Omics and systems view of innate immune pathways

Deepali Rathore¹, Matthew Marino¹, and Aleksandra Nita-Lazar¹

¹National Institutes of Health

February 15, 2023

Abstract

Multiomics approaches to studying systems biology are very powerful tools that can elucidate changes in the genomic, transcriptomic, proteomic, and metabolomic levels within a particular cell type in response to an infection. These approaches are valuable for understanding the mechanisms behind disease pathogenesis, and specifically how the immune system responds to being challenged. With the emergence of the COVID-019 pandemic, now more than ever, the importance and utility of these tools has become evident in garnering a better understanding of the systems biology within the innate and adaptive immune response and for developing treatments and preventative measures for new and emerging pathogens that pose a threat to human health. In this review we focus on the various state of the art “omics” technologies used within the scope of innate immunity.

OMICS AND SYSTEMS VIEW OF INNATE IMMUNE PATHWAYS

Deepali Rathore¹, Matthew Marino¹, Aleksandra Nita-Lazar¹*

¹Functional Cellular Networks Section, Laboratory of Immune Systems Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States

*Corresponding author: nitalazarau@niaid.nih.gov

ABSTRACT:

Multiomics approaches to studying systems biology are very powerful tools that can elucidate changes in the genomic, transcriptomic, proteomic, and metabolomic levels within a particular cell type in response to an infection. These approaches are valuable for understanding the mechanisms behind disease pathogenesis, and specifically how the immune system responds to being challenged. With the emergence of the COVID-019 pandemic, now more than ever, the importance and utility of these tools has become evident in garnering a better understanding of the systems biology within the innate and adaptive immune response and for developing treatments and preventative measures for new and emerging pathogens that pose a threat to human health. In this review, we focus on the various state of the art “omics” technologies within the scope of innate immunity.

KEYWORDS: Innate immunity, mass spectrometry, cellular signaling, omics, systems immunology

INTRODUCTION:

Host-Pathogen Interactions in Innate Immune Pathways

Host-pathogen interactions describe complex and dynamic processes that include all stages of pathogenic infection, from invasion to dissemination. Both host cells and pathogens have evolved to adopt a wide array of strategies to interact and survive, adding to the traditional views of host-pathogen interactions. These new insights shed light on the intracellular signaling pathways that lead to the innate immune response. Central to the new studies was the discovery of pattern recognition receptors (PRRs) on the innate immune cells (dendritic cells and macrophages) that can recognize pathogen-associated molecular patterns (PAMPs)
derived from various microbes. Among several classes of these PRRs including RIG-I-like receptors (RLRs) and Nod-like receptors (NLRs), the toll-like receptors (TLRs) are best characterized and are specific in their response.\textsuperscript{1} Human TLRs can be classified as either cell-surface TLRs (i.e., TLR1, TLR2, TLR4, TLR5, TLR6, and TLR10) or intracellular TLRs that are localized in the endosome (i.e., TLR3, TLR7, TLR8, TLR9, TLR11, TLR12, and TLR13). Each TLR is composed of leucine-rich repeats (LRRs) that recognize PAMPs, a transmembrane domain, and a cytoplasmic Toll/IL-1 receptor (TIR) domain that triggers downstream signaling, resulting in the production of effector molecules, including cytokines, chemokines, and antimicrobial proteins, to combat the invading pathogens.\textsuperscript{2} Signaling pathways associated with different TLRs can also vary, since they respond to different stimuli and recruit different adaptor proteins in different temporal patterns. For example, TLR4 recognizes lipopolysaccharide (LPS), an integral component of the outer membranes of Gram-negative bacteria that causes endotoxic shock and activates the MyD88/IRAK signaling pathways.\textsuperscript{3} TLR3 is localized in the endosome and recognizes double-stranded RNA (dsRNA), a viral replication intermediate, initiating the downstream signaling for interferon production.\textsuperscript{4} TLR signaling occurs via two major pathways, MyD88 dependent and TRIF dependent, named after the adaptor proteins initiating signal transduction pathways. MyD88 and TRIF activate NF-κB, IRFs (interferon regulatory factors) and MAPK (mitogen-activated protein kinase) to regulate the expression of inflammatory cytokines and type I IFNs.\textsuperscript{5}

Aberrations in the PAMP recognition or mutations in molecules involved in TLR signaling pathway can cause autoimmune, inflammatory and allergic diseases. Post translational modifications (PTMs) on the signaling molecules such as ubiquitination and phosphorylation play critical roles in the activation of TLR signaling, and hence their characterization is highly informative in decoding the mechanism of signaling. Efforts have been made to identify molecules (proteins, transcripts, metabolites and lipids), through integrated approaches, at both cell and tissue level.\textsuperscript{6}

Here, we provide a comprehensive overview of the current and emerging tools for systems level analysis of innate immunity. The next sections highlight the prominent contribution of “omics” methodologies in understanding the innate immune signaling pathways at systems level. Along with the description of techniques, case studies are also presented to elaborate on the suitability and applicability of each technique, depth and breadth of information it provides, and the data analysis needed to process the information. Figure 1. presents a brief timeline of the various “omics” technologies in terms of their first appearance and the application in context.

**MULTIOMICS IN INNATE IMMUNITY:**

\subsection*{2.1 Transcriptomics}

A common tool used for most systems biology analysis is RNA sequencing. RNA sequencing is used to study the transcriptome of a particular cell or organism. The core workflow of any RNA sequencing experiment is the extraction of the RNA, followed by enrichment of the subtype of RNA to be analyzed and depletion of the RNA subtypes that are not of experimental interest. This is followed by the preparation of an adapter ligated cDNA library, amplification of the constructed library, and high throughput sequencing of the library between 10 and 30 million reads per sample. Next step is the computational analysis of the sequenced library. This involves alignment to a reference transcriptome, quantification of overlapping sequences, data normalization between samples and pre-processing, and statistical modeling which is typically done through a surplus of different coding languages and software packages. Majority of RNA sequencing experiments are done with short read sequencing instruments, but recent advances in long read sequencing and direct read sequencing offer new approaches and methodologies to tackle questions not answerable within short read sequencing alone. Each of these approaches come with their own limitations. For short read sequencing a major limitation involves biases that are introduced during sample preparation and downstream computational analysis. These biases can affect the quantification of gene isoforms (especially regarding longer transcripts), and how the multi-mapped reads are processed. Long read sequencing and direct read sequencing aim to overcome these limitations, yet they also have limitations of their own. However, regardless of the type of approach used, the primary application of RNA sequencing remains to assess differential gene expression.
2.2 Proteomics

Another commonly employed set of methodologies for systems analysis in innate immunity are proteomics-based methods. Bottom-up proteomics using mass spectrometry (MS), commonly coupled with liquid chromatography (LC-MS), has been traditionally used to assay changes in the proteomic, phosphoproteomic, kinomic and secretomic innate immune landscape. Various omics strategies offer diverse biological conclusions and when combined and can elucidate the complex signaling environment that occurs in host-pathogen interactions. In the simplest of terms, bottom-up proteomics refers to the digestion of proteins into peptide components using suitable enzymes before analysis with LC-MS. Bottom-up proteomics is also the most widely used approach of proteomics-based methodologies since it provides information on protein identity and post-translational modifications (PTMs), both qualitatively and quantitatively. Bottom-up proteomics can be performed in numerous ways; data independent acquisition (DIA), data dependent acquisition (DDA) such as the case of shotgun MS proteomic profiling, and targeted LC-MS. DDA is usually utilized as a baseline experiment, or a first approach, in what is known as discovery level proteomics. It involves a semi-random selection of target analytes for fragmentation based on set cutoffs for relative intensities, which can be very useful for identifying unknown proteins without any prior knowledge about them. In DIA experiments, non-random selection of precursor ions is curated with a wide precursor ion isolation window. This accounts for more specificity and quantification accuracy for more peptides in comparison to DDA, with the tradeoff of a more challenging spectral analysis. In targeted LC-MS, the parameters describing target analytes to be fragmented are set by the user and include precursor mass to charge ratio \((m/z)\), LC elution time, and collision energy needed to drive fragmentation for each analyte. This makes targeted MS a useful tool for validation experiments or absolute quantification. Targeted proteomics has two main variations: selected reaction monitoring (SRM) (also known as multiple reaction monitoring or MRM when applied to more than one precursor and fragment ions) and parallel reaction monitoring (PRM). SRM allows for precise quantification across multiple sample groups and is typically used with LC-MS instruments that have a triple quadrupole (QQQ) where the first quadrupole (Q1) act as a mass filter and selectively monitors analyte precursor, fragmentation is carried out in the second quadrupole Q2, and guided to the third quadrupole Q3 for analysis. Untargeted peptides and their fragment ions are discarded leaving only the ions coming from the analyte of interest to be quantified by the detector over time. Like SRM, PRM also uses MS2 data as input, but rather than selectively monitoring a single target analyte’s fragmentated ions, the full MS2 spectra is scanned at high resolution and the intensity of multiple fragment ions is monitored in parallel.

