National Cancer Institute Think-Tank Meeting Report on Proteomic Cartography and Biomarkers at the Single-Cell Level: Interrogation of Premalignant Lesions


Cite This: J. Proteome Res. 2020, 19, 1900−1912

ABSTRACT: A Think-Tank Meeting was convened by the National Cancer Institute (NCI) to solicit experts' opinions on the development and application of multiomic single-cell analyses, and especially single-cell proteomics, to improve the development of a new generation of biomarkers for cancer risk, early detection, diagnosis, and prognosis as well as to discuss the discovery of new targets for prevention and therapy. It is anticipated that such markers and targets will be based on cellular, subcellular, molecular, and functional aberrations within the lesion and within individual cells. Single-cell proteomic data will be essential for the establishment of new tools with searchable and scalable features that include spatial and temporal cartographies of premalignant and malignant lesions. Challenges and potential solutions that were discussed included (i) The best way/s to analyze single-cells from fresh and preserved tissue; (ii) Detection and analysis of secreted molecules and from single cells, especially from a tissue slice; (iii) Detection of new, previously undocumented cell type/s in the premalignant and early stage cancer tissue microenvironment; (iv) Multimomic integration of data to support and inform proteomic measurements; (v) Subcellular organelles—identifying abnormal structure, function, distribution, and location within individual premalignant and malignant cells; (vi) How to improve the dynamic range of single-cell proteomic measurements for discovery of differentially expressed proteins and their post-translational modifications (PTM); (vii) The depth of coverage measured concurrently using single-cell techniques; (viii) Quantitation - absolute or semi quantitative? (ix) Single methodology or multiplexed combinations? (x) Application of analytical methods for identification of biologically significant subsets; (xi) Data visualization of N-dimensional data sets; (xii) How to construct intercellular signaling networks in individual cells within premalignant tumor microenvironments (TME); (xiii) Associations between intrinsic cellular processes and extrinsic stimuli; (xiv) How to predict cellular responses to stress-inducing stimuli; (xv) Identification of new markers for prediction of progression from precursor, benign, and localized lesions to invasive cancer, based on spatial and temporal changes within individual cells; (xvi) Identification of new targets for immunoprevention or immunotherapy—identification of neoantigens and surfactome of individual cells within a lesion.

KEYWORDS: single-cell proteomics, single-cell mass spectrometry, targeted proteomics, precursor lesion, precancer, tumorigenic lesion, lesion's heterogeneity, clonal evolution, spatial and temporal cartography, biomarkers, early detection, targets for prevention and therapy

INTRODUCTION

On April 23−24, 2019, the National Cancer Institute, Division of Cancer Prevention, held a Think-Tank meeting on “Proteomic Cartography and Biomarkers at the Single-Cell Level: Interrogation of Premalignant and Early Stage Lesions” to solicit experts’ opinions on application of multiomic single-cell analyses and especially proteomics based single-cell analyses for discovery of a new generation of cancer biomarkers, for risk, early detection, diagnosis, early prognosis, and for identification of new targets for cancer prevention.

Currently, the gold standard for diagnosis and classification of premalignant and cancerous lesions is based on histopathological examination of hematoxylin and eosin (H&E) stained slides from cancer biopsies and resected tumors. More detailed...
targeted examination is often based on immunohistochemical examination of a limited number of markers or examination of a limited number of targeted gene expression by in situ hybridization. Although such assays are very valuable in determining the grade and stage of the tumor, they provide limited information on the molecular and cellular content of the tumorigenic lesion, the physiological state of the cells within the lesion, and tumor heterogeneity. However, in many instances, the molecular information is pivotal for proper diagnosis and prognosis. For example, the t(9;22)(q34;q11) chromosomal translocation and the bcr-abl fusion transcript are the hallmarks of chronic myeloid leukemia (CML). Furthermore, the abl gene product is a therapeutic target for Dasatinib, which competitively targets the abl kinase domain in CML with bcr-abl chimeric proteins. A deletion of a tumor suppressor gene (TSG) PTEN is a bad prognostic marker in prostate cancer. Mutations in driver genes, such as EGFR, are considered prognostic markers in non-small-cell lung carcinoma (NSCLC) and targets for therapy, and the detection of fusion transcripts of ALK-EML4 in NSCLC are important prognostic markers for successful application of targeted therapy.

Applications of genomic, transcriptomic, and epigenomic deep sequencing analyses identified additional driver events, which resulted in aberrant biochemical pathways and biological programs that can be used to subclassify specific tumors and identify unique vulnerabilities for targeted therapies. Such analyses were based on bulk tumor tissue analysis, which frequently fails to capture the tumor heterogeneity; the molecular and cellular spatial and temporal composition; aberrant structures within lesions (e.g., vascular mimicry); aberrant composition and spatial distribution of organelles within individual cells (cancerous and noncancerous cells); and the aberrant location, distribution or composition of molecular complexes and their functional interactions within individual cells, all of which could be important cofactors in driving tumor evolution. Identification of such changes at the cellular level is essential for discovery of new generation biomarkers and design of new targeted preventive and interventional therapies.

Recent developments in single-cell genomic and transcriptomic technologies enable a better understanding of the cellular content of normal, tumorigenic, and metastatic lesions, including the precise spatial location of different cell types (e.g., cancerous and normal epithelial cells, cancer stem cells, mesenchymal cells, immune cells [myeloid, T and B cells], pericytes, fibroblasts), deciphering the cell’s physiological state, detection of aberrant gene expression indicative of the emergence of new, more aggressive or drug resistant clones.

Moreover, single-cell proteomics is emerging as another powerful approach for phenotypic characterization of individual cell types and their physiological state, protein quantitative measurements, and the detection of cancer related post-translational modifications (PTM) that cannot be predicted by genomic/transcriptomic analysis. Interestingly, comparison between transcriptomic and proteomic analysis of The Cancer Genome Atlas (TCGA) samples revealed that proteome profiling outperforms transcriptome profiling for coexpression-based gene function prediction. Perhaps one of the most exciting applications of single-cell proteomic technologies is for quantitative detection of targets for immune intervention including inhibitory receptors for cytotoxic T cell response, also known as immune checkpoints (e.g., PD1 and PDL1, CTLA-4) and the identification of neoantigens and MHC-associated neoantigens.

