Article Watch: September, 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

Ding J, Adiconis X, Simmons S K, Kowalczyk M S, Hession C C, Marjanovic N D, Hughes T K, Wadsworth M H, Burks T, Nguyen L T, Kwon J Y H, Barak B, Ge W, Kedaigle A J, Carroll S, Li S, Hacohen N, Rozenblatt-Rosen O, Shalek A K, Villani A-C, Regev A, Levin J Z. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nature Biotechnology* 38;2020:737-746.

Mereu E, Lafzi A, Moutinho C, Ziegenhain C, McCarthy D J, Álvarez-Varela A, Batlle E, Sagar, Grün D, Lau J K, Boutet S C, Sanada C, Ooi A, Jones R C, Kaihara K, Brampton C, Talaga Y, Sasagawa Y, Tanaka K, Hayashi T, Braeuning C, Fischer C, Sauer S, Trefzer T, Conrad C, Adiconis X, Nguyen L T, Regev A, Levin J Z, Parekh S, Janjic A, Wange L E, Bagnoli J W, Enard W, Gut M, Sandberg R, Nikaido I, Gut I, Stegle O, Heyn H. Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nature Biotechnology* 38;2020:747-755.

Single-cell RNA-sequencing (scRNA-seq) and single-nucleus RNA-sequencing (snRNA-seq) are fundamentally changing the way we think about cell differentiation, cellular composition and organization of tissues, and molecular basis of cellular physiology in health and disease. Large-scale projects are under way to document the emerging knowledge, *e.g.*, the Human Cell Atlas project. The snRNA-seq methodology is continually evolving, and many choices of protocol now exist for practitioners to choose between. Two groups now undertake the complex but essential task of documenting the comparative capabilities of the principal contemporary methods. The investigators observe significant variation between protocols in efficiency of RNA capture and, hence, in sequence library complexity. This leads to variation in sensitivity of identification of different transcripts and resulting differences in the ability to discriminate and enumerate different cell types or cell states in tissues. Mereu *et al.* emphasize that comparison of results acquired in different studies or in different laboratories depends upon benchmarking, standardization, and quality control. These 2 papers help define criteria upon which the effort will be based.

GLYCANS

Wu X, Delbianco M, Anggara K, Michnowicz T, Pardo-Vargas A, Bharate P, Sen S, Pristl M, Rauschenbach S, Schlickum U, Abb S, Seeberger P H, Kern K. Imaging single glycans. *Nature* 582; 2020:375-378.

Glycan structures are noted both for the conformational flexibility of their polymer chains and for the variability of their covalent stereochemistry. Imaging of single biopolymer molecules, by cryo-electron microscopy (cryo-EM) for example, is contributing richly to knowledge of protein structure. Wu *et al.* here take initial steps toward the imaging of single glycan polymers. They employ mass selection in a mass spectrometer to prepare glycans for imaging. They produce gas-phase glycan ions by electrospray ionization, select them according to mass, and deliver them to a target surface under high vacuum. To prevent disruption of covalent bonds when the ions impinge on the target, the ions are decelerated to low kinetic energy (5 eV). This allows a soft landing. The glycans are then imaged *in situ* by scanning tunneling microscopy. For this purpose, thermal motion on the target is minimized by maintenance of low temperature: 120 K during ion deposition and 4.5 K during imaging. The researchers image 3 linear synthetic glycans and 3 branched glycans with this technology. Monosaccharide units are resolved as topographic protrusions, and polymer chains are frozen into varying shapes. The stereochemistry of linkages affects the apparent center-to-center distance between monosaccharide units. The α1–2 linkages are distinguished from α1–6 on this basis. The images help define the position of branch points. Conformational flexibility is visualized as variation in chain shape.

https://doi.org/10.7171/jbt.20-3103-006

Weiss G L, Stanisich J J, Sauer M M, Lin C-W, Eras J, Zyla D S, Trück J, Devuyst O, Aebi M, Pilhofer M, Glockshuber R. Architecture and function of human uromodulin filaments in urinary tract infections. *Science* 2020:eaaz9866.

