Functional questions.

- Describe 3 quantitative methods: Silac (stable isotope labeling of amino acids in cell colture, Icat (isotope coded affinity tags), Itraq (isobaric tag for relative and absolute quantitation)

- Metabolic - Silac: amino acids are modified with stable isotopes introduced in the cell. Metabolic labeling starts at an early stage and the label is included directly on the protein, usually on the Arginine and modified by 13C isotope. We have 2 groups, A and B, cells are mixed and the proteins are extracted and purified. The workflow is performed joining the 2 conditions, LC/MS-MS in the end to identify and quantify the proteins. The relative intensity of the peaks comparing light and heavy peptides gives us information about the quantity.
  We can mark the proteins very early removing some technical variabilities, but the drawback is the fact that we can use it only in models.

- Chemical - ICAT: it uses an engineered tag with different functionalities, at one end we have iodoacetyl that reacts with the SH of cysteines, then a linker region with deuterium or 13C isotope to make it heavy, and a biotin tag to allow the purification of the peptides by AC. Again we have 2 cell states (or control and disease), we extract the proteins and we label them with light and heavy ICAT labels, then mix together and purify on AC. Again we compare the relative M/Z peaks ratios. This method is compatible with a lot of body fluids and cells and the alkylation reaction with the cysteine is very specific and it reduces the complexity of the sample, since we can isolate only cysteine-containing peptides. The problems are that it fails completely for protein that have no cysteines, plus the modification is a rather large label of around 500Da, if our peptide is 1000Da we truly change the condition of ionization.

- Chemical - iTRAQ: it uses a diminished tag in respect of ICAT, the weight is not changing, is perfectly co-eluted and ionized in the same way. The molecule is engineered with a peptide reactive group to tag the epsilon amino groups of the N-terminal domain and an isobaric tag composed by a reporter and a balancer with a total mass of 305Da. We can multiplex up to 8 targets and follow the quantities in time. The workflow is pretty similar with the other but the different thing is that the tag is fragmented during the first tandem MS, yielding a single low mass reporter ion that is used then to calculate the relative abundance of each peptide in the next MS. The remaining peaks are used for peptide identification

- Methods to define complex composition:
  - Complex composition is obtained by using affinity tools, to look at the proteins and molecular interaction between them. We can use ImmunoPrecipitation, Co-IP, Affinity Purification like TapTag (either single or dual tag). We also can use labeling methods either based on Silac or chemical labelling, like PAM AP-QMS, MAP AP-QMS, Label Free AP-QMS e QTAX (in vivo cross linking using formaldehyde).
    The analysis of complex PPi, which are the key of all biological processes and clinical conditions, are studied through quantitative MS, comparