47 |  961.6160  393.49 |  960.9623  383.00 |
```

140 m/z Values and Intensities:

```
  100.8270   36.49 |  101.0728   38.97 |  101.8457   33.87 |  110.0699   33.82 |  110.0734   57.70 |
  115.7450   41.03 |  116.0190   47.07 |  117.9770   33.38 |  120.0421   33.56 |  120.6567   31.55 |
  123.1840   33.48 |  129.1006   40.01 |  133.0401  345.99 |  133.0465  333.55 |  136.0733   97.54 |
  136.0789  109.84 |  144.7398   39.66 |  157.0945   48.40 |  164.2668   38.42 |  171.1095   55.08 |
```

(Text File) Scan:53151 RT:107.326 PrecursorScan:53123? nMSN:45049 MonoisoPrMZfromRa

human_hcd_selected_2016_best; human_hair_selected_with_gvps_passed; 496201 total spectra

▲Scan:53151 RT:107.326 PrecursorScan:53123

Head to Tail MF=232 RMF=252

▼TYVIAASTMSVCSSDVGR/2_2(8,M,Oxidation

232 252R 100.0P

#	Library	Score	Dot Pro...	Prob. (%)	Name
1	human_hair_selected_with_gvps_p...	232	252	100.0	TYVIAASTMSVCSSDVGR/2_2(8,M,Oxidation)(11,C,CAM) 30eV

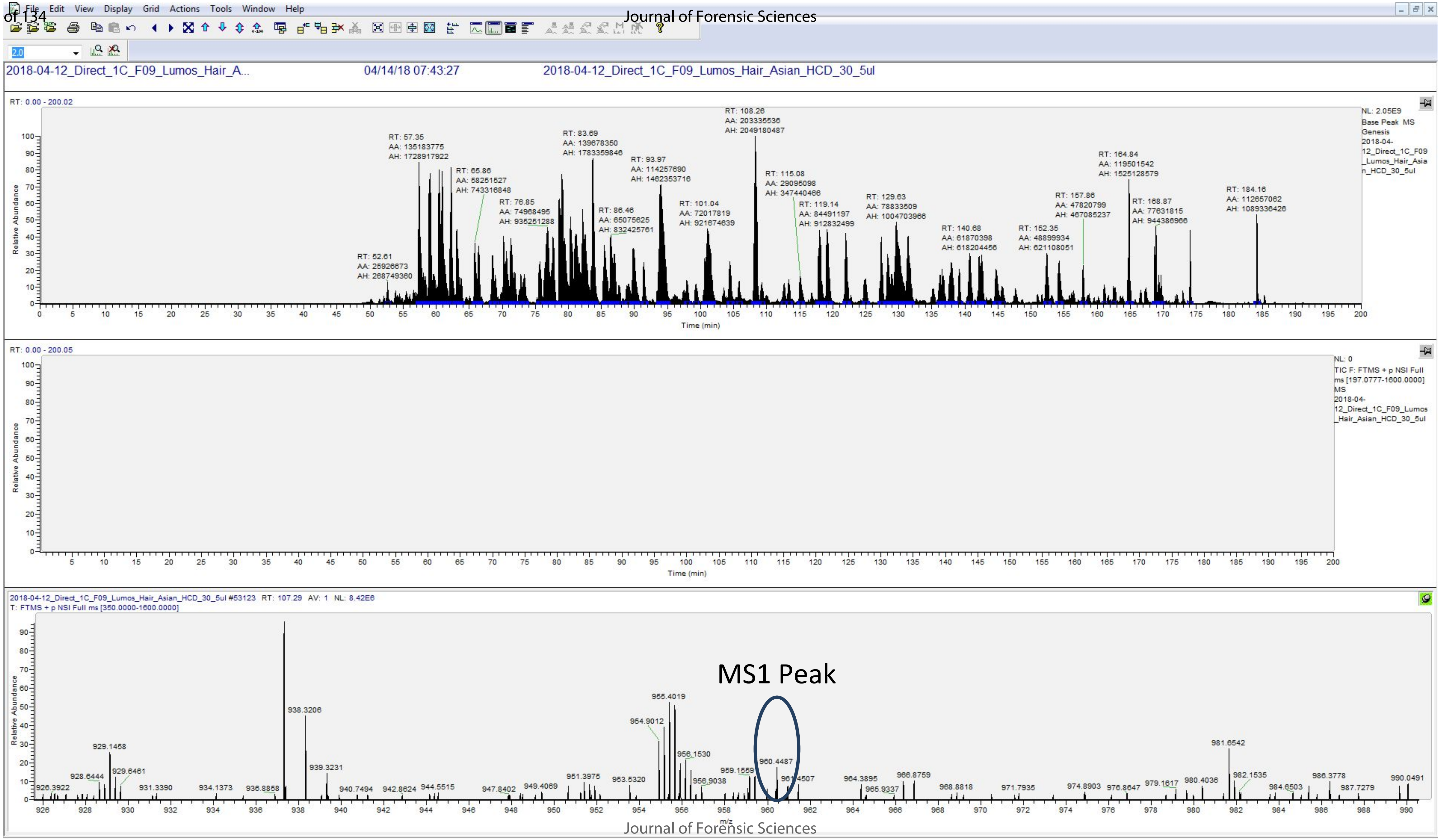
Name: TYVIAASTMSVCSSDVGR/2_2(8,M,Oxidation)(11,C,CAM) 30eV
MW: 1920 Exact Mass: 1920.891 ID#: 5602 DB: human_hair_selected_with_gvps_passed
Comment: Single Pep=Tryptic Mods=2(8,M,Oxidation)(11,C,CAM) Fullname=-.TYVIAASTMSVCSSDVGR.V/2 Charge=2 Parent=960.4455 Mz_diff=3.2ppm HCD=30.00% Scan=43744 Origfile="2018-04-12_10_F01_10d_Lumos_Hair_Asian_HCD_30_5ul.raw.FT.hcd.ch.MGF" Sample="10_hairs" Protein="ref|NM_198695.2.v3|GN=KRTAP10-8_GVP_H26R_R(pre=-,post=V)" Unassign_all=0.2260 Unassigned=0.0588 max_unassigned_ab=0.18 num_unassigned_peaks=2/20 FTResolution=30000 ms2IsolationWidth=2.00 ms1PrecursorAb=7100341.60 Precursor1MaxAb=7192899.45 Filter="FTMS + p NSI d Full ms2 961.4504@hcd30.00 [100.0000-1933.0000]"
10 largest peaks:

