Article Watch: December, 2020

This column highlights recently published articles that are of interest to the readership of this publication. We encourage ABRF members to forward information on articles they feel are important and useful to Clive Slaughter, MCG-UGA Medical Partnership, 1425 Prince Ave., Athens GA 30606, USA. Tel: (706) 713-2216; Fax: (706) 713-2221; E-mail; cslaught@uga.edu, or to any member of the editorial board. Article summaries reflect the reviewer's opinions and not necessarily those of the Association.

NUCLEIC ACID SEQUENCING

Miga K H, Koren S, Rhie A, Vollger M R, Gershman A, Bzikadze A, Brooks S, Howe E, Porubsky D, Logsdon G A, Schneider V A, Potapova T, Wood J, Chow W, Armstrong J, Fredrickson J, Pak E, Tigyi K, Kremitzki M, Markovic C, Maduro V, Dutra A, Bouffard G G, Chang A M, Hansen N F, Wilfert A B, Thibaud-Nissen F, Schmitt A D, Belton J-M, Selvaraj S, Dennis M Y, Soto D C, Sahasrabudhe R, Kaya G, Quick J, Loman N J, Holmes N, Loose M, Surti U, Risques R a, Graves Lindsay T A, Fulton R, Hall I, Paten B, Howe K, Timp W, Young A, Mullikin J C, Pevzner P A, Gerton J L, Sullivan B A, Eichler E E, Phillippy A M. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 585;2020:79-84.

Although 20 yr have elapsed since the initial release of the human genome sequence, the human genome remains unfinished. The unresolved areas are repetitive regions that include centromeric regions, the short arms of acrocentric chromosomes, and regions containing segmental duplications. Their absence from the reference genome discourages their genetic study and results in erroneous conclusions such as misinterpretation of paralogs as allelic variants and misassignment of human sequences to bacterial databases. In the latest release from the Genome Reference Consortium, GRCh38, no chromosome is yet complete. Contig lengths have greatly improved, but the need to assemble relatively short sequence reads continues to be a serious barrier. Miga *et al.* now supply the first gapless assembly of a human chromosome, the X chromosome. This is achieved with high-coverage, ultra-long nanopore sequence reads. The data also permit mapping of methylation patterns in complex tandem repeats and satellite regions. The authors use an effectively haploid cell line for their purpose. It is derived from a hydatidiform mole, whose 46,XX karyotype results from postmeiotic chromosome duplication from a single sperm. This study represents a further step toward truly finishing the human genome sequence.

GLYCANS

Riley N M, Malaker S A, Driessen M D, Bertozzi C R. Optimal dissociation methods differ for *N*- and *O*-glycopeptides. *Journal of Proteome Research* 19; 2020:3286-3301.

A variety of tandem mass spectrometric methods have become available for glycopeptide analysis. They broadly fall into 2 categories: methods based on collisional activation [higher-energy collisional dissociation (HCD) on Orbitrap systems] and methods based on electron transfer dissociation (ETD). Variants within each category of method exist, including stepped collision energy HCD (sceHCD), and ETD with supplemental HCD activation (EThcD), both of which have been used for large-scale proteomic surveys. Each regime has strengths and weaknesses. The authors therefore perform a systematic comparison of these various methods, including platform-dependent protocols, to evaluate their suitability for proteomic study of both *N*- and *O*-glycopeptides. They evaluate identification rates, spectral quality, and suitability for the different glycopeptide classes. Although it would be convenient to be able to identify a single best technique for general application, the authors conclude that different methods are optimal for *N*- and for *O*-glycopeptides. HCD and sceHCD are adequate for *N*-glycopeptide identification, although sceHCD provides somewhat better spectral quality. In this application, the more rapid scan rates favor HCD and sceHCD over EThcD in numbers of identifications. For *O*-glycopeptides, however, HCD-based methods are inadequate for localization of glycosylation sites. Instead, EThcD is the method of choice. The authors' data set will be useful to investigators in choosing for their particular application both mass spectral instrument platform and the details of experimental design. The results of the study also indicate a need for a continued effort to improve the software for interpretation

https://doi.org/10.7171/jbt.2020-3104-009

of both HCD- and ETD-based spectra in glycoproteomic analysis.

MASS SPECTROMETRY

Peters-Clarke T M, Schauer K L, Riley N M, Lodge J M, Westphall M S, Coon J J. Optical fiber–enabled photoactivation of peptides and proteins. *Analytical Chemistry* 92;2020:12363-12370.