Uromodulin (also known as Tamm-Horsfall protein), a product of renal tubular cells, is the most abundant protein in human urine. It has antibacterial properties and is believed to function as an antimicrobial molecule in protection from urinary tract infection. Uromodulin forms homopolymeric filaments of length ~2.5 μm. This paper illustrates the use of imaging technology to discover how the antibacterial function is accomplished. The researchers identify multiple sites of *N*-glycosylation, some occupied by high-mannose–type glycans and some by di-, tri-, or tetra-antennary types. The domain structure of the uromodulin filament is then deduced by cryo-electron tomography, a technique in which a series of cryo-EM projection images is acquired at defined tilt angles to reconstruct a 3-dimensional (3-D) structure. The filament is found to exist as a zigzag-shaped backbone with laterally protruding, flexible arms. The localization of glycan side chains is deduced from the resulting domain structure. Biochemical experiments show that high-mannose *N*-glycans on uromodulin filaments bind to an adhesin, FimH, on type 1–piliated, uropathic *Escherichia coli*. The researchers are able to determine which of the high-mannose chains on uromodulin is responsible. Cryo-electron tomography then reveals that uromodulin filaments form a tight meshwork around bacteria, causing bacterial cells to clump. These aggregates are presumably cleared in the urine. This study provides an example of the contribution of imaging to the analysis of glycan function.

FUNCTIONAL GENOMICS AND PROTEOMICS

Villaseñor R, Pfaendler R, Ambrosi C, Butz S, Giuliani S, Bryan E, Sheahan T W, Gable A L, Schmolka N, Manzo M, Wirz J, Feller C, von Mering C, Aebersold R, Voigt P, Baubec T. ChromID identifies the protein interactome at chromatin marks. *Nature Biotechnology* 38;2020:728-736.

Villaseñor *et al.* establish a new system for exploring how epigenetic marks on nucleosomes or DNA affect recruitment of regulatory proteins to the marked sites in living cells. The researchers select domains from a variety of natural chromatin-binding proteins for recognition of particular marks: methyl-cytosine, H3K9me3, H3K4me3, and H3K27me3. When expressed in cells of interest, these recognition domains act as "engineered chromatin reader domains" (eCRs), which replace exogenous antibodies used formerly in many studies. The domains are expressed either singly or as dual domains that can bind to 2 marks simultaneously. The expression cassettes include a nuclear localization signal and a biotin acceptor site for purification. They are integrated into a single defined site in the mouse genome by recombinase-mediated cassette exchange. When coupled to enhanced green fluorescent protein for live-cell imaging, eCRs with 2 binding domains are observed to undergo specific localization, whereas eCRs with 1 domain do not. This indicates that bivalent interaction with pairs of modifications on the same nucleosome is required for detection of the interaction. The researchers validate the expected specificity and binding patterns of the eCRs and demonstrate that eCR expression does not interfere with normal cellular behavior. The researchers then detect proteins recruited to genomic sites of particular marks using expression of eCRs coupled to a nonspecific biotin ligase for proximity biotinylation. Proteins within a radius of action of ~35 nm are susceptible to biotinylation, allowing recruited proteins to be purified for identification by liquid chromatography–tandem mass spectrometry. The researchers provide lists of proteins specifically interacting with epigenetic marks. The methodology is expected to yield information about dynamic cellular processes and developmental changes through deployment for comparison of different cell types.

Gérard A, Woolfe A, Mottet G, Reichen M, Castrillon C, Menrath V, Ellouze S, Poitou A, Doineau R, Briseno-Roa L, Canales-Herrerias P, Mary P, Rose G, Ortega C, Delincé M, Essono S, Jia B, Iannascoli B, Richard-Le Goff O, Kumar R, Stewart S N, Pousse Y, Shen B, Grosselin K, Saudemont B, Sautel-Caillé A, Godina A, McNamara S, Eyer K, Millot G A, Baudry J, England P, Nizak C, Jensen A, Griffiths A D, Bruhns P, Brenan C. High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics. *Nature Biotechnology* 38;2020:715-721.

There has long been interest in the comprehensive sequence analysis of the repertoire of antibody heavy (H) and light (L) chain variable regions (V_H and V_L) that are elicited in response to infection or immunization. The aim is partly to understand the maturation of immune responses and partly to obtain antibodies for experimental or therapeutic applications. Gérard *et al.* describe methodology for high-throughput selection of millions of nonimmortalized plasma cells or plasmablasts according to antigen

specificity and then sequence their V_H and V_L mRNA. They use microfluidics technology to create picoliter-volume droplets containing single IgG-secreting cells and sort the droplets according to antigen-binding properties of their contents at rates up to 600/s. The sorted cells are re-compartmentalized into subnanoliter droplets together with a hydrogel bead that bears barcoded sequencing primers. The cells are lysed within the droplets, and the V_H and V_L mRNAs are reverse transcribed for high-throughput sequencing of the cDNAs. The cDNAs incorporate the barcodes, which allow cognate V_H and V_L sequences to be correctly paired. The researchers utilize this system to characterize the immune responses to 2 soluble protein antigens, the vaccine target tetanus toxoid, and the enzyme glucose-6-phosphate isomerase. They also investigate the immune response to a cellular-based antigen, an integral membrane protein, tetraspanin-8, which is a potential tumor-specific antigen involved in tumor proliferation and invasion. The methodology allows rapid characterization of the antigen-specific plasma cell/plasmablast repertoire within 3.5 d. Subsequent gene synthesis, cloning, production, and validation of the properties of the selected antibodies may be accomplished in just 17 d.

MACROMOLECULAR CHARACTERIZATION

Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown P M G E, Grigoras I T, Malinauskaite L, Malinauskas T, Miehling J, Yu L, Karia D, Pechnikova E V, de Jong E, Keizer J, Bischoff M, McCormack J, Tiemeijer P, Hardwick S W, Chirgadze D Y, Murshudov G, Aricescu A R, Scheres S H W. Single-particle cryo-EM at atomic resolution. *bioRxiv* 2020:2020.2005.2022.110189.

Yip K M, Fischer N, Paknia E, Chari A, Stark H. Breaking the next cryo-EM resolution barrier – atomic resolution determination of proteins! *bioRxiv* 2020:2020.2005.2021.106740.

Since 2013, single-particle cryo-EM for the study of protein 3-D structure has undergone rapid improvement in electron detection and image processing, enabling progressive increases in resolution. Although hitherto unable to match the resolution attainable by single crystal X-ray diffraction, the application of cryo-EM to proteins that cannot be crystallized in the resolution range of 3–4 Å has contributed greatly to knowledge of protein structure. Two groups now report further improvements in methodology that mark an important milestone in the evolution of the technique: attainment of true atomic resolution. Both groups measure performance of their methodology using the benchmark protein mammalian apoferritin, a favorably rigid and symmetrical protein. Yip *et al.* achieve 1.25 Å resolution, which provides twice the information content of the previous record of 1.54 Å. They use an energy filter to reduce the energy spread of the electron beam and an aberration corrector to reduce distortion in the images. Nakane *et al.* achieve 1.2 Å resolution using a prototype "cold" field emission gun to reduce energy spread of their electron beam. They further employ an energy filter to remove electrons inelastically scattered by the sample and a new high-performance camera. Nakane *et al.* also construct a 1.7-Å–resolution model of a more difficult protein, the human membrane protein β3 $GABA_A$ receptor homopentamer. These new studies demonstrate the capability of cryo-EM to visualize individual atoms in a protein and even to reveal densities for hydrogen atoms (which X-ray diffraction at comparable resolution cannot do).

Kappel K, Zhang K, Su Z, Watkins A M, Kladwang W, Li S, Pintilie G, Topkar V V, Rangan R, Zheludev I N, Yesselman J D, Chiu W, Das R. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nature Methods* 17;2020:699-707.

Despite their diverse functions, the number of 3-D structures that have been determined for RNAs remains small, in part because of the technical difficulties of applying conventional X-ray diffraction and NMR-based methods to RNA molecules. Kappel *et al.* now investigate the utility of cryo-EM for RNA structure analysis. They select for analysis a benchmark set of 18 RNAs of previously unknown structure ranging in size from 65 to 288 nt (21–126 kDa). The set includes ribozymes, riboswitches, and computationally designed RNAs (*e.g.*, aptamers). Density maps are successfully determined for 11 of the molecules, with resolution ranging from 4.7 to 11 Å. The remainder failed to form sharp bands on native gel electrophoresis and were therefore anticipated to be conformationally too flexible. The density maps are combined with secondary structure information obtained by a chemical mapping method called "mutate-and-map read out through next-generation sequencing" (M2-seq). The combined data are used to build atomic models with the Rosetta modeling program auto-DRRAFTER. The pipeline produces models rapidly. The accuracy of the models is assessed on the basis of modeling convergence. The methodology provides information precise enough to define global features of tertiary structure and to detect structural rearrangements upon target binding.

SYNTHETIC BIOLOGY AND MACROMOLECULAR SYNTHESIS

Qian J, Lu Z-x, Mancuso C P, Jhuang H-Y, del Carmen Barajas-Ornelas R, Boswell S A, Ramírez-Guadiana F H, Jones V, Sonti A, Sedlack K, Artzi L, Jung G, Arammash M, Pettit M E, Melfi M, Lyon L, Owen S V, Baym M, Khalil A S, Silver P A, Rudner D Z, Springer M. Barcoded microbial system for high-resolution object provenance. *Science* 368;2020: 1135-1140.

Qian *et al.* construct DNA-barcoded microbes for use in tracing where items such as food substances have come from and where they have been. The system exploits the natural persistence of microbial spores and utilizes the high sensitivity with which they can be detected and the high specificity with which they can be distinguished from irrelevant microbial species. The researchers introduce nonredundant DNA barcodes into the genomes of *Bacillus subtilis* and *Saccharomyces cerevisiae*, 2 species of widespread environmental presence. Both species are amenable to culturing for manufacture at scale. The combinatorial deployment of cells with different barcodes creates a practically unlimited repertoire of unique sequence signatures. Target sequences in microbial lysates are detected with high sensitivity by the SHEROLCK system (the name stands for "specific high sensitivity enzymatic reporter unlocking"), in which hybridization of the microbial target sequence to a guide RNA activates a Cas13 enzyme. The activated Cas13 then generates a fluorescent product by cleavage of a labeled RNA sensor. Spores are applied to objects by spraying. To ensure environmental safety, spores are disabled from germinating by auxotrophy. Additionally, *B. subtilis* is disabled by deletion of genes for germinant receptors and enzymes required to lyse spore cell walls, and *S. cerevisiae* is killed by heating. The researchers demonstrate persistence of spores on sand, soil, carpet, and wood for at least 3 mo, and spores remain detectable after simulation of wind, rain, vacuuming, and sweeping. They are resistant to secondary spreading onto other objects. The system is envisaged initially as a method for monitoring food provenance, but a broad variety of further applications in agriculture, commerce, and forensics is conceivable.

