Article Watch: April 2014

Clive A. Slaughter

GRU-UGA Medical Partnership, Athens, GA

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, Georgia Regents University/University of Georgia Medical Partnership, 1425 Prince Ave., Athens, GA 30606, USA (Phone: 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.

DNA SEQUENCING AND CHARACTERIZATION

Steijger T, Abril J F, Engstrom P G, Kokocinski F, The R C, Hubbard T J, Guigo R, Harrow J, Bertone P. Assessment of transcript reconstruction methods for RNA-seq. *Nature Methods* 10;2013:1177–1184.

Engström P G, Steijger T, Sipos B, Grant G R, Kahles A, The R C, Ratsch G, Goldman N, Hubbard T J, Harrow J, Guigo R, Bertone P. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods* 10;2013:1185–1191.

This pair of papers documents the results of a quality assessment for RNA-seq algorithms that has been conducted by the RNA-seq Genome Annotation Assessment Project Consortium. Steijger et al. evaluate computational methods for exon identification, transcript reconstruction, and expression-level quantification, whereas Engström et al. scrutinize how well partial transcript reads can be aligned to a reference genome. Algorithms for identification of exons deliver high success rates, but assembly and quantification vary widely from method to method. Alignment methods also vary widely in major respects when tested on human and mouse data sets. Given these limitations, investigators using these procedures are encouraged to exercise caution in interpreting RNA-seq data.

Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. *Nature Biotechnology* 31;2013:1009–1014.

Although contemporary methods for short-read RNA sequencing are providing penetrating insight into gene expression, exhaustive cataloging of differentially spliced isoforms cannot be accomplished by reassembling short reads. In the present study, Sharon et al. test the capabilities of the Pacific Biosciences sequencing platform to provide long, error-corrected, circular consensus reads of entire single cDNA molecules. The authors analyze pooled RNA from 20 human organs and tissues without fragmentation or amplification, producing a total of 476,000 reads. The majority contains all of the splice sites from the original transcripts. Substantial numbers of previously unannotated splice isoforms are identified, some of which represent candidates for long noncoding RNAs with few introns, and others represent the products of known, protein-coding genes with additional introns. The results indicate that deep sequencing of full-length RNA at the single-molecule level is practicable and yields new information about alternatively spliced isoforms.

MACROMOLECULAR SYNTHESIS AND SYNTHETIC BIOLOGY

Wang P, Dong S, Shieh J-H, Peguero E, Hendrickson R, Moore M A S, Danishefsky S J. Erythropoietin derived by chemical synthesis. *Science* 342;2013:1357–1360.

Wang et al. here announce the total chemical synthesis of the glycoprotein hormone erythropoietin (EPO). The project was initiated in 2002. Recombinant EPO is a highly successful, biologic pharmaceutical consisting of a 166-residue polypeptide chain to which four complex carbohydrate chains—three *N*-linked and one *O*-linked—are attached. The biologic product is, of course, heterogeneous in numerous respects, notably in the structure of its carbohydrate side-chains. The goal of total chemical synthesis was to produce a chemically homogeneous entity, whose glycan chains mimic the complexity of EPO synthesized in vivo. The polypeptide portion was assembled by sectional synthesis using the native chemical ligation methodology pioneered by Steve Kent. This process requires a cysteine

doi: 10.7171/jbt.14-2501-004

residue at the ligation site between the peptides, but the present work adapted desulfurization methods to convert the cysteine residues to alanine to overcome unfavorable placement of cysteines in the native structure of EPO. The *N*-linked glycan chains were synthesized as a consensus branch-chain structure containing 12 monosaccharide units. The *O*-linked chain was a glycophorin tetrasaccharide. The *N*-linked glycans were attached to the polypeptide by combining the carboxylate group of an aspartic acid with an amine group on the terminal glycan, a process that yields an asparagine linkage. To form disulfide bridges, cysteine/cysteine was used as a redox-shuffling reagent. The biological activity of folded molecule is closely similar to that of the recombinant EPO, Procrit. This work provides methodology that can be used to prepare other glycopeptides of defined structure. The availability of such defined molecules enables the functional correlates of their structural features to be elucidated. Such homogeneous, defined molecules also represent "gold standards", against which the properties of biologic preparations may be compared.