Single-cell proteomic data will be essential for the establishment of new tools such as searchable and scalable features that include spatial and temporal cartographies of premalignant and malignant lesions, starting with high resolution three-dimensional (3D) organ structure and function through 3D tissue structural and cellular organization to characterization of individual cell-types, with their unique structures and subcellular organelle organization, and finally to detailed high resolution quantitative molecular and functional characterization. It is currently believed that these tools will enable the development of a new generation of cancer markers.

**Spatial Characterization of Human Tumors**

It is thought that tumors frequently evolve from benign precancerous precursor lesions. Spatial and temporal cartography of precancerous and tumorigenic lesions, based on multidisciplinary approaches including pathology, multomic analysis, imaging, and computational modeling are likely to identify new aberrant structures, and better define the composition, location, and molecular and cellular interactomes; identify new macromolecular complexes within individual cells and organelles; identify new aberrations in cells and organelles within those cells; identify secreted mediators within the tumor microenvironment (TME), some of which may be codrivers of tumor evolution. Furthermore, spatial and temporal cartography may allow better understanding of how invasive protrusions are developing and provide new targets for reversion or inhibition of tumor invasion. The molecular and cellular drivers of these processes could serve as cancer markers and targets for prevention and therapy.

Dr. Jessica L. Riesterer from Dr. Joe Gray’s laboratory presented their studies on the molecular and spatial characterization of human tumors with the emphasis on microscale 3D electron microscopy imaging. Interrogation of cellular ultrastructure of cancer cells and the microenvironment within tumors is becoming increasingly important with respect to therapeutic targeting and understanding what role individual cells play in tumor development. For example, cell-cell interactions found in breast cancer have shown important signaling mechanisms occurring on filopodia-like protrusions (FLPs) that “reach out and touch” neighboring cells, potentially influencing cancer proliferation. Nuclear invaginations observed in cultured cells and tumor tissue indicate that cancer cells have a distinct nuclear morphology different from normal tissue. These described behaviors, and other biological relationships present at a microto-nanoscale range, are best understood in the realm of three-dimensional interactions to truly model in vivo microenvironments. Three-dimensional electron microscopy (3DEM) via scanning electron microscopy techniques (SEM) have recently become a mature technology within the life sciences. High resolution 3DEM data sets have been collected via focused ion beam—and serial block face—scanning electron microscopy (FIB-SEM and SBF-SEM, respectively) on human cancer specimens and have aided in the elucidation of ultrastructural changes in individual cells and their surrounding microenvironment. High resolution 3DEM imaging can be applied to cancer studies via cell cultures, xenografts, and organoids in addition to early lesions and advanced metastases found from studying human biopsies. When volume imaging is combined with correlative light and electron microscopy (CLEM), biomarkers and ultrastructure can be viewed.
simultaneously to explain unique observations that can be subsequently used to formulate and answer new questions regarding the natural history of cancer progression. Future directions include but are not limited to investigations concerning observed ultrastructural changes; correlation of findings with multiomic analysis; improvements in tissue processing to enhance high resolution 3DEM; address bottlenecks such as data visualization and understanding functional relationships; application of the developing technologies to analyze additional tumor types.

Lessons Learned from Application of Recent Single-Cell Genomic and Transcriptomic Technologies

Recent applications of single-cell genomic and transcriptomic technologies revolutionized the analysis of complex tissues and disease disorders including cancer. Single-cell sequencing revealed earlier detection of tumor heterogeneity and identified new rare cell types within tissues and cancerous lesions. Furthermore, new molecular features were correlated with tumor aggressiveness and clinical outcomes. However, by dissociating individual cells and extracting RNA, there is a loss of information on spatial context. Spatial context, both at the tissue and intracellular level, can be clinically important, and new measurements offer new avenues for tumor stratification, diagnosis, prognosis, and the development of targets for prevention and therapy.

Image-based approaches to single-cell transcriptomics are emerging as powerful complements to single-cell RNA sequencing, in part, because these techniques preserve the native spatial context of RNAs within cells and tissues. Dr. Moffitt described the Multiplexed Error Robust Single-Molecule Fluorescence In Situ Hybridization (MERSFISH), a technique capable of imaging thousands of different RNAs simultaneously in fixed cells, and its use for the discovery and mapping of cell types within intact tissues. Because this technique is so highly sensitive and efficient in its detection efficiency, natural variations in gene expression from cell to cell are likely to reflect true biological features. However, as the think-tank discussed, potential challenges include an inability to identify new isoforms, which were formed due to alternative splicing or the expression of cell-infected viral RNA. Dr. Moffitt stated that their team plans to increase the content of their libraries of tagged-probes to increase the coverage of all known transcripts and potential splice variants with multiple probes for each of the transcribed genes; also, they will increase the coverage of viral RNA that may have infected certain cells.

Another attractive technology, fluorescent in situ sequencing (FISSEQ), combines the spatial context of RNA-FISH and the global transcriptome profiling, and an indirect methodology to view global gene expression patterns 3D was presented by Dr. Je Hyuk Lee. The technology is amenable for scRNA-seq. Dr. Je Hyuk Lee also discussed a new strategy for screening functional de novo mutations using an RNA-templated NGS method using programmable k-mers that represent deleterious codon mutation types (i.e., missense, nonsense, indels) to label single molecules or cells. This platform provides an opportunity to screen, sort, or image rare cells directly based on the presence of functionally deleterious de novo mutations in an allelic-agnostic manner.

Dr. Orit Rozenblatt-Rosen discussed the application of single-cell genomic and transcriptomic technologies to better understand the complexity of a tumorigenic lesions and to create spatial cell-type specific tumor atlases. Advances in genomics have resulted in many translational discoveries, but our understanding has been hindered because the methods used were typically applied to bulk tissue thereby masking intricate functional complexities and confounding an understanding of underlying physiology. Thus, there is an enormous need to build an atlas of the cells that compose a tumor, their spatial organization, and functional relationships. Recent advances in single-cell and spatial genomics now provide an extraordinary opportunity for building a Tumor Cell Atlas. This will lead to comprehensive identification of cells and their states and genetic clones in a tumor and relate them spatially to each other and to the overall tumor and assess their impact on resistance or response to therapy. For example, by leveraging single-cell RNA sequencing from melanoma tumors, their team identified a resistance program expressed by malignant cells that is associated with T cell exclusion and immune evasion. This program predicts clinical responses to anti-PD-1 therapy, and CDK4/6-inhibition represses in vitro and reduces melanoma tumor outgrowth in mouse models in vivo when given in combination with immunotherapy. To harness this new opportunity and demonstrate the feasibility and value of building cell atlases, the team has to (1) apply single-cell and spatial genomics to diverse tumor samples collected in a clinical setting; (2) develop and disseminate robust standard operating protocols, SOPs, that can be deployed across collection sites; (3) develop experimental design strategies, supported by appropriate power analysis, to determine the number of specimens, cells, and regions to be analyzed; (4) establish experimental and computational strategies to integrate across cellular and spatial data; and (5) develop robust and effective data sharing approaches. Taken together, the development of these capabilities will ensure a strong foundation for obtaining reproducible results and lay the foundation for future larger scale atlas efforts.

Single-cell RNA-seq methods are being increasingly applied in complex study designs that involve measurements of many samples commonly spanning multiple individuals, a variety of conditions and tissue compartments. Combined analysis of such extensive and often heterogeneous sample collections requires a way of identifying and tracking recurrent cell subpopulations across an entire collection. Dr. Peter Vasili Kharchenko described a flexible approach, called Conos (Clustering on Network of Samples), that relies on multiple plausible intersample mappings to construct a global graph connecting all measured cells. The graph can then be used to propagate information between samples and to identify cell communities that show consistent grouping across broad subsets of the collected samples. Conos results enable investigators to balance between resolution and breadth of the detected subpopulations. In this way, it is possible to focus on the fine-grained clusters appearing within more similar subsets of samples, or analyze coarser clusters spanning broader sets of samples in the collection. Such multiresolution joint clustering provides an important basis for downstream analysis and interpretation of sizable multisample single-cell studies and atlas-scale collections.

Dr. Eytan Ruppin presented his team’s efforts for identifying novel genetic interactions occurring between cancer genes that may open new drug treatment opportunities across the whole cancer genome. They have used the strategy of “synthetic lethality and rescue” interactions by directly mining patients’ tumor genomic data. Preliminary data showed that synthetic lethal interactions facilitated the prediction of a patient’s drug
response and provided new selective drug target candidates, while synthetic rescue interactions take part in mediating the resistance emerging to both targeted and immunotherapy.52 His laboratory intends to expand this strategy to single-cell DNA and RNA-seq analyses of tumors to improve the detection of additional targets that may exist in heterogeneous polyclonal populations of cancerous cells within tumors.

Dr. Robert Moritz summarized the discussions and highlighted the great progress made in genomic and transcriptomic technologies that are used to interrogate single cells in normal and neoplastic tissues. He emphasized the remaining challenges to the development of a holistic approach to study complex systems such as dynamic changes in individual cells within preneoplastic and neoplastic lesions. Incorporation of single-cell proteomic analysis is a major challenge due to limitations in sample preparation and sensitivity of mass spectrometric technologies. The new generation of technological developments are likely to improve the development of new markers and targets for prevention.

Single Cell Proteomics

Part I. Targeted Proteomics. Targeted proteomics is an approach that is focused on the detection of proteins of interest with high sensitivity, and most often, with quantitative accuracy and reproducibility. Most often, targeted proteomic methods are focused on a small subset of proteomes and frequently have high-throughput capabilities. The discussions in this section were focused on antibody based, mass spectrometry based, and a combination of both approaches in the phenotypic characterization of tissues, specific cell-types within the tissue, and cell specific organelles. The targeted approach is not a discovery approach because it may miss dynamic changes in the proteome including the expression of new or newly modified proteins in different physiological states and in the disease state. The discussion also focused on spatial and temporal cartography of normal, premalignant, and malignant lesions.

The Human Protein Atlas (HPA) is an international project that was initiated in 2003. The goal of the HPA is to map all the human proteins in cells, tissues, and organs using integration of various omics technologies, including genomics, transcriptomics, antibody-based imaging, mass spectrometry-based proteomics, and systems biology.53,54 The current version (www.proteinatlas.org) consists of three separate parts, each focusing on a particular aspect of the genome-wide analysis of the human proteins: (1) the Tissue Atlas showing the distribution of the proteins across all major tissues and organs in the human body,55,56 (2) the Cell Atlas showing the subcellular localization of proteins in single cells,57 and (3) Pathology Atlas showing the impact of protein levels for survival of patients with cancer.55

Much of the protein cellular and subcellular maps were developed using immunohistochemistry or immunofluorescence staining. Protein expression results were supplemented by gene expression analysis of the bulk tissue using RNA-seq. The laboratory generated ~50 000 human recombinant proteins and a similar number of antibodies. Dr. Fredrik Edfors discussed the application of the integrated omics approach based on transcriptomics analysis and antibody-based mapping in human, mouse, and porcine tissues to generate a detailed multilevel molecular view of the human proteome. Soon their team and international collaborators will launch a new version of the Blood Atlas, which will focus on the genome-wide expression analysis of single-cell types in human blood including various T-cells, B-cells, granulocytes, monocytes, and dendritic cells.58 Some limitations of their approach of cataloging differentially expressed proteins in cancer, especially for early detection and diagnosis of cancer, were discussed. For example, the comparison between normal hepatic tissue and liver cancer should be expanded to include fibrotic and cirrhotic liver tissues, since most liver cancers progress from fibrosis to cirrhosis to cancer.59

Dr. Fiona Ginty discussed the development of the highly multiplexed immunofluorescence platform (MxIF) by the GE Global Research Center.60 Some of the key drivers for this development have been a deepening interest in the biological and clinical significance of tumor heterogeneity, spatial and temporal cell biology, cell-to-cell interactions, and elucidation of the varied immune cell responses to tumors. Over the last 10 years her team has developed and validated this tissue multiplexing platform (Cell DIVE and MultiOmx) that generates multiplexed images of up to 60 proteins in a single FFPE tissue section via an iterative sequence of staining, imaging, and dye inactivation. Single-cell segmentation algorithms conserve spatial coordinates of the cells and provide potentially millions of biomarker data points coupled to unique cell identification numbers, IDs. The application of this high-dimensional imaging approach is revealing new insights into tumor biology, cellular behavior, and cellular interactions including immune response at the earliest stages of breast cancer (ductal carcinoma in situ (DCIS))61 tumor heterogeneity,62–64 breast cancer metabolism phenotypes,65 and characterization of immunotherapy response.66 There are many technical challenges that need to be considered when generating high resolution, highly multiplexed imaging data. For example, the generation of thousands of images and data integration at a cohort level require stringent and automated quality control measures before, during, and postimaging. Selective use of positive and negative reference controls is essential for data normalization and quantitation. Finally, translation to meaningful mechanistic insights requires deep multidisciplinary collaboration between researchers, pathologists, algorithm and imaging scientists, as well as access to high quality, well-annotated tissue samples.61,63

Dr. Yury Goltsve, from Dr. Garry Nolan’s laboratory at Stanford University, described a recently developed, a highly multiplexed cytometric imaging approach termed CO-Detection by indEXing (CODEX). CODEX is a unique technology that provides spatial and quantitative analysis of up to 50 biomarkers of individual cells in complex tissues. They have used this platform to characterize the microenvironment of normal and tumorigenic tissues.68 Examples included the visualization of invasion of lymphatic vessels by metastatic melanoma,69 visualization of tumor microenvironment (neighborhoods), and changes detected in tumorigenic lesions pre- and postadjuvant treatment of human breast cancer. A special set of paradigms for quantifying pairwise and combinatorial cell-to-cell interactions was established. Numerous foundational principles of tissue architecture (e.g., the prevailing homotypic adhesion) were observed in the data. There was a strong correlation between the composition of the cellular neighborhood and the expression of certain protein receptors on the immune cells within the lesion. Finally, multidimensional staining data provide a rich source for training and testing the diagnostic performance of unsupervised machine learning techniques such as convolutional neural networks. Altogether the fidelity of multiplexed spatial cytometry by CODEX allowed a reliable quantitative systematization of tissue...
architecture in normal and clinically aberrant samples. The discussion centered on how to set the thresholds for each cell-type in the neighborhood, the size of the neighborhoods, how many antibodies are needed to bind to the appropriate antigens to distinguish a signal from noise, and whether they could detect cytokine secretion. Dr. Goltsev responded that they are identifying each cell-type by phenotypic clustering. Each neighborhood gets an independent statistical score, and for effective visualization the antibodies must detect at least 1000 identical target antigens with a single cell.

Dr. Mark Chee from Encodlia Inc. described a rough draft of a concept for a highly scalable digital protein analysis technology with potential for single-cell analysis. The approach is based on a multistep process in which peptides or proteins are probed for identifying information. The peptides are made as chimeric constructs between a peptide and cDNA sequence (peptide-cDNA) by in vitro transcription/translation from pools of DNA templates generated by microarray-based synthesis. The utility of a library of such constructs pools was demonstrated in two activity-based assays designed to discover protease and kinase substrates. In the protease assay, cleaved peptide substrates were separated from uncleaved and identified by digital sequencing of their cognate cDNAs. The approach is designed to be compatible with high-throughput multisample workflows and could be broadly applicable. Similar approaches were developed by other companies as well.

Dr. Hui Zhang from Johns Hopkins University proposed a single-molecule protein sequencing approach for single cell proteomics. The currently available protein sequencing techniques are limited. The ability to perform protein sequence analysis at a single-molecule level could transform the field of proteomics by enabling the detection of low-abundance proteins and single-cell proteomics with ultrahigh sensitivity. Thus, the selection of suitable reporters for sequencing results is crucial. Ideally, these reporters should be specific to each amino acid, although signal boosting strategies should also be taken into consideration. Sequencing techniques that can precisely read the entire protein sequence are yet to be developed. Dr. Zhang proposed to use the 20 aminoacyl-tRNA synthetases (asRSs), which are responsible for high-fidelity translation of triplet nucleotide codons in mRNA to amino acid sequences. In the translation process, aminoacyl-tRNA synthetases utilize ATP to activate amino acids and form aminoacyl-tRNA Synthetase-aminoacyl-adenylate complexes releasing inorganic pyrophosphates. The activated aminoacyl is then transferred to tRNAs to form aminoacyl-tRNAs. In sequencing the entire protein, the released amino acid will be read by its specific aminoacyl-tRNA synthetases. This could be detected by the binding of each released amino acid to the specific aminoacyl-tRNA synthetase, taking ATP to the synthetase, releasing inorganic pyrophosphates from the specific aminoacyl-tRNA synthetase, or reading aminoacyl-tRNAs conjugates. This has the potential to serve as a high-throughput method for protein sequencing.

To gain insight into the detailed phenotype, functional state, and location of cells in complex tissues, including the tumor microenvironment in diverse organ locations, Dr. Ronald N. Germain’s laboratory developed and employed novel immunohistochemical methods called histo-cytometry and Ce3D. These methods and associated with IBEX computational and analytical tools that permit highly multiplexed quantitative analysis of surface, cytoplasmic and nuclear proteins including transcription factors, post-translationally modified molecules such as pSTATs, extracellular matrix components and simultaneous detection of RNA species of mammalian and bacterial origin in both 2D sections and 3D tissue volumes. Basically, histo-cytometry is an analytical microscopy method for visualizing and directly tagging surface, cytoplasmic, and nuclear proteins within tissue sections. Histo-cytometry can simultaneously use 8–12 different antibodies not only as surface markers, but also to identify phosphoproteins and cytokines in each cycle of imaging. Together with methods for rapid iteration of staining and analysis (IBEX), these approaches permit imaging of >30 target proteins in a single tissue slice or 3D biopsy sample. Recent advances include combining multiplex RNA FISH with antibody-based staining to take advantage of RNA-seq data. High-resolution volumetric microscopy with Ce3D enables quantitative spatial visualization and mapping of cellular neighborhoods and the potential analysis of cell–cell interactions in a variety of tissues.

The cell surface proteome (surface-ome) is the primary hub for cells to communicate with the outside world. Oncogenes are known to cause huge changes in cells, and Dr. James Wells has hypothesized that transformation will lead to changes in the cancer surface-proteome. To test this hypothesis, Dr. Wells’ team have looked at the quantitative cell-surface proteome signature of KRASG12V in the MCF10A cell line, which is driven by MAPK pathway signaling. Simultaneously through a CRISPRi screening approach, they have identified integrin and Wnt signaling proteins as critical to RAS-transformed cells. CDCP1, which was found by both methods to be an upregulated cell-surface protein, and is frequently upregulated in other RAS-transformed cell lines, was used as a target for immune cytotoxic treatment and as a cell-surface indicator for mutant RAS. The team generated a toolkit of recombinant antibodies to seven of the RAS-induced proteins and found that five of these proteins are broadly distributed on cancer cell-lines harboring the RAS mutation. The strategy of identification and targeting upregulated cell-surface proteins and the development of recombinant antibodies to target them was also applied to target an acute myeloid leukemia (AML) cell line.

Dr. John LaCava discussed an approach to study specific protein interactions in the context of large dynamic macro-molecular assemblies (interactome). His laboratory combines protein mass spectrometry with affinity capture to study the interactome associated with expression of LINE-1 retrotransposon (L1). L1 sequences comprise ~17% of the human genome, are frequently expressed in cancerous cells, and are epigenetically suppressed in normal cells. Currently, his technical approach requires a large number of cells and is not amenable for a single-cell analysis.

Dr. Robert Moritz and Dr. Jacob Kagan summarized the session and reiterated the need for centralized effort for the development and characterization of high quality, cost-effective monoclonal antibodies, including recombinant antibodies, affibodies, and other forms of recombinant antibody derivatives that will be directed by several different epitopes of each protein. Such reagents would be pivotal for single-cell targeted proteomics and would be used for multiple purposes including better phenotypic characterization of the molecular and cellular microenvironment; detection of gradients of secreted proteins within the TME; exosomes; characterization of cell-type specific dynamic interactomes; and intracellular targeting of specific
molecules. Practical considerations for multiplexing of antibodies for single-cell proteomics included questions on how many antibodies could be multiplexed at one time. Additional discussions focused on protein-specific antibodies versus antibodies directed at protein PTM. Most often, PTM antibodies are not as specific, and the interaction of the antibody with specific PTM sites will depend on the sequence and the 3D structure of the target.

**Part 2. Interrogation of a Single-Cell Proteome.**

Antibody-based flow cytometry and mass cytometry are the predominant technologies for targeted proteomic analysis of single cells. However, they share common shortcomings with other antibody-based methods (e.g., low-multiplexing potential, the need for high-quality antibodies, and unavailability of antibodies for new proteins). Furthermore, the techniques cannot provide absolute protein concentrations. Mass spectrometry (MS)-based targeted proteomics has emerged as a promising alternative for antibody-free high-multiplexed quantification of target proteins. However, it has not been suitable for the analysis of single cells or small numbers of cells due to ineffective processing of such small samples for MS analysis. To tackle this challenge, Dr. Tujin Shi, Dr. Richard Smith, and their teams recently developed two convenient carrier-assisted targeted MS approaches, cPRISM-SRM (carrier-assisted high-pressure, high-resolution separations with intelligent selection and multiplexing coupled to selected reaction monitoring)\(^{82}\) and cLC-SRM (carrier-assisted LC-SRM),\(^{83}\) for enabling in-depth proteomics analysis of small numbers of mammalian cells. Both cPRISM-SRM and cLC-SRM capitalize on the addition of exogenous carrier proteins to assist the effective processing of low numbers of cells with minimal loss. cPRISM-SRM has sufficient sensitivity to quantify the 2500 most abundant proteins in a single human cell and target proteins at ≥3000 copies per cell in 100 human cells. However, a key drawback of cPRISM-SRM is its moderate sample throughput due to the need for high-resolution PRISM to reduce the increased dynamic concentration range originating from the addition of tremendous amounts of carrier proteins (∼50 μg). To address this issue, they have recently developed an alternative nanoscale targeted proteomic approach cLC–SRM, which processes small numbers of cells in a single tube at low volume (i.e., single-tube digestion) with an excessive amount of digestion enzyme (∼0.25 μg) serving as the carrier protein. cLC-SRM allows for sensitive and reproducible detection of most EGFR pathway proteins expressed at levels ≥30000 and ≥3000 copies per cell for 10 and 100 mammalian cells, respectively. Furthermore, the single-tube digestion method was also demonstrated to be highly effective for preparation and enrichment of protein posttranslational modifications (e.g., phosphorylation) in small numbers of human cells for proteomics analysis.\(^{82,83}\)

It is anticipated that the two complementary targeted proteomic methods would be broadly applied to biomedical research and systems biology for absolute protein quantification in small numbers of mammalian cells as well as precious mass-limited clinical specimens with the potential of moving toward precision medicine.