Although protein secondary structure impedes the dissociation of precursors with low charge density during ETD, substantial success in improving dissociation has been demonstrated by concurrent irradiation with infrared photons in a process known as activated ion-ETD (AI-ETD). The photon flux is adjusted to cleave peptide bonds without losing labile post-translational modifications. The authors here describe modifications to an Orbitrap Fusion Lumos mass spectrometer that provide a safe, robust platform for IR photoactivation, as well as for infrared multiphoton dissociation. Photons of 10.6 μm in wavelength are transmitted to the instrument *via* a hollow-core fiber-optic cable to avoid mirror-based implementation. Laser alignment and power are adjusted by monitoring the decay rate of a standard peptide precursor ion. The system is shown to support comprehensive sequence coverage of diverse peptides and small proteins. It is hoped that this work will facilitate the commercial implementation of AI-ETD.

FUNCTIONAL GENOMICS AND PROTEOMICS

Abascal F, Acosta R, Addleman N J, Adrian J, Afzal V, Aken B, Akiyama J A, Jammal O A, Amrhein H, Anderson S M, Andrews G R, Antoshechkin I, Ardlie K G, Armstrong J, Astley M, Banerjee B, Barkal A A, Barnes I H A, Barozzi I, Barrell D, Barson G, Bates D, Baymuradov U K, Bazile C, Beer M A, Beik S, Bender M A, Bennett R, Bouvrette L P B, Bernstein B E, Berry A, Bhaskar A, Bignell A, Blue S M, Bodine D M, Boix C, Boley N, Borrman T, Borsari B, Boyle A P, Brandsmeier L A, Breschi A, Bresnick E H, Brooks J A, Buckley M, Burge C B, Byron R, Cahill E, Cai L, Cao L, Carty M, Castanon R G, Castillo A, Chaib H, Chan E T, Chee D R, Chee S, Chen H, Chen H, Chen J-Y, Chen S, Cherry J M, Chhetri S B, Choudhary J S, Chrast J, Chung D, Clarke D, Cody N A L, Coppola C J, Coursen J, D'Ippolito A M, Dalton S, Danyko C, Davidson C, Davila-Velderrain J, Davis C A, Dekker J, Deran A, DeSalvo G, Despacio-Reyes G, Dewey C N, Dickel D E, Diegel M, Diekhans M, Dileep V, Ding B, Djebali S, Dobin A, Dominguez D, Donaldson S, Drenkow J, Dreszer T R, Drier Y, Duff M O, Dunn D, Eastman C, Ecker J R, Edwards M D, El-Ali N, Elhajjajy S I, Elkins K, Emili A, Epstein C B, Evans R C, Ezkurdia I, Fan K, Farnham P J, Farrell N, Feingold E A, Ferreira A-M, Fisher-Aylor K, Fitzgerald S, Flicek P, Foo C S, Fortier K, Frankish A, Freese P, Fu S, Fu X-D, Fu Y, Fukuda-Yuzawa Y, Fulciniti M, Funnell A P W, Gabdank I, Galeev T, Gao M, Giron C G, Garvin T H, Gelboin-Burkhart C A, Georgolopoulos G, Gerstein M B, Giardine B M, Gifford D K, Gilbert D M, Gilchrist D A, Gillespie S, Gingeras T R, Gong P, Gonzalez A, Gonzalez J M, Good P, Goren A, Gorkin D U, Graveley B R, Gray M, Greenblatt J F, Griffiths E, Groudine M T, Grubert F, Gu M, Guigó R, Guo H, Guo Y, Guo Y, Gursoy G, Gutierrez-Arcelus M, Halow J, Hardison R C, Hardy M, Hariharan M, Harmanci A, Harrington A, Harrow J L, Hashimoto T B, Hasz R D, Hatan M, Haugen E, Hayes J E, He P, He Y, Heidari N, Hendrickson D, Heuston E F, Hilton J A, Hitz B C, Hochman A, Holgren C, Hou L, Hou S, Hsiao Y-H E, Hsu S, Huang H, Hubbard T J, Huey J, Hughes T R, Hunt T, Ibarrientos S, Issner R, Iwata M, Izuogu O, Jaakkola T, Jameel N, Jansen C, Jiang L, Jiang P, Johnson A, Johnson R, Jungreis I, Kadaba M, Kasowski M, Kasparian M, Kato M, Kaul R, Kawli T, Kay M, Keen J C, Keles S, Keller C A, Kelley D, Kellis M, Kheradpour P, Kim D S, Kirilusha A, Klein R J, Knoechel B, Kuan S, Kulik M J, Kumar S, Kundaje A, Kutyavin T, Lagarde J, Lajoie B R, Lambert N J, Lazar J, Lee A Y, Lee D, Lee E, Lee J W, Lee K, Leslie C S, Levy S, Li B, Li H, Li N, Li X, Li Y I, Li Y, Li Y, Li Y, Lian J, Libbrecht M W, Lin S, Lin Y, Liu D, Liu J, Liu P, Liu T, Liu X S, Liu Y, Liu Y, Long M, Lou S, Loveland J, Lu A, Lu Y, Lécuyer E, Ma L, Mackiewicz M, Mannion B J, Mannstadt M, Manthravadi D, Marinov G K, Martin F J, Mattei E, McCue K, McEown M, McVicker G, Meadows S K, Meissner A, Mendenhall E M, Messer C L, Meuleman W, Meyer C, Miller S, Milton M G, Mishra T, Moore D E, Moore H M, Moore J E, Moore S H, Moran J, Mortazavi A, Mudge J M, Munshi N, Murad R, Myers R M, Nandakumar V, Nandi P, Narasimha A M, Narayanan A K, Naughton H, Navarro F C P, Navas P, Nazarovs J, Nelson J, Neph S, Neri F J, Nery J R, Nesmith A R, The E P C. Perspectives on ENCODE. *Nature* 583;2020:693-698.

