

Statistically Significant Differences in Composition of Petroleum Crude Oils Revealed by Volcano Plots Generated from Ultrahigh Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectra

Manhoi Hur,^{†,‡} Rebecca L. Ware,[§] Junkoo Park,^{||} Amy M. McKenna,[⊥] Ryan P. Rodgers,^{§,⊥} Basil J. Nikolau,[‡] Eve S. Wurtele,^{†,‡} and Alan G. Marshall^{*,§,⊥}

[†]Center for Metabolic Biology, Iowa State University, 3254 Molecular Biology Building, Ames, Iowa 50011, United States

[‡]Department of Genetics, Development, and Cell Biology, Iowa State University, 2310 Pammel Drive, Ames, Iowa 50011, United States

[§]Department of Chemistry and Biochemistry, Florida State University, 95 Chieftain Way, Tallahassee, Florida 32306, United States

^{||}School of Science and Technology, Georgia Gwinnett College, 1000 University Center Lane, Lawrenceville, Georgia 30043, United States

[⊥]National High Magnetic Field Laboratory, Florida State University, 1800 East Paul Dirac Drive, Tallahassee, Florida 32310, United States

Supporting Information

ABSTRACT: A “volcano” plot provides a visual means for identifying statistically significant differences between two populations. Here, we introduce the volcano plot as a means for simple, visual identification and statistical ranking of compositional differences between petroleum crude oils. Ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry yields the relative abundances of ions in each spectrum that contains up to tens of thousands of elemental compositions ($C_xH_yN_nO_oS_s$). From that data, a volcano plot may be generated by plotting statistical significance (p -value, obtained from t test) versus \log_2 (relative abundance ratio). The volcano plot data may be color-coded to highlight differences in heteroatom class ($N_nO_oS_s$), double bond equivalents (DBE = number of rings plus double bonds to carbon), and/or carbon number. The volcano plot may be used either directly or as a “filter” for including only the most statistically significant differences for data entered into more conventional analyses based on DBE vs carbon number, van Krevelen diagram, and Kendrick mass defect plots. In each case, the volcano plot provides statistically significant criteria, rather than visual grouping.

INTRODUCTION

The need for energy is rapidly growing, as every nation in the world seeks to modernize and industrialize, and most energy sources still derive from fossil fuels. However, the global supply of high-quality and low-sulfur crude oils is decreasing, whereas the proportion of heavy crude oils is increasing. To utilize the heavy crude oil more effectively, technological advances in production and refining are needed, along with improved understanding of the detailed chemical composition of heavy crude oil. Crude oil analysis is rendered difficult by its enormous compositional complexity (e.g., tens of thousands of elemental compositions), requiring ultrahigh mass resolution and mass measurement accuracy to separate the chemical components. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS)¹ offers mass-resolving power higher than 1 400 000 at 500 Da² and has been successfully applied to reveal detailed chemical characteristics of heavy and unconventional crude oils.^{3–5}

Each high-resolution mass spectrum of crude oil routinely yields tens of thousands of elemental compositions, which in turn yield a distribution of heteroatom class species (i.e., molecular formulas with a common $N_nO_oS_s$ composition), Kendrick mass series,⁶ double bond equivalents (DBE = number of rings plus double bonds to carbon) vs carbon

number, and H:C vs O:C ratio (van Krevelen plot^{7,8}). The plots provide means for identifying compositional patterns for many species (e.g., all members of a given heteroatom class) from a single graphical display. Statistical analysis methods such as principal component analysis (PCA)⁹ and correlation analysis¹⁰ can be particularly helpful in highlighting differences and relations between different samples (e.g., oil spills from different sources).¹¹ These graphical and statistical tools are essential to interpret high-resolution mass spectra of oils. Development of additional interpretation methods is very important for further development of oil research.⁵ For an example, tools for statistical and quantitative evaluation of data are needed.^{12,13}

The “volcano plot”, a kind of scatter plot of univariate analysis, has recently been successfully applied to interpret multidimensional metabolomics¹⁴ as well as microarrays and RNA-sequence data.^{15,16} A volcano plot is generated by plotting a measure of statistical significance (e.g., p -value representing the repeatability in the abundance in experimental replicates) determined from statistical tests such as t test or

Received: October 11, 2017

Revised: January 4, 2018

Published: January 22, 2018

ANOVA versus fold-change between samples. Thus, differences in chemical composition between two petroleum samples, even for large FT-ICR MS data sets, may be clearly visualized in a statistically significant way. Here, we demonstrate the applicability of volcano plot for analysis of ultrahigh-resolution FT-ICR mass spectra from petroleum crude oils. As described in detail below, we compare Arab Heavy and National Institute of Standards and Technology (NIST) samples as an example of a common regional difference. In addition, we compare Arab Heavy and Arab Heavy 593 °C as an example of compositional differences between the oil and its distillate residue.