MASS SPECTROMETRY

Hebert A S, Merrill A E, Stefely J A, Bailey D J, Wenger C D, Westphall M S, Pagliarini D J, Coon J J. Amine-reactive neutron-encoded labels for highly plexed proteomic quantitation. *Molecular & Cellular Proteomics* 12;2013: 3360–3369.

Richards A L, Vincent C E, Guthals A, Rose C M, Westphall M S, Bandeira N, Coon J J. Neutron-encoded signatures enable product ion annotation from tandem mass spectra. *Molecular & Cellular Proteomics* 12;2013:3812–3823.

Ulbrich A, Merrill A, Hebert A, Westphall M, Keller M, Attie A, Coon J. Neutron-encoded protein quantification by peptide carbamylation. *Journal of the American Society for Mass Spectrometry* 25;2014:6–9.

This series of papers charts the development and application of a novel labeling strategy for quantification of proteins in mass spectral proteomic analysis. It makes use of the high resolution available with modern, commercially available Fourier transform ion cyclotron resonance and Orbitrap mass spectrometers, which allows these instruments to discriminate mass defects—the small mass differences between atoms that result from variation in nuclear-binding energies among common stable isotopes. For example, replacement of a single ^{14}N by ^{15}N, while concomitantly replacing a ^{13}C by a ^{12}C, produces a mass change of 6.3 mDa. With the readily attainable resolving power of 30,000, this shift can be detected. Hebert et al. show that amino acids or oligopeptides with suitable isotopic replacements may be used as amine-reactive labels to create chemically identical isotopologs that are indistinguishable when analyzed at low resolution but sufficiently separated at high resolution to enable their abundances to be extracted for quantifying peptide analytes across diverse experimental conditions. Multiplexing up to 12-plex is demonstrated with encoded tags, differing by successive increments of 12.6 mDa. The labels permit quantification from precursor ion scans—an important feature that avoids the distortion of fold-change values as a result of precursor co-isolation that is seen with isobaric labeling strategies, which rely on product ion scanning for quantification. Richards et al. deploy the mass-defect principle for 2-plex stable isotope labeling of amino acids in cell culture, where two isotopologs of lysine, differing by 36 mDa, are used for labeling. In tandem mass spectrometry scans, acquired at resolution of 50,000 or greater, product ions containing a C-terminal lysine appear as a doublet, allowing *y*-ions to be readily identified. This capability augments accuracy of peptide identification. Ulbrich et al. perform labeling by carbamylation with isotopically labeled forms of urea. A mass difference of 6.3 mDa is thereby introduced and is used for duplex analysis at a resolving power of 50,000. This new general approach provides a basis for improvements in the accuracy and the productivity of studies in proteomic quantification.

Hebert A S, Richards A L, Bailey D J, Ulbrich A, Coughlin E E, Westphall M S, Coon J J. The one hour yeast proteome. *Molecular & Cellular Proteomics* 13;2014:339–347.

The capabilities in proteomic analysis of a new generation of Orbitrap hybrid mass spectrometers are outlined in this publication. The work tests a new instrument consisting of a mass-resolving quadrupole, a collision cell, a high-field Orbitrap analyzer, and a dual linear ion trap analyzer. The configuration of this Orbitrap fusion instrument is Q-OT-qIT, in contrast to the preceding Orbitrap Elite—the configuration of which is qIT-OT. The additional dual cell linear ion trap has a very fast scan speed, and the new instrument incorporates a control system that parallelizes ion injection, precursor isolation, fragmentation, and mass analysis to achieve a tandem mass spectrometry (MS^2) acquisition rate of 20 Hz—a 2× improvement compared with the Orbitrap Elite. A tryptic digest of the yeast proteome is subjected to liquid chromatography-MS analysis. The new instrument identifies ~8 peptides/s, sometimes rising to 19 peptides/s. During a 1.3-h chromatographic method, it performs an average of 13,447 MS^1 scans and 80,460 MS^2 scans. These yield 43,400 spectral matches and result in identification of 34,255 different peptides. The number of protein identifications at 1% false discovery rate is 3977. This performance represents unprecedented depth of coverage in such a short time-frame and encour-

