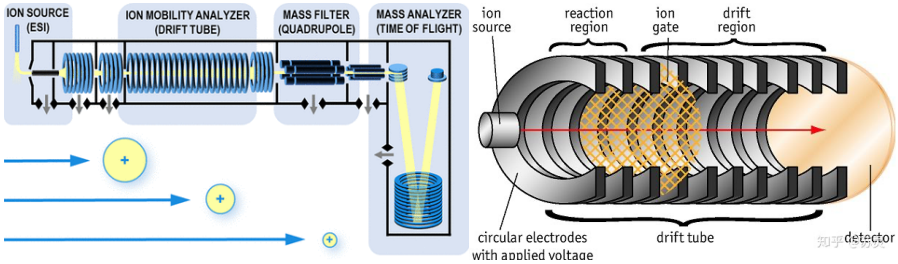
Short answer questions:

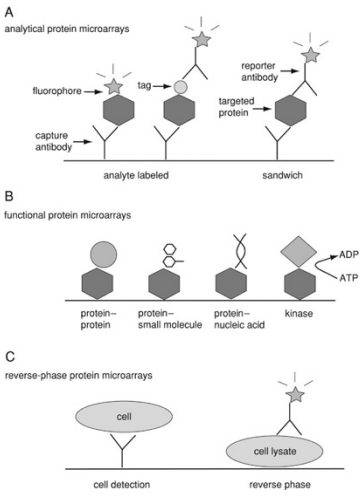
1. The working principle of IMS based mass spectrometer is as shown in the figure above. The inlet ion funnel is focused and enters the trap. After accumulating a certain number of ions, the ion gate is opened and ions are injected into the drift tube. The ions pass through the drift tube in turn, enter the exit ion funnel, focus again, and enter the post-quadrupole mass selector.



The working mode of **DTIMS** is very similar to that of TOF. It works intermittently. There is a pulsed ion gate at the front end of the drift tube, and a certain amount of ions is put in each time. After the complete scanning is completed, the next working cycle is entered. The mobility of the ions can be obtained from the time that the ions pass through the drift tube, which is related to the collision cross section of the ions. Generally speaking, the smaller the cross-sectional area and the higher the charge, the faster the migration speed.In IMS, the force exerted by an **electric field** on an analyte ion is exactly balanced by friction with the buffer gas, yielding a steady-state analyte velocity v_d.

CCS is the momentum transfer between ion and gas particles averaged over all gas-ion relative thermal velocities. Ion mobility separates ions based on their collision cross sections, i.e., their shapes. If a molecule or molecular complex has multiple conformations, these can often be resolved from one another. Thus, changes in protein folding or differences in glycan branching can be detected and closely related molecules can be separated based on the variations in their shapes.**4D** – in addition to retention time, precursor mass, isotopic pattern and MS/MS spectrum. This allows CCS values to be utilized across the timsTOF community to increase confidence by reducing ambiguities - resulting in fewer false positive annotations.

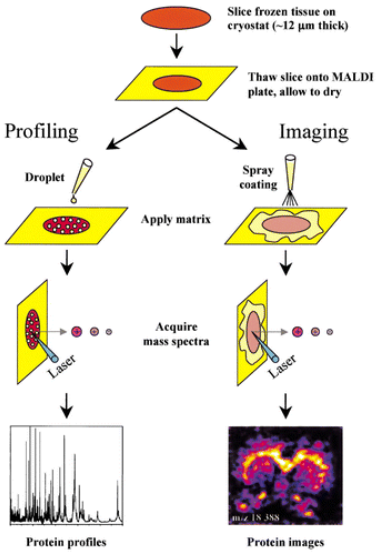
2. **Analytical microarrays** take advantage of highly specific antigen-antibody recognition to build a protein detection system. The expansion of the capability of conventional immunoassays into antibody array applications enabled a parallel and multiple detection system using a small amount of sample. Moreover, this technology has high sensitivity and good reproducibility in quantitative assays. Targeted proteins can be detected either by direct labeling or using a reporter antibody in sandwich assay format.



Functional protein microarrays are made by spotting all of the proteins encoded by an organism and therefore are useful for the characterization of protein functions, such as protein-protein binding, biochemical activity, enzyme-substrate relationships, and immune responses.

Reverse-phase protein microarrays providing an alternative format to analytical microarrays in which tissue/cell lysates (or fractionated lysates) are used to form such an array. They provide a different array format by immobilizing many different lysate samples on the same chip.