```
  265.1189  999.00 |   364.1875  793.20 |   237.1239  680.03 |   331.2095  333.42 |  1302.5519  310.47 |
  780.3329  246.17 |   232.1411  244.69 |   365.1908  236.52 |  1309.5928  209.88 |   781.3345  209.87 |
```

196 m/z Values and Intensities:

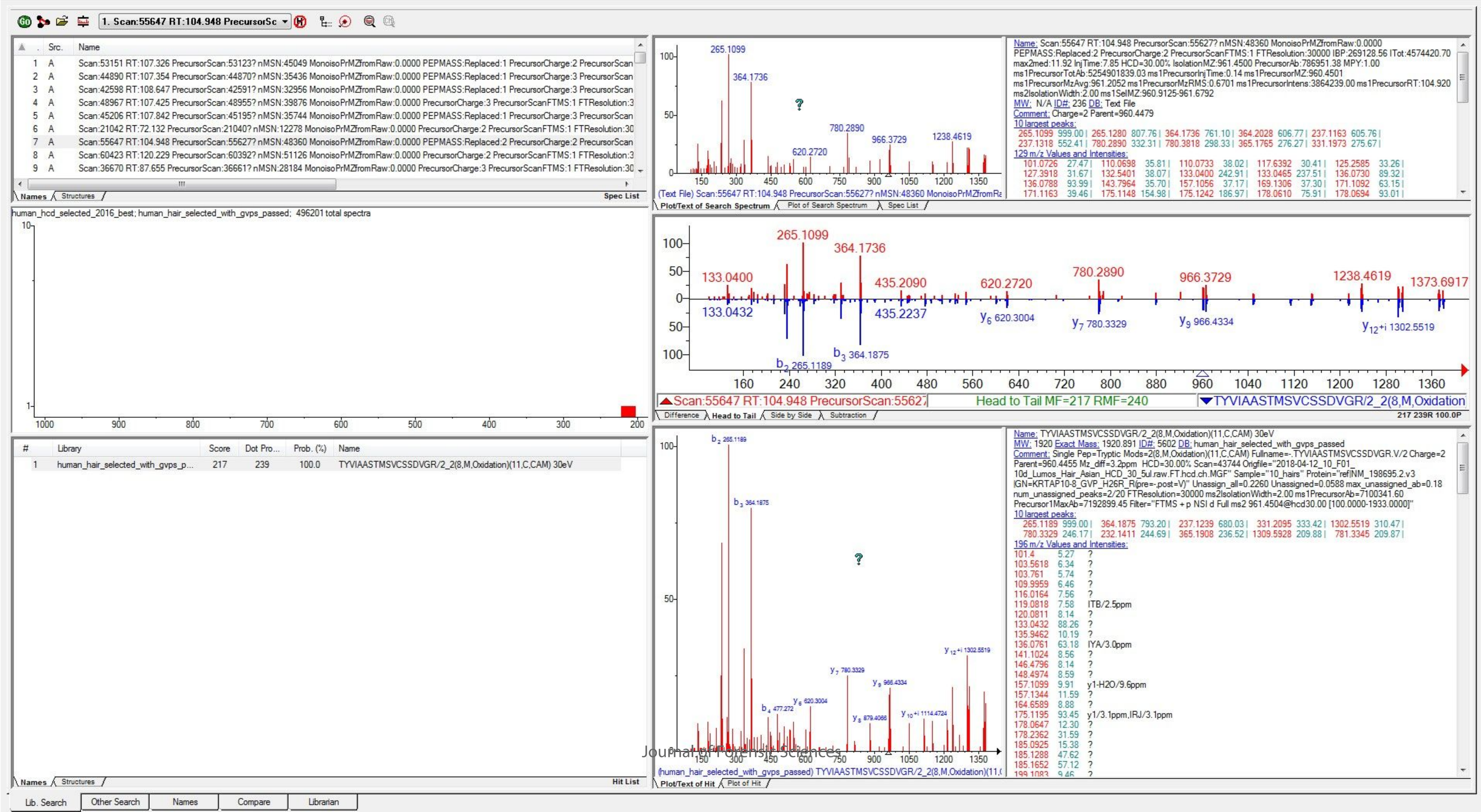
```
101.4      5.27   ?
103.5618   6.34   ?
103.761    5.74   ?
109.9959   6.46   ?
116.0164   7.56   ?
119.0818   7.58   ITB/2.5ppm
120.0811   8.14   ?
133.0432   88.26  ?
135.9462   10.19  ?
136.0761   63.18  IYA/3.0ppm
141.1024   8.56   ?
146.4796   8.14   ?
148.4974   8.59   ?
157.1099   9.91   y1-H2O/9.6ppm
157.1344   11.59  ?
164.6589   8.88   ?
175.1195   93.45  y1/3.1ppm,IRJ/3.1ppm
178.0647   12.30  ?
178.2362   31.59  ?
185.0925   15.38  ?
185.1288   47.62  ?
185.1652   57.12  ?
199.1083   9.46   ?
```

(human_hair_selected_with_gvps_passed) TYVIAASTMSVCSSDVGR/2_2(8,M,Oxidation)(11,(


File Edit View Display Grid Actions Tools Window Help
2.0
2018-04-12_Direct_1C_F09_Lumos_Hair_A...
04/14/18 07:43:27
2018-04-12_Direct_1C_F09_Lumos_Hair_Asian_HCD_30_5ul
RT: 0.00 - 200.02
NL: 2.05E9
Base Peak MS
Genesis
2018-04-12_Direct_1C_F09_Lumos_Hair_Asian_HCD_30_5ul
RT: 52.61 AA: 25926673 AH: 268749360
RT: 57.35 AA: 135183775 AH: 1728917922
RT: 65.86 AA: 58251527 AH: 743316848
RT: 76.85 AA: 74968495 AH: 935251288
RT: 83.69 AA: 139678350 AH: 1783359846
RT: 86.46 AA: 65075625 AH: 832425761
RT: 93.97 AA: 114257690 AH: 1462353716
RT: 101.04 AA: 72017819 AH: 921674639
RT: 108.26 AA: 203335536 AH: 2049180487
RT: 115.08 AA: 29095098 AH: 347440466
RT: 119.14 AA: 84491197 AH: 912832499
RT: 129.63 AA: 78833509 AH: 1004703966
RT: 140.68 AA: 61870398 AH: 618204456
RT: 152.35 AA: 48899934 AH: 621108051
RT: 157.86 AA: 47820799 AH: 467085237
RT: 164.84 AA: 119501542 AH: 1525128579
RT: 168.87 AA: 77631815 AH: 944386966
RT: 184.16 AA: 112657062 AH: 1089336426
Relative Abundance
Time (min)
RT: 0.00 - 200.05
NL: 0
TIC F: FTMS + p NSI Full ms [197.0777-1600.0000] MS
2018-04-12_Direct_1C_F09_Lumos_Hair_Asian_HCD_30_5ul
Relative Abundance
Time (min)
2018-04-12_Direct_1C_F09_Lumos_Hair_Asian_HCD_30_5ul #53123 RT: 107.29 AV: 1 NL: 8.42E6
T: FTMS + p NSI Full ms [350.0000-1600.0000]
Relative Abundance
m/z
926.3922
928.6444
929.1458
929.6461
931.3390
934.1373
936.8858
938.3206
939.3231
940.7494
942.8624
944.5515
947.8402
949.4069
951.3975
953.5320
954.9012
955.4019
956.1530
956.9038
959.1559
960.4487
961.4507
964.3895
965.9337
966.8759
968.8818
971.7935
974.8903
976.8647
979.1617
980.4036
981.6542
982.1535
984.6503
986.3778
987.7279
990.0491

MS1 Peak

KRTAP10-8_H26R_R: TYVIAASTMSVCSSDVGR in Fraction 10 confirmed



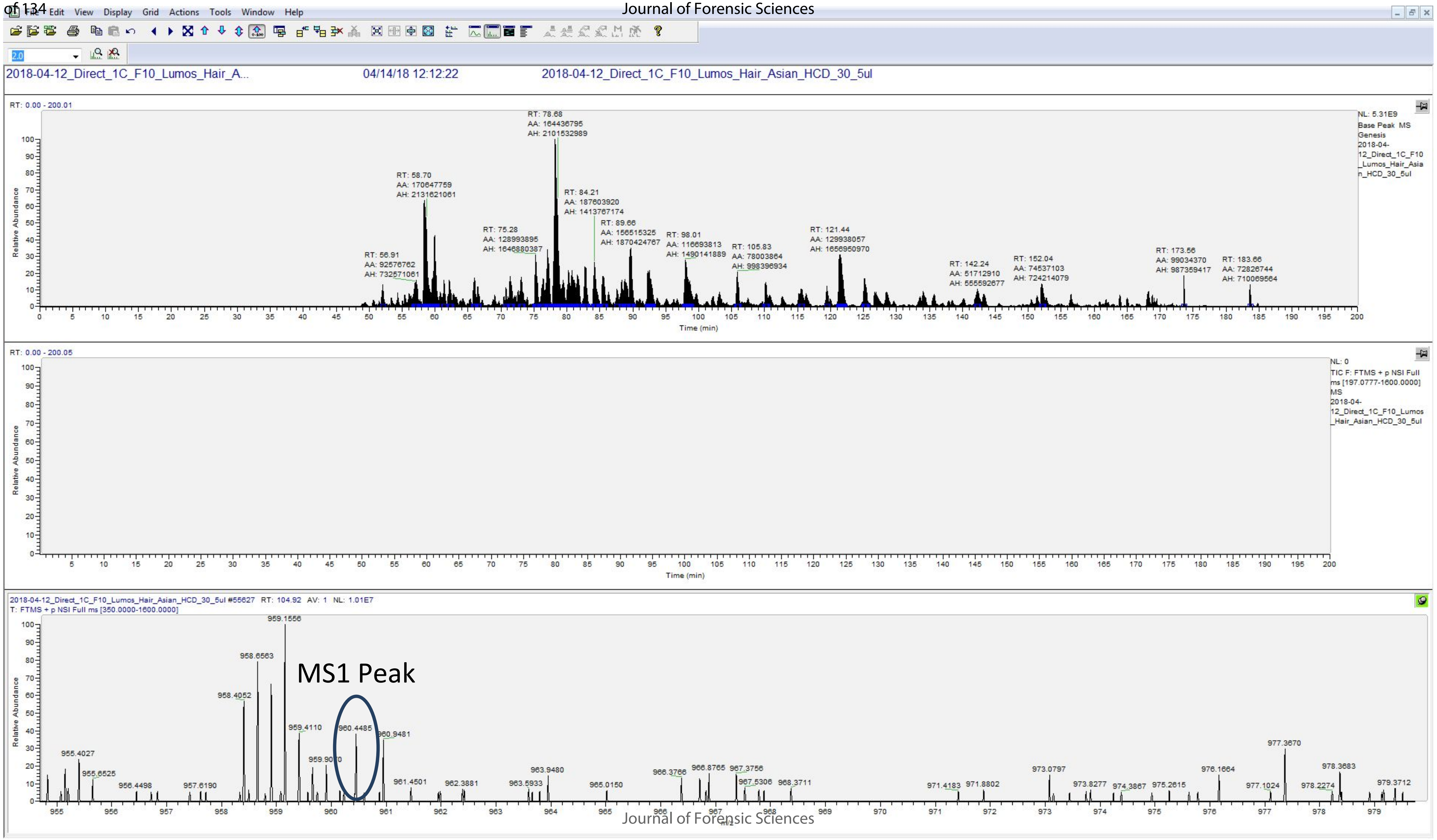
▲	.	Src.	Name
	1	A	Scan:53151 RT:107.326 PrecursorScan:53123? nMSN:45049 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:2 PrecursorScan
	2	A	Scan:44890 RT:107.354 PrecursorScan:44870? nMSN:35436 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
	3	A	Scan:42598 RT:108.647 PrecursorScan:42591? nMSN:32956 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
	4	A	Scan:48967 RT:107.425 PrecursorScan:48955? nMSN:39876 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:3
	5	A	Scan:45206 RT:107.842 PrecursorScan:45195? nMSN:35744 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
	6	A	Scan:21042 RT:72.132 PrecursorScan:21040? nMSN:12278 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30
	7	A	Scan:55647 RT:104.948 PrecursorScan:55627? nMSN:48360 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:2 PrecursorCharge:2 PrecursorScan
	8	A	Scan:60423 RT:120.229 PrecursorScan:60392? nMSN:51126 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:3
	9	A	Scan:36670 RT:87.655 PrecursorScan:36661? nMSN:28184 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30

human_hcd_selected_2016_best; human_hair_selected_with_gvps_passed; 496201 total spectra

#	Library	Score	Dot Pro...	Prob. (%)	Name
1	human_hair_selected_with_gvps_p...	217	239	100.0	TYVIAASTMSVCSSDVGR/2_2(8,M,Oxidation)(11,C,CAM) 30eV

```
Name: Scan:55647 RT:104.948 PrecursorScan:55627? nMSN:48360 MonoisoPrMZfromRaw:0.0000
PEPMASS:Replaced:2 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30000 IBP:269128.56 ITot:4574420.70
max2med:11.92 InjTime:7.85 HCD=30.00% IsolationMZ:961.4500 PrecursorAb:786951.38 MPY:1.00
ms1PrecursorTotAb:5254901839.03 ms1PrecursorInjTime:0.14 ms1PrecursorMZ:960.4501
ms1PrecursorMzAvg:961.2052 ms1PrecursorMzRMS:0.6701 ms1PrecursorIntens:3864239.00 ms1PrecursorRT:104.920
ms2IsolationWidth:2.00 ms1SelMZ:960.9125-961.6792
MW:  N/A ID#: 236 DB: Text File
Comment: Charge=2 Parent=960.4479
10 largest peaks:
  265.1099  999.00 |  265.1280  807.76 |  364.1736  761.10 |  364.2028  606.77 |  237.1163  605.76 |
  237.1318  552.41 |  780.2890  332.31 |  780.3818  298.33 |  365.1765  276.27 |  331.1973  275.67 |
129 m/z Values and Intensities:
   101.0726   27.47 |   110.0698   35.81 |   110.0733   38.02 |   117.6392   30.41 |   125.2585   33.26 |
   127.3918   31.67 |   132.5401   38.07 |   133.0400  242.91 |   133.0465  237.51 |   136.0730   89.32 |
   136.0788   93.99 |   143.7964   35.70 |   157.1056   37.17 |   169.1306   37.30 |   171.1092   63.15 |
   171.1163   39.46 |   175.1148  154.98 |   175.1242  186.97 |   178.0610   75.91 |   178.0694   93.01 |
```

```
Name: TYVIAASTMSVCSSDVGR/2_2(8,M,Oxidation)(11,C,CAM) 30eV
MW: 1920 Exact Mass: 1920.891 ID#: 5602 DB: human_hair_selected_with_gvps_passed
Comment: Single Pep=Tryptic Mods=2(8,M,Oxidation)(11,C,CAM) Fullname=-.TYVIAASTMSVCSSDVGR.V/2 Charge=2
Parent=960.4455 Mz_diff=3.2ppm  HCD=30.00% Scan=43744 Origfile="2018-04-12_10_F01_
10d_Lumos_Hair_Asian_HCD_30_5ul.raw.FT.hcd.ch.MGF" Sample="10_hairs" Protein="ref|NM_198695.2.v3
|GN=KRTAP10-8_GVP_H26R_R(pre=-,post=V)" Unassign_all=0.2260 Unassigned=0.0588 max_unassigned_ab=0.18
num_unassigned_peaks=2/20 FTResolution=30000 ms2IsolationWidth=2.00 ms1PrecursorAb=7100341.60
Precursor1MaxAb=7192899.45 Filter="FTMS + p NSI d Full ms2 961.4504@hcd30.00 [100.0000-1933.0000]"
10 largest peaks:
   265.1189  999.00 |   364.1875  793.20 |   237.1239  680.03 |   331.2095  333.42 |  1302.5519  310.47 |
   780.3329  246.17 |   232.1411  244.69 |   365.1908  236.52 |  1309.5928  209.88 |   781.3345  209.87 |
196 m/z Values and Intensities:
101.4      5.27    ?
103.5618   6.34    ?
103.761    5.74    ?
109.9959   6.46    ?
116.0164   7.56    ?
119.0818   7.58    ITB/2.5ppm
120.0811   8.14    ?
133.0432   88.26   ?
135.9462   10.19   ?
136.0761   63.18   IYA/3.0ppm
141.1024   8.56    ?
146.4796   8.14    ?
148.4974   8.59    ?
157.1099   9.91    y1-H2O/9.6ppm
157.1344   11.59   ?
164.6589   8.88    ?
175.1195   93.45   y1/3.1ppm,IRJ/3.1ppm
178.0647   12.30   ?
178.2362   31.59   ?
185.0925   15.38   ?
185.1288   47.62   ?
185.1652   57.12   ?
199.1083   9.46    ?
```


File Edit View Display Grid Actions Tools Window Help
2.0
RT: 0.00 - 200.01
NL: 5.31E9
Base Peak MS
Genesis
2018-04-12_Direct_1C_F10_Lumos_Hair_Asian_HCD_30_5ul
RT: 78.68 AA: 164436795 AH: 2101532989
RT: 58.70 AA: 170647759 AH: 2131621061
RT: 84.21 AA: 187603920 AH: 1413767174
RT: 89.66 AA: 156515325 AH: 1870424767
RT: 75.28 AA: 128993895 AH: 1646880387
RT: 98.01 AA: 116693813 AH: 1490141889
RT: 56.91 AA: 92576762 AH: 732571061
RT: 105.83 AA: 78003864 AH: 998396934
RT: 121.44 AA: 129938057 AH: 1656950970
RT: 142.24 AA: 51712910 AH: 555592677
RT: 152.04 AA: 74537103 AH: 724214079
RT: 173.56 AA: 99034370 AH: 987359417
RT: 183.66 AA: 72826744 AH: 710069564
Relative Abundance
Time (min)
RT: 0.00 - 200.05
NL: 0
TIC F: FTMS + p NSI Full ms [197.0777-1600.0000] MS
2018-04-12_Direct_1C_F10_Lumos_Hair_Asian_HCD_30_5ul
Relative Abundance
Time (min)
959.1556
958.6563
958.4052
959.4110
960.4485
960.9481
959.9070
955.4027
955.6525
956.4498
957.6190
961.4501
962.3881
963.5933
963.9480
965.0150
966.3766
966.8765
967.3756
967.5306
968.3711
971.4183
971.8802
973.0797
973.8277
974.3867
975.2615
976.1664
977.1024
977.3670
978.2274
978.3683
979.3712
Relative Abundance
m/z

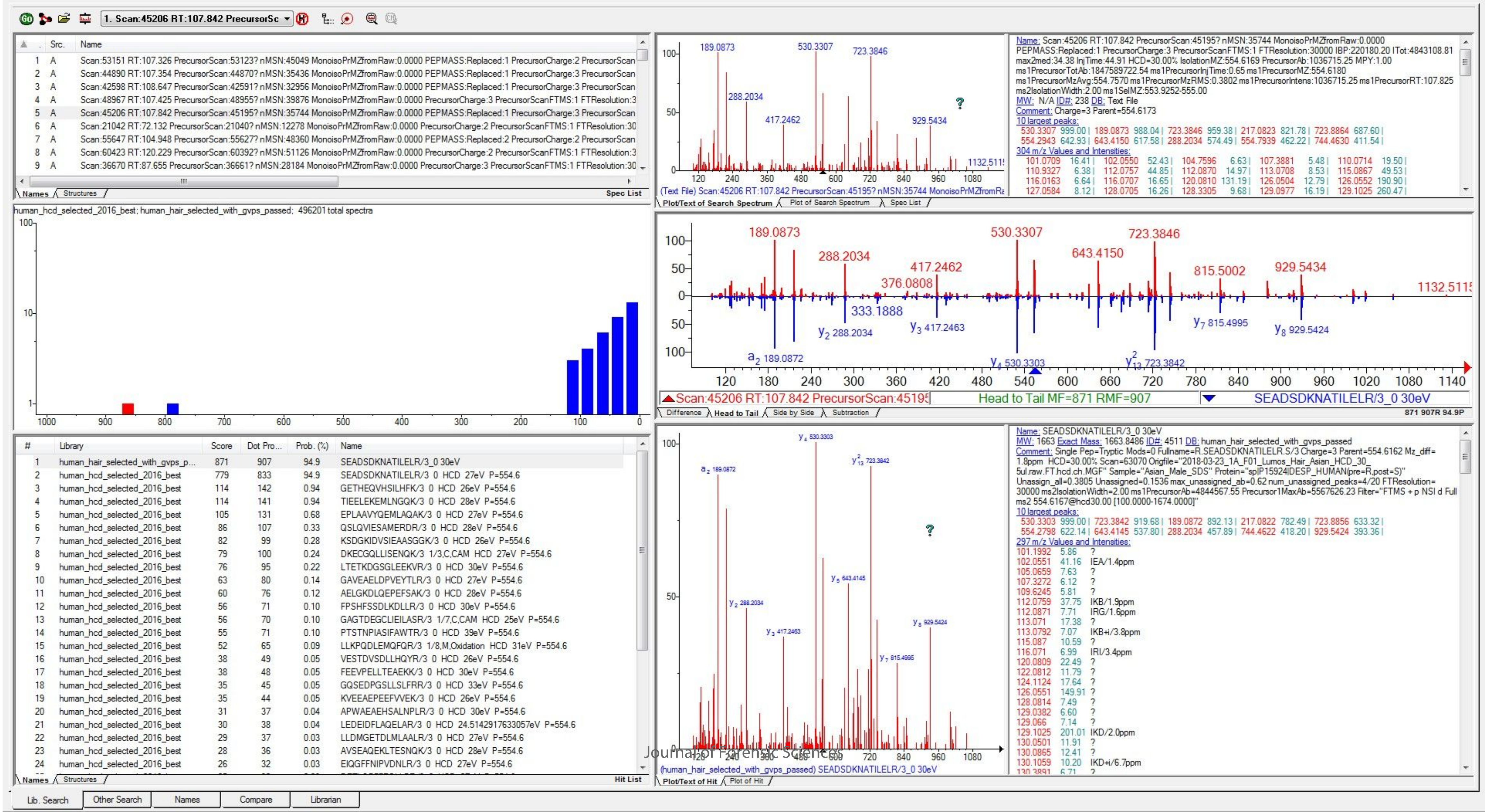
2018-04-12_Direct_1C_F10_Lumos_Hair_A... 04/14/18 12:12:22 2018-04-12_Direct_1C_F10_Lumos_Hair_Asian_HCD_30_5ul

2018-04-12_Direct_1C_F10_Lumos_Hair_Asian_HCD_30_5ul #55627 RT: 104.92 AV: 1 NL: 1.01E7
T: FTMS + p NSI Full ms [350.0000-1600.0000]

MS1 Peak

Check Low Abundance Regular Forms (if applicable)

DSP_R1738Q_Q: (R)SEADSDKNATILELR in Fraction 2 confirmed



	Src.	Name
1	A	Scan:53151 RT:107.326 PrecursorScan:53123? nMSN:45049 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:2 PrecursorScan
2	A	Scan:44890 RT:107.354 PrecursorScan:44870? nMSN:35436 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
3	A	Scan:42598 RT:108.647 PrecursorScan:42591? nMSN:32956 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
4	A	Scan:48967 RT:107.425 PrecursorScan:48955? nMSN:39876 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:3
5	A	Scan:45206 RT:107.842 PrecursorScan:45195? nMSN:35744 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
6	A	Scan:21042 RT:72.132 PrecursorScan:21040? nMSN:12278 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30
7	A	Scan:55647 RT:104.948 PrecursorScan:55627? nMSN:48360 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:2 PrecursorCharge:2 PrecursorScan
8	A	Scan:60423 RT:120.229 PrecursorScan:60392? nMSN:51126 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:3
9	A	Scan:36670 RT:87.655 PrecursorScan:36661? nMSN:28184 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30

Name: Scan:45206 RT:107.842 PrecursorScan:45195? nMSN:35744 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30000 IBP:220180.20 ITot:4843108.81 max2med:34.38 InjTime:44.91 HCD=30.00% IsolationMZ:554.6169 PrecursorAb:1036715.25 MPY:1.00 ms1PrecursorTotAb:1847589722.54 ms1PrecursorInjTime:0.65 ms1PrecursorMZ:554.6180 ms1PrecursorMzAvg:554.7570 ms1PrecursorMzRMS:0.3802 ms1PrecursorIntens:1036715.25 ms1PrecursorRT:107.825 ms2IsolationWidth:2.00 ms1SelMZ:553.9252-555.00
MW: N/A ID#: 238 DB: Text File
Comment: Charge=3 Parent=554.6173
10 largest peaks:

```
 530.3307  999.00 |  189.0873  988.04 |  723.3846  959.38 |  217.0823  821.78 |  723.8864  687.60 |
 554.2943  642.93 |  643.4150  617.58 |  288.2034  574.49 |  554.7939  462.22 |  744.4630  411.54 |
```

304 m/z Values and Intensities:

```
  101.0709   16.41 |  102.0550   52.43 |  104.7596    6.63 |  107.3881    5.48 |  110.0714   19.50 |
  110.9327    6.38 |  112.0757   44.85 |  112.0870   14.97 |  113.0708    8.53 |  115.0867   49.53 |
  116.0163    6.64 |  116.0707   16.65 |  120.0810  131.19 |  126.0504   12.79 |  126.0552  190.90 |
  127.0584    8.12 |  128.0705   16.26 |  128.3305    9.68 |  129.0977   16.19 |  129.1025  260.47 |
```

(Text File) Scan:45206 RT:107.842 PrecursorScan:45195? nMSN:35744 MonoisoPrMZfromRa

human_hcd_selected_2016_best; human_hair_selected_with_gvps_passed; 496201 total spectra

Scan:45206 RT:107.842 PrecursorScan:4519! Head to Tail MF=871 RMF=907 SEADSDKNATILELR/3_0 30eV

871 907R 94.9P

#	Library	Score	Dot Pro...	Prob. (%)	Name
1	human_hair_selected_with_gvps_p...	871	907	94.9	SEADSDKNATILELR/3_0 30eV
2	human_hcd_selected_2016_best	779	833	94.9	SEADSDKNATILELR/3 0 HCD 27eV P=554.6
3	human_hcd_selected_2016_best	114	142	0.94	GETHEQVHSILHFK/3 0 HCD 26eV P=554.6
4	human_hcd_selected_2016_best	114	141	0.94	TIEELEKEMLNGQK/3 0 HCD 28eV P=554.6
5	human_hcd_selected_2016_best	105	131	0.68	EPLAAVYQEMLAQAK/3 0 HCD 27eV P=554.6
6	human_hcd_selected_2016_best	86	107	0.33	QSLQVIESAMERDR/3 0 HCD 28eV P=554.6
7	human_hcd_selected_2016_best	82	99	0.28	KSDGKIDVSIEAASGGK/3 0 HCD 26eV P=554.6
8	human_hcd_selected_2016_best	79	100	0.24	DKECGQLLISENQK/3 1/3,C,CAM HCD 27eV P=554.6
9	human_hcd_selected_2016_best	76	95	0.22	LTETKDGSGLEEKVR/3 0 HCD 30eV P=554.6
10	human_hcd_selected_2016_best	63	80	0.14	GAVEAELDPVEYTLR/3 0 HCD 27eV P=554.6
11	human_hcd_selected_2016_best	60	76	0.12	AELGKDLQEPEFSAK/3 0 HCD 28eV P=554.6
12	human_hcd_selected_2016_best	56	71	0.10	FPSHFSSDLKDLLR/3 0 HCD 30eV P=554.6
13	human_hcd_selected_2016_best	56	70	0.10	GAGTDEGCLIEILASR/3 1/7,C,CAM HCD 25eV P=554.6
14	human_hcd_selected_2016_best	55	71	0.10	PTSTNPIASIFAWTR/3 0 HCD 39eV P=554.6
15	human_hcd_selected_2016_best	52	65	0.09	LLKPQDLEMQFQR/3 1/8,M,Oxidation HCD 31eV P=554.6
16	human_hcd_selected_2016_best	38	49	0.05	VESTDVSDLLHQYR/3 0 HCD 26eV P=554.6
17	human_hcd_selected_2016_best	38	48	0.05	FEEVPELLTEAEKK/3 0 HCD 30eV P=554.6
18	human_hcd_selected_2016_best	35	45	0.05	GQSEDPGSLLSLFRR/3 0 HCD 33eV P=554.6
19	human_hcd_selected_2016_best	35	44	0.05	KVEEAEPEEFVVEK/3 0 HCD 26eV P=554.6
20	human_hcd_selected_2016_best	31	37	0.04	APWAEAEHSALNPLR/3 0 HCD 30eV P=554.6
21	human_hcd_selected_2016_best	30	38	0.04	LEDEIDFLAQELAR/3 0 HCD 24.5142917633057eV P=554.6
22	human_hcd_selected_2016_best	29	37	0.03	LLDMGETDLMLAALR/3 0 HCD 27eV P=554.6
23	human_hcd_selected_2016_best	28	36	0.03	AVSEAQEKLTESNQK/3 0 HCD 28eV P=554.6
24	human_hcd_selected_2016_best	26	32	0.03	EIQGFFNIPVDNLR/3 0 HCD 27eV P=554.6

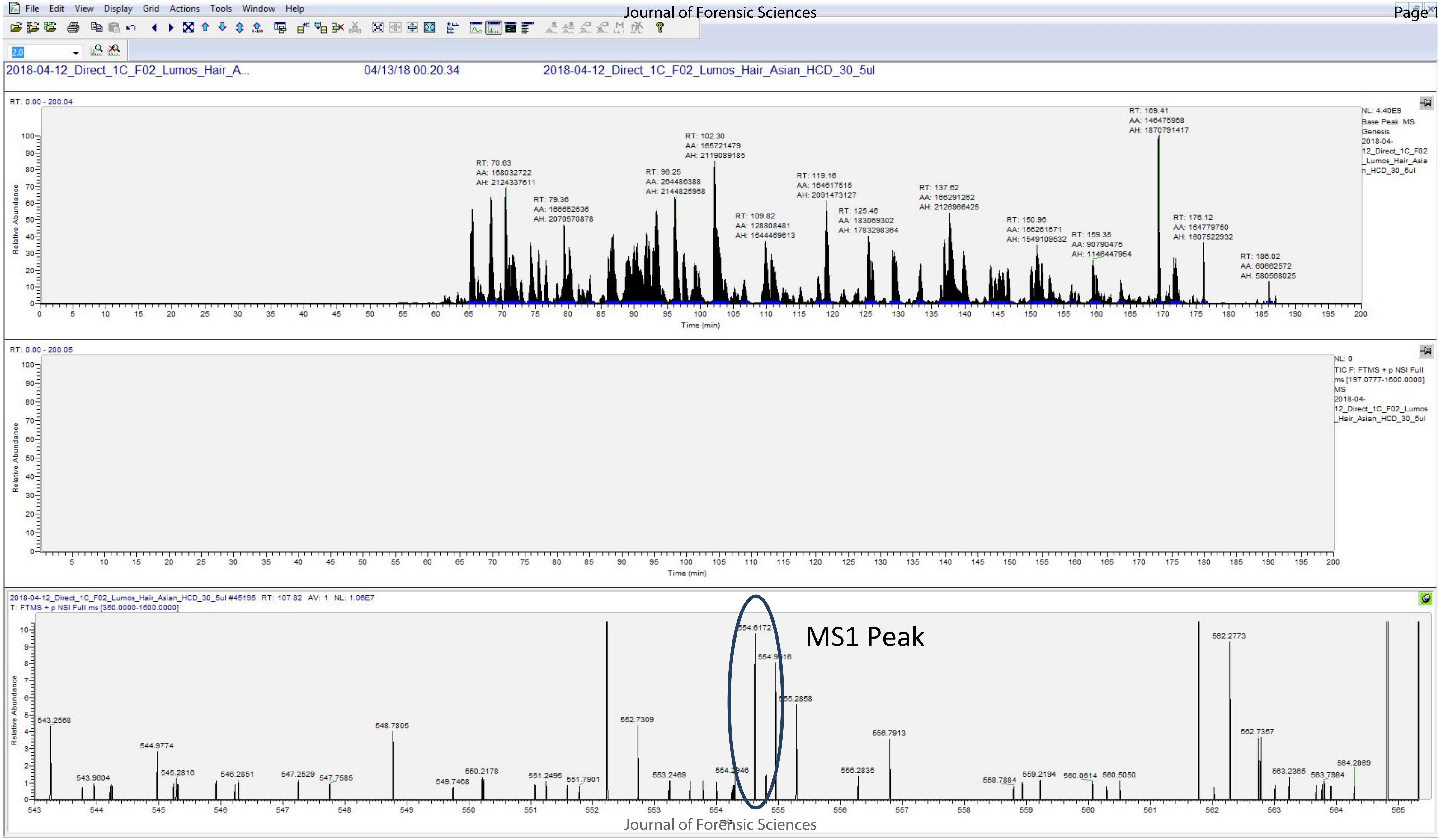
(human_hair_selected_with_gvps_passed) SEADSDKNATILELR/3_0 30eV

Name: SEADSDKNATILELR/3_0 30eV
MW: 1663 Exact Mass: 1663.8486 ID#: 4511 DB: human_hair_selected_with_gvps_passed
Comment: Single Pep=Tryptic Mods=0 Fullname=R.SEADSDKNATILELR.S/3 Charge=3 Parent=554.6162 Mz_diff=1.8ppm HCD=30.00% Scan=63070 Origfile="2018-03-23_1A_F01_Lumos_Hair_Asian_HCD_30_5ul.raw.FT.hcd.ch.MGF" Sample="Asian_Male_SDS" Protein="sp|P15924|DESP_HUMAN(pre=R,post=S)" Unassign_all=0.3805 Unassigned=0.1536 max_unassigned_ab=0.62 num_unassigned_peaks=4/20 FTResolution=30000 ms2IsolationWidth=2.00 ms1PrecursorAb=4844567.55 Precursor1MaxAb=5567626.23 Filter="FTMS + p NSI d Full ms2 554.6167@hcd30.00 [100.0000-1674.0000]"
10 largest peaks:

```
 530.3303  999.00 |  723.3842  919.68 |  189.0872  892.13 |  217.0822  782.49 |  723.8856  633.32 |
 554.2798  622.14 |  643.4145  537.80 |  288.2034  457.89 |  744.4622  418.20 |  929.5424  393.36 |
```

297 m/z Values and Intensities:

```
101.1992   5.86    ?
102.0551   41.16   IEA/1.4ppm
105.0659   7.63    ?
107.3272   6.12    ?
109.6245   5.81    ?
112.0759   37.75   IKB/1.9ppm
112.0871   7.71    IRG/1.6ppm
113.071    17.38   ?
113.0792   7.07    IKB+i/3.8ppm
115.087    10.59   ?
116.071    6.99    IRI/3.4ppm
120.0809   22.49   ?
122.0812   11.79   ?
124.1124   17.64   ?
126.0551   149.91  ?
128.0814   7.49    ?
129.0382   6.60    ?
129.066    7.14    ?
129.1025   201.01  IKD/2.0ppm
130.0501   11.91   ?
130.0865   12.41   ?
130.1059   10.20   IKD+i/6.7ppm
130.3891   6.71    ?
```



2018-04-12_Direct_1C_F02_Lumos_Hair_A... 04/13/18 00:20:34 2018-04-12_Direct_1C_F02_Lumos_Hair_Asian_HCD_30_5ul

RT: 0.00 - 200.04

NL: 4.40E9
Base Peak MS
Genesis
2018-04-12_Direct_1C_F02_Lumos_Hair_Asian_HCD_30_5ul

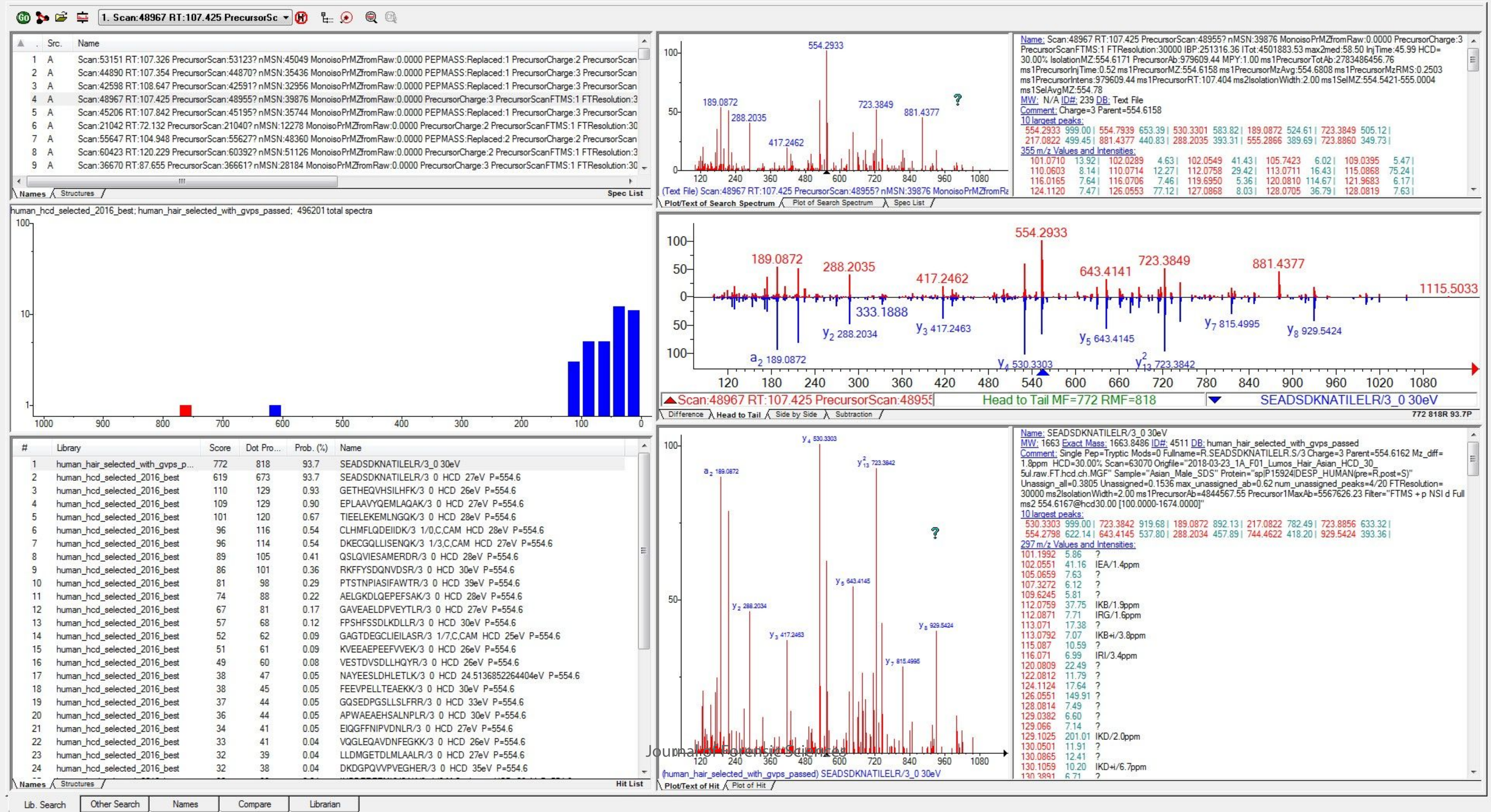
RT: 0.00 - 200.05

NL: 0
TIC F: FTMS + p NSI Full ms [197.0777-1600.0000] MS
2018-04-12_Direct_1C_F02_Lumos_Hair_Asian_HCD_30_5ul

2018-04-12_Direct_1C_F02_Lumos_Hair_Asian_HCD_30_5ul #45195 RT: 107.82 AV: 1 NL: 1.06E7
T: FTMS + p NSI Full ms [350.0000-1600.0000]

MS1 Peak

DSP_R1738Q_Q: (R)SEADSDKNATILELR in Fraction 3 confirmed



▲	.	Src.	Name
	1	A	Scan:53151 RT:107.326 PrecursorScan:53123? nMSN:45049 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:2 PrecursorScan
	2	A	Scan:44890 RT:107.354 PrecursorScan:44870? nMSN:35436 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
	3	A	Scan:42598 RT:108.647 PrecursorScan:42591? nMSN:32956 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
	4	A	Scan:48967 RT:107.425 PrecursorScan:48955? nMSN:39876 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:3
	5	A	Scan:45206 RT:107.842 PrecursorScan:45195? nMSN:35744 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:1 PrecursorCharge:3 PrecursorScan
	6	A	Scan:21042 RT:72.132 PrecursorScan:21040? nMSN:12278 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30
	7	A	Scan:55647 RT:104.948 PrecursorScan:55627? nMSN:48360 MonoisoPrMZfromRaw:0.0000 PEPMASS:Replaced:2 PrecursorCharge:2 PrecursorScan
	8	A	Scan:60423 RT:120.229 PrecursorScan:60392? nMSN:51126 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:3
	9	A	Scan:36670 RT:87.655 PrecursorScan:36661? nMSN:28184 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30

(Text File) Scan:48967 RT:107.425 PrecursorScan:48955? nMSN:39876 MonoisoPrMZfromRa

Name: Scan:48967 RT:107.425 PrecursorScan:48955? nMSN:39876 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30000 IBP:251316.36 ITot:4501883.53 max2med:58.50 InjTime:45.99 HCD=30.00% IsolationMZ:554.6171 PrecursorAb:979609.44 MPY:1.00 ms1PrecursorTotAb:2783486456.76 ms1PrecursorInjTime:0.52 ms1PrecursorMZ:554.6158 ms1PrecursorMzAvg:554.6808 ms1PrecursorMzRMS:0.2503 ms1PrecursorIntens:979609.44 ms1PrecursorRT:107.404 ms2IsolationWidth:2.00 ms1SelMZ:554.5421-555.0004 ms1SelAvgMZ:554.78
MW: N/A ID#: 239 DB: Text File
Comment: Charge=3 Parent=554.6158
10 largest peaks:

```
 554.2933  999.00 |  554.7939  653.39 |  530.3301  583.82 |  189.0872  524.61 |  723.3849  505.12 |
 217.0822  499.45 |  881.4377  440.83 |  288.2035  393.31 |  555.2866  389.69 |  723.8860  349.73 |
```

355 m/z Values and Intensities:

```
  101.0710   13.92 |  102.0289    4.63 |  102.0549   41.43 |  105.7423    6.02 |  109.0395    5.47 |
  110.0603    8.14 |  110.0714   12.27 |  112.0758   29.42 |  113.0711   16.43 |  115.0868   75.24 |
  116.0165    7.64 |  116.0706    7.46 |  119.6950    5.36 |  120.0810  114.67 |  121.9683    6.17 |
  124.1120    7.47 |  126.0553   77.12 |  127.0868    8.03 |  128.0705   36.79 |  128.0819    7.63 |
```

human_hcd_selected_2016_best; human_hair_selected_with_gvps_passed; 496201 total spectra

#	Library	Score	Dot Pro...	Prob. (%)	Name
1	human_hair_selected_with_gvps_p...	772	818	93.7	SEADSDKNATILELR/3_0 30eV
2	human_hcd_selected_2016_best	619	673	93.7	SEADSDKNATILELR/3 0 HCD 27eV P=554.6
3	human_hcd_selected_2016_best	110	129	0.93	GETHEQVHSILHFK/3 0 HCD 26eV P=554.6
4	human_hcd_selected_2016_best	109	129	0.90	EPLAAVYQEMLAQAK/3 0 HCD 27eV P=554.6
5	human_hcd_selected_2016_best	101	120	0.67	TIEELEKEMLNGQK/3 0 HCD 28eV P=554.6
6	human_hcd_selected_2016_best	96	116	0.54	CLHMFLQDEIIDK/3 1/0,C,CAM HCD 28eV P=554.6
7	human_hcd_selected_2016_best	96	114	0.54	DKECGQLLISENQK/3 1/3,C,CAM HCD 27eV P=554.6
8	human_hcd_selected_2016_best	89	105	0.41	QSLQVIESAMERDR/3 0 HCD 28eV P=554.6
9	human_hcd_selected_2016_best	86	101	0.36	RKFFYSDQNVDSR/3 0 HCD 30eV P=554.6
10	human_hcd_selected_2016_best	81	98	0.29	PTSTNPIASIFAWTR/3 0 HCD 39eV P=554.6
11	human_hcd_selected_2016_best	74	88	0.22	AELGKDLQEPEFSAK/3 0 HCD 28eV P=554.6
12	human_hcd_selected_2016_best	67	81	0.17	GAVEAELDPVEYTLR/3 0 HCD 27eV P=554.6
13	human_hcd_selected_2016_best	57	68	0.12	FPSHFSSDLKDLLR/3 0 HCD 30eV P=554.6
14	human_hcd_selected_2016_best	52	62	0.09	GAGTDEGCLIEILASR/3 1/7,C,CAM HCD 25eV P=554.6
15	human_hcd_selected_2016_best	51	61	0.09	KVEEAEPEEFVVEK/3 0 HCD 26eV P=554.6
16	human_hcd_selected_2016_best	49	60	0.08	VESTDVSDLLHQYR/3 0 HCD 26eV P=554.6
17	human_hcd_selected_2016_best	38	47	0.05	NAYEESLDHLETLK/3 0 HCD 24.5136852264404eV P=554.6
18	human_hcd_selected_2016_best	38	45	0.05	FEEVPELLTEAEKK/3 0 HCD 30eV P=554.6
19	human_hcd_selected_2016_best	37	44	0.05	GQSEDPGSLLSLFRR/3 0 HCD 33eV P=554.6
20	human_hcd_selected_2016_best	36	44	0.05	APWAEAEHSALNPLR/3 0 HCD 30eV P=554.6
21	human_hcd_selected_2016_best	34	41	0.05	EIQGFFNIPVDNLR/3 0 HCD 27eV P=554.6
22	human_hcd_selected_2016_best	33	41	0.04	VQGLEQAVDNFEGKK/3 0 HCD 26eV P=554.6
23	human_hcd_selected_2016_best	32	39	0.04	LLDMGETDLMLAALR/3 0 HCD 27eV P=554.6
24	human_hcd_selected_2016_best	32	38	0.04	DKDGPQVVPVEGHER/3 0 HCD 35eV P=554.6

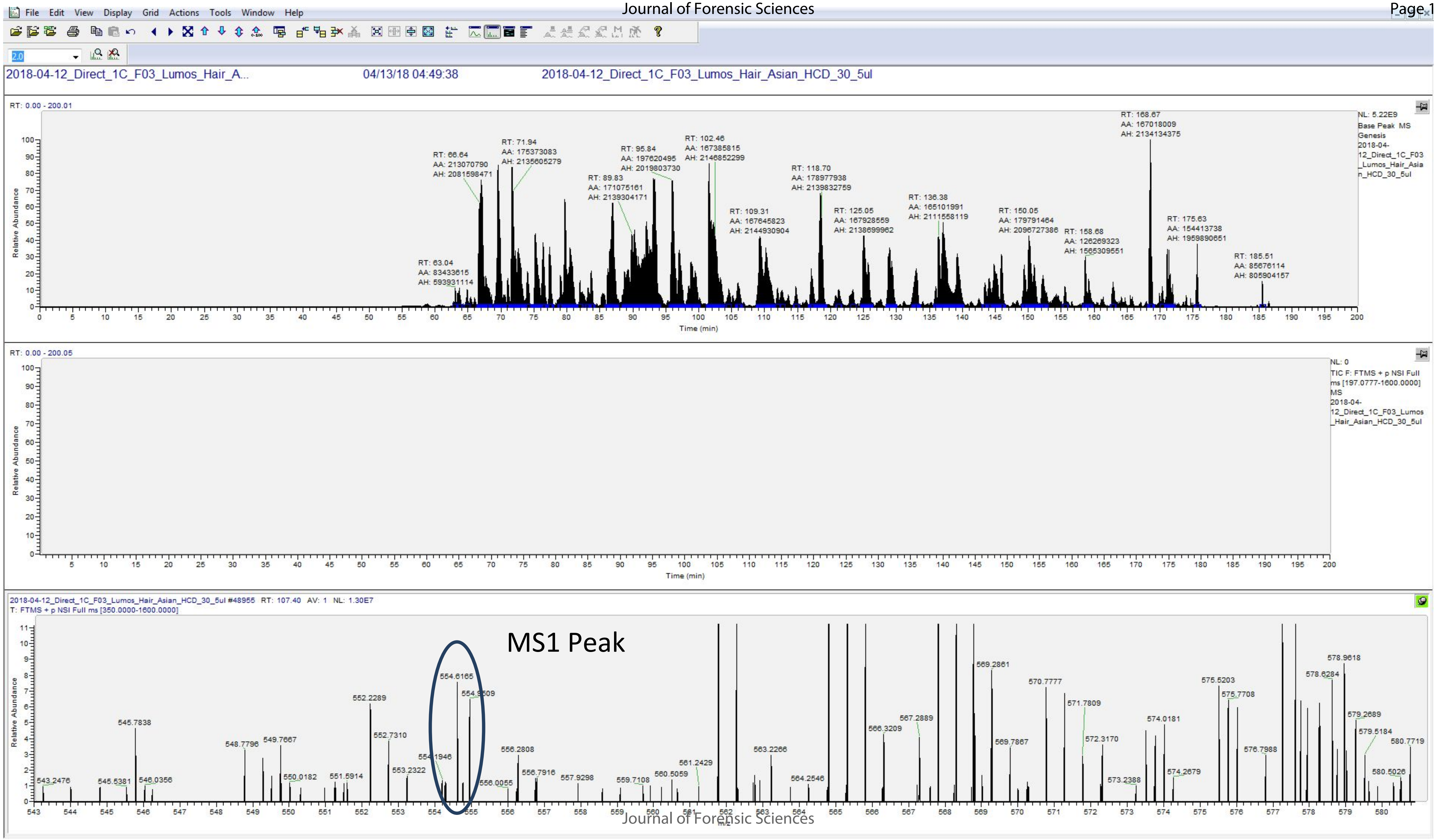
(human_hair_selected_with_gvps_passed) SEADSDKNATILELR/3_0 30eV

Name: SEADSDKNATILELR/3_0 30eV
MW: 1663 Exact Mass: 1663.8486 ID#: 4511 DB: human_hair_selected_with_gvps_passed
Comment: Single Pep=Tryptic Mods=0 Fullname=R.SEADSDKNATILELR.S/3 Charge=3 Parent=554.6162 Mz_diff=1.8ppm HCD=30.00% Scan=63070 Origfile="2018-03-23_1A_F01_Lumos_Hair_Asian_HCD_30_5ul.raw.FT.hcd.ch.MGF" Sample="Asian_Male_SDS" Protein="sp|P15924|DESP_HUMAN(pre=R,post=S)" Unassign_all=0.3805 Unassigned=0.1536 max_unassigned_ab=0.62 num_unassigned_peaks=4/20 FTResolution=30000 ms2IsolationWidth=2.00 ms1PrecursorAb=4844567.55 Precursor1MaxAb=5567626.23 Filter="FTMS + p NSI d Full ms2 554.6167@hcd30.00 [100.0000-1674.0000]"
10 largest peaks:

```
 530.3303  999.00 |  723.3842  919.68 |  189.0872  892.13 |  217.0822  782.49 |  723.8856  633.32 |
 554.2798  622.14 |  643.4145  537.80 |  288.2034  457.89 |  744.4622  418.20 |  929.5424  393.36 |
```

297 m/z Values and Intensities:

```
101.1992   5.86    ?
102.0551   41.16   IEA/1.4ppm
105.0659   7.63    ?
107.3272   6.12    ?
109.6245   5.81    ?
112.0759   37.75   IKB/1.9ppm
112.0871   7.71    IRG/1.6ppm
113.071    17.38   ?
113.0792   7.07    IKB+i/3.8ppm
115.087    10.59   ?
116.071    6.99    IRI/3.4ppm
120.0809   22.49   ?
122.0812   11.79   ?
124.1124   17.64   ?
126.0551   149.91  ?
128.0814   7.49    ?
129.0382   6.60    ?
129.066    7.14    ?
129.1025   201.01  IKD/2.0ppm
130.0501   11.91   ?
130.0865   12.41   ?
130.1059   10.20   IKD+i/6.7ppm
130.3891   6.71    ?
```


File Edit View Display Grid Actions Tools Window Help
2.0

2018-04-12_Direct_1C_F03_Lumos_Hair_A... 04/13/18 04:49:38 2018-04-12_Direct_1C_F03_Lumos_Hair_Asian_HCD_30_5ul

RT: 0.00 - 200.01

NL: 5.22E9 Base Peak MS Genesis 2018-04-12_Direct_1C_F03_Lumos_Hair_Asian_HCD_30_5ul

RT: 63.04 AA: 83433615 AH: 593931114
RT: 66.64 AA: 213070790 AH: 2081598471
RT: 71.94 AA: 175373083 AH: 2135605279
RT: 89.83 AA: 171075161 AH: 2139304171
RT: 95.84 AA: 197620495 AH: 2019803730
RT: 102.46 AA: 167385815 AH: 2146852299
RT: 109.31 AA: 167645823 AH: 2144930904
RT: 118.70 AA: 178977938 AH: 2139832759
RT: 125.05 AA: 167928559 AH: 2138699962
RT: 136.38 AA: 165101991 AH: 2111558119
RT: 150.05 AA: 179791464 AH: 2096727386
RT: 158.68 AA: 126269323 AH: 1565309551
RT: 168.67 AA: 167018009 AH: 2134134375
RT: 175.63 AA: 154413738 AH: 1959890651
RT: 185.51 AA: 85676114 AH: 805904157
Relative Abundance
0 10 20 30 40 50 60 70 80 90 100
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200
Time (min)

RT: 0.00 - 200.05

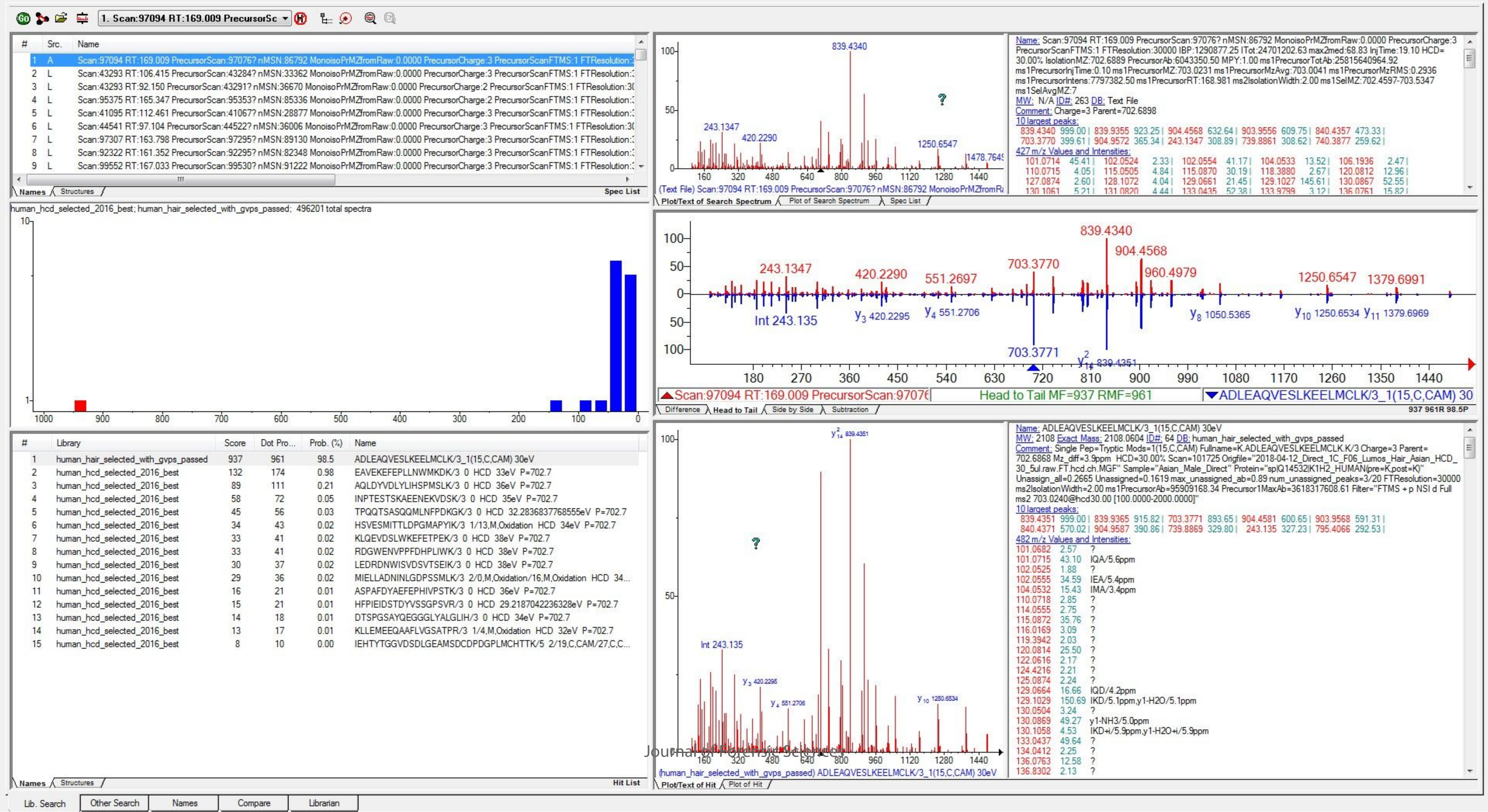
NL: 0 TIC F: FTMS + p NSI Full ms [197.0777-1600.0000] MS 2018-04-12_Direct_1C_F03_Lumos_Hair_Asian_HCD_30_5ul

Relative Abundance
0 10 20 30 40 50 60 70 80 90 100
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200
Time (min)

2018-04-12_Direct_1C_F03_Lumos_Hair_Asian_HCD_30_5ul #48955 RT: 107.40 AV: 1 NL: 1.30E7
T: FTMS + p NSI Full ms [350.0000-1600.0000]

MS1 Peak

KRT32_S222Y_S: ADLEAQVE(S)LKEELMCLK in Fraction 7 confirmed



#	Src.	Name
1	A	Scan:97094 RT:169.009 PrecursorScan:97076? nMSN:86792 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:
2	L	Scan:43293 RT:106.415 PrecursorScan:43284? nMSN:33362 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:
3	L	Scan:43293 RT:92.150 PrecursorScan:43291? nMSN:36670 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:30
4	L	Scan:95375 RT:165.347 PrecursorScan:95353? nMSN:85336 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:2 PrecursorScanFTMS:1 FTResolution:
5	L	Scan:41095 RT:112.461 PrecursorScan:41067? nMSN:28877 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:
6	L	Scan:44541 RT:97.104 PrecursorScan:44522? nMSN:36006 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30
7	L	Scan:97307 RT:163.798 PrecursorScan:97295? nMSN:89130 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:
8	L	Scan:92322 RT:161.352 PrecursorScan:92295? nMSN:82348 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:
9	L	Scan:99552 RT:167.033 PrecursorScan:99530? nMSN:91222 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:

Name: Scan:97094 RT:169.009 PrecursorScan:97076? nMSN:86792 MonoisoPrMZfromRaw:0.0000 PrecursorCharge:3 PrecursorScanFTMS:1 FTResolution:30000 IBP:1290877.25 ITot:24701202.63 max2med:68.83 InjTime:19.10 HCD=30.00% IsolationMZ:702.6889 PrecursorAb:6043350.50 MPY:1.00 ms1PrecursorTotAb:25815640964.92 ms1PrecursorInjTime:0.10 ms1PrecursorMZ:703.0231 ms1PrecursorMzAvg:703.0041 ms1PrecursorMzRMS:0.2936 ms1PrecursorIntens:7797382.50 ms1PrecursorRT:168.981 ms2IsolationWidth:2.00 ms1SelMZ:702.4597-703.5347 ms1SelAvgMZ:7
MW: N/A ID#: 263 DB: Text File
Comment: Charge=3 Parent=702.6898
10 largest peaks:

```
 839.4340  999.00 | 839.9355  923.25 | 904.4568  632.64 | 903.9556  609.75 | 840.4357  473.33 |
 703.3770  399.61 | 904.9572  365.34 | 243.1347  308.89 | 739.8861  308.62 | 740.3877  259.62 |
```

427 m/z Values and Intensities:

```
  101.0714   45.41 |  102.0524    2.33 |  102.0554   41.17 |  104.0533   13.52 |  106.1936    2.47 |
  110.0715    4.05 |  115.0505    4.84 |  115.0870   30.19 |  118.3880    2.67 |  120.0812   12.96 |
  127.0874    2.60 |  128.1072    4.04 |  129.0661   21.45 |  129.1027  145.61 |  130.0867   52.55 |
  130.1061    5.21 |  131.0820    4.44 |  133.0435   52.38 |  133.9799    3.12 |  136.0761   15.82 |
```

human_hcd_selected_2016_best; human_hair_selected_with_gvps_passed; 496201 total spectra

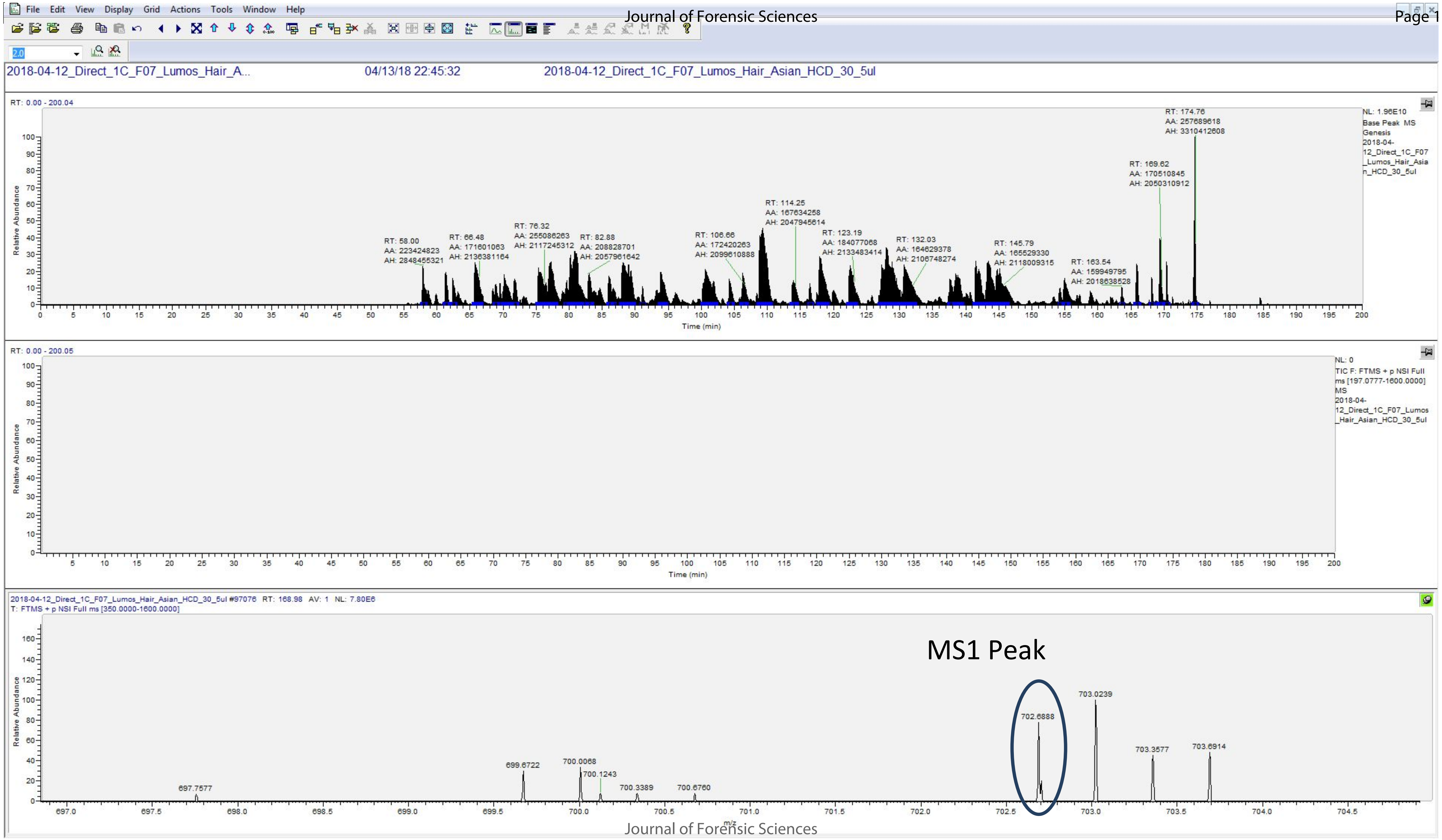
#	Library	Score	Dot Pro...	Prob. (%)	Name
1	human_hair_selected_with_gvps_passed	937	961	98.5	ADLEAQVESLKEELMCLK/3_1(15,C,CAM) 30eV
2	human_hcd_selected_2016_best	132	174	0.98	EAVEKEFEPLLNWMKDK/3 0 HCD 33eV P=702.7
3	human_hcd_selected_2016_best	89	111	0.21	AQLDYVDLYLIHSPMSLK/3 0 HCD 36eV P=702.7
4	human_hcd_selected_2016_best	58	72	0.05	INPTESTSKAEENEKVDSK/3 0 HCD 35eV P=702.7
5	human_hcd_selected_2016_best	45	56	0.03	TPQQTSASQQMLNFPDKGK/3 0 HCD 32.2836837768555eV P=702.7
6	human_hcd_selected_2016_best	34	43	0.02	HSVESMITTLDPGMAPYIK/3 1/13,M,Oxidation HCD 34eV P=702.7
7	human_hcd_selected_2016_best	33	41	0.02	KLQEVDSLWKEFETPEK/3 0 HCD 38eV P=702.7
8	human_hcd_selected_2016_best	33	41	0.02	RDGWENVPPFDHPLIWK/3 0 HCD 38eV P=702.7
9	human_hcd_selected_2016_best	30	37	0.02	LEDRDNWISVDSVTSEIK/3 0 HCD 38eV P=702.7
10	human_hcd_selected_2016_best	29	36	0.02	MIELLADNINLGDPSSMLK/3 2/0,M,Oxidation/16,M,Oxidation HCD 34...
11	human_hcd_selected_2016_best	16	21	0.01	ASPAFDYAEFEPHIVPSTK/3 0 HCD 36eV P=702.7
12	human_hcd_selected_2016_best	15	21	0.01	HFPIEIDSTDYVSSGPSVR/3 0 HCD 29.2187042236328eV P=702.7
13	human_hcd_selected_2016_best	14	18	0.01	DTSPGSAYQEGGGLYALGLIH/3 0 HCD 34eV P=702.7
14	human_hcd_selected_2016_best	13	17	0.01	KLLEMEEQAAFLVGSATPR/3 1/4,M,Oxidation HCD 32eV P=702.7
15	human_hcd_selected_2016_best	8	10	0.00	IEHTYTGGVDSDLGEAMSDCDPDGPLMCHTTK/5 2/19,C,CAM/27,C,C...

Name: ADLEAQVESLKEELMCLK/3_1(15,C,CAM) 30eV
MW: 2108 Exact Mass: 2108.0604 ID#: 64 DB: human_hair_selected_with_gvps_passed
Comment: Single Pep=Tryptic Mods=1(15,C,CAM) Fullname=K.ADLEAQVESLKEELMCLK.K/3 Charge=3 Parent=702.6868 Mz_diff=3.9ppm HCD=30.00% Scan=101725 Origfile="2018-04-12_Direct_1C_F06_Lumos_Hair_Asian_HCD_30_5ul.raw.FT.hcd.ch.MGF" Sample="Asian_Male_Direct" Protein="sp|Q14532|K1H2_HUMAN(pre=K,post=K)" Unassign_all=0.2665 Unassigned=0.1619 max_unassigned_ab=0.89 num_unassigned_peaks=3/20 FTResolution=30000 ms2IsolationWidth=2.00 ms1PrecursorAb=95909168.34 Precursor1MaxAb=3618317608.61 Filter="FTMS + p NSI d Full ms2 703.0240@hcd30.00 [100.0000-2000.0000]"
10 largest peaks:

```
 839.4351  999.00 | 839.9365  915.82 | 703.3771  893.65 | 904.4581  600.65 | 903.9568  591.31 |
 840.4371  570.02 | 904.9587  390.86 | 739.8869  329.80 |  243.135  327.23 | 795.4066  292.53 |
```

482 m/z Values and Intensities:

```
101.0682   2.57    ?
101.0715   43.10   IQA/5.6ppm
102.0525   1.88    ?
102.0555   34.59   IEA/5.4ppm
104.0532   15.43   IMA/3.4ppm
110.0718   2.85    ?
114.0555   2.75    ?
115.0872   35.76   ?
116.0169   3.09    ?
119.3942   2.03    ?
120.0814   25.50   ?
122.0616   2.17    ?
124.4216   2.21    ?
125.0874   2.24    ?
129.0664   16.66   IQD/4.2ppm
129.1029   150.69  IKD/5.1ppm,y1-H2O/5.1ppm
130.0504   3.24    ?
130.0869   49.27   y1-NH3/5.0ppm
130.1058   4.53    IKD+i/5.9ppm,y1-H2O+i/5.9ppm
133.0437   49.64   ?
134.0412   2.25    ?
136.0763   12.58   ?
136.8302   2.13    ?
```


File Edit View Display Grid Actions Tools Window Help
2.0
RT: 0.00 - 200.04
Relative Abundance
Time (min)
RT: 58.00 AA: 223424823 AH: 2848455321
RT: 66.48 AA: 171601063 AH: 2136381164
RT: 76.32 AA: 255086263 AH: 2117245312
RT: 82.88 AA: 208828701 AH: 2057961642
RT: 106.66 AA: 172420263 AH: 2099610888
RT: 114.25 AA: 167634258 AH: 2047945614
RT: 123.19 AA: 184077068 AH: 2133483414
RT: 132.03 AA: 164629378 AH: 2106748274
RT: 145.79 AA: 165529330 AH: 2118009315
RT: 163.54 AA: 159949795 AH: 2018638528
RT: 169.62 AA: 170510845 AH: 2050310912
RT: 174.76 AA: 257689618 AH: 3310412608
RT: 0.00 - 200.05
Relative Abundance
Time (min)
Relative Abundance
697.7577
699.6722
700.0068
700.1243
700.3389
700.6760
702.6888
703.0239
703.3577
703.6914
m/z

2018-04-12_Direct_1C_F07_Lumos_Hair_A... 04/13/18 22:45:32 2018-04-12_Direct_1C_F07_Lumos_Hair_Asian_HCD_30_5ul

NL: 1.96E10
Base Peak MS
Genesis
2018-04-12_Direct_1C_F07_Lumos_Hair_Asian_HCD_30_5ul

NL: 0
TIC F: FTMS + p NSI Full ms [197.0777-1600.0000]
MS
2018-04-12_Direct_1C_F07_Lumos_Hair_Asian_HCD_30_5ul

2018-04-12_Direct_1C_F07_Lumos_Hair_Asian_HCD_30_5ul #97076 RT: 168.98 AV: 1 NL: 7.80E6
T: FTMS + p NSI Full ms [350.0000-1600.0000]

MS1 Peak

GVP sites are summarized from all 10 fractions:

Asian_1hair_5cm	DSP	GSDMA	KRT31	KRT32	KRT33A	KRT33B	KRT35	KRT35	KRT81	KRT82	KRT83	KRT83	KRTAP10-8	TGM3
	R1738Q_Q	V128L_L	A82V_V	S222Y_Y	A270V_V	V279L_L	P443A_A	S36P_P	S13R_R	T458M_M	G362S_S	I279M_M	H26R_R	T13K_K
D_LG_F1_TO_F10_R1	X		X	?*	X			X	X		X	X	X	X

*note for "?": It means we cannot confirm its identification at this time with a borderline intensity and lack of MS1 peak. However, some of the major peaks still match well and it showed up from the expected fraction. For such case, we put "hold" to be confirmed.

- Analyses above led to several general findings:
  - High abundance GVP analysis is very convincing –
    - ✓ with its regular non-variant form presenting in all 10 fractions
    - ✓ with convincing nistms_metrics information:
      - ❖ Abundance (log10)
      - ❖ Match Factor (MF)
      - ❖ Retention Time (RT)
  - Low abundance GVP analysis is harder, but confidence can be increased by at least one of the following –
    - ✓ from expected gel bands (based on molecular weight of its protein)
    - ✓ with the presence of its regular non-variant form
    - ✓ with convincing nistms_metrics information:
      - ❖ MF
      - ❖ RT
      - ❖ MS1 Peak

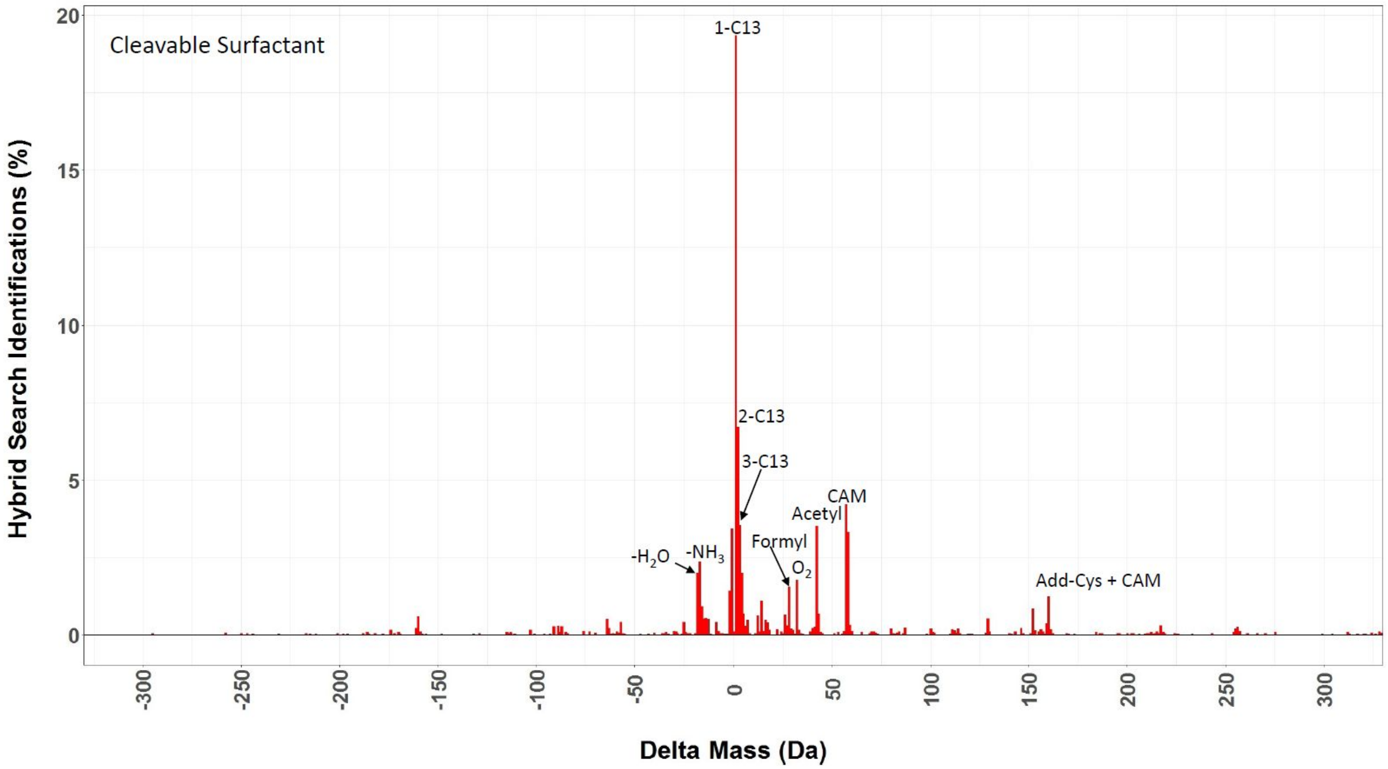
Supporting Information:

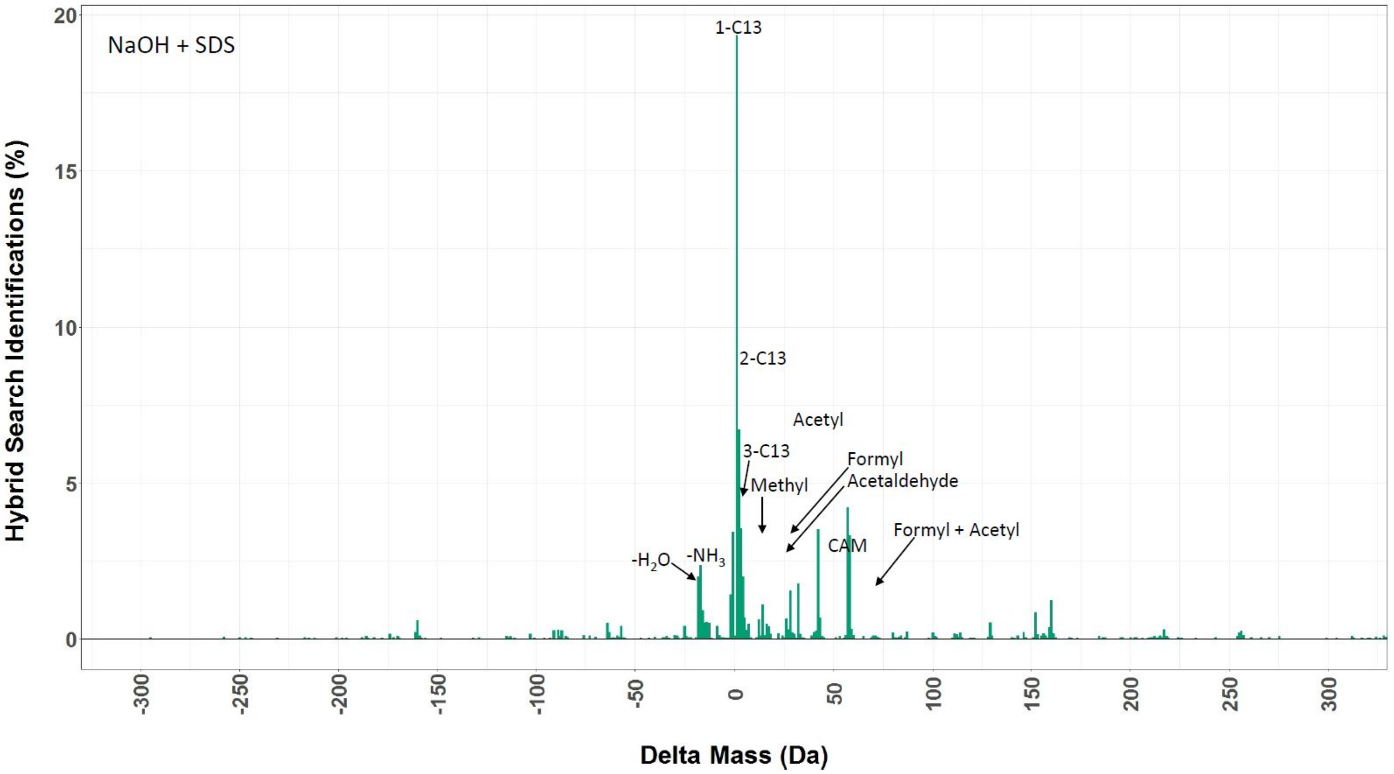
Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Keratin

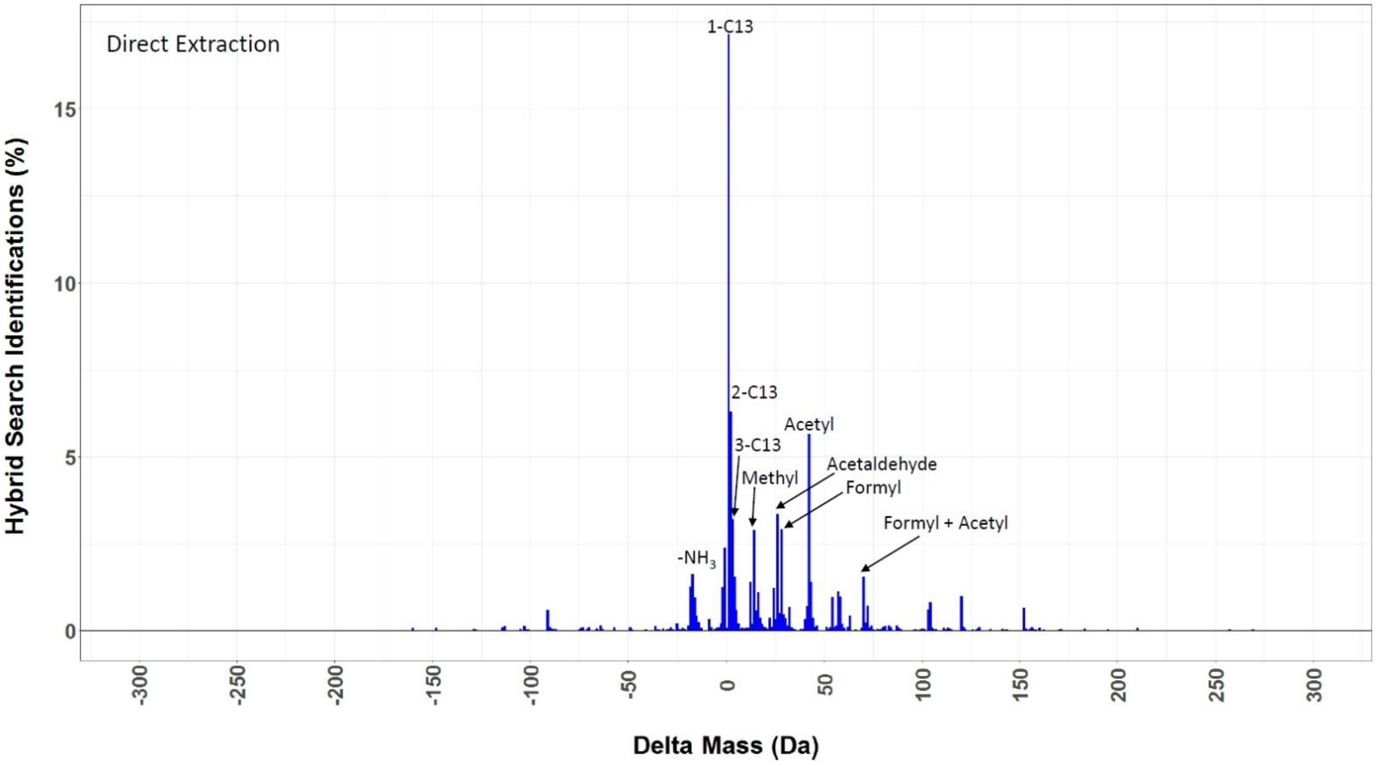
Authors: Zheng Zhang, Meghan C. Burke, William E. Wallace, Yuxue Liang, Sergey L. Sheetlin, Yuri A. Mirokhin, Dmitrii V. Tchekhovskoi, Stephen E. Stein

Mass Spectrometry Data Center, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, United States

Supplemental Document S4




NaOH + SDS
Hybrid Search Identifications (%)
20
15
10
5
0
1-C13
2-C13
3-C13
Acetyl
Methyl
Formyl
Acetaldehyde
Formyl + Acetyl
CAM
$-H_2O$
$-NH_3$
-300
-250
-200
-150
-100
-50
0
50
100
150
200
250
300
Delta Mass (Da)


Direct Extraction
Hybrid Search Identifications (%)
15
10
5
0
1-C13
2-C13
3-C13
Acetyl
Methyl
Acetaldehyde
Formyl
Formyl + Acetyl
$-NH_3$
-300
-250
-200
-150
-100
-50
0
50
100
150
200
250
300
Delta Mass (Da)

Supplementary Document S4. Distribution of DeltaMass values obtained from hybrid search identifications of hair-derived peptides (hair shaft length of 5 cm) extracted by the cleavable surfactant (red), NaOH+SDS (green), and direct (blue) method above a spectral match score threshold of 500. The major labeled peaks in each panel are correspond to those in Table 5.

Reviewer(s)' Comments to Author(s):
Reviewer: 1

Comments to the Author
This manuscript, which contains much valuable information, is improved. However, some fundamental conceptual problems remain. The authors need not defend these inaccuracies, since they are not the central focus of the manuscript, providing they are openly acknowledged. The manuscript should be recast to emphasize what can be done with the gel approach without trying to present this as a general method for hair proteomic analysis. That the manuscript has strong aspects in its present form will not exonerate it from misleading other investigators on this point.

1. The estimate of a maximal 75% yield of protein from hair shafts using their treatment method is welcome. Their observation that an "inability to digest substantial portions of the proteome is common" is well taken, but this is highly method dependent, a take home lesson of the manuscript. Moreover, the statement that "we find no reason to assume that such crosslinked, insoluble material might yield undetected GVPs" is at variance with the literature they cite (ref 7) and appears to be an ignorance is bliss approach. That the crosslinked material is readily digestible with trypsin (90% solubilized) and contains a wealth of identifiable nonkeratin and keratin proteins was reported well over a decade ago. Since, by analyzing only the proteins solubilized from the hair shaft, the authors are focusing on keratins, the title should be modified to "Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Keratin".

Response to reviewer's comment 1: We thank the reviewer's comment. We followed the reviewer's suggestion to change the sentence to "In case 1 and 2, substantial portions of the hair undigested although it is method dependent" on page 9 to make it clearer. We added a sentence "According to reference 7, the insoluble, crosslinked portion has a higher content of non-keratin proteins and may contain additional non-keratin-GVP identifications" on page 12 to clarify this point. We also followed the reviewer's suggestion to modify the title to "Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Keratin" on title page (separated from the main document) and all the related places if the title is mentioned to better indicate that we are mainly focusing on hair keratins in this manuscript.

2. The authors counter the critical opinion above by pointing out that the solubilized proteins appear to contain some cross-linked material. They suggest on this basis that "the insoluble, crosslinked, portion of the hair protein may not contain additional GVP identifications." This supposition is totally unwarranted because the cross-linked material has a much higher content of nonkeratin proteins, some of which are found only there. Other laboratories digesting the entire hair shaft report GVPs in numerous proteins enriched in the insoluble crosslinked fraction.

Response to reviewer's comment 2: We agree with the reviewer's comment and changed the sentence to "the insoluble, crosslinked portion of the hair protein may not contain additional keratin-GVP identifications" on page 12 to clarify it. As mentioned above, we also added a sentence "According to reference 7, the insoluble, crosslinked portion has a higher content of non-keratin proteins and may contain additional non-keratin-GVP identifications" on page 12 to make it clearer.