IMAGING

Botvinik-Nezer R, Holzmeister F, Camerer C F, Dreber A, Huber J, Johannesson M, Kirchler M, Iwanir R, Mumford J A, Adcock R A, Avesani P, Baczkowski B M, Bajracharya A, Bakst L, Ball S, Barilari M, Bault N, Beaton D, Beitner J, Benoit R G, Berkers R M W J, Bhanji J P, Biswal B B, Bobadilla-Suarez S, Bortolini T, Bottenhorn K L, Bowring A, Braem S, Brooks H R, Brudner E G, Calderon C B, Camilleri J A, Castrellon J J, Cecchetti L, Cieslik E C, Cole Z J, Collignon O, Cox R W, Cunningham W A, Czoschke S, Dadi K, Davis C P, Luca A D, Delgado M R, Demetriou L, Dennison J B, Di X, Dickie E W, Dobryakova E, Donnat C L, Dukart J, Duncan N W, Durnez J, Eed A, Eickhoff S B, Erhart A, Fontanesi L, Fricke G M, Fu S, Galván A, Gau R, Genon S, Glatard T, Glerean E, Goeman J J, Golowin S A E, González-García C, Gorgolewski K J, Grady C L, Green M A, Guassi Moreira J F, Guest O, Hakimi S, Hamilton J P, Hancock R, Handjaras G, Harry B B, Hawco C, Herholz P, Herman G, Heunis S, Hoffstaedter F, Hogeveen J, Holmes S, Hu C-P, Huettel S A, Hughes M E, Iacovella V, Iordan A D, Isager P M, Isik A I, Jahn A, Johnson M R, Johnstone T, Joseph M J E, Juliano A C, Kable J W, Kassinopoulos M, Koba C, Kong X-Z, Koscik T R, Kucukboyaci N E, Kuhl B A, Kupek S, Laird A R, Lamm C, Langner R, Lauharatanahirun N, Lee H, Lee S, Leemans A, Leo A, Lesage E, Li F, Li M Y C, Lim P C, Lintz E N, Liphardt S W, Losecaat Vermeer A B, Love B C, Mack M L, Malpica N, Marins T, Maumet C, McDonald K, McGuire J T, Melero H, Méndez Leal A S, Meyer B, Meyer K N, Mihai G, Mitsis G D, Moll J, Nielson D M, Nilsonne G, Notter M P, Olivetti E, Onicas A I, Papale P, Patil K R, Peelle J E, Pérez A, Pischedda D, Poline J-B, Prystauka Y, Ray S, Reuter-Lorenz P A, Reynolds R C, Ricciardi E, Rieck J R, Rodriguez-Thompson A M, Romyn A, Salo T, Samanez-Larkin G R, Sanz-Morales E, Schlichting M L, Schultz D H, Shen Q, Sheridan M A, Silvers J A, Skagerlund K, Smith A, Smith D V, Sokol-Hessner P, Steinkamp S R, Tashjian S M, Thirion B, Thorp J N, Tinghög G, Tisdall L, Tompson S H, Toro-Serey C, Torre Tresols J J, Tozzi L, Truong V, Turella L, van 't Veer A E, Verguts T, Vettel J M, Vijayarajah S, Vo K, Wall M B, Weeda W D, Weis S, White D J, Wisniewski D, Xifra-Porxas A, Yearling E A, Yoon S, Yuan R, Yuen K S L, Zhang L, Zhang X, Zosky J E, Nichols T E, Poldrack R A, Schonberg T. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582;2020:84-88.

Functional magnetic resonance imaging (fMRI) provides data for identification of brain regions that contribute to particular tasks either by activation or inhibition of signaling. The fMRI datasets are large, complex, and noisy: their interpretation requires extensive signal processing. In

this study, a dataset is acquired for 108 experimental subjects performing decision-making under risk. The experiment was originally undertaken to determine whether the magnitude of the subjects' potential loss in deciding wrongly is encoded by regions of the brain that are also involved with negative emotions, *e.g.*, the amygdala. The data are here used to test a series of predefined hypotheses as to whether particular brain regions are, or are not, involved in a particular part of the task. The dataset is supplied to 70 independent laboratories to test these hypotheses using data processing methods of their own preference. No 2 laboratories chose identical analysis protocols. Their conclusions as to the validity of several hypotheses are consequently disparate, although meta-analysis does yield a consensus. Among the parameters most influential in determining differences in the conclusions reached are degree of spatial smoothness, the software package employed, and the method of multiple testing correction. The authors recommend that researchers share their unthresholded activity maps so that the community may undertake consensus building by meta-analysis. The authors also recommend that researchers share their data processing code so that precise replication of processing conditions may be accomplished among teams and suggest that researchers should process data by multiple methods to assess the robustness of their conclusions.