EXPERIMENTAL SECTION

Mass Spectrometry. Three oil samples, Arab Heavy, Arab Heavy 593 °C,¹⁷ and NIST 2717a (residual fuel oil), were analyzed to highlight compositional differences in a single crude oil due to distillation (Arab Heavy) as well as a whole crude (Arab Heavy) and a commercial fuel oil from a different source (NIST 2717a). Each sample was dissolved in HPLC-grade toluene (Sigma-Aldrich Chemical Co., St. Louis, MO) to yield a final concentration of 100–250 µg/mL prior to atmospheric pressure photoionization (APPI) mass spectral analysis. Samples were delivered to the ionization source via a syringe pump at a rate of 50 µL/min into the mass spectrometer under typical APPI conditions. An APPI source (Thermo-Fisher Scientific, San Jose, CA) was coupled to the first pumping stage of a custom-built FT-ICR mass spectrometer (see below) through a custom-built interface.¹⁸ The tube lens was set to 50 V (to minimize ion fragmentation), and the heated metal capillary current was 4.5 A. A Hamilton gas-tight syringe (2.5 mL) and syringe pump were used to deliver the sample (50 µL/min) to the heated vaporizer region (300 °C) of the APPI source, where N₂ sheath gas (50 psi) facilitates nebulization, while the auxiliary port remains plugged. Gas-phase molecules flow out of the heated vaporizer in a confined jet, and photoionization is initiated by a krypton vacuum ultraviolet gas discharge lamp (10 eV photons, 120 nm), where photoionization occurs. Toluene increases the ionization efficiency for nonpolar aromatic compounds through dopant-assisted APPI¹⁹ via charge exchange²⁰ and proton transfer reactions²¹ that occur between ionized toluene molecules and neutral analyte molecules at atmospheric pressure. Protonated ions exhibit half-integer DBE values (DBE = $c - h/2 + n/2 + 1$), calculated from the ion elemental composition, C_cH_hN_nO_oS_s,²² and may thus be distinguished from radical cations with integer DBE values.

Samples were analyzed with a custom-built FT-ICR mass spectrometer²³ equipped with a 9.4 T horizontal 225 mm bore diameter superconducting solenoid magnet operated at room temperature, and a modular ICR data acquisition system (PREDATOR) facilitated instrument control, data acquisition, and data analysis.²⁴ Ions generated at atmospheric pressure were accumulated in an external linear octopole ion trap for 50–1000 ms and transferred by rf-only octopoles (2.0 MHz and 240 V_{p-p} amplitude) to the ICR cell. ICR time-domain transients were collected from a 7 segment open cylindrical cell with capacitively coupled excitation electrodes based on the Tolmachev configuration.^{25,26} Multiple (200) individual time-domain ~6 s transients were coadded, half-Hanning-apodized, zero-filled, and fast Fourier-transformed prior to frequency conversion to mass-to-charge ratio and phase-corrected to yield absorption-mode spectra.²⁷ ICR frequencies were converted to ion masses based on the quadrupolar trapping potential approximation.^{28,29}

Mass Spectral Calibration. Internal calibration of each mass spectrum was achieved by use of a highly abundant homologous C_cH_hS₁⁺ ion series (i.e., S₁ heteroatom class), based on a “walking calibration”.²⁵ IUPAC mass was converted to Kendrick mass

$$\text{Kendrick mass} = \text{IUPAC mass} \times (14/14.01565) \quad (1)$$

to sort compounds that differ in IUPAC mass by 14.015 65 Da (mass of CH₂), corresponding to 14.0000 Kendrick mass difference.⁶ Elemental compositions (C_cH_hN_nO_oS_s, with c unlimited, h unlimited,

$0 \leq n \leq 5$, $0 \leq o \leq 5$, $0 \leq s \leq 4$) were assigned for mass spectral peaks with signal magnitude greater than six times the baseline root-mean-square (rms) noise level by use of PetroOrg[®] software.³⁰ For each elemental composition, the number of rings plus double bonds to carbon, DBE, was calculated from eq 2.²²

$$\text{DBE} = c - h/2 + n/2 + 1 \quad (2)$$

Statistical Analysis and Its Visualization from Volcano Plots.

Statistical t test and p -values for volcano plots were generated from the ALGLIB C# library (ALGLIB project, <http://www.alglib.net/>). Volcano plots were generated by use of custom software, implemented by C/C++ and C# programming language (Microsoft, Cambridge, MA). Isoabundance-contoured DBE vs carbon number⁵ and Kendrick mass defect⁶ quantitative comparison (QC) plots^{12,13} and van Krevelen⁸ QC diagrams were produced by use of custom software.

Treatment of Missing Values by Use of Minimum Limit of Detection (minLOD). When spectra obtained from different samples are compared, some peaks are found only in one sample but not in others because those species are not present or their magnitudes fall below the detection threshold. In either case, statistical analysis cannot be performed on the peaks with missing values. If the missing data are simply omitted, then subsequent statistical analysis may fail to fully reflect the difference between the samples. To solve the problem, each missing peak is assigned a magnitude less than or equal to the detection threshold (minimum limit of detection (minLOD)).^{31,32} However, since use of minLOD can skew the results, the value should be chosen carefully. The same approach has been used in a previous study.¹⁴

Construction and Data Interpretation of Volcano Plots. As noted above, a volcano plot can be generated by plotting statistical significance vs degree of change, fold-change (FC).^{33,34} The degree of change may be the ratio of mass spectral peak magnitudes for two different samples. For the pooled t test analysis,^{35,36} the degree of change is generally reported on a log₂ scale to minimize skewness³⁷ from abundant compounds. If x_i is the magnitude of the i th spectral peak for n_0 repeated measurements of one sample and y_i is the magnitude of the i th spectral peak for n_1 repeated measurements of another sample, then the sample mean values, \bar{x}_i and \bar{y}_i , may be expressed as follows:

$$\bar{x}_i = \frac{1}{n_0} \sum_{j=1}^{n_0} x_{ij} \text{ and } \bar{y}_i = \frac{1}{n_1} \sum_{k=1}^{n_1} y_{ik} \quad (3)$$