Dr. Bogdan Budnik and his team introduced another technique, SCoPE-MS,\(^{84}\) that allows measurements of thousands of proteins in single mammalian cells by introducing a carrier channel proteome consisting of 100–200 cells for a better protein identification rate. A major limitation of applying quantitative mass spectrometry to a small amount of sample, such as a single mammalian cell, is the loss during sample cleanup. To overcome this limitation, they developed a minimal proteomic sample Preparation (mPOP) method for culture-grown mammalian cells. mPOP obviates cleanup and thus eliminates cleanup related losses while simplifying and expediting sample preparation for analysis of single cells. Using mPOP with cell-sorting and liquid handling, one can lyse thousands of single cells in minutes and quantify more than 2500 proteins across hundreds of single cells with a throughput of 700 single cells per week of mass spectrometer instrument time. Likewise, mPOP with SCoPED MS enables protein quantification in 100 and 1000 cell samples with a unique depth and throughput.

Dr. John Yates argued that at early time points of disease progression, molecular changes may be small and difficult to identify because they are hidden by the overwhelming static proteome. He proposed a strategy based on bio-orthogonal chemistry to overcome this difficulty. The procedure uses metabolic insertion of specific molecules into biologics such as a protein or carbohydrate. Azidohomoalanine (AHA) is a Met analogue that is used by the Met tRNA synthetase that can be inserted into proteins in vivo. Azidohomoalanine containing proteins can be enriched by covalently linking azidohomoalanine to a biotin alkyne through click chemistry. Thus, AHA proteins or peptides can be enriched and efficiently separated from the whole proteome through avidin bead enrichment. When pulsed into cells for a short period of time, only newly synthesized proteins (NSP) are labeled. This method was applied to cells from worms, flies, tadpoles, mammalian tissue slices, and mice.\(^{85–88}\) Several uses of the method were illustrated, and the ideas for use in single cell analysis were discussed. Dr. Yates plans to apply this strategy to examine targeted newly synthesized proteins at the single-cell level.

Dr. Brian T. Chait described his collaborations with Dr. Andrew Krutchinsky to engineer mass spectrometry for single-cell proteome analysis. Proteome analysis of single cells poses significant methodological challenges since currently available mass spectrometric technologies can be inefficient and therefore lack the necessary dynamic range, sensitivity, and desired signal-to-noise specifications for the desired task. Dr. Chait’s team argued that to achieve the desired goals will require:

- **Transferring 100% of the ions produced during the electrospray process into the mass spectrometer**
- **Filtering all background ions, including the so-called chemical noise, from the useful protein-derived mass spectra**
- **Measuring the useful protein-derived ions by mass spectrometry with efficiencies that approach unity**

To obtain these goals, one may effectively need up to 1000 independently coupled mass spectrometers with the capability to interrogate all ions continuously over time. One would like to look in detail at both the most and the least abundant species present. This will require tremendous dynamic range capability. Dr. Chait’s team also wants to look comprehensively across all proteins with full MS/MS coverage. To accomplish this task, they have developed a multi-quadrupole ion trap, which in their most recent prototype is made up of ∼500 quadrupoles. Each of these quadrupoles can serve as either an input or output for ions, and each can be tuned to a different m/z range. They started with a computer simulation that demonstrated the feasibility of their approach. One demonstrated use for this device is in removing background ions, especially when the ions of interest are rare—as in single-cell proteomics. Here, the tryptic ions of
interest generally have ≥2 charges, while it turns out that the background is dominated by singly charged species. In this case, it would be useful to filter all the singly charged ions from the spectrum. Another potentially interesting application of the multi-quadrupole ion trap is filtering unmodified peptide background in chemical cross-linking experiments. In this example, the ions of interest generally have ≥4 charges, so one wants to get rid of the much more abundant unmodified 2+ and 3+ ion species, which in this case make up the “noise” background.

To accomplish such filtering, Dr. Chait’s team operates the multi-quadrupole ion trap in what they term “z mode”, where the injected ions are rapidly thermalized while they diffuse around the volume of the trap. To perform this filtering, they placed a small DC potential difference of ~−0.03 V (~−3/2kT, where k is the Boltzmann constant, and T is the temperature) created between all but one quadrupole and its extraction electrode. This small potential difference creates a potential barrier that does not prevent thermalized singly charged ions with kinetic energy ~−3/2kT from exiting the trap. However, higher charged ions “see” a much higher effective repelling potential — proportional to their charge — making it considerably harder for them to exit the trap. This effect (as well as the more rapid rate for cooling of the lower charged species) causes preferential “evaporation” of singly charge ions from the trap along some 500 exits. Having so many exits allows the singly charged ions to escape quickly. This is basically a diffusion process, which is similar to the problem solved by the array of nuclear pore complexes that study the nuclei within eukaryotic cells. In these cells, rapid and efficient transport into and out of the nucleus is accomplished by surprisingly low numbers of pores. Dr. Chait demonstrated a spectrum that contains minute amounts of five peptide species. In the face of the chemical noise that was present, largely as single charged ions, none of the peptide species could be discerned in the mass spectrum. By simply flipping a switch on their device, the previously invisible peptide ions were clearly revealed and easily identified. Getting rid of noise is important when you have such small numbers of ions of interest. They also envisage using their device for the analysis of chemical cross-linking by mass spectrometry. By adjusting the potential upward on the extraction electrodes, they found that it is possible to preferentially filter out ions with 2 and 3 charges, while retaining ions with ≥4 charges. Thus, Drs. Chait and Krutchinsky think they have a robust method for controlling noise using their multi-quadrupole ion trap, which should be pluggable into the front end of most mass spectrometers. Next, they plan to work on the division of the input beam into their multi-quadrupole ion trap into 10 subbeams with nonoverlapping m/z ranges. Such a device has the potential for increasing the sensitivity, speed of analysis, and dynamic range of proteomic analysis by a factor of 10.