Bottom-up proteomics approaches also allow for both relative and absolute quantification of a target analyte across multiple biological sample groups using different labeling techniques. Relative quantification can also be achieved with a label free approach (LFQ), where LC-MS is done on each sample separately. When
labeling is done prior to LC-MS, such as stable isotope labeling, samples can be pooled together (both labeled and unlabeled) and run in a single LC-MS run to quantify the target analyte based on the ratio of labeled to unlabeled forms of the analyte. Absolute quantification can be done by comparing to known standards of the stable isotopes. There are many different types of labeling strategies available that can be employed for metabolic labeling; stable isotope labeling by amino acids in cell culture (SILAC) using heavy isotopes of carbon and nitrogen is frequently used. There is also an abundance of chemical labels that are utilized such as tandem mass tags (TMT), isobaric tags, O labels, dimethyl labeling, and others. All of these labeling methods, when compared to known peptide standards, provide highly accurate and reproducible workflows for both relative and absolute quantification of target analytes across multiple biological sample groups. It should also be noted that bottom-up proteomics workflows can be adapted for the detection of PTMs. PTMs are chemical modifications on the side chains of amino acids within a peptide. Thus, detection methods rely on tandem MS to compare the mass shift of the unmodified tryptic peptide to the modified peptide. The immune system signaling is tightly controlled by changes in post translation modifications of protein complexes, which are highly dynamic and often labile. Lower abundance of modified peptides compared to the non-modified pool adds another layer of complexity in the analysis of PTMs. Therefore, enrichment of post translationally modified peptides is almost always needed for their detection by LC-MS. Some of the most common PTM enrichment techniques include immobilized metal affinity chromatography (IMAC) and metal oxide affinity chromatography (MOAC) for the detection of phosphorylated sites, diglycine remnant (KGG) immunospecific enrichment for ubiquitination site identification, and hydrophilic interaction chromatography (HILIC) for glycopeptide enrichment. Quantitative measurement of PTMs can be performed along with their identification by mass spectrometry, as discussed here for the phosphoproteome analysis in immune cell signaling. As with any quantification method, data analysis software is essential for making useful interpretations of the data. Comprehensive overview of the software available for targeted proteomics analysis and the advantages and disadvantages to using each application can be found elsewhere. Bottom-up proteomics methodologies are widely used within systems biology for their ability to investigate functional signaling dynamics and represent a fundamental area of research in the battles against new and emerging diseases. For example, in a study conducted by Wendisch et al., using a multiomics approach including shotgun-based proteomics, it was found that monocyte derived macrophages accumulate in the lungs during acute respiratory distress syndrome (ARDS) leading to a phenotype like the one observed in patients with pulmonary fibrosis. Their data supports the hypothesis that SARS-CoV-2 induces a profibrotic transcriptome and proteome profile within macrophages. Usually, these damage repair pathways when activated within macrophages help to control inflammatory mediated tissue damage. However, when left unchecked, these pathways can also lead to dysregulated fibroproliferation as well as protracted respiratory failure. This study is one example of how proteomics-based methods can further our understanding of host-pathogen interactions and demonstrates the need to develop more proteomics-based workflows that will broaden our knowledge of immune cell pathogen interactions, and aid in the development of effective treatments to limit the emerging viral diseases on human health.

While perhaps the most common approach, bottom-up proteomics is not the only set of proteomics-based methodologies used to quantify relative and absolute abundance of target analytes within a given proteome. Both top-down and middle-up/down proteomics can provide valuable information on the higher order structure and intact topology of the target protein, where bottom-up falls short owing to a completely different sample processing approach. Top-down proteomics permits analysis of intact proteins and even native-like protein complexes, without the need for proteolysis, and is useful for characterizing functional and genetic isoforms of target proteins. Since the proteins stay undigested, top-down MS also preserves PTM sites and allows for the characterization of entire proteoforms. With sufficient number of useful fragments, top-down can provide complete primary sequence characterization and at the same time reveal modifications on the protein. The sample preparation methods are however, not as high throughput or standardized as in bottom up proteomics, high-resolution mass analyzers (also more expensive and somewhat less user friendly compared to the bench-top instruments) are needed to resolve charges on high mass precursors, and the data analysis is needs specialized knowledge, which keeps top-down rather out of reach, although recent developments are promising. There have also been many attempts at adapting bottom-up targeted proteomics
labeling techniques to top-down workflows, such as TMT and SILAC labeling. These labeling techniques can be used for absolute quantification of target peptides or for entire proteoforms and negates the protein inference problem in bottom-up proteomics. Conclusions for protein identification in a bottom-up experiment rely on inferring the identification of a protein based on the peptides that compose the protein. In top-down proteomics, the data is representative of the parent protein which allows for direct protein and proteoform quantification, and when combined with the appropriate fragmentation technique, has the potential to achieve near complete identification of individual proteoforms. In concordance with the developments in targeted top-down proteomics, untargeted approaches with LFQ have also been developed and are especially useful for characterizing changes in proteoform abundance between different samples. For a more comprehensive overview on the quantitative proteomics-based methodologies for quantification of target proteins, Neagu et al. have provided a review summarizing various MS-based approaches and applications of tandem MS for protein analysis in biomedical research.

2.3 Metabolomics

Another tool used for systems wide analysis of innate immune pathways is metabolomics, which includes the profiling and quantification of metabolites predicated with methods such as mass spectrometry (MS) and nuclear magnetic resonance (NMR). A metabolite can be defined as an intermediate or terminal product of metabolism. They are essential in the production of ATP that provides energy to cells through catabolism and can also directly impact cellular signaling dynamics in various pathways, especially those that are involved in immune cell signaling. The study of the role of metabolites within the immune system makes up the field known as immunometabolism, which is now emerging as a key field of study within metabolomics, and it has been shown that metabolic rewiring of immune cell populations can regulate their activation. Fluxes in the metabolic cycle within a cell, as well as the metabolites themselves, can modulate cellular signaling dynamics directly and indirectly through regulation of post translational modifications. For example, macrophages that are stimulated with lipopolysaccharide (LPS) have been shown to undergo metabolic reprogramming leading to a pro-inflammatory phenotype. Notably, they underwent a switch from oxidative phosphorylation-based ATP production to glycolysis, leading to a higher reactive oxygen species production (ROS) by repurposing their mitochondria to produce ROS in a diethyl succinate dependent manner. Elevated production of succinate within the LPS stimulated macrophages also directly impacts production of interleukin-1B by modulating HIF-1$\alpha$ activity via ROS dependent oxidation, independent of NF-$\kappa$B signaling. Elevated succinate levels inhibit the production of anti-inflammatory cytokines such as IL-1RA and IL-10. Suggesting that upon macrophage activation, the accumulation of succinate directly enhances endogenous pro-inflammatory gene activation and inhibits anti-inflammatory gene expression. Like a proteomics workflow, a metabolomic LC-MS experiment can be conducted in a targeted or untargeted fashion, the experimental design is dependent on the chemical composition of the metabolites to be analyzed as well as metabolite absolute abundance. Due to the differences in structure, chemical composition and mass of peptides and metabolites, experimental parameters required for their analysis also differ to ensure proper fragmentation for the analyte. Untargeted or global approaches offer a wider detection window and do not require a pre-defined number of metabolites to be screened, unlike the targeted approach. Targeted metabolomics is usually geared toward a subset of metabolites within a biological pathway of interest. Both NMR and LC-MS techniques can be used to perform targeted metabolomics. The spectra obtained from $^1$H-NMR are usually compared to the spectra of a known chemical standard, as in a targeted LC-MS experiment, where quantification is based on the ratio of intensities for the detected metabolites, within the sample groups that match the pre-defined standards. Untargeted metabolomics offers an approach aimed at simultaneously measuring the most metabolites within a given sample with as little bias as possible. Each peak in an untargeted LC-MS run corresponds to a unique mass-to-charge (m/z) ratio and retention time known as a metabolite feature. The data sets that are generated are gigabytes in size when high resolution instrumentation is utilized, and a particular metabolite may have multiple unique metabolic features as well. This complicates data analysis and makes manual inspection impractical. However, developments in bioinformatics software designed for this specific purpose have revolutionized the way in which biological conclusions can be drawn from global scale metabolic experiments. A more comprehensive review on
analysis of metabolomics data with systems biology approaches can be accessed here. For quantification of metabolites, and for measuring metabolic fluxes or the rate at which they occur, label-based LC-MS approaches prove to be robust. Quantification of metabolites is an informative approach to metabolomics; however, it does not consider the non-linear relationship of the production and consumption of metabolites, i.e., the metabolic flux, which dictates metabolic pathway activities. Metabolic flux analysis (MFA) helps elucidate the rates at which metabolites interconvert by analyzing the patterns associated with labeling frequencies of isotope tracers and their target metabolites to infer fluxes in the metabolic network. In this approach, the rate at which the metabolite is labeled corresponds to the rate at which flux occurs.