3. **Tissue profiling:** In general, there are two basic modes of data acquisition: profiling and imaging. In the profiling experiment, one is interested in comparing protein patterns from a discrete number of spots or areas. Although there is no spot limitation, typically this is done for 5–20 regions across a given tissue section. Matrix is applied, as discrete droplets (spots) to the regions of interest, and the sample is placed into the MS source. In the imaging experiment, the goal is to display a detailed molecular image of an entire tissue section or a specific subregion. In this case, matrix needs to be homogeneously deposited across the section without generating any major lateral protein migration.



In-situ drug quantification: Imaging drugs by mass spectrometry offers significant advantages over traditional imaging techniques (such as autoradiography and fluorescence spectroscopy). Because intact drugs may be desorbed directly from tissue surfaces, no additional syntheses are required, thus allowing this type of localization analysis to occur earlier in the drug discovery process with less cost. In addition, any confounding pharmacological effects due to a bulky label are eliminated, as are the environmental issues associated with radioactivity. Finally, the molecular specificity of mass spectrometry allows the intact drug to be distinguished from its metabolites that differ in mass.

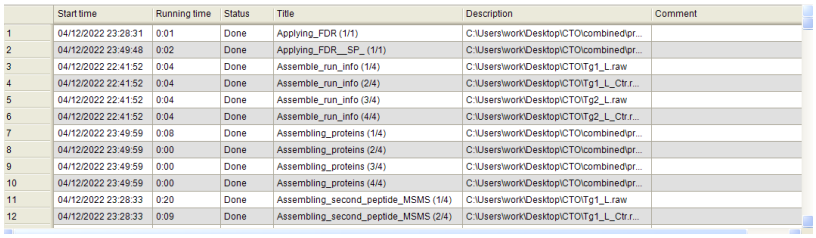
Western blotting & immunohistochemistry (IHC)

Immunohistochemistry is the most common of immunostaining techniques. It visualizes the presence of targeted antigens within a tissue sample by using antibodies that attach themselves to the antigens, having a catalytic effect, and splitting the molecule into identifiable compounds. Visualizing this interaction can be achieved by using an enzyme (producing a chromogenic signal) or a fluorophore (creating a fluorescent signal). Western blot works on the same principle. **Preparation differs:** For **immunohistochemistry**, the tissue samples are sliced or left whole, depending on their size. The limit for the size of samples generally ranges from 3 µm to 5 µm. Next, samples are embedded into a medium, usually paraffin wax or cryomedia. **Western blot** requires that samples are first separated by electrophoresis and then immobilized in a blotting membrane. **Sample staining: Immunohistochemistry** adds either polyclonal or monoclonal antibodies. **Western blot** method sees a fluorescent dye added to the sample, followed by exposure to a light source of the appropriate wavelength to excite the molecules of the fluorophores. **Applications: Immunohistochemistry** is mostly relied on as a diagnostic tool for numerous cancers. **Western blot** is mostly used for the detection of autoimmune diseases, allergies, and infectious diseases. It is widely used in the fields of molecular biology, biochemistry, and cell biology, with its most notable applications being used as a diagnostic tool for HIV and BSE.

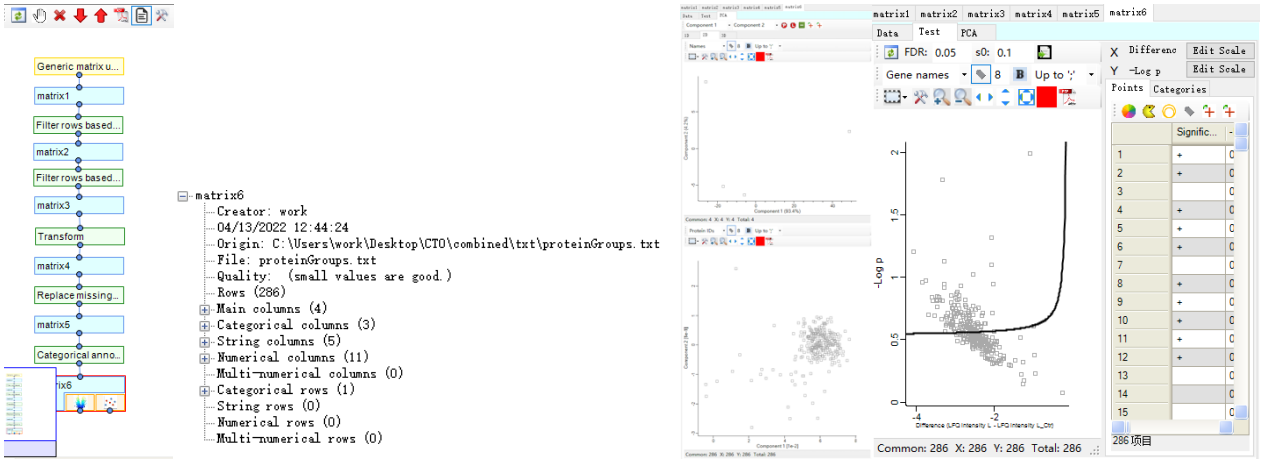
Computational questions:

1. Using the semi-trypsin option means the samples were digested by this enzyme. The modifications can be set in group parameter option, we can choice Carbamidomethyl(C) in Fixed modifications, then Variable modifications is above that where you can find Acetyl(K) or citrullination(0.98) (but this one I did not find in my software). In instrument type option, we select Orbitrap according to our experiment selection, and set first search peptide tolerance at 20; Main search peptide tolerance at 4.5; and peptide unit is ppm. The raw data from MS is RAW file.

2. screen dumps:



	Start time	Running time	Status	Title	Description	Comment
1	04/12/2022 23:28:31	0:01	Done	Applying_FDR (1/1)	C:\Users\work\Desktop\CTO\combined\pr...	
2	04/12/2022 23:49:48	0:02	Done	Applying_FDR__SP_ (1/1)	C:\Users\work\Desktop\CTO\combined\pr...	
3	04/12/2022 22:41:52	0:04	Done	Assemble_run_info (1/4)	C:\Users\work\Desktop\CTO\Tg1_L.raw	
4	04/12/2022 22:41:52	0:04	Done	Assemble_run_info (2/4)	C:\Users\work\Desktop\CTO\Tg1_L_Ctr.r...	
5	04/12/2022 22:41:52	0:04	Done	Assemble_run_info (3/4)	C:\Users\work\Desktop\CTO\Tg2_L.raw	
6	04/12/2022 22:41:52	0:04	Done	Assemble_run_info (4/4)	C:\Users\work\Desktop\CTO\Tg2_L_Ctr.r...	
7	04/12/2022 23:49:59	0:08	Done	Assembling_proteins (1/4)	C:\Users\work\Desktop\CTO\combined\pr...	
8	04/12/2022 23:49:59	0:00	Done	Assembling_proteins (2/4)	C:\Users\work\Desktop\CTO\combined\pr...	
9	04/12/2022 23:49:59	0:00	Done	Assembling_proteins (3/4)	C:\Users\work\Desktop\CTO\combined\pr...	
10	04/12/2022 23:49:59	0:00	Done	Assembling_proteins (4/4)	C:\Users\work\Desktop\CTO\combined\pr...	
11	04/12/2022 23:28:33	0:20	Done	Assembling_second_peptide_MSMS (1/4)	C:\Users\work\Desktop\CTO\Tg1_L.raw	
12	04/12/2022 23:28:33	0:09	Done	Assembling_second_peptide_MSMS (2/4)	C:\Users\work\Desktop\CTO\Tg1_L_Ctr.r...	

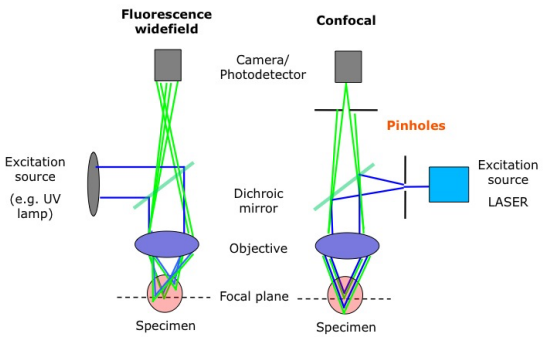


The PCA plot is quiet success where PC1 illustrate 95% of the result, but the Volcano plot cannot bring a good illustrate the separation of patients.