The Encyclopedia of DNA Elements (ENCODE) project began in 2003 with pilot studies of transcribed

segments of the human genome (both coding and non-coding) and of *cis*-regulatory elements. The studies relied upon array-based assays. A second phase followed in 2007–2012 in which sequencing-based technologies such as chromatin immunoprecipitation with sequencing and RNA sequencing were used to search an extended set of cell lines for functional elements encompassing the whole genome. The project was also extended to the mouse genome to gain evolutionary perspective. Data collected in the project's third phase, 2012–2017, are now analyzed in a collection of 8 papers in *Nature* accompanied by the *Perspective* piece cited here. They describe cell and tissue repertoires of RNA transcription, chromatin structure and modifications, DNA methylation, chromatin looping, and DNA occupancy by transcription factors and RNA-binding proteins. The work also begins to build a picture of spatiotemporal changes during fetal development. Of particular methodological interest, phase 3 of the project has added global mapping of 3-D chromatin interactions and RNA-binding regions to its technical repertoire. New data standards for replicability and antibody specificity have been applied, and uniform protocols for data processing have been adopted. These improvements have stimulated other large-scale projects to do likewise. The ENCODE data are made publicly accessible as a resource for the scientific community. Next stages in the work of annotating DNA elements are expected to include extension of the description of tissues to the single-cell level, identification of transcription factors that interact with particular sequences, and analysis of repetitive sequences.

MACROMOLECULAR CHARACTERIZATION

Lodge J M, Schauer K L, Brademan D R, Riley N M, Shishkova E, Westphall M S, Coon J J. Top-down characterization of an intact monoclonal antibody using activated ion electron transfer dissociation. *Analytical Chemistry* 92;2020:10246-10251.

The structural characterization of monoclonal antibodies for the purposes of pharmaceutical quality control is complicated by the large size of the proteins, the heterogeneity introduced by post-translational modification, and the presence of regions that persistently resist analysis though the particular characteristics of their covalent and noncovalent structure. Lodge *et al.* here deploy ETD and AI-ETD for fragmentation of the intact protein to demonstrate the comprehensive characterization of antibody structure by a combination of these 2 techniques. Using direct infusion of the intact protein, they screen different laser powers in AI-ETD for photon-induced vibrational excitation of the precursor to disrupt noncovalent interactions to varying degrees. They also screen different ETD reaction times to vary the disruption of disulfide connectivity. These maneuvers provide controlled changes in the yield of fragments from disulfide bonded regions to increase sequence coverage. This work demonstrates improvement in capability for assessing the integrity of mAb preparations by top-down mass spectral methods.

IMAGING

Chen W-W, Lemieux G A, Camp C H, Chang T-C, Ashrafi K, Cicerone M T. Spectroscopic coherent Raman imaging of *Caenorhabditis elegans* reveals lipid particle diversity. *Nature Chemical Biology* 16; 2020:1087-1095.

Chen *et al.* demonstrate the use of Raman spectroscopy for the imaging of lipid distribution and turnover in a live animal. Their methodology is a coherent Raman imaging technique previously developed by the group in which broadband coherent anti-Stokes Raman scattering is monitored for acquisition of both CH-stretch data for discrimination of water, protein and lipid, and fingerprint data with chemical specificity for inference of detailed chemical structure. An image pixel is acquired with each laser pulse. The methodology achieves 500-nm spatial resolution at 3.5 ms per pixel and provides a high-quality, intrinsically calibrated Raman spectrum in the frequency range of 600 cm^{-1} to >3100 cm^{-1}. Metabolic flux can be studied in pulse-chase experiments by taking advantage of the capability to distinguish between C-H and C-D resonances. The authors apply this methodology to study the lipid content of *Caenorhabditis elegans*. *C. elegans* lacks adipocytes but stores lipid droplets in intestine and epidermal cells. The epidermal cells provide nutrients for developing oocytes in the hermaphrodite gonad. They find that epidermal lipid droplets are relatively homogeneous in size, are triglyceride-rich and protein poor, and turn over slowly. The intestinal lipid droplets are more heterogeneous. Some are transitory and protein-rich. Those droplets behave as lipoprotein particles for transfer of lipids to developing oocytes. The results of the study illustrate how Raman imaging may be used for biochemical and physiologic studies of living tissues.

CELL BIOLOGY

Mok B Y, de Moraes M H, Zeng J, Bosch D E, Kotrys A V, Raguram A, Hsu F, Radey M C, Peterson S B, Mootha V K, Mougous J D, Liu D R. A bacterial

cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* 583;2020:631-637.

The existence of disease-causing mutations in the mitochondrial genome creates a need for editing mitochondrial DNA (mtDNA) for construction of disease model systems and, ultimately, for the repair of mtDNA in affected patients. Unfortunately, RNA-guided methods such as CRISPR-Cas9 cannot be used for editing mitochondrial DNA because mitochondria lack mechanisms for import of RNA. Mitochondrion-targeted nucleases have been used to destroy target sequences in mtDNA, but the elimination of a target sequence is only helpful if the cell contains some copies lacking the target sequence beforehand; otherwise, all the mitochondria would be destroyed. In any case, the method cannot be used to introduce a new mutation into mtDNA. Mok *et al.* now introduce the first base editing system for circumventing this impasse. It utilizes a cytidine deaminase called DddA, which converts cytidine (C) in a C•G base pair to uracil (U) in double-stranded DNA without causing double-stranded breaks. Their construct incorporates a uracil glycosylase inhibitor that prevents repair of the U back to C. In the ensuing DNA replication, U pairs with adenine (A), eventually to create an A•T base pair. The system is capable of efficiencies up to 50%, although not higher because it mutates only 1 strand of the double-stranded mtDNA. The cytidine deaminase is rendered sequence specific without recourse to RNA templates by incorporating amino acid sequences from transcription activator-like nucleases (TALENs) engineered to bind to specific target DNA sequences. Because cytidine deaminases are cytotoxic, the enzyme is introduced into cells split into 2 catalytically inactive halves, which become active only when they reach proximity on the target mtDNA sequence to which they are guided by their TALEs. The authors test the final cytosine base editor (DdCBE) construct by creating models of known human mitochondrial disease mutations and demonstrating the expected deficiency in respiration rates and oxidative phosphorylation without off-target mutation.

Enache O M, Rendo V, Abdusamad M, Lam D, Davison D, Pal S, Currimjee N, Hess J, Pantel S, Nag A, Thorner A R, Doench J G, Vazquez F, Beroukhim R, Golub T R, Ben-David U. Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nature Genetics* 52;2020:662-668.

Enache *et al.* provide evidence that expression of the Cas9 endonuclease consistently upregulates the p53 DNA repair pathway to a small extent in human cancer cell lines. The effect is associated with modestly elevated levels of the p53 and p21 mRNA and protein. Cas9 is frequently introduced into cell lines for the purpose of genome editing with the CRISPR-Cas9 system, but the effects studied here are due to the expression of Cas9 itself. Reasoning that activation of the p53 DNA repair pathway might represent a selective disadvantage to cells in which it occurs, the authors assessed processes inactivating the pathway. They observe an elevated frequency of *TP53* inactivating mutations and overgrowth of Cas9 cell lines with inactivated *TP53* when cocultured with wild-type Cas9 cells. These effects may be understood as responses to increased levels of DNA damage resulting from Cas9 expression. Investigators using CRISPR-Cas9-based screening will wish to be mindful of these effects. They should also bear in mind the authors' ancillary observation that Cas9 is less active in *TP53*-wild-type cells than in *TP53* mutant cells.