**SINGLE CELL OMICS:**

The field of multiomics has seen consistent progress over the years as new “omics” technologies continue to be added to the arsenal. Most recently, with the development of the single cell techniques, it is now possible to explore the heterogeneity between cells of the same population. Macrophages act as the primary regulators of inflammation in the innate immune response. They can adopt stimulus induced phenotypes because of their functional plasticity. This allows macrophages to appropriately respond to diverse pathogens and aid in tissue repair following acute damage. M0 monocytes represent the common progenitor lineage for the pro-inflammatory M1 phenotype and anti-inflammatory M2 phenotype of macrophages. This differentiation is dependent on induction with polarizing cytokines. The M1 phenotype is a result of stimulation with interferon gamma (IFN-γ) and granulocyte macrophage colony stimulating factor (GM-CSF). M1 macrophages secrete pro-inflammatory cytokines such as IL-1β, TNF-α, and IL-12, all of which recruit immune cells to the site of infection and aid in pathogen clearance. However, this classification has been proposed to be less rigid as there are cells displaying characteristics of both and the macrophages have been reported to transition from one phenotype to another. This would normally be overlooked during global omics-based analyses, as the target molecule is pooled together with many different cells. For example, during the global transcriptomic analyses, cells are lysed together, and the RNA extract is pooled. This relies on the assumption that the cell population of interest is homogenous in nature. However, that is not the case, and with single cell resolution, one can assess quantifiable changes from cell to cell in the population of interest. Recent developments have now made it possible to extract multiple layers of omics data from a single cell. For example, obtaining the transcriptomic profile and doing targeted proteomics on the same cell. Single cell techniques can also provide insight on the spatial and temporal organization, as well as the population architecture of each compartment of the cell, tissue, or the organ. This is a powerful tool for studying systems biology, especially in the context of innate immunity, as spatio-temporal organization within cells is imperative for relaying cellular messages, leading to activation of immune signaling pathways. Isolation of single cells is the biggest limitation when it comes to any of the various downstream omics analyses. This presents a constant battle between the loss of cellular material during the isolation process and conservation of the spatial temporal information. Low throughput methods for single cell sorting such as laser capture microdissection or manual micromanipulation are costly, with output below a thousand cells for a given study and are time consuming, but they provide the advantage of decent spatio-temporal resolution. High throughput methods such as fluorescent-activated cell sorting (FACS) are less costly, automated, and provide a high yield of isolated cells. However, they do not retain spatio-temporal resolution since the tissues are homogenized during the pooling of cell suspensions. Thus, choosing the appropriate isolation method for the single cell omics analyses is a key part of a successful experiment.

Barcoding of cells for library preparation also represents a key component of single cell omics experiments, it allows for libraries generated by each single cell to be pooled and sequenced together which saves time and lowers the cost. However, this requires downstream bioinformatics analysis to be able to distinguish barcoded single cells from one another with a high confidence interval. An overview on the multiomics based single cell approaches with detailed insights into each category can be accessed here.

### 3.1 Single cell transcriptomics

Analysis of the transcriptome profile for a single cell is done using single cell RNA sequencing (scRNA-seq). Most scRNA-seq workflows involve the capture of mRNA based upon separation via the 3’polyadenylated...
region of the transcript. However, many different protocols exist to isolate other forms of RNA, such as ribosomal RNA (rRNA), or long non-coding RNA (lncRNA) and are available in both low and high throughput manners. Generally, library construction for scRNA-seq is relatively similar regardless of the type of RNA to be isolated. Cells must be lysed with the appropriate lysis buffer concentration and volume as to expel the contents of the cell without denaturing the RNA to be isolated followed by the “pulling down” of the RNA to be isolated. Traditionally, this is done by using an oligonucleotide sequence which is complementary to the poly adenylated tail of the transcript. Since the polyadenylated region is not protein coding, this essentially preserves the region of the transcript which would later be translated into protein. Some of the highest throughput methods for capture of polyadenylated RNAs involve bead-based separation, where the capture bead is coated with the complementary oligonucleotide sequence, a barcode that tags the transcript with a cell-identifier, and a second barcode which serves as a unique molecular identifier (UMI) for each transcript within the cell. After the library preparation, a third barcode is enzymatically added that allows for differentiation of different sample groups; this is especially important when multiple samples are sequenced at the same time. This method is employed in “droplet-based platforms” where the entire workflow is carried out within an oil droplet. Droplet based workflows allow for accurate identification and quantification of transcripts from a single cell and elucidate heterogeneity within a given cell population.

A study by Kong et al., investigating the changes in the single cell transcriptomic profiles within monocytes after induction with the widely used Bacillus Calmette-Guerin (BCG) vaccine, found that the amount of systemic inflammation and differential response to LPS upon secondary immune stimulation were markedly reduced. BCG has been used for decades to confer protection against the deadly mycobacterium tuberculosis (TB). However, recent literature has supported the idea that BCG confers immunity against other pathogens as well, including viruses, and the phenomena has been dubbed “trained immunity.” It has been hypothesized that the decline in the rate of BCG vaccination (owing to the lack of TB infections worldwide), might have contributed to the increase in COVID-19 mortality. In the study, an elevated production of key pro-inflammatory cytokines such as IL-1β, TNFα, and IL-6 was observed upon secondary challenge with LPS, following induction with the BCG vaccine. Differential expression of inflammatory mediating chemokines such as CCL3 and CCL4 was also observed, along with a strong correlation between IL-1β and CCL3 and CCL4 before training. Both CCL3 and CCL4 are regulators of inflammation in an inter-cellular signaling dependent manner, highlighting the ability of scRNA-seq to shed light upon the spatio-temporal indications provided in the data. After training, or transcriptional reprogramming induced by BCG vaccination, this correlation dampened or disappeared, suggesting that the reprogramming event lowered sensitivity of the monocytes to pro-inflammatory signals. It should be noted that this trend was found for both IL-6 and TNFα along with their correlated but differentially responding chemokine pairs. These observations are in concordance with the hypothesis that BCG induction can lower systemic inflammation upon secondary challenge with LPS. This study represents an effective use of scRNA-seq to evaluate the transcriptional re-programming of monocytes upon BCG induction and touches on the possible indications of spatial dynamics within the analyzed monocyte populations.