Thus, the degree of change of the i th compound, log₂FC _{i} , on a log₂ scale

$$\log_2 \text{FC}_i = \log_2(\bar{y}_i) - \log_2(\bar{x}_i) \text{ or } \log_2 \text{FC}_i = \log_2\left(\frac{\bar{y}_i}{\bar{x}_i}\right) \quad (4)$$

and FC _{i} takes the form

$$\text{FC}_i = 2^{\log_2 \text{FC}_i} \quad (5)$$

For hypothesis testing, null and alternative hypotheses, H_0 and H_a , may be stated in the following form

$$H_0 : \mu_{xi} = \mu_{yi}, H_a : \mu_{xi} \neq \mu_{yi} \quad (6)$$

in which μ_{xi} and μ_{yi} are the population mean of the magnitude of the spectral peak x_i and y_i for repeated measurements of the given sample. In the hypothesis testing, it is always assumed that $\mu_{xi} = \mu_{yi}$, and the sampling distribution is examined on the basis of this assumption. In other words, a hypothesis test is begun by assuming that the equality condition in the null hypothesis is true. Within this sampling distribution, one will determine whether or not a sample exhibits unusual statistics. Here, it is assumed that population standard deviations, σ_{xi} and σ_{yi} , of the magnitude of the spectral peaks, x_i and y_i , have the same value; therefore, the sample variances, s_{xi}^2 and s_{yi}^2 , of the magnitude of the spectral peak x_i and y_i for repeated measurements of the given sample can be pooled to obtain an estimate of the common population variance σ_i^2 of the magnitude of the i th spectral peak. The

pooled estimate of σ_i^2 is denoted by $s_{p,i}^2$ and is a weighted average of s_x^2 and s_y^2 . The standardized test criterion can be then calculated from

$$t_i = \frac{(\bar{y}_i - \bar{x}_i) - (\mu_{\bar{y}_i} - \mu_{\bar{x}_i})}{\sqrt{s_{p,i}^2 \left(\frac{1}{n_1} + \frac{1}{n_0} \right)}} \quad (7)$$

in which $s_{p,i}^2$ is the pooled sample variance for the spectral peaks, x_i and y_i in eq 3, and n_0 and n_1 are the number of x_i and y_i replicate spectra obtained from each sample for the i th spectral peak. Then, the p -value plotted as the ordinate (typically the $-\log_{10}$ of the p -value) is calculated from

$$p\text{-value} = 2P(T_i > |t_i|) \quad (8)$$

in which T_i is a random variable, t_i is the standardized test statistic obtained from eq 7, and P is the probability that t_i will fall in the given interval under t distribution. One can find the probability by calculating the area under the t distribution for the given interval. The hypothesis test in eq 6 is a two-tailed test, and each tail has the same area. Therefore, the p -value in eq 8 is twice the area under the t distribution to the right of the positive standardized test statistic. In the volcano plot shown in Figure 1, a smaller p -value indicates greater statistical significance. In other words, the smaller the p -value of the test, the more evidence there is to reject the null hypothesis. In eq 7, the larger the difference between \bar{x}_i and \bar{y}_i , the larger will be the absolute value of the test statistic t_i . As a result, the p -value will be smaller, and there is a higher probability to reject the null hypothesis. Also, the larger the pooled variance, $s_{p,i}^2$, the smaller the test statistic t_i , resulting in a larger p -value and a higher probability to fail to reject the null hypothesis. As a consequence, a smaller p -value ($p < 0.05$) represents small variation in the abundance of the chemical compounds identified with similar values in experimental replicates, and chemical compounds with the most statistically significant differences are located in the upper part of the volcano plot. The abscissa, $(\pm)\log_2 FC$ (eq 4) represents the degree of change.

RESULTS AND DISCUSSION

Volcano Plot for Various Heteroatom Class Species Abundances. Figure 1 is a volcano plot generated for

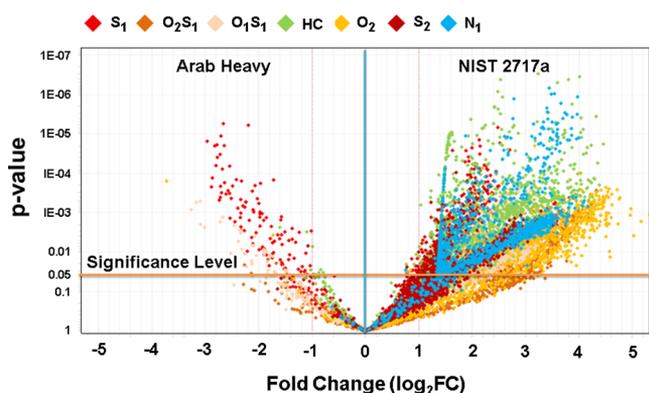


Figure 1. Volcano plot of p -value (\log_2 scale) vs fold-change (positive if peak magnitude is higher for NIST 2717a than for Arab Heavy) generated for members of seven major heteroatom classes in ultrahigh-resolution positive APPI 9.4 T FT-ICR mass spectra of Arab Heavy and NIST 2717a crude oils. Each dot represents an elemental composition, and each heteroatom class is color-coded (see top of Figure).

comparison of members of seven major 7 heteroatom classes (S_1 , O_2S_1 , O_1S_1 , hydrocarbons (HC), O_2 , S_2 , and N_1) from Arab Heavy and NIST 2717a crude oils. Each dot represents an elemental composition and is color-coded according to its chemical class. The central blue vertical line distinguishes the