ages the hope that routine, comprehensive analysis of mammalian proteomes will eventually become a reality.

Wang D, van Amerom F H W, Evans-Nguyen T. High-speed digital frequency scanning ion trap mass spectrometry. *Analytical Chemistry* 85;2013:10935–10940.

Wang et al. explore ways of increasing the scan rate of three-dimensional ion traps to better match the very sharp peaks produced in contemporary chromatographic techniques. Conventionally, ion traps are operated by ramping the fundamental radio-frequency (rf) trapping voltage to destabilize the orbits of ions with successively higher mass so that they are ejected into a detector. In the present work, the rf of the ring electrode is instead subject to scanning. This is done with high-speed, high-voltage switches, driven by low-voltage, square waveforms. The normal pulsed injection is replaced by a continuous mode of injection. The authors achieve acquisition rates of up to 1000 Hz and scan rates of 400,000 Th/s. Although such fast scan rates are attained at the expense of resolution, the results suggest that for some applications, perhaps including certain metabolomic analyses, an appropriate compromise between scan rate and resolution would prove more useful than presently available methods.

PROTEINS: PURIFICATION AND CHARACTERIZATION

Liu W, Wacker D, Gati C, Han G W, James D, Wang D, Nelson G, Weierstall U, Katritch V, Barty A, Zatsepin N A, Li D, Messerschmidt M, Boutet S, Williams G J, Koglin J E, Seibert M M, Wang C, Shah S T A, Basu S, Fromme R, Kupitz C, Rendek K N, Grotjohann I, Fromme P, Kirian R A, Beyerlein K R, White T A, Chapman H N, Caffrey M, Spence J C H, Stevens R C, Cherezov V. Serial femtosecond crystallography of G protein-coupled receptors. *Science* 342;2013:1521–1524.

This paper exemplifies ongoing progress in the use of serial femtosecond (fs) crystallography for determining protein structure. This technique provides a way to obtain high-resolution X-ray diffraction data at room temperature (i.e., without cryo-cooling) from micrometer- and submicrometer-size protein crystals. The importance of the approach lies in the difficulty of obtaining larger crystals with many proteins. The methodology uses high-intensity X-ray free-electron lasers to irradiate crystals in very short pulses of $<$50 fs. High-resolution diffraction data are collected from a single crystal in a single burst before the crystal is destroyed by the incident energy. A stream of crystals is passed continuously through the X-ray beam in random orientations to build up enough "snapshot" data to solve the structure. Liu et al. study the serotonin 5-hydroxytryptamine receptor 2B, bound to its agonist ergotamine. This G-protein coupled receptor represents a difficult protein to crystalize. Sub-10-μm crystals are obtained in this study using the membrane mimetic environment of a lipidic cubic phase. Data collection methodology is refined to enable a structure to be calculated at 2.8 Å resolution, using only 300 μg protein. Phasing is accomplished by molecular replacement. With continued development, the new technology is expected to open the door to structural study of many previously intractable proteins.

Kirkwood K J, Ahmad Y, Larance M, Lamond A I. Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. *Molecular & Cellular Proteomics* 12; 2013:3851–3873.

Gordon S M, Deng J, Tomann A B, Shah A S, Lu L J, Davidson W S. Multi-dimensional co-separation analysis reveals protein–protein interactions defining plasma lipoprotein subspecies. *Molecular & Cellular Proteomics* 12;2013:3123–3134.