### 3.2 Single cell proteomics

As scRNA-seq allows to examine the heterogeneity of the transcriptome between cells, single cell proteomics enables quantification of proteomic heterogeneity. In contrast to scRNA-seq, where RNA can be amplified, proteins cannot be, which leads to lesser starting material in single cell proteomics experiments. There are many methodologies that attempt to overcome this limitation by absolute quantification of a small number of proteins or through multiplexed measurements. Due to the wide variety in methodologies, there is no centralized workflow for exploring the single cell proteome. Instead, the workflow chosen by the researcher should be specific to the research question being addressed. As such, the field of single cell proteomics can be broken down into two subsections—‘absolute quantification’ which involves targeted and untargeted methodologies, and ‘multiplexing based methodologies’ to carry out multiple protein measurements within the same experiment. As is the case with all quantification methodologies, there are tradeoffs in single cell proteomics as well. While one technique may have a wider dynamic range, it can suffer from low resolution, or a versatile multiplexing assay may lose sensitivity for the protein targets. Quantification with SCP also
immunoassays to accommodate single cell methodologies, bottom-up proteomics workflows have also been developed which most of the targeted single cell proteomics approaches rely upon. Similar to the miniaturization of ELISA, which is a real limitation, especially considering non-specific antibody binding as well as antibody availability. In the direction of for miniaturizing these immunoassays, microfluidic technologies have emerged at the forefront for absolute quantification-based techniques. Microfluidics now allow for experiments to be run in the microliter and even picolitre volumes, which is essential for detecting the low protein concentrations released from a single cell. For single cell western blotting (scWestern), Hughes et al. have developed a method that has been multiplexed for 11 protein targets and supports up to 1000 concurrent blots in only four hours. When a low starting number of cells is used (<200), FACS sorting can be applied in tandem to improve upon resolution for single cells. The overall workflow of a typical scWestern blot is housed in a 30 μm thick photoactive polyacrylamide gel which rests atop a glass microscope slide with an array of 6270 wells. Every step of the procedure from single cell sorting, cell lysis, target protein capture, and fluorescence detection are compiled within this apparatus. The group concluded that scWesterns represent a viable singe cell protein assay capable of high throughput quantitative analysis coupled with multiplexing ability. The technique can address target molecular mass via protein electrophoresis, as well as subsequent high affinity antibody probe binding. The information when combined, generates a high confidence protein identification and specificity, representing a powerful diagnostic tool for single cell proteomics analysis. A miniaturized version of ELISA was presented by Shirai et al., where they created a single-molecule ELISA apparatus using micro/nanofluidic technology. The device can use sample volumes in the picolitre (pL) range and has the capability to chemically process and capture single molecular targets. Both techniques represent significant stepping stones in the fields of microfluidics and single cell proteomics with vast applications in medicine and systems biology research. Even though most immunoassays rely upon fluorescent-based readouts, there are newer methods such as single-cell barcode chip (SCBC) which do not require fluorescence. Absolute quantification of proteins from a single cell can also be conducted in an untargeted manner. This bypasses the need for pre-defined protein targets and their corresponding high affinity antibody conjugates, which most of the targeted single cell proteomics approaches rely upon. Similar to the miniaturization of immunoassays to accommodate single cell methodologies, bottom-up proteomics workflows have also been adapted to the single cell scale with different considerations during the workflow. For example, when extracting the proteins from single cells, it is imperative to limit the non-specific absorption from materials that are housing the sample. Another consideration comes in with the use of enzyme trypsin. Since trypsin follows Michaelis-Menten Kinetics, the digestion rate is significantly hampered when using small substrate amounts (i.e., total protein within a single cell) when compared with the amount of substrate in a global bottom-up proteomics workflow (i.e., total protein from many cells pooled together). The ramifications of both these factors on the quantitative readout can be controlled by minimizing sample volumes. A more in-depth review on sample processing and emerging technologies in nano proteomics can be found here.

Specht et al. describe that using SCoPE-MS, they were able to detect the underlying heterogeneity in macrophage populations differentiated from monocytes using the agonist phorbol-12-myristate-13-acetate (PMA). They concluded that macrophage heterogeneity exists within each cell’s proteome independent of cytokine induced differentiation processes. The group assessed whether cell type can be assigned based on the abundance of proteins specific to monocytes or macrophages by color-coding cells based on the median abundance of differentially abundant proteins. They also assessed the protein fold changes, by averaging the fold changes of protein abundance in single cells (monocytes and macrophages) and comparing that to the fold changes in bulk samples for each cell type (mixing the lysates of single cells). This supports the utility of SCoPE2 as an accurate method for determining protein fold changes in single cells. Their findings also support the known biological functionality of M1 and M2 - polarized macrophages cell types suggesting that SCoPE2 provides an excellent framework for quantifying relative protein abundance at the single cell level and facilitates identification of cell-type specific proteins to assess population heterogeneity. This is one example of how single cell proteomics workflows can provide meaningful biological insight into heterogeneity, which is an important aspect to consider for any systems biology approach as heterogeneity is not accounted for in the omics approaches at global level.

Single cell analytics also involves the use of multiplexing methodologies allowing multiple analyses to run
simultaneously. This is important for systems biology research as cellular signaling dynamics are realized by the concerted functioning of many proteins. Some of the techniques can even retain spatial information, which is another fundamental aspect of cell signaling. Multiplexing has been traditionally limited by spectral overlap of fluorophores that are conjugated to the detection antibodies used for reading out data in a highly multiplexed experimental workflow. This limitation is bypassed using antibodies conjugated to metal isotopes which do not have spectral overlap, in a technique known as mass cytometry (CyTOF). This technique involves the vaporization of a tissue sample region using a laser beam, where the ejection or plume released by contact with the laser is aerosolized, atomized, ionized, and entered into a time-of-flight mass spectrometer to quantify the isotope abundance. Using the quantified isotope abundance, the “spots” of the vaporized material can then be mapped back to the original coordinates on the tissue section, with the generation of high dimensional map. This provides targeted proteomic information mediated by the detection antibodies while maintaining spatial resolution. The technique can be used with up to 40 unique isotope labeled antibodies to detect up to 40 different proteins. Since multiplexed mass cytometry-based imaging (IMC) provides information regarding cell composition, phenotype, and spatial organization, all of which dictate cellular signaling, IMC represents an important area of research for systems biology.

3.3 Single Cell Metabolomics

Single cell metabolomics (SCM) is designed to evaluate the metabolic profiles of cells with single cell resolution. Like many other single cell omics approaches, SCM provides a means of assessing heterogeneity between single cells. Majority of SCM pipelines rely on mass spectrometry-based methods to quantify the metabolomic profiles. As with all single cell methodologies, the first step is to isolate single cells. This can be accomplished while keeping the morphology intact using techniques such as FACS or microfluidic arrays, or by atomic force microscopy (AFM) which completely isolates the single cell’s metabolites in a probe. The next logical step is to quench all metabolic activity within the single cells. This is accomplished with the use of organic acids or solvents to denature enzymes and impede further conversion of metabolites, or by a technique known as snap freezing using liquid nitrogen to derail metabolic activity and promote membrane lysis. Snap freezing limits the use of reagents that could be considered contaminants in the downstream MS analysis but requires further workup to obtain pure metabolic profiles. The use of organic solvents in the quenching phase is beneficial because it also aids in metabolite extraction from the lysate. Depending on the type of MS used, different combinations of organic solvents are utilized. Guo et al. provide an excellent review on the appropriate use of solvent mixtures for each type of MS. After quenching metabolic activity and extracting the metabolites, samples are ionized. Ionization can be broken down into two distinct mechanisms, vacuum based or ambient methods. The review by Liu et al. provides an excellent summary of these sub variations including applications for each technique and data preprocessing/analysis strategies for the obtained MS spectra.

M1 macrophages undergo metabolic reprogramming from oxidative phosphorylation to glycolysis upon phenotypical differentiation. Sustained M1 macrophage activation leads to sustained inflammation and can cause tissue damage if not appropriately regulated. The M2 phenotype is acquired after polarization with IL-4 and macrophage colony stimulating factor (M-CSF). M2 macrophages mediate the inflammatory response by secreting anti-inflammatory cytokines such as IL-10 and TGFβ. This gives them an important role in tissue repair after the host inflammatory response. Unlike the M1 phenotype, M2 macrophages rely upon the citric acid cycle (TCA) to support their production of ATP through oxidative phosphorylation. Phenotypical assays to assess macrophage population heterogeneity rely upon detection of cytokines or membrane bound surface markers. However, cytokines and surface markers are expressed in both phenotypes, thus presenting the need to develop a more specific phenotypical classification assay. To reliably quantify phenotypical differences, metabolic profiling with an LC-MS approach can be used, specifically, time of flight SIMS MS (TOF-SIMS) coupled with an Orbitrap analyzer (3D OrbiSIMS) which in contradiction to ordinary SIMS or TOF-SIMS, allows for MS/MS of the metabolite. TOF-SIMS provides highly localized spatial resolution of cell surfaces and does not require extensive sample preparation compared to most other LC-MS techniques. Historically, TOF-SIMS approach has not been reliably used to document endogenous metabolic profiles due to the poor mass resolving power. However, when coupled with the high mass resolving power
and mass accuracy of an Orbitrap, characterizing metabolic profiles at the single cell level is possible. Using a targeted approach to analyze the lipid palette (matching lipid ion peaks to LIPID MAPS database), it was found that M1 macrophages had the highest lipid counts and different lipid composition compared to M2 or M0. The amino acid composition and other metabolites showed notable differences as well. Overall, the study represents a novel approach for in-situ characterization of metabolic profiles to assess phenotypical differences between closely related cell types with single cell resolution.

HIGH THROUGHPUT IMAGING:

High throughput imaging (HTI) represents a robust set of methodologies which provide information on cellular morphology and includes large scale automated sample preparation and image analysis. All HTI workflows are based upon a targeted approach, where a dye, fluorescent reagent, fluorophore conjugated antibodies/oligonucleotides, or genetic construct which expresses fluorescent protein are used to label a particular component of the cell. This can include proteins, nucleic acids, or specific organelles within the cell. The workflow begins with perturbing cells from steady state, by the addition of ligands to activate cellular signaling pathways and thus inducing differential gene expression or by utilizing short hairpin RNA (shRNA) and clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) which can respectively knockdown and knockout genes in a targeted manner, giving scientists the ability to diagnose the role of the labeled target in response to cellular perturbation. The ligand used to disrupt cellular steady state dictates the cellular pathways involved in regulating the change in phenotype for the target of interest.

Traditionally, drug discovery methodologies are target based, they rely on pre-cognition of a target molecule (interacting protein, or specific receptor), and assess the structural dynamics of “hit/lead” compounds that can bind the target and modulate its effector function. This approach typically involves screening the target against a library of compounds predicted to have high binding affinity potential based on knowledge of the binding pocket and or molecular docking simulations and assessing the mechanism of action for each compound in further downstream analysis experiments. The structure-based drug discovery is rapidly developing as artificial intelligence and deep learning algorithms progress. It should be noted that there is inherent bias with target-based screening methodologies, such that hit compounds designed for an individual target may have off-target effects (polypharmacology). To resolve this bias, phenotypic based drug discovery (PDD) methodologies start by looking at a cellular phenotype and attempt to correlate hit compounds that can modulate the phenotypic readout. Lin et al. provide a review that encompasses the details of workflow design and data analysis for each of these screening methodologies in greater detail.

HTI can also be used to profile compounds or libraries in a multiparameter fashion based on statistical analysis and clustering of compounds that elicit a similar phenotypic readout. For example, a multiplexed image-based assay termed “Cell Painting” can measure 1,500 morphological features pertaining to different combinations of size, shape, texture, staining intensity, and so on, in response to multiple perturbations. These perturbations can be chemically induced or could be a result of genetic manipulation (knockdown/knockout). The technique is capable of sensing subtle phenotypical differences and it groups perturbing agents (compounds/genes) based on similar pathways that they affect, thus shedding light on the markers of disease. The basic workflow involves genetic or chemical perturbation to a target cell line, staining cells with fluorescent dyes that label 8 different compartments within the cell (nucleus, F-Actin, plasma membrane, mitochondria, etc.), microscopy imaging, and image analysis where the morphological features taken from the image correlate to a profile that reflects the phenotypic state of the cell. Comparing profiles taken at different time points or with different perturbing agents can elucidate the mechanism of action and cellular signaling pathways pertaining to a particular phenotype.

Lastly, HTI methodologies can be classified into a third category, used for deep imaging which combines the use of fully automated high-resolution microscopy and sophisticated computational analysis of many images. Due to the massive amount of cellular input, this technique can assess rare cellular phenotypes that may only be present at low probabilities. This niche application of HTI is relatively unexplored, perhaps due to the limited availability of cost prohibitive high throughput imaging systems. However, even with the limited number of studies utilizing the deep imaging approach, the potential of HTI to identify
rare cellular phenotypes in response to a small subset of perturbing agents represents an important area of study in systems biology, that is, phenotypical profiling of biomarkers of disease pathogenesis for rare and understudied host-pathogen interactions.

Nuclear factor kappa-light-chain enhancer of activated B cells (NF-κB) is a family of transcription factors that regulate innate and adaptive immune responses, cellular differentiation, proliferation, and apoptosis. The mammalian NF-κB family is dimeric in nature and includes five different protein monomers (p65/RelA, RelB, cRel, p50/105, and p52/100) that form either homo or heterodimers and each dimer differentially binds to DNA. All the monomers have a conserved N-terminal domain known as the Rel homology domain (RHD), which is essential for DNA binding, dimerization, nuclear localization, and inhibitor binding. NF-κB p105 and p100 proteins contain a IκB inhibitory domain which contains, multiple copies of the ankyrin repeat (ANK) at the C-terminus. Both NF-κB subfamily proteins (p105 and p100) undergo proteasome-dependent partial proteolysis to their active DNA binding forms (p50 and p52, respectively). NF-κB dimers reside in the cytoplasm bound to inhibitor proteins of the IκB family, where upon degradation of the inhibitor (phosphorylation of IκB-by-IκB kinase (IKK) followed by ubiquitylation and proteasomal degradation), NF-κB translocated into the nucleus where it binds DNA and stimulates transcription of target genes. Importantly, one of the target genes is the inhibitor itself which provides a mode of NF-κB signaling regulation via a negative feedback loop. With constant stimulation, the degradation of the inhibitor as well as NF-κB re-synthesis leads to oscillations of NF-κB nuclear translocation. Oscillations in NF-κB translocation are signal dependent, for example, sustained stimulation of cells with tumor necrosis factor α (TNFα) leads to oscillations, whereas a short one-time stimulation with TNFz leads to only one sharp peak of NF-κB translocation/activation. Induction of cells with LPS leads to an entirely different response, where NF-κB has been found to translocate in either one cycle of translocation, persistent translocation, or oscillations in the patterns of translocation. Since the translocation of NF-κB has been found to be stimulus-dependent, it is important to consider the kinetic and spatio-temporal landscape of NF-κB when studying pertinent signaling dynamics. These studies highlight the importance of high throughput imaging techniques for studying cellular signaling dynamics with real time quantification and showcase how HTI workflows can be utilized to elucidate the mechanisms underlying host-pathogen interactions.

**COMPUTATIONAL SIMULATION AND MODELING:**

Computational biology involves the assessment of complex biological systems through the development of computational models and simulations which can be used to develop predictive models of the factors involved in disease pathogenesis. The field is rapidly progressing with developments in computer hardware, software, and experimental methods, lowering the computational efforts required to produce these models. Computational models can be separated into 2 sub groups, quantitative and logical models. A quantitative model utilizes sets of differential equations to define the dynamics of the model which are typically non-linear. It requires pre-defined knowledge of details regarding the pathway or cellular event under study and is thus limited to modeling small portions of a well classified pathway. A logical model is based upon a Boolean system and qualitatively defines the dynamics of the model. It does not require a pre-defined knowledge of the system to be analyzed, and thus can be applied to large scale systems.

A sub field of computational biology utilizes both modeling approaches and resides at the intersection of systems biology and traditional bioinformatics, known as systems bioinformatics. The field of systems bioinformatics can be defined as the framework for integrating the multiomics landscape traditionally used in systems biology approaches, to provide insight into each individual omics layer and the cumulative interactions between them. In this way, systems bioinformatics provides methodologies capable of assessing the biological mechanisms of the entire interwoven system rather than the summation of each individual component or omics layer. The generation of this field is based around systems theory which is holistic in nature, and it’s use in systems bioinformatics is dependent on graph theory, network science, and other mathematical approaches which facilitate the analysis of complex networks derived from the system of interest.

Mathematical models are fundamental for analyzing network topology and kinetics. As multiomics based quantification methods advance in both high throughput ability and sensitivity, more accurate parameters can
be fed into models and provide more accurately quantifiable simulations of signaling dynamics. Networks can also be applied to qualitative models of pathway modeling. Many open access platforms are available for that purpose. These software work by taking an input list of gene symbols or protein names and assessing their gene ontology (GO) to map them. For example, Cytoscape facilitates the visualization of complex biological networks with annotated gene symbols and expression data. Reactome enhanced pathway visualization is another peer reviewed alternative. There are several pathway databases which allow for the visualization of signaling components based upon GO terms. A review comparing some of the most widely used databases can be found here. Another detailed review on construction and analysis of biological pathways can be found here.

Construction of networks to showcase biological pathways can incorporate both mathematical modeling functions and qualitative visualization to help researchers fully understand the molecular dynamics involved in cell signaling. One such software is Simmune, that generates computational models incorporating the spatially resolved reaction-diffusion networks. Simmune utilizes rule-based approaches to lower computational complexity of simulations. This involves pre-programming of the simulation with fundamental signaling components (important proteins) and their pair-wise interactions which allows the computer to assemble the complexes that constitute the signaling network. This rule-based approach was incorporated as a response to one of the most traditional challenges in pathway modeling- combinatorial explosion. Combinatorial explosion can occur when there is excessively high number of alternative interactions arising from a network consisting of many different signaling components, or individual components that have multiple binding sites and thus many possible interactions. Simmune works by generating a local network in a multi-step fashion. The first step involves the construction of a non-spatial network that includes every possible molecular interaction for each fundamental signaling component. This ‘template network’ is then adjusted to reflect the local molecular environment which lowers computational extensivity of the simulation. Simmune is also able to account for morphologically dynamic models which usually requires rebuilding the network every time the cellular morphology changes, to account for spatial constraints concerning the receptor ligand interactions during membrane fluctuation. In-silico approaches for assessing pathway dynamics represent an important stepping stone in systems bioinformatics and all related disciplines to either validate or predict experimental findings. In a study put forth by Manes et al., the chemo sensing pathway Sphingosine-1-phosphate (S1P) was explored with a combinatorial approach of RNA sequencing, targeted proteomics, and Simmune based modeling for computational validation. This highlights the importance of in silico based computational models in providing insights into molecular mechanics of complex signaling networks and validating the experimental findings.

Protein folding simulations such as AlphaFold2 are also an informative tool for predicting and assessing protein structure and can be used to diagnose the protein’s effector function. AlphaFold2 can construct a 3-dimesnional representation of how a protein will fold, based upon its primary sequence. The software uses a deep learning-based algorithm with multi-sequence alignment which incorporates both physical and biological knowledge regarding protein structure. In addition to the software’s abilities, the AlphaFold team has now released accurate structure predictions for human proteome in a freely available database. The availability of a database with highly accurate protein structures that are continuously updated is a major step forward for the field of structural biology as it takes away the burden of generating these structural models from scratch. As AlphaFold2 evolves along with the database, we can expect to see more structural predictions that are publicly available and provide researchers with tools that can be exploited for drug discovery, investigating heteromeric protein-protein interactions, and creating simulations of pathway dynamics where each component of the signaling pathway includes a highly accurate 3-dimensional model of its native conformational shape. This will greatly benefit the field of systems biology as better structural predictions of the human proteome can help researchers assess all the possible functions of a protein and build more complex models regarding their kinetics.

INTEGRATING OMICS APPROACHES:

Case studies
Often overlooked, the microbiome constitutes an important mechanism involved in both gut-mucosal and systemic immunity. 70-80% of all immune cells are present in the gut. There are intricate interplays associated with the intestinal microbiota and local mucosal immune system which shape the immune response to invasive pathogens. The gut microbiota influences many aspects of host physiology such as the development of the immune system, drug metabolism, regulation of inflammatory diseases, overall nutritional status, and drug metabolism. For assessing the impact of the microbiota on host physiology, it is important to have a model that facilitates drawing meaningful biological conclusions. One such model is the use of germfree (GF) mice which have dramatic impairments on their metabolism and immune system.\(^8\) In the study put forth by Manes et al., comparisons are presented between the proteomic and transcriptomic expression profiles within the terminal ileum (a part of the small intestine with a high concentration of commensal microbes) of GF and conventional mice. The germfree status was treated as the perturbed state. Upregulated (up) corresponded to GF/C ratio values >1 and downregulated (down) corresponded to GF/C ratios <1. The first part of their analysis involved the global transcriptomic data of the ilea (total RNA-Sequencing) to assess transcripts affected by germ status. Hierarchical cluster analysis (HCA) of the data revealed two distinct clusters for up and down regulated genes. It should be noted that different strains of mice were used, and the germ status had a greater effect on the C57BL/10A mice than on the BALB/c mice. Ileum proteome data was then analyzed to assess the impact of germ status on protein expression. Like the trend in the transcriptome data, mouse strains also had a stronger effect on C57BL/10A mice than BALB/c, but this observation was milder in the proteomics data set. Using HCA on the data, 63 proteins were found to be affected by mouse strain. An HCA of these proteins was performed, and they were partitioned into up and down regulated groups. Highly expressed genes were identified more at the proteome level, however those that were more affected by germ status were classified at both the transcriptomic and proteomic levels. Next the group compared significantly affected genes based on germ status in both datasets to assess their statistically significant correlations. This was done by tallying the list of genes and analyzing them with HCA. Most of the genes which were significantly affected within one dataset were unaffected in the other, showing a level of underlying discordance between the two omics layers. However, those genes which were significantly affected in both showed concordance more so than discordance. Thus, the concordant and discordant genes between the two omics layers were used to construct subsets for each. These subsets were then put through ingenuity pathway analysis to uncover the enriched biological pathways associated with them. Presence of the microbiota in conventional mice was linked top upregulation of immune system pathways for the gene subsets which were transcriptome-proteome concordant. The group hypothesized that this was due to migration of immune cells to the ileum as shown in previous studies.\(^9\) Also, metabolic pathways were downregulated in conventional mice, and this was representative for both transcriptome-proteome concordant (concordant[T,P]) and discordant (discordant[T,P]) gene sets. For pathway analysis on the transcriptome enriched and proteome enriched genes, 22 pathways were downregulated in both and almost all of them were immune related such as antigen presentation. 43 pathways were upregulated in either or and they were almost all metabolism related. Only three pathways were upregulated in both, and they were the glutathione mediated detoxification, serotonin degradation, and xenobiotic metabolism signaling pathways. Network analysis was performed on the proteome subset and concordant[T,P] subset showing downregulation of immune system pathways and upregulated metabolic pathways, respectively. These pathways may be representative to those that of which are normally resistant to perturbation but affected by microbiota presence. Four pathways were upregulated in the discordant[T,P] gene set and two of them highly overlapped, mTOR and eukaryotic initiation factor 2 (eIF2) both of which contain protein translation machinery. Using ingenuity biofunctions, genes annotated with “RNA processing” or “Translation” and showed intra-transcriptome concordance and transcriptome-proteome discordance indicating mechanisms involved in posttranscriptional regulation. In fact, all the ribosomal proteins (RP) and eIF genes are known to be post transcriptionally regulated. However, this example is representative of only a subset of the multiomics workflows, and reveals the need to interrogate multiple omics layers to draw meaningful biological conclusions. Due to transcriptomic-proteomic discordance caused by the microbiota, neither of the layers alone is sufficient to produce a complete picture of the complex influence the microbiota has on the host’s physiology, emphasizing the need for integrative omics-based approaches to fully elucidate the host-microbiota interactions.
interactions and their molecular underpinnings. An overview of the discussed “omics” approaches and their application in analyzing various biomolecules (transcripts, proteins, metabolites) within a cell is presented in Figure 2.