Dr. Richard D. Smith argued that the ability to achieve in-depth proteome coverage for broad quantitative proteomics measurements, at the single-cell level, would be important for both basic and applied aspects of cancer research as well as biomarker development for early diagnosis. Rapidly evolving mass spectrometry (MS)-based proteomics methodologies are approaching the broad proteome coverage desired as well as developing the potential for detection of modified proteins. However, the technical challenges for pushing MS-based proteomics to the level of single cells while still achieving reasonable depth of coverage, are daunting, as evidenced by the early efforts aimed at these challenges. These challenges include the requirements for efficient nanoscale cell selection and isolation, processing and manipulation, high peak capacities for peptide and/or protein separation(s), and efficient ionization (typically using nanoelectrospray ionization) followed by MS-based analysis. Importantly, such analyses must be able to be conducted robustly and must achieve reasonable high throughput because many single cells typically need to be studied. Dr. Smith discussed approaches being developed at Pacific Northwest National Laboratory (PNNL) that aim to address these challenges including robotic nanoscale sample preparation and manipulations in nanoPOTS82,89,90 for very small numbers of human cells including single cells, high ion utilization efficiency nanoelectrospray ionization for ultrahigh sensitivity, and their combination with advanced ion mobility-MS based upon the use of Structures for Lossless Ion Manipulations (SLIM) for higher throughput and improved sensitivity.91 Nanodroplet Processing in One Pot for Trace Samples (nanoPOTS) is a platform for proteomic analysis of a very small number of cells (10−150). The platform minimizes surface losses through improvement of sample processing and recovery by reducing the sample processing to ~200 nL. A combination of nanoPOTS platform for sample processing with ultrasensitive liquid chromatography-MS resulted in detection of approximately 1500−3000 proteins from 10−150 cells, respectively. Consistent detection of close to 3000 proteins from as few as 10 cells was achieved using MaxQuant’s “Match Between Runs algorithm”. This technology was also applied to detect ~2400 proteins from 10 μm-thick cross sections of individual human pancreatic islet cells.90

### CONCLUDING REMARKS

Obtaining proteomic information from single cells is still a significant challenge. Mass spectrometry analysis, which is used to capture and quantify entire proteomes, are usually carried out with tens of thousands of mammalian cells. Interrogation of a single cell by mass spectrometry has been possible for only the most abundant proteins. Currently, single-cell proteomic analysis relies primarily on a targeted approach, which is dependent on antibodies directed at limited number of defined proteins. The targeted proteomic approach was used to analyze the type of cells and their spatial location within tumor microenvironments and for visualization and characterization of cellular compartments and distribution of known proteins in normal and disease states.38,61,63,64,73,74,92−94 The number of tagged antibodies that could be multiplexed without interference is still a challenge. Also, some investigators argued that quite often antibodies directed at PTM are not as specific, and the interaction of the antibody with specific PTM sites is dependent on the sequence context and the 3D structure of the target. Additional limitations include limited options for multiplexing and enormous difficulties in the generation of high-quality antibodies.95 Drs. Tujin Shi and Richard Smith argued that mass spectrometry (MS)-based proteomics is a promising alternative for quantitative single-cell proteomics because it is antibody-free and has high specificity and ultrahigh multiplexing capability.96 Nevertheless, MS-based single-cell proteomics is still at the infancy stage.97 In the last three years, significant progress in sample preparation has been made for MS-based single-cell proteomics by either significantly reducing sample processing volume (e.g., a nanoPOTS platform downsizing the processing volume to ~200 nL98 or by using an excessive amount of carrier proteins or peptides (e.g., the

---

**Journal of Proteome Research**

**Number of unique words:** 2042

**Number of sentences:** 29

**Number of paragraphs:** 10

**Average sentence length:** 44 words

**Average paragraph length:** 134 words

---

**Citation:**


**DOI:** 10.1021/acs.jproteome.0c00021
addition of exogenous proteins as a carrier \cite{32,85} or tandem mass tag (TMT)-labeled 100s of cell lysate digests as a carrier/boosting channel \cite{84}). In spite of these accomplishments, single-cell MS proteomics still does not meet the threshold for rapid, comprehensive, quantitative analysis of single mammalian cells in terms of proteome coverage, measurement accuracy, and sample throughput. Another unmet technical challenge concerns the quantitative analysis of protein post-translational modifications (PTMs) due to low stoichiometry of PTMs \cite{e.g., ~1% of phosphorylation in the entire protein amount}.\cite{18} With highly effective sample processing, the most sensitive single-cell MS published platform \cite{i.e., nanoPOTS-MS} enables label-free quantification of only \( \sim 670-870 \) proteins from single mammalian cells \( \sim 5\% \) of the total proteome) with \( \sim 8-10 \) samples per day.\cite{90,99} When combined with multiplexed TMT labeling to boost single-cell signal, the nanoPOTS-MS platform permits high-throughput quantification of \( \sim 1200 \) proteins from single mammalian cells with \( \sim 80 \) samples per day.\cite{100} Further optimization of LC (e.g., LC flow rates) and MS parameters \cite{e.g., automatic gain control settings and ion injection times} can further increase the proteome coverage while maintaining a good “balance” between the quantitation quality and proteome coverage.\cite{100,101} However, the quantification accuracy of the TMT-based approach is still significantly affected by ratio compression or distortion from coeluting interferences. \cite{101,102} Future developments in single-cell MS proteomics will focus on significant improvements in MS detection sensitivity and sample throughput as well as single-cell sample preparation. Enhancing detection sensitivity can be achieved by effective integration of ultra-low flow-high resolution RPLC \cite{103–107} or capillary electrophoresis (CE) \cite{108} and a high-efficiency ion source/ion transmission interface \cite{e.g., multimeritter SPIN source/dual ion funnel interfaces} with the most advanced MS platform \cite{e.g., Orbitrap Eclipse or timsTOF}. The detection sensitivity may also be improved by reducing surface adsorption loss \cite{e.g., systematic evaluation of different types of MS-friendly surfactants} and increasing reaction kinetics through further reducing processing volume \cite{e.g., low or sub-nL processing volume}. Sample throughput can be increased by using ultrafast high-resolution ion mobility-based gas-phase separation \cite{e.g., SLIM} \cite{112–114} to replace current slow liquid-phase (LC or CE) separation, and effective integration of liquid- and gas-phase separations \cite{e.g., SLIM} \cite{112–115} or FAIMS \cite{115,116} for greatly reducing separation time but without trading off separation resolution. Alternatively, sample hyper-multiplexing with isobaric barcoding and implementation of a multiple LC column system can also be considered to greatly increase sample throughput. All these improvements could lead to a more powerful single-cell MS platform and will certainly close the gap between single-cell proteomics and single-cell transcriptomics or genomics. Promising protein or peptide sequencing technologies are rapidly emerging and include nanopore sensors \cite{119} and a modification of a traditional Edman degradation sequencing for low abundant peptides. The technology is based on fluorescently labeling selected amino acids (lysine and cysteine) and by monitoring the decrease in fluorescence during consecutive rounds of Edman degradation. The obtained sequence is then compared to a reference protein database.\cite{117}