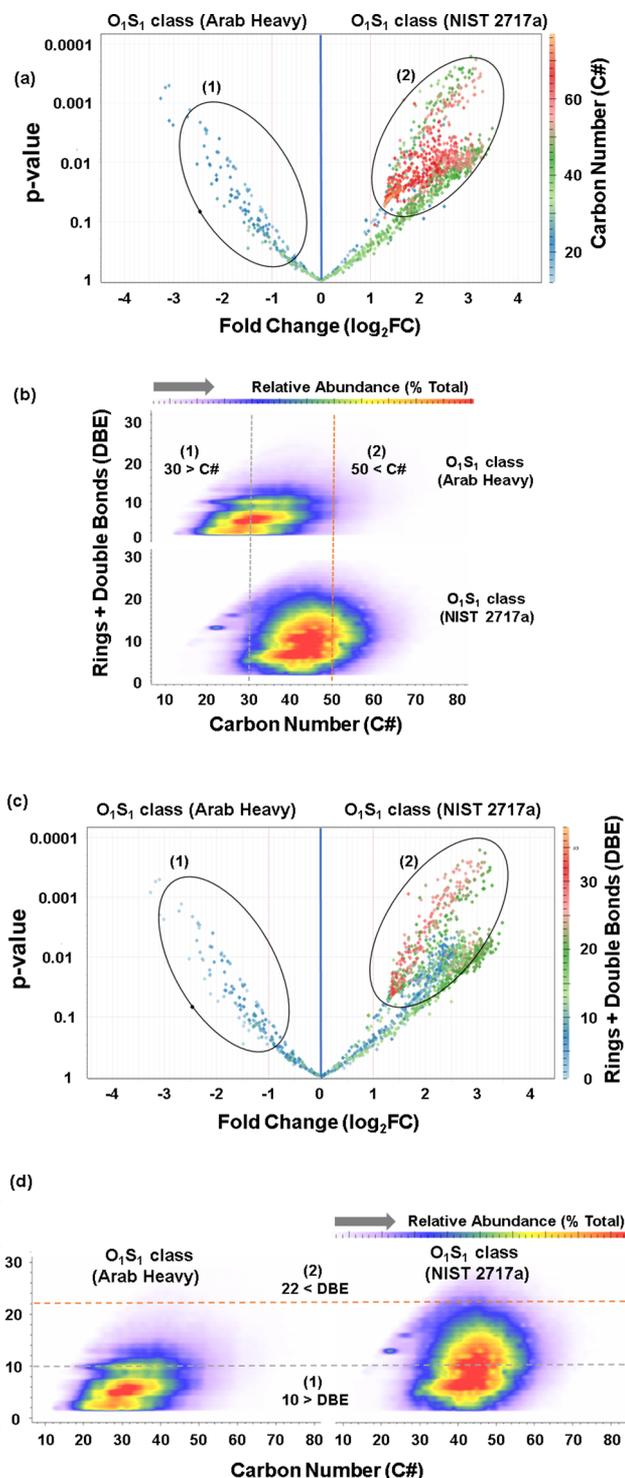


Figure 2. Volcano plots based on carbon number (a) and double bond equivalents (c) for members of the O_1S_1 heteroatom class from Arab Heavy and NIST 2717a crude oils. DBE vs carbon number plots for the same data, showing ranges of overlap for carbon number (b) and DBE value (d). Plots (a) and (c) were color-coded by carbon number and double bond equivalents; (b) and (d) are color-coded by isoabundance (see text).

two samples from each other. The compounds more abundant in Arab Heavy are located to the left of the vertical line, and those more abundant in NIST 2717a are located to the right of the line. The orange horizontal line on the plot shows the significant level (p -value of 0.05).³⁸ Compounds with statisti-

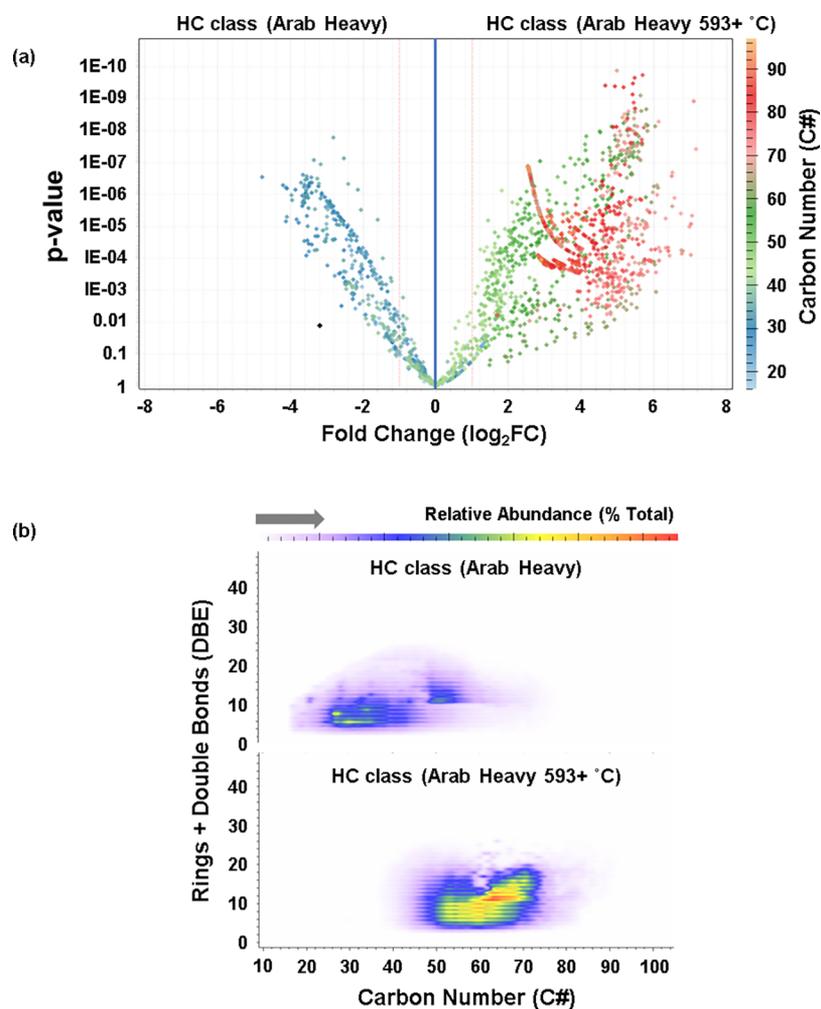


Figure 3. Volcano plot (a) and carbon number vs DBE distribution plots (b) for HC class compounds from positive APPI 9.4 T FT-ICR mass spectra of Arab Heavy and its high-boiling distillate cut. Plots (a) and (b) are color-coded by carbon number and isoabundance (see text).

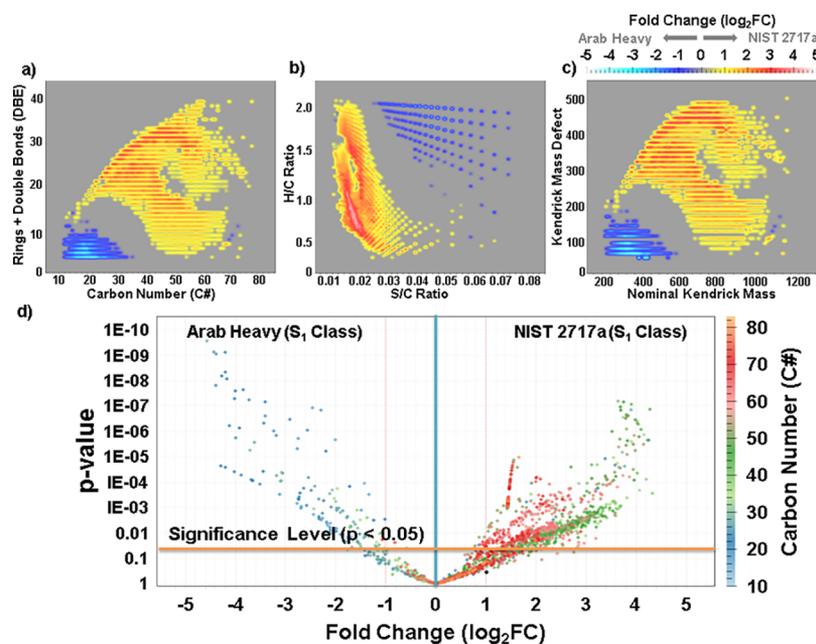


Figure 4. Volcano plot (d) for S₁ class compounds from positive APPI 9.4 T FT-ICR mass spectra of Arab Heavy and NIST 2717a crude oils. DBE vs C# QC plot (a), van Krevelen QC plot (b), and Kendrick mass defect QC plot (c) were generated for the statistically most different data (i.e., $p < 0.05$) in the volcano plot. Plots (a), (b), and (c) are color-coded by log₂ FC value; (d) was color-coded by carbon number (see text).

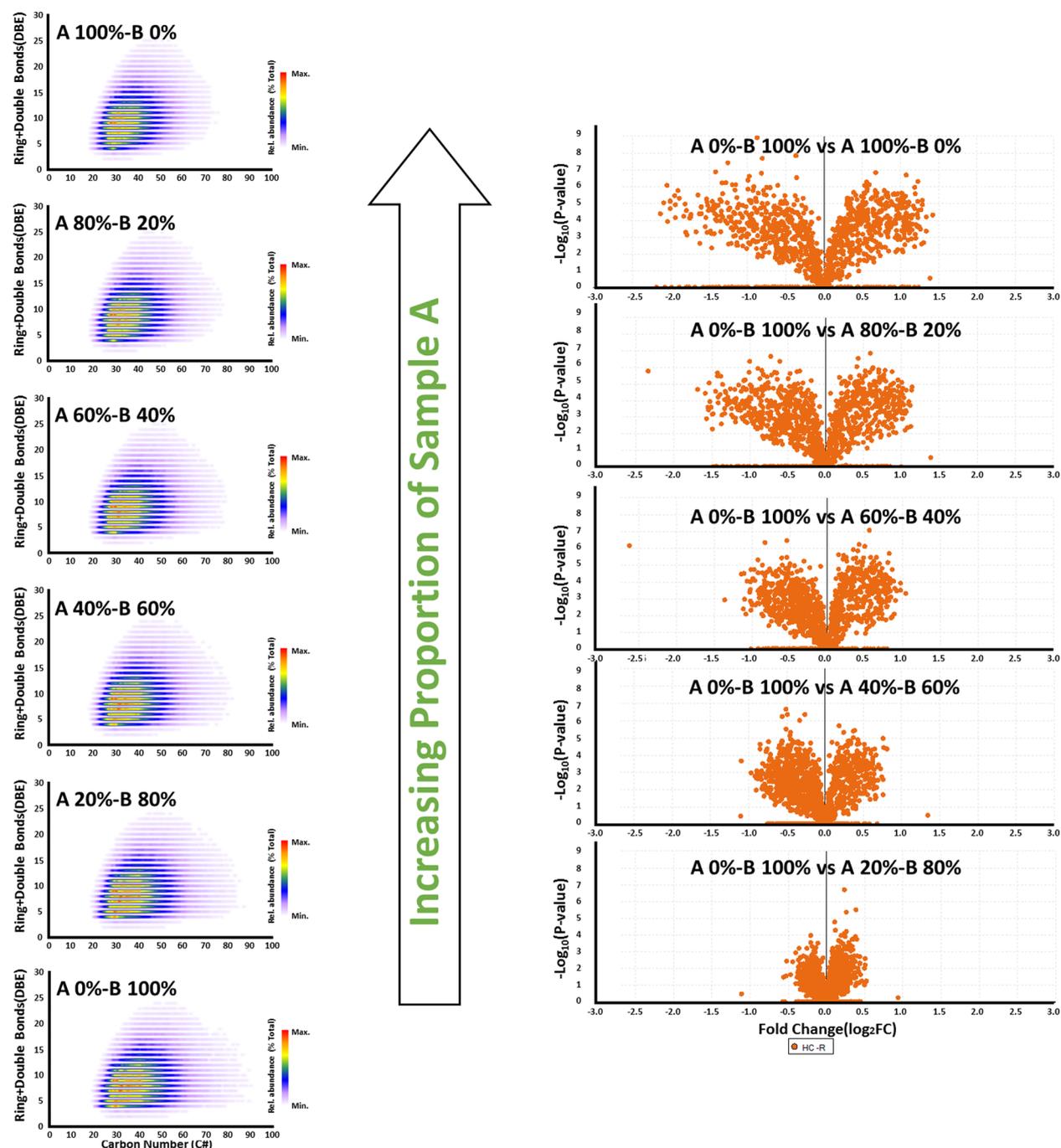


Figure 5. Left: Isoabundance-contoured plots of double bond equivalents vs. carbon number for two similar North Sea crude oils (samples A and B from ref. 39) mixed in ratios ranging from 0% to 100% of A (or B) in 20% increments. The plots are virtually indistinguishable. Right: Volcano plots for the same samples. The five volcano plots are clearly distinguishable, based on fold-change and p -value, even for these mixtures of two highly similar crude oils.

cally significant difference in abundance between samples are located above the orange horizontal line. The two vertical dashed lines enclose points with a fold change of less than 2 between Arab Heavy (left line) and NIST 2717a (right line): the horizontal line separates statistically significant points (p -value < 0.05 , above) from nonsignificant points (p -value > 0.05 , below).

Volcano Plots Based on Carbon Number and Double Bond Equivalents. The volcano plots can be constructed for other important parameters such as carbon number (C#) and double bond equivalents (DBE = number of rings plus double

bonds to carbon), as shown in Figures 2a and 2c for members of the O_1S_1 class species. Figure 2 shows the relationship between volcano plots (2a and 2c) and DBE vs carbon number plots (2b and 2d). The circled region (1) with the smaller carbon number in plot 2a corresponds to carbon numbers (C# < 30) located to the left of the light green dashed line (1) in DBE vs carbon number plot 2b. The circled region (2) with larger carbon numbers in the volcano plot 2a corresponds to carbon numbers (C# > 50) located to the right of the red dashed line (2) in DBE vs carbon number plot 2b. Therefore, the circled regions in each volcano plot highlight compounds

whose carbon number or DBE value are significantly larger or smaller for NIST 2717a than for Arab Heavy crude oil. Note that compounds with DBE values larger than 10 are significantly more abundant in NIST 2717a than in Arab Heavy (2d). It is clear that the same conclusions could be drawn from either the volcano plot or the DBE vs carbon number plot. However, the advantage of the volcano plot over DBE vs carbon number plot is that the chemical compounds responsible for the significant differences can be statistically evaluated and ranked. The compounds exhibiting statistically significant differences are listed in the Supporting Information (Table S1).

Volcano Plot To Characterize Distillate Fractions.

Volcano plots can be used for more quantitative comparison of processed oil samples, as shown in Figure 3a for Arabian Heavy oil and its 593+ °C distillate cut, generated from their positive APPI FT-ICR mass spectra. Color-coding is as for Figure 2a. Most of the peaks with large \log_2 FC values (\log_2 FC > 4) correspond to larger carbon numbers (carbon number > 65). Therefore, high carbon number compounds are more abundant in the high-boiling fraction. Conventional DBE vs carbon number plots (Figure 3b) show the same qualitative trends but without the statistical measure provided by the volcano plot.

Quantitative Comparison Plot: Combination of Volcano Plot with Conventional Visualization Methods.

Quantitative comparison (QC) plots^{12,13} are generated by combining the volcano plot with conventional visualization methods such as Kendrick mass defect (KMD) plot, DBE vs C# plot, and van Krevelen diagrams. For example, the volcano plot for S₁ class compounds in Arabian heavy oil and NIST 2717a, generated from positive APPI 9.4 T FT-ICR mass spectra, is shown in Figure 4d. The volcano plot *p*-value can be used as a filter; e.g., only the peaks exhibiting statistically significant magnitude difference ($p < 0.05$) were selected for a DBE vs C# QC plot (Figure 4a, top left), van Krevelen QC plot (Figure 4b, top middle), and Kendrick mass defect QC plot (Figure 4c, top right). The data in Figure 4 (top) are color coded according to their \log_2 FC values. Color-coding for Figure 4d is as for Figure 2a. Thus, Figure 4a (top left) clearly shows that the NIST 2717a sample contains more high carbon number and DBE compounds than Arabian heavy oil. The QC plots in Figure 4 (top) show that combining volcano and conventional plots is beneficial by displaying only the more statistically significant compositional differences.

CONCLUSION

The volcano plot comparisons reported here are for crude oils whose composition differs sufficiently to distinguish them (qualitatively) by visual images of isoabundance-contoured plots of double bond equivalents (DBE) versus carbon number. However, we were provided with previously published positive atmospheric pressure photoionization FT-ICR mass spectral data for two similar North Sea crude oils (samples A and B) mixed in ratios ranging from 0% to 100% of A (or B) in 20% increments.³⁹ Although the DBE vs carbon number plots for the hydrocarbon class were essentially indistinguishable (Figure 5, left), the corresponding volcano plots (Figure 5, right) showed clear differences based on fold-change and *p*-value.

We have demonstrated that the volcano plot is a useful comparative Petroleomics tool, either alone or in combination with existing data visualization tools (DBE vs carbon number, Kendrick mass defect, and van Krevelen diagram), as QC plots

and provides statistically valid quantitation of compositional differences between petroleum crude oils and their products. We look forward to future applications of the volcano plot to improve understanding of the compositional basis for distillation fractionation, chromatographic separation, corrosion, emulsions, process chemistry, etc.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.energyfuels.7b03061.

Chemical composition of ions exhibiting statistically significant FT-ICR MS abundance differences (Table S1) (XLSX)

AUTHOR INFORMATION

Corresponding Author

*Phone: +1 850-644-0529. Fax: +1 850-644-0133. E-mail: marshall@magnet.fsu.edu.

ORCID

Amy M. McKenna: 0000-0001-7213-521X

Ryan P. Rodgers: 0000-0003-1302-2850

Alan G. Marshall: 0000-0001-9375-2532

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Sungwhan Kim and Yunju Cho at Kyungpook National University in Korea for providing data for statistical modeling and initial writing assistance. M.H. thanks David and Isaac Hur for their help with testing of statistical software. The authors thank Johan P. Quinn and Gregory T. Blakney for design and maintenance of the 9.4 T instrument and data station. A portion of this work was performed at the National High Magnetic Field Laboratory, which is supported by the National Science Foundation Division of Materials Research through DMR-11-57490 and the State of Florida. This material is based in part on work supported by the National Science Foundation under award number EEC-0813570 (E.S.W. and B.J.N.) and IOS 1546858 (E.S.W.).

REFERENCES

- (1) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998**, *17* (1), 1–35.
- (2) McKenna, A. M.; Williams, J. T.; Putman, J. C.; Aeppli, C.; Reddy, C. M.; Valentine, D. L.; Lemkau, K. L.; Kellermann, M. Y.; Savory, J. J.; Kaiser, N. K.; Marshall, A. G.; Rodgers, R. P. Unprecedented ultrahigh resolution FT-ICR mass spectrometry and parts-per-billion mass accuracy enable direct characterization of nickel and vanadyl porphyrins in petroleum from natural seeps. *Energy Fuels* **2014**, *28* (4), 2454–2464.
- (3) Rodgers, R. P.; Schaub, T. M.; Marshall, A. G. Petroleomics: MS returns to its roots. *Anal. Chem.* **2005**, *77* (1), 20A–27A.
- (4) Marshall, A. G.; Rodgers, R. P. Petroleomics: Chemistry of the underworld. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (47), 18090–18095.
- (5) Cho, Y.; Ahmed, A.; Islam, A.; Kim, S. Developments in FT-ICR MS instrumentation, ionization techniques, and data interpretation methods for petroleomics. *Mass Spectrom. Rev.* **2015**, *34* (2), 248–263.
- (6) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. Kendrick mass defect spectrum: a compact visual analysis

for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* **2001**, *73* (19), 4676–4681.

(7) Kim, S.; Kramer, R. W.; Hatcher, P. G. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal. Chem.* **2003**, *75* (20), 5336–5344.

(8) Wu, Z.; Rodgers, R. P.; Marshall, A. G. Two- and three-dimensional van Krevelen diagrams: A graphical analysis complementary to the Kendrick mass plot for sorting elemental compositions of complex organic mixtures based on ultrahigh-resolution broadband Fourier transform ion cyclotron resonance mass measurements. *Anal. Chem.* **2004**, *76* (9), 2511–2516.

(9) Hur, M.; Yeo, I.; Park, E.; Kim, Y. H.; Yoo, J.; Kim, E.; No, M.-h.; Koh, J.; Kim, S. Combination of statistical methods and Fourier transform ion cyclotron resonance mass spectrometry for more comprehensive, molecular-level interpretations of petroleum samples. *Anal. Chem.* **2010**, *82* (1), 211–218.

(10) Hur, M.; Yeo, I.; Kim, E.; No, M.-h.; Koh, J.; Cho, Y. J.; Lee, J. W.; Kim, S. Correlation of FT-ICR mass spectra with the chemical and physical properties of associated crude oils. *Energy Fuels* **2010**, *24* (10), 5524–5532.

(11) Corilo, Y. E.; Podgorski, D. C.; McKenna, A. M.; Lemkau, K. L.; Reddy, C. M.; Marshall, A. G.; Rodgers, R. P. Oil spill source identification by principal component analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **2013**, *85* (19), 9064–9069.

(12) Hur, M. Application of volcano plots for quantitative visualization and comparison of a set of two spectra obtained by high-resolution mass spectrometric analysis of crude oils. In *PacificChem, 2015 International Chemical Congress of Pacific Basin Societies*, Honolulu, HI, Dec. 15–20, 2015.

(13) Hur, M.; Kim, S.; Hsu, C. S. Petroinformatics. In *Springer Handbook of Petroleum Technology*; Hsu, C. S., Robinson, P., Eds.; Springer: New York, 2017 (in print).

(14) Hur, M.; Campbell, A. A.; Almeida-de-Macedo, M.; Li, L.; Ransom, N.; Jose, A.; Crispin, M.; Nikolau, B. J.; Wurtele, E. S. A global approach to analysis and interpretation of metabolic data for plant natural product discovery. *Nat. Prod. Rep.* **2013**, *30* (4), 565–583.

(15) Cui, X.; Churchill, G. A. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* **2003**, *4* (4), 210.

(16) Li, W. Volcano plots in analyzing differential expressions with mRNA microarrays. *J. Bioinf. Comput. Biol.* **2012**, *10* (06), 1231003.

(17) McKenna, A. M.; Donald, L. J.; Fitzsimmons, J. E.; Juyal, P.; Spicer, V.; Standing, K. G.; Marshall, A. G.; Rodgers, R. P. Heavy petroleum composition. 3. Asphaltene aggregation. *Energy Fuels* **2013**, *27* (3), 1246–1256.

(18) Purcell, J. M.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G. Atmospheric pressure photoionization Fourier transform ion cyclotron resonance mass spectrometry for complex mixture analysis. *Anal. Chem.* **2006**, *78* (16), 5906–5912.

(19) Robb, D. B.; Blades, M. W. Factors affecting primary ionization in dopant-assisted atmospheric pressure photoionization (DA-APPI) for LC/MS. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (2), 130–138.

(20) Purcell, J. M.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G. Speciation of nitrogen containing aromatics by atmospheric pressure photoionization or electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (7), 1265–1273.

(21) Purcell, J. M.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G. Atmospheric pressure photoionization proton transfer for complex organic mixtures investigated by Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (9), 1682–1689.

(22) McLafferty, F. W.; Tureček, F. *Interpretation of mass spectra*; University Science Books: South Orange, NJ, 1993.

(23) Kaiser, N. K.; Quinn, J. P.; Blakney, G. T.; Hendrickson, C. L.; Marshall, A. G. A novel 9.4 T FTICR mass spectrometer with

improved sensitivity, mass resolution, and mass range. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (8), 1343–1351.

(24) Blakney, G. T.; Hendrickson, C. L.; Marshall, A. G. Predator data station: A fast data acquisition system for advanced FT-ICR MS experiments. *Int. J. Mass Spectrom.* **2011**, *306* (2), 246–252.

(25) Kaiser, N. K.; Savory, J. J.; McKenna, A. M.; Quinn, J. P.; Hendrickson, C. L.; Marshall, A. G. Electrically compensated Fourier transform ion cyclotron resonance cell for complex mixture mass analysis. *Anal. Chem.* **2011**, *83* (17), 6907–6910.

(26) Tolmachev, A. V.; Robinson, E. W.; Wu, S.; Kang, H.; Lourette, N. M.; Paša-Tolić, L.; Smith, R. D. Trapped-ion cell with improved DC potential harmonicity for FT-ICR MS. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (4), 586–597.

(27) McKenna, A. M.; Blakney, G. T.; Xian, F.; Glaser, P. B.; Rodgers, R. P.; Marshall, A. G. Heavy petroleum composition. 2. Progression of the Boduszynski model to the limit of distillation by ultrahigh-resolution FT-ICR mass spectrometry. *Energy Fuels* **2010**, *24* (5), 2939–2946.

(28) Shi, S. D.-H.; Drader, J. J.; Freitas, M. A.; Hendrickson, C. L.; Marshall, A. G. Comparison and interconversion of the two most common frequency-to-mass calibration functions for Fourier transform ion cyclotron resonance mass spectrometry. *Int. J. Mass Spectrom.* **2000**, *195*, 591–598.

(29) Ledford, E. B., Jr; Rempel, D. L.; Gross, M. Space charge effects in Fourier transform mass spectrometry. II. Mass calibration. *Anal. Chem.* **1984**, *56* (14), 2744–2748.

(30) Corilo, Y. E. *PetroOrg software*; Florida State University: Tallahassee, FL, 2015.

(31) Helsel, D. R. More than obvious: better methods for interpreting nondetect data. *Environ. Sci. Technol.* **2005**, *39* (20), 419A–423A.

(32) Proctor, C. H. A simple definition of detection limit. *J. Agric. Biol. Environ. Stat.* **2008**, *13* (1), 99–120.

(33) Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet.* **2002**, *32*, 496–501.

(34) Mariani, T. J.; Budhraj, V.; Mecham, B. H.; Gu, C. C.; Watson, M. A.; Sadovsky, Y. A variable fold change threshold determines significance for expression microarrays. *FASEB J.* **2003**, *17* (2), 321–323.

(35) Boneau, C. A. The effects of violations of assumptions underlying the t test. *Psychol. Bull.* **1960**, *57* (1), 49.

(36) Julious, S. A. Two-sided confidence intervals for the single proportion: comparison of seven methods by Robert G. Newcombe, *Statistics in Medicine* 1998; 17:857–872. *Stat. Med.* **2005**, *24* (21), 3383–3384.

(37) Bennett, B. Note on power transformations which minimize skewness. *Biom. Z.* **1967**, *9* (2), 73–75.

(38) Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D. R.; Pimentel, H.; Salzberg, S. L.; Rinn, J. L.; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* **2012**, *7* (3), 562–578.

(39) Witt, M.; Timm, W. Determination of Simulated Crude Oil Mixtures from the North Sea Using Atmospheric Pressure Photoionization Coupled to Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy Fuels* **2016**, *30* (5), 3707–3713.