These papers represent the work of teams investigating different problems but using the common approach of physical fractionation of a mixture of native protein complexes, followed by proteomic analysis of the fractions to make deductions about the composition and function of the complexes. Kirkwood et al. perform native, size-exclusion chromatography to characterize soluble protein complexes in human osteosarcoma cells. By monitoring co-separation across the elution profile, the authors are able to conclude that some proteins associate with multiple independent complexes and that distinct post-translationally modified forms of certain proteins associate with distinct subsets of protein complexes. Gordon et al. set out to study the heterogeneity of high-density lipoprotein (HDL) particles in blood plasma. They separate plasma by size exclusion and ion-exchange chromatography and by isoelectric focusing and track the co-separation of 76 lipid-associated proteins. The data indicate the existence of distinct, stable subspecies of HDLs. These two studies demonstrate that proteomic analysis of fractionated complexes of native proteins can provide important insight into protein function.

PROTEOMICS

Lambert J-P, Ivosev G, Couzens A L, Larsen B, Taipale M, Lin Z-Y, Zhong Q, Lindquist S, Vidal M, Aebersold R, Pawson T, Bonner R, Tate S, Gingras A-C. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nature Methods* 10;2013:1239–1245.

Collins B C, Gillet L C, Rosenberger G, Rost H L, Vichalkovski A, Gstaiger M, Aebersold R. Quantifying protein

interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nature Methods* 10;2013: 1246–1253.

This pair of papers documents quantitative changes in protein–protein interactions, elicited by a variety of biological interventions. Lambert et al. study changes in interactions of heat shock protein 90 (HSP90), as a result of an HSP90 inhibitor, and changes in interactions of the human kinase, cyclin-dependent kinase 4 (CDK4), as a result of melanoma-associated mutations in CDK4. Collins et al. study a time-course of changes in interactions of the 14-3-3β scaffold protein, following stimulation of the insulin-PI3K-AKT pathway. Both studies use affinity chromatography to purify interacting proteins and quantify the proteins by mass spectrometry (MS). The novel feature of both studies is the methodology used for mass spectrometric quantification, a procedure named sequential window acquisition of all theoretical spectra, which is a recently described implementation of data-independent acquisition of product ion spectra. Unlike the standard multiple reaction-monitoring approach to data-independent acquisition, product ion spectral information is collected in toto without preselection. Unlike standard data-dependent (shotgun) schema, it avoids the stochastic variation, inherent in precursor ion detection that makes the absence of peptides in a dataset so difficult to substantiate. The procedure involves consecutive acquisition of high-resolution, accurate-mass product ion spectra by repeatedly stepping through a set of discrete, broad isolation windows of width 25 Da, known as swathes. The product ions in these swathes are connected with their various precursor ions by reference to a spectral library. Such spectral libraries may be derived from publically available resources but in the present studies, are generated from data-dependent (shotgun) analyses of samples within each experimental set. A complete experimental pipeline for deployment of this methodology is provided in these two papers, including statistical procedures for determining fold changes with associated confidence intervals. The methodology is expected to provide a robust, label-free approach for using MS to compare complex biological samples that overcomes many of the technical challenges encountered with current methods.

Bronsema K J, Bischoff R, van de Merbel N C. High-sensitivity LC-MS/MS quantification of peptides and proteins in complex biological samples: the impact of enzymatic digestion and internal standard selection on method performance. *Analytical Chemistry* 85;2013:9528–9535.

Bronsema et al. describe the effects of tryptic digestion and various kinds of quantification standards on the sensitivity and specificity with which an example small protein can be quantified by liquid chromatography-tandem mass spectrometry. The protein chosen for study is salmon calcitonin, a 32-aa peptide of molecular mass 3431.9 Da. This protein can be quantified in plasma after solid-phase extraction, without digestion. The measurement gives a lower limit of quantification of 10 pg/mL. After digestion with trypsin and quantification, based on selected reaction monitoring of a signature peptide, however, the best lower limit of quantification achieved was 50 pg/mL. The poorer performance is attributed to the added complexity of the mixture being analyzed when a trypsin digestion step is introduced. The performance degradation is a result of interferences from structurally similar peptides created by the digestion. The study goes on to compare the use of various analog and stable isotope-labeled internal standards for quantification and concludes that the best standard for use in analyses that include or omit digestion is a stable, isotope-labeled analyte. Although the optimal methods may vary from one peptide or protein to another, this work will be of interest to investigators embarking on any new peptide or protein quantification study using MS. The work provides a model of the consideration that must be given to the numerous details of the protocol under development.

FUNCTIONAL GENOMICS AND PROTEOMICS

Maeder M L, Angstman J F, Richardson M E, Linder S J, Cascio V M, Tsai S Q, Ho Q H, Sander J D, Reyon D, Bernstein B E, Costello J F, Wilkinson M F, Joung J K. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nature Biotechnology* 31;2013:1137–1142.

Mendenhall E M, Williamson K E, Reyon D, Zou J Y, Ram O, Joung J K, Bernstein B E. Locus-specific editing of histone modifications at endogenous enhancers. *Nature Biotechnology* 31;2013:1133–1136.

Konermann S, Brigham M D, Trevino A E, Hsu P D, Heidenreich M, Le C, Platt R J, Scott D A, Church G M, Zhang F. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500;2013:472–476.

Experiments correlating various kinds of epigenetic modification with gene-expression patterns and experiments involving knockout of chromatin-modifying factors have provided important information about epigenomic effects. But, the phenomena elucidated by such studies are genome-wide changes as a result of epigenetic modification, not the particular effects at specific gene loci. The three present studies break new ground by targeting epigenetic marks at specific genes. The importance of this can be appreciated when it is realized that DNA modifications, such as methylation or acetylation, may have different

effects, depending on whether they occur in an enhancer, promoter, insulator, or gene body. All three studies target their interventions to particular DNA sequences by fusing chromatin modifiers to transcription activator-like effector (TALE) proteins, customized for binding to the appropriate target DNA sequence. Maeder et al. fuse TALEs to the catalytic domain of ten–eleven translocation 1, a protein that demethylates 5-methylcytosine at CpG islands. They target the human *KLF4, RHOXF2* and globin genes and show specifically increased expression—an effect that is strongest for CpG targets upstream of the gene. Mendenhall et al. study histone H3 lysine methylation by fusing TALEs with lysine-specific demethylase 1, which demethylates histone H3 monomethyl/dimethyl Lys4 (H3K4me1 and H3K4me2, respectively). They target several putative enhancer sites in K562 erythroleukemia cells. Demethylation of H3 at some of these sites causes down-regulation of associated genes, confirming functional enhancer status. Konermann et al. combine chromatin-modifying activities with optogenetic control to study the effect of repressing specific genes reversibly in primary mouse neurons and in the brains of freely behaving mice. They use a variety of histone effector domains, including histone deacetylases (HDACs), methyltransferases (HMTs), acetyltransferase inhibitors, and HDAC- and HMT-recruiting proteins, and demonstrate repression of recruiting proteins, and demonstrate repression of *Grm2* and *Neurog2* transcription. These studies presage the ability to alter expression of specific target genes by modifying associated epigenetic marks and to probe the function of specific epigenetic modifications.

Shalem O, Sanjana N E, Hartenian E, Shi X, Scott D A, Mikkelsen T S, Heckl D, Ebert B L, Root D E, Doench J G, Zhang F. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343;2014:84–87.

Sung Y H, Kim J M, Kim H-T, Lee J, Jeon J, Jin Y, Choi J-H, Ban Y H, Ha S-J, Kim C-H, Lee H-W, Kim J-S. Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. *Genome Research* 24;2014:125–131.

Wang T, Wei J J, Sabatini D M, Lander E S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343;2014:80–84.

These three studies illustrate the incipient impact of the still-new clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 system of RNA-guided endonucleases for inducing gene-specific knockout in cells and whole animals. Shalem et al. construct a genome-scale CRISPR-Cas9 library, targeting 18,080 genes. Lentiviral delivery of this library is used to identify genes for cell viability in cancer and pluripotent stem cells and genes involved in resistance to the RAF inhibitor vemurafenib. Sung et al. establish gene knockout mice and gene knockout zebrafish by injecting CRISPR-Cas9 constructs into one-cell stage embryos of both species and show that the resulting mutations are germline-transmissible. Wang et al. use a library of 73,000 RNA guides, delivered via lentivirus, for a pooled loss-of-function screen. Expression of the pool is followed by massively parallel sequencing. The system is used to identify genes involved in resistance to the nucleotide analog 6-thioguanine. The scope and speed with which these studies have been completed inspire optimism for future productivity of the technology.

Auer T O, Duroure K, De Cian A, Concordet J-P, Del Bene F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Research* 24;2014:142–153.

Auer et al. demonstrate the feasibility of site-directed gene knockin with the CRISPR/Cas9 system. With the use of the double-stranded breaks, introduced by CRISPR/Cas9, they show knockin of plasmid DNA as large as 5.7 kb in zebrafish DNA. The approach is similar to that used successfully by others for knockin with zinc finger and TALE nucleases: the donor plasmid is co-injected into target cells with a short-guide RNA and Cas9 nuclease mRNA. Cleavage of the donor plasmid DNA, concurrent with the selected chromosomal receptor site, results in efficient integration of donor DNA into that site. The process is mediated by nonhomologous recombination and is found to be substantially more efficient than methods relying on homologous recombination.

CELL BIOLOGY AND TISSUE ENGINEERING

Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour A A, Caspi I, Krupalnik V, Zerbib M, Maza I, Mor N, Baran D, Weinberger L, Jaitin D A, Lara-Astiaso D, Blecher-Gonen R, Shipony Z, Mukamel Z, Hagai T, Gilad S, Amann-Zalcenstein D, Tanay A, Amit I, Novershtern N, Hanna J H. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 502;2013:65–70.

Reprogramming of differentiated cells into induced pluripotent stem cells can be achieved by ectopic expression of just four transcription factors: Oct4, Sox2, Klf4, and Myc. However, this reprogramming remains inefficient, asynchronous, and apparently stochastic. The epigenetic chromatin de-repression required for successful reprogramming is apparently countermanded by the nucleosome remodeling and acetylation repressor complex. Interestingly, sustained expression of the reprogramming factors recruits this inhibitory complex, ironically impeding the trajectory toward pluripotency. Rais et al. perform a loss-of-function screen to

reveal that by depleting methyl-CpG-binding domain protein 3, a core component of this complex, reprogramming becomes synchronous and deterministic, reaching nearly 100% efficiency. This discovery suggests how to overcome a key impediment to the production of pluripotent stem cells for medical research and provides key insight into the molecular events that determine cell fate, both in normal development and potentially, in abnormal states, such as cancer.

Potter N E, Ermini L, Papaemmanuil E, Cazzaniga G, Vijayaraghavan G, Titley I, Ford A, Campbell P, Kearney L, Greaves M. Single-cell mutational profiling and clonal phylogeny in cancer. *Genome Research* 23;2013:2115–2125.

Cancers are well-known to display complex patterns of acquired mutations that may differ among cancer subtypes, among patients sharing the same subtype, and among clonal lineages within the same tumor. Mutational patterns also change as the disease progresses. Mutational profiling is used increasingly for diagnosis, prognosis, and treatment planning. Knowledge of mutational complexity is relevant for each of these clinical activities. The authors here undertake single-cell genetic profiling of cancer cells. In a proof-of-principle study of acute lymphoblastic leukemia, they demonstrate that DNA from flow-sorted single cells may be screened by multiplexed, high-throughput, targeted quantitative PCR in a commercially available microfluidic platform. Gene fusion, copy number, and single nucleotide variants are all detectable in single cells, and 200–300 leukemic cells are analyzed from individual patients to help ensure representative sampling. With the use of these data, the clonal architecture and phylogeny of cells within tumors are deduced. Given the presence of two rare variants within the population of tumor cells, for example, it is now possible to tell whether the two mutations occur in the same or in different cells in the population. The data from the method are expected to be useful for prognosis and for choosing therapeutic targets, for example, by focusing on mutations present in all cells.

IMAGING

Wu Y, Wawrzusin P, Senseney J, Fischer R S, Christensen R, Santella A, York A G, Winter P W, Waterman C M, Bao Z, Colon-Ramos D A, McAuliffe M, Shroff H. Spatially isotropic four-dimensional imaging with dual-view plane illumination microscopy. *Nature Biotechnology* 31;2013:1032–1038.

Confocal fluorescence microscopy has long been the standard method for imaging genetically expressed markers in live samples. It achieves high axial resolution, despite unfocused laser illumination, by masking the out-of-focus light with a pinhole before detection, but the "unused" illumination still contributes to phototoxicity and photobleaching of the sample. Selective plane illumination microscopy (SPIM) offers a way around this problem by illuminating the sample with a thin sheet of light in a direction orthogonal to the detection path so that only the portion of the sample in the focal plane of the detector is illuminated. But, depth (z-axis) resolution remains poorer than x- and y-axis resolution. In a refinement of the SPIM method by Wu et al., two orthogonal objectives, positioned above the sample, alternate between excitation and detection in a rapid duty cycle. The illuminating objective scans a light-sheet through the depth of the sample, while the detecting objective acquires an image volume sharply focused in the x- and y- direction but poorly in the z-direction. Then, the two objectives switch function, and a second, orthogonal image volume is acquired. A joint deconvolution algorithm then merges the two image volumes into a single image volume with good resolution along all three dimensions (isotropic resolution of 330 nm). The entire sample volume ($70\times70\times50$ μm) is imaged in a 0.5-s duty cycle, which is $10\times$ faster than the fastest comparable method. Wu et al. use this methodology to follow microtubule dynamics in live cells, to image nuclei over a 14-h period of nematode embryogenesis, and to image neuronal wiring development in nematodes over a 5-h period, all with minimal phototoxicity. The methodology is expected to provide a robust, practical replacement for confocal fluorescence microscopy.

COMPUTATIONAL BIOLOGY

Denny J C, Bastarache L, Ritchie M D, Carroll R J, Zink R, Mosley J D, Field J R, Pulley J M, Ramirez A H, Bowton E, Basford M A, Carrell D S, Peissig P L, Kho A N, Pacheco J A, Rasmussen L V, Crosslin D R, Crane P K, Pathak J, Bielinski S J, Pendergrass S A, Xu H, Hindorff L A, Li R, Manolio T A, Chute C G, Chisholm R L, Larson E B, Jarvik G P, Brilliant M H, McCarty C A, Kullo I J, Haines J L, Crawford D C, Masys D R, Roden D M. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature Biotechnology* 31;2013:1102–1111.

Phenome-wide association is a recently developed method for exploring the influence of genotype on disease susceptibility. Whereas genome-wide association begins with a phenotype and seeks to identify genetic markers associated with that phenotype, phenome-wide association begins with genetic markers and seeks associated phenotypes. In principle, the strength of phenome-wide association studies is that they are capable of detecting pleiotropic effects of genes if they detect association between a given gene and hitherto unconnected phenotypic effects. The

present study expands the scope of the methodology by screening a very large number of genetic markers: 3144 single nucleotide polymorphisms. The authors look for associations between each of these genes and 1358 disease phenotypes among 13,835 individuals of European descent, recorded mainly as diagnostic codes in electronic medical records. They compare the detected associations with those discovered by genome-wide association and conclude that their method replicates 66% of those associations, appropriately filtered using stated criteria. The strongest novel associations putatively of a pleiotropic nature are validated in a separate population. Differences in pleiotropy among markers linked to a common disease are of interest, as they suggest differences in mechanism of action. The methodology is expected to increase the speed and efficiency of detecting pleiotropic associations. More generally, the work pioneers new procedures for mining electronic medical records for genetic and clinical research.

Zhao S, Kumar R, Sakai A, Vetting M W, Wood B M, Brown S, Bonanno J B, Hillerich B S, Seidel R D, Babbitt P C, Almo S C, Sweedler J V, Gerlt J A, Cronan J E, Jacobson M P. Discovery of new enzymes and metabolic pathways by using structure and genome context. *Nature* 502; 2013:698–702.

Bastard K, Smith A A T, Vergne-Vaxelaire C, Perret A, Zaparucha A, De Melo-Minardi R, Mariage A, Boutard M, Debard A, Lechaplais C, Pelle C, Pellouin V, Perchat N, Petit J-L, Kreimeyer A, Medigue C, Weissenbach J, Artiguenave F, De Berardinis V, Vallenet D, Salanoubat M. Revealing the hidden functional diversity of an enzyme family. *Nature Chemical Biology* 10;2014:42–49.

These two studies use recent developments in methods for discovering the functions of the 50–60% of bacterial genes presently lacking reliable annotation. Zhao et al. begin with a structurally characterized protein of unknown function. With the rationalization that bacterial metabolic pathway members are often encoded within operons, they perform docking studies of their target protein and its genomic neighbors and their homologs. They predict that the target protein catalyzes a previously unknown reaction—2-epimerization of a hydroxyproline betaine—and confirm this prediction experimentally. They then use metabolomics to deduce a catabolic pathway, in which the enzyme participates, and confirm the predictions by in vitro assays and genetic analyses. Bastard et al. begin with an enzyme involved in lysine fermentation. They establish that it catalyzes a generic β-keto acid hydrolysis reaction. They then align all sequence homologs of the enzyme and choose a set of sequences representative of different bacterial species of origin, different sequence subclusters, and different regional genomic contexts. Expression of these representatives is then followed by high-throughput biochemical testing of substrate specificity and homology modeling and computational substrate docking to deduce structural elements responsible for the observed activities. They finally study genomic and metabolic contexts to determine physiologic roles of the broad family of activities that they discover. These studies show how computational and biochemical strategies may be combined to discover the function of unannotated bacterial proteins.

POLICY

McShane L M, Cavenagh M M, Lively T G, Eberhard D A, Bigbee W L, Williams P M, Mesirov J P, Polley M-Y C, Kim K Y, Tricoli J V, Taylor J M G, Shuman D J, Simon R M, Doroshow J H, Conley B A. Criteria for the use of omics-based predictors in clinical trials. *Nature* 502;2013:317–320.

This paper summarizes criteria for deciding whether tests based on high-dimensional molecular measurements, derived from research in disciplines such as genomics, transcriptomics, proteomics, metabolomics, and epigenomics, are ready for use in guiding decisions on patient care in clinical trials. The criteria were drawn up at the instigation of the National Cancer Institute by scientists and other stakeholders. They cover issues related to the specimens required, the assay methods to be deployed, the models to be used for predicting clinical outcome, the design of the clinical trial, and ethical, legal, and regulatory considerations. The paper also provides examples of instances where adequate criteria were not met. It is hoped that the criteria will be useful for evaluating manuscripts submitted for publication and reaching funding decisions on research proposals, as well as helping guide the advancement of translational research. Several of the criteria also apply to more conventional tests using single biomarkers or limited panels of biomarkers.