The Think-tank Chair, Dr. Moritz, indicated that to understand the biological function and organization of molecular systems, high-throughput strategies based on individual omics components have provided insight, albeit at the limitation of comprehensiveness. As a mainstay, genomic and transcriptomic data, produced using amplification strategies, have allowed the dissection of biological systems including the interpretation and trajectories of molecular profiles of cancer cells. Stepping beyond genomics, the task of providing molecular data at the single-cell and massively parallelized level is harder due to the inherent sensitivity barriers from omics capabilities devoid of amplification and multiplexing strategies. However, the need for the integration of “omics,” including epigenomics, transcriptomics, proteomics, and metabolomics, into physiological and clinical studies is paramount. The development of single-cell interrogation technologies, especially at the protein level, can provide novel insights into the mechanism of tumor evolution and identify new markers and novel targets for preventive and therapeutic interventions.

Drs. Kagan and Srivastava reiterated the challenges in integration of multiomic data and spatial visualization of preneoplastic lesions at the single-cell level. The combination of microscale 3D electron microscopy imaging and confocal microscopy is a powerful approach that revealed new subcellular structures/compartments and abnormal distribution of organelles within individual cells, depending on the spatial context of a particular cell within the tissue microenvironment and the cell type. Tumorigenic cells at the tip of invasive protrusion of a normal tissue show different characteristics. The context is very important and should integrate clinical data and parameters such as the physiological state of the cells, the effect of hypoxia on blood values, and multiomic molecular and cellular characterization. At the present time, it is believed much of the 3D visual observations should be used to direct the development of multiomic data analysis going forward.

**THINK-TANK RECOMMENDATIONS**

- Further improve and apply single-cell proteome interrogation technologies to acquire novel insights into the mechanism of tumor evolution from pre-neoplastic lesions to invasive tumors. Interrogate a lesion’s cellular neighborhood, molecular and cellular interactomes including gradients of secreted microvesicles and cellular mediators (i.e., TGFβ).
  - Apply single-cell proteomic technologies to identify new candidate markers for early detection, diagnosis and prognosis, and targets for prevention and therapy.
- Apply single-cell proteomic technologies to interrogate and map intracellular sub-microscopic structures, organelles, unique compartments, and interactomes. Identify their abnormal location, differential distribution, and function.
- Apply single-cell proteomics to interrogate dynamic, functional, and structural changes within a benign lesion (e.g., DCIS or benign prostate cancer), based on comparison of repeat biopsies over time from the same lesion of the same individual who is managed through active surveillance.
- Apply computational tools to integrate multiomic single-cell analyses data with pathological and clinical data.
- Develop a data repository of integrated massive parallel multiomic data, 2D and 3D visualization data. Data should include cellular networks, secreted microvesicles and molecules (i.e., cartography of gradients of secreted molecules such as TGFβ) by certain cells within a
neighborhood) and the composition of the extracellular matrix.

- Develop a centralized effort for the development, characterization, and validation of high-quality cost-effective monoclonal antibodies and other forms of recombinant antibodies derivatives for interrogation of every protein and its proteoforms.

## AUTHOR INFORMATION

### Corresponding Authors

Jacob Kagan — Cancer Biomarkers Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland, United States; orcid.org/0000-0001-5267-1672; Email: kaganj@nih.gov

Robert L. Moritz — Institute for Systems Biology, Seattle, Washington, United States; Email: Robert.Moritz@systemsbiology.org

### Authors

Richard Mazurchuk — Cancer Biomarkers Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland, United States

Je Hyuk Lee — Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States

Peter Vasilii Kharchenko — Blavatnik Institute for Biomedical Information, Harvard Medical School, Boston, Massachusetts, United States

Orit Rozenblatt-Rosen — Klarman Cell Observatory, Broad Institute, Boston, Massachusetts, United States

Eytan Ruppin — Cancer Data Science Laboratory, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland, United States

Fredrik Edfors — Science for Life Laboratory, KTH - Royal Institute of Technology, SE-171 21 Stockholm, Sweden; orcid.org/0000-0002-0017-7987

Fiona Ginty — Life Sciences and Molecular Diagnostics Laboratory, GE Global Research Center, Niskayuna, New York, United States

Yury Golstev — Department of Microbiology and Immunology, Baxter Laboratory in Stem Cell Biology, Stanford University, Stanford, California, United States

James A. Wells — Department of Pharmaceutical Sciences, University of California, San Francisco, California, United States; orcid.org/0000-0001-8267-5519

John LaCava — Laboratory of Celluluar and Structural Biology, Rockefeller University, New York, New York, United States

Jessica L. Riesterer — Center for Spatial Systems Biomedicine, Oregon Health and Science University, Portland, Oregon, United States

Ronald N. Germain — Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland, United States

Tujin Shi — Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, United States; orcid.org/0000-0002-5592-3588

Mark S. Chee — Enodia, Inc, San Diego, California, United States

Bogdan A. Budnik — Faculty of Arts & Sciences, Division of Science, Harvard University, Boston, Massachusetts, United States

John R. Yates, III — Department of Molecular Medicine, Scripps Research Institute, La Jolla, California, United States; orcid.org/0000-0001-5267-1672

### REFERENCES


(88) Zhu, Y.; Clair, G.; Chrisler, B. W.; Shen, Y.; Zhao, R.; Shukla, A. K.; Moore, R. J.; Misra, R. S.; Pryhuber, G. S.; Smith, R. D.; Ansong, C.;