The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused the disease COVID-19, a global health crisis. Mutated variants of the SARS-CoV-2 genome are responsible for enhancing the pathogenicity of the virus and thus the overall impact of the pandemic. Using a multiomics approach, Thorne et al. demonstrated that SARS-CoV-2 affects more than just the adaptive immune response; the innate immune system is implicated as well in the severity and transmission of the virus. They first assessed the multiplicity of infection for the first wave isolates (early lineage) and Alpha variant isolates in Calu-3 human epithelial cells by measuring intracellular copies of envelope (E), Nuclear capsid (N), and virion production. It should be noted that viral dsRNA is classified as a PAMP which are recognized by RNA sensing adaptors such as Mitochondrial antiviral-signaling protein (MAVS). Interestingly, they noticed 6 hours post infection (hpi) that the total area of dsRNA decreased for the Alpha isolates even though replication was comparable between alpha and first wave isolates, using single cell immunofluorescence. They hypothesize that this could be due to two factors. One being that the Alpha N protein can contribute to innate immune evasion by sequestering dsRNA and thus induce epitope masking, the other being the transposable elements contributing to the reduction of endogenous dsRNA production. It has been shown that Alpha infection leads to lower expression and secretion of interferon-β (IFNβ) - a prominent marker for innate immune activation. At early time points (24 hpi), Alpha was shown to induce less expression of IFNβ and interferon stimulated genes (ISGs) when compared to an early lineage variant B.1.13 hCoV-19/England/IC19/2020 (IC19), using RNA-Seq. This suggested enhanced innate immune evasion potential of the Alpha variant. To further characterize innate immune antagonism by Alpha, they compared global host responses through mass spectrometry-based protein abundance and phosphopeptide enrichment as well as total RNA-seq in Calu-3 cells (10 and 24 hpi). Notably, changes in RNA abundance and protein phosphorylation seem to be infection driven yet the changes in overall protein abundance are less drastic. It was also noticed that there is no clear correlation between protein/mRNA abundance and the levels of protein phosphorylation, indicating that enhanced phosphorylation may be driven by another mechanism independent of the levels of protein abundance. Gene set enrichment analysis comparing Alpha to early lineage variants highlights differences in pathways related to the innate immune system. The highest scoring terms, namely IFNα, IFNβ, and cytokine/chemokine signaling are the most enriched in the RNA and phosphorylation datasets. Also, the RNA-seq and protein abundance datasets show reduced expression of ISGs driven by Alpha infection which agrees with Alpha driven reduction of IFNβ production. They discovered that the innate immune activation is antagonized through up and downstream modulation of TBK1 (a kinase involved in nucleic acid sensing) related pathways, and due to decreased phosphorylation at early time points, suggesting that delayed activation of viral recognition signaling pathways is characteristic of the Alpha variant and not the variants of earlier lineage. To further assess the differences in host response between the different variants, viral RNA-seq and proteomics data were analyzed in parallel, and it was uncovered that the innate immune antagonist Orf9b was substantially increased in Alpha relative to earlier lineage variants. Looking further into the impact of Orf9b on the host response to the Alpha variant, RNA-seq data was used to map target genes to their corresponding transcriptional regulators to estimate transcription factor (TF) activities. By extracting significantly regulated transcription factors from the enriched terms in the RNA-seq pathway analysis, it became clear that IRF and STAT family TFs are less activated in Alpha infected cells than by early lineage variants. To corroborate this finding, single-cell immunofluorescence showed reduced IRF3 nuclear translocation after infection with Alpha in comparison to VIC. STAT1, STAT2, and IRF9 are downstream of the type I IFN receptor, and they are inhibited by Orf6 in Alpha infected cells. This inhibition leads to the inability of STAT1 and IRF3 to undergo nuclear translocation. In conclusion, Orf9b is regulated by host phosphorylation and suppresses MAVS downstream signaling by targeting TOM70 binding pocket (Ser50 and Ser53). This study demonstrates how multiomics approaches can be used to study changes within SARS-COV-2 variants and that the enhanced innate immune evasion is an important mechanism for enhanced pathogenicity of the Alpha variant.
INNATE IMMUNE PATHWAYS IN COVID-19 PATHOGENESIS:

The COVID-19 pandemic caused by the SARS-CoV-2 virus has been known to lead to immune dysregulation and potentially lead to acute respiratory distress syndrome (ARDS). ARDS can be caused by hypercytokinemia, otherwise known as a “cytokine storm.” This is characteristic of a hyperactivated and unregulated pro-inflammatory cytokine response, which can lead to inflammatory induced tissue damage for the host. Typically, viral infection is mediated by type-I IFN gene induction leading to the expression of interferon stimulated genes (ISG) that exhibit anti-viral mechanisms of action. SARS-CoV-2 has developed mutations that are structural changes in the proteome and can antagonize the host IFN response. Thus, studying the host immune response to SARS-CoV-2 represents a fundamental step for diagnosing phenotypes indicative of a dysregulated immune response and may provide insight on how to treat the disease. In the study put forth by Zhou et al., bronchoalveolar lavage fluid (BALF) was collected from 8 COVID-19 patients (SARS2), 146 community-acquired pneumonia patients (CAP), and 20 healthy controls (Healthy). Meta-transcriptomic analysis revealed 56% alignment to the human genome among all samples, also captures the transcriptomic information for microbes found in the samples. The differentially expressed gene (DEG) population for the SARS2 vs healthy (SARS2-H) comparison was markedly higher than the rest, showing that SARS-CoV-2 infection causes significant perturbations from homeostasis in the host lung tissue. Some key upregulated DEGs in SARS2-H included pro-inflammatory cytokine and chemokine genes (IL-1β, CXCL17, CXCL8, CCL2), anti-viral ISGs (IFIT, IFITM family genes), and calgranulin genes (S100A8, S100A9, S100A12). Downregulated DEGs included genes involved in morphogenesis and cellular migration (NCKAP1L, DOCK2, SPN, DOCK10). Among the cell signaling pathways, “interferon signaling” was most enriched in the SARS2 group and was also enriched to a lesser extent in the Virus-like CAP samples. This indicates an IFN-driven response to SARS-CoV-2 infection. Many of the enriched pathways for the SARS2 group are classified as innate immune pathways such as NF-κB, TNF, and the IL-17 pathways. Network analysis showed a densely connected subnetwork between the ribosome and chemokine signaling pathway. Upon assessment of the cytokine profiles, it was observed that the pro-inflammatory cytokine expression dissipates over time, suggesting that inflammation during COVID-19 is resolved over time and unquenched inflammation may lead to detrimental outcomes regarding disease progression. CXCL17, which has a role in neutrophil recruitment in the lung was the most upregulated chemokine for SARS2. Monocyte attractors, such as CCL2 and CCL7 were upregulated as well. Data indicated a correlation between viral load and chemokine production, where higher viral load corresponds to higher chemokine gene expression. IL1RN and ILB interleukin genes were enriched specific to SARS2, validated by their quantification of cognate protein products (IL-1Ra and IL-1β respectively) in COVID-19 patient plasma. IL-1Ra is the inhibitor to IL-1β, and IL-1β has been previously reported to be the driving factor of the proinflammatory response during ARDS, this suggests a potential biomarker for COVID-19 severity based upon the ratio of IL-1Ra and IL-1β. The IFN response was then examined where SARS2 showed elevated expression of ISGs (83 significantly upregulated) compared to CAP subgroups which diminished over time for each patient. These included IFIT and IFITM genes with broad antiviral functionality (IFIT1,2,3 and IFITM2,3); these have also been shown to inhibit viral cellular entry of SARS-CoV-2. Also, key innate immunity associated transcription factors, namely IRF7 and STAT1 were markedly upregulated which could further potentiate the IFN response. SARS2 also displays higher neutrophil population compared to the pneumonia group, and less diversity in the T and B cell populations when compared to the innate cell populations. In a previous study, neutrophil to lymphocyte ratio was characteristic of disease severity in COVID-19 cases, indicating yet another biomarker for disease severity. This study showcases the capacity of transcriptomics to detect phenotypes associated with disease severity at different expression levels, such as DGE analysis of cytokine/chemokine, ISG, and even broad morphogenic cell population determination. Several potential crosstalk points between the innate and adaptive immune system also bring to light the innate immune pathways which affect the pro-inflammatory response to SARS-CoV-2 pathogenesis. Study of host-pathogen interactions at both the innate and adaptive immune levels is required to decode the early and late biomarkers associated with pathogenicity and virulence and to develop effective treatment options in the clinical setting.

CONCLUSIONS:
Multiomics approaches are powerful tools for studying host-pathogen interactions and provide a comprehensive view of the innate immune system. Transcriptomics and proteomics offer valuable information in several areas such as differential gene expression, protein levels, site of modification, proteoform profiling, higher order structure, and protein interactions, both qualitatively and quantitatively, and with high accuracy. Considering the role of metabolites in the regulation of immune cell signaling; immunometabolism has gained more interest as well. It can predict how precise changes in the metabolites involved in the signaling pathways shape the immune response. Single cell measurements on the other hand, have made it possible to observe the heterogeneity between cells of the same population, whether it be the heterogeneity between transcriptome, proteome or metabolome or the architecture of a population of cells. Single cell omics has completely revolutionized the field, as a tool to study systems biology. As is the case with any technique, limitations regarding cost, highly specialized instrumentation, trained personnel, or availability remain for omics methods. Table 1 summarizes prominent features, limitations and executability of various “omics” technologies.

The development of advance instrumentation has truly empowered the omics approaches and has reduced our reliance on high affinity antibodies for exploring the immune system. High throughput imaging is highly informative in exploring cellular morphology and has been used to profile large compound libraries for drug discovery. Assessing molecular interactions over time in order to predict disease pathogenesis has become easier with the use of computational models and molecular dynamics simulations. Latest deep learning-based tools such as AlphaFold have established their value in structural predictions, investigating protein-protein interactions, and simulating pathway dynamics more accurately. Multi-omics analysis platforms have also revolutionized the diagnosis and pathogenesis of Covid-19, the biggest pandemic of our era, as well as aided in the identification of potential therapeutic treatments. Taken together, or even in part, omics tools are highly powerful in investigating systems biology closely, holistically, and multidimensionally, providing a larger and more complete picture of the innate immune landscape.

ACKNOWLEDGEMENTS:
This research was sponsored by the Intramural Research Program of NIAID, NIH.

AUTHORS’ DISCLOSURE STATEMENT:
The authors declare no conflict of interest.

REFERENCES:


(60) Tamahill, G. M.; Curtis, A. M.; Adamik, J.; Palsson-McDermott, E. M.; McGettirc, A. F.; Goel, G.; Frezza, C.; Bernard, N. J.; Kelly, B.; Foley, N. H.; et al. Succinate is an inflammatory signal that induces IL-1β through HIF-1α. *Nature* 2013, 496 (7444), 238-242. DOI: 10.1038/nature11986 From NLM.


(63) Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Audricopulo, A. D. Molecular docking and structure-based drug design strategies. *Molecules* 2015, 20 (7), 13384-13421. DOI: 10.3390/molecules200713384 From NLM.


(68) Roukos, V.; Misteli, T. The biogenesis of chromosome translocations. *Nature Cell Biology* 2014, 16 (4), 293-300. DOI: 10.1038/ncb2941.


(87) Bryant, P.; Pozzati, G.; Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* 2022, 13 (1), 1265. DOI: 10.1038/s41467-022-28865-w From NLM.


---

**Figure 1.** A timeline of major developments in the “omics” history.
Figure 2. An overview of the various applications of “omics” approaches in analyzing biomolecules within a cell as well as interactomics and morphology and at both the cell and tissue level. The instrumentation and data types are presented to emphasize the level of information provided by “omics” technologies.

Table 1. A summary of important features for each “omics” technique.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Prime Features</th>
<th>Limitations</th>
<th>Ease of execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteomics</td>
<td>Quantitative Ability to identify 1000s of proteins in single experimental run. Native (e.g., protein complexes) and non-native (e.g., peptide level) states can be examined. Lower sample amounts required (nanograms). Ability to assess PTMs and their distribution. Characterization of binding sites that modulate protein effector function.</td>
<td>Data analysis is customized based on the experiment and need knowledge of a variety of software. Isobaric labeling approaches for targeted proteomics may suffer from incomplete incorporation in cell culture. Membrane proteins are hard to isolate and digest due to difficulty of precipitation because of aggregation.</td>
<td>High-resolution instruments are cost prohibitive; specialized personnel needed to perform advanced workflows. Proteomics based MS workflows are well documented in literature, which makes execution of these methods easy to follow.</td>
</tr>
<tr>
<td>Technology</td>
<td>Prime Features</td>
<td>Limitations</td>
<td>Ease of execution</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Transcriptomics</td>
<td>Accurate quantification of transcript abundance. Alignment to a reference genome allows for identification of target genes. Can be used to assess differential gene expression after perturbation from homeostasis.</td>
<td>cDNA library construction can lead to dimerization of primers making datasets noisy. During cDNA library construction spurious second strand cDNA artefacts can be generated, which can confound sense vs antisense transcripts.</td>
<td>Next generation sequencing instruments are expensive. However, sending samples to specialized sequencing facilities is cost effective and often a better alternative. Protocols for sequencing, alignment, and statistical analysis are well defined making RNA-Seq practical for the masses.</td>
</tr>
<tr>
<td>Metabolomics</td>
<td>Metabolic profiling in quantitative manner. Ability to assess metabolic flux in response to perturbation from homeostasis.</td>
<td>Real time metabolic profiling is difficult. Lipids can be particularly hard to ionize using MS based approaches.</td>
<td>MS based approaches for metabolite quantification are well defined in the literature. Difficult to execute without access to a high-resolution Mass spectrometer.</td>
</tr>
<tr>
<td>Single Cell omics</td>
<td>Deeper analysis of rare cell types. Recent advances have increased throughput and reproducibility. Phenotypical classification by assessing intrapopulation heterogeneity at each omics layer.</td>
<td>Sample preparation, isolation of single cells and throughput are limitations for all single cell approaches. Targeted approaches are reliant on antibody availability. With new improvements coming up rapidly, reproducibility could be an issue.</td>
<td>Due to the abundance of methodologies for both targeted and untargeted approaches, it is now relatively easy to incorporate single cell workflows into any workflow. Some approaches rely on expensive instrumentation (MS, NGS), others, for example, miniaturized immunoassays for scProteomics are more affordable.</td>
</tr>
<tr>
<td>Technology</td>
<td>Prime Features</td>
<td>Limitations</td>
<td>Ease of execution</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>High Throughput Imaging</strong></td>
<td>- Localization of surface molecules, subcellular localization of organelles and macromolecules. Phenotypical characterizations based on expression of specific markers in response to homeostatic perturbations. Accurate characterization of cellular morphology. Assessment of intrapopulation heterogeneity based on expression of specific markers or overall cellular morphology.</td>
<td>The availability of antibodies specific to the markers of interest.</td>
<td>High resolution imaging instruments are expensive.</td>
</tr>
<tr>
<td><strong>Computational Modeling</strong></td>
<td>- Ability to profile the kinetics of molecular interactions. In silico validation of experimental findings can help predict the outcome of experiments. Highly accurate 3-dimensional modeling of protein structures. Crucial for early stages of drug discovery.</td>
<td>Specific parameters need to be set from experimental results for the simulation to mimic endogenous conditions. Assumptions are required at the software end that may not accurately describe the environment, for example, assuming the reaction volume to be homogeneous. AlphaFold multimer does not accurately account for PTMs and not all PPIs are accurately mapped.</td>
<td>Most software packages that can be used to model molecular dynamics, pathways, or protein modeling are open source and do not require a commercial license. This makes <em>in-silico</em> computational models highly accessible and easy to use. Some modeling software are computationally costly and require the use of high-performance clusters.</td>
</tr>
</tbody>
</table>