The sources of **missing values** range from tryptic miscleavages to ion suppression in the mass spectrometer, and improper MS/MS fragmentation. Further data processing use 'Perseus v1.6.14.0'. Protein group LFQ intensities are log2-transformed to reduce the effect of outliers. To overcome the obstacle of missing LFQ values, missing values are imputed before fit the models. Hierarchical clustering is performed on Z-score normalized, log2-transformed LFQ intensities. Log ratios are calculated as the difference in average log2 LFQ intensity values between experimental and control groups. Two-tailed, Student's t test calculations are used in statistical tests. A protein is considered statistically significant if its fold change is ≥ 2 and FDR ≤ 0.01.

Assay question:

If we want to see the location of different proteins within a cell, we can use **Confocal microscope**. In this kind of light microscope, light is emitted from the excitation source and hits the dichroic mirror, then dichroic mirror will only allow some wavelengths to pass through while other wavelengths are reflected. The excitation light is let through the dichroic mirror and then hits the specimen that has been "stained" with fluorophore. Finally, the fluorophore will re-emit light that will pass through the dichroic mirror to the photodetector. The laser is focused on the specimen one point at a time until the whole specimen has been scanned. **The central steps** for detection of the ACE2 or S-protein usually contains: Specimen collection, Fixation, Embedding, Sectioning, Staining, Microscopy, and Report.



We can use a combination of targeted lipidomics and untargeted metabolomics optimized in-house for screening human plasma samples. Then, accomplish structural confirmation based on tandem mass

spectrometry (MS/MS) spectra, and quantitate these metabolites (lipids or polar metabolites) using internal standards. **Targeted Lipidomics**: Prior to analysis, plasma lipid extracts should be resuspended in 100 μL of chloroform: methanol 1:1 (v/v) spiked with appropriate concentrations of internal standards. While **Untargeted Metabolomics** Prior to analysis, aqueous extracts should be resuspended in 100 μL of 2% acetonitrile in water. Chromatographic separation could perform on a reversed-phase ACQUITY UPLC HSS T3 1.8 μm column (i.d. 3.0 × 100 mm) (Waters) using an UPLC system (Agilent 1290 Infinity II; Agilent Technologies).

Protein array: In pathogen-borne diseases, protein array can be used for proteome-wide surveys to study humoral immune responses to discover multiple isotypes of antibodies against multiple antigens. In this paper, by using arrays printed with overlapping peptides from the SARS-CoV-2 proteome or antibodies, researchers have probed COVID-19 to investigate fundamental protein−protein interactions. In analytical protein microarrays, targeted proteins can be detected either by direct labeling or using a reporter antibody in sandwich assay format. **Mass spectrometry-based technologies:** iTRAQ and TMT technologies use multiple isotopic labels, which can be linked with amino groups to achieve the characterization and quantification of multiple sample proteomes. Label Free technology does not rely on isotope labeling. It can analyze protein enzymatic peptide fragments by LC/MS, analyze mass spectrometry data generated during large-scale identification of proteins. Data-independent acquisition (DIA) is a technique that significantly increases the throughput and reproducibility of proteomics research. The classic DIA process usually relies on the construction of the DDA spectral library, which needs to consume a large number of samples, and has a long cycle and high cost. Targeted proteomic quantification can make up for the deficiencies of non-targeted quantitative omics quantification, and has the advantages of high throughput, high accuracy, and reproducibility. Based on proteome/transcriptome/genome data, in large biological sample volumes, validate these markers. Targeted proteomics quantification mainly includes MRM/SRM and PRM. Targeted proteomics quantitative technology can be used for signal transduction pathway detection, tumor marker research and post-translational modification research.

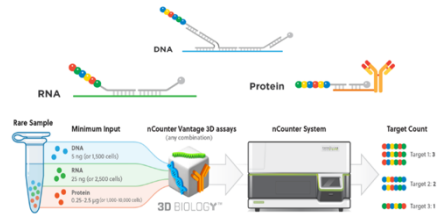
Using **mRNA-protein data** could find correlation between proteome and transcriptome data of variant genes. We can compare protein versus mRNA abundance of all genes or only variant genes (like the genes detected at the protein level tended to have a higher mRNA expression level as compared to genes not detected at protein levels, in both variant genes and whole genes). Total RNAs can be extracted using RNeasy® Mini kit (Qiagen). Libraries can be constructed using Illumina TruSeq RNA Sample Prep Kit v2-Set A and sequencing by using the Illumina Genome Analyzer IIx.

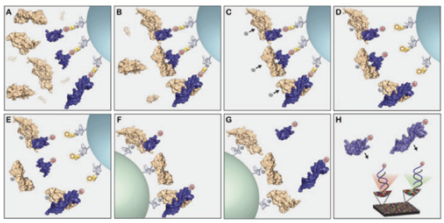
Long-read sequencing technologies are capable of reading longer lengths, between 5,000 and 30,000 base pairs. They sequence a single molecule, eliminating amplification bias, and generate a reasonable length to overlap a sequence for better sequence assembly. There are two predominant long-read sequencing technologies. Pacific Biosciences have developed a Single Molecule Real-Time (SMRT) sequencer that generates reads that can exceed 10,000 bases in less than two hours. Oxford Nanopore Technologies' platform relies on changes in the ion flow as nucleotides pass through a nanopore. A downside to long-read sequencing is that the accuracy per read can be much lower than that of short-read sequencing. The high error rate of nanopore technology is largely due to the inability to control the speed of the DNA molecules through the pore – these are systematic errors. Errors in SMRT sequencing is completely random. **Short-read technologies** carry out sequencing by synthesis or ligation. Each strategy uses DNA polymerase or ligase enzymes, respectively, to extend numerous DNA strands in parallel. Nucleotides can either be provided one at a time, or

they can be modified with identifying tags. They can be further categorized as either single molecule-based, involving the sequencing of a single molecule, or ensemble-based, which is the sequencing of multiple identical copies of a DNA molecule that have usually been amplified together on isolated beads. While short-read sequencing technologies contain: Illumina, 454 pyrosequencing, Ion Torrent, SOLiD, cPAL.

Wetlab: For **RNA isolation**, samples can be diluted 1:1 in Zymo DNA/RNA Shield (Zymo Research, cat. R1100-250) and frozen at −80°C until RNA isolation. 500 µL of sample were isolated using the Qiagen RNeasy Mini Kit (Qiagen, cat. 74106), following the manufacturer's protocol. Then **Shotgun Metagenomics Sequencing:** The cDNA can reverse transcribe (without DNase treatment) from 5 µL of RNA per sample using the Ovation® RNA-Seq System kit (Tecan, cat. 7102).

The basic idea of **PASEF** is to utilize the accumulation and ordered "serial" release feature of TIMS to increase the efficiency of MS/MS experiments. We achieved this by first accumulating ions "in parallel" to their mobility analysis so as to avoid ion losses and then by synchronizing the precursor selection in the analytical quadrupole with the elution of mobility-separated ions from the TIMS device. Rather than selecting only a single precursor per TIMS scan, PASEF selects multiple precursors one after another, that is, serially during each TIMS scan. In **DIA**, groups of ions are recursively isolated by the quadrupole and concurrently fragmented to generate convoluted fragment ion spectra composed of fragments from many different precursors, and the sampled fraction of the ion current was approximately threefold higher than in DDA. To make use of the correlation between the ion mobility and the m/z of peptides, precursors are trapped and then released in synchronization with the quadrupole position in **diaPASEF** scheme, which results in almost complete sampling of the precursor ion beam. We can apply diaPASEF to increase proteome coverage of COVID-19 samples, or try a compact and robust quadrupole-orbitrap mass spectrometer equipped with a front-end High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) Interface.





Quantitative reverse transcription PCR (RT-qPCR) involves the detection and quantification of RNA. The process is performed by reverse transcription of total RNA or mRNA to complementary DNA (cDNA) by the enzyme reverse transcriptase, followed by amplification and detection of specific targets of this cDNA using a technique called quantitative PCR (qPCR) or real-time PCR. **NanoString nCounter** fluorescent molecular tag technology is a new digital nucleic acid and protein quantification technology that captures signals through molecular barcode and single-molecule imaging technology, and counts the number of specific target molecules in the reaction system through molecular tags. This technology uses target-specific fluorescent probes to bind target molecules (DNA, RNA or protein), and can achieve quantification of target molecules without tedious enzymatic reaction steps such as library building, amplification, and reverse transcription. This technolgy has **no reverse transcription, no amplification, no technical duplication**, and directly quantitatively analyze the target nucleic acid in the sample, with **less external interference**. **SOMAmer** reagents are single stranded DNA-based protein affinity reagents. The assay offers exceptional dynamic range, quantifying proteins that span over 8 logs in abundance (from femtomolar to micromolar), with excellent reproducibility (4.6 median %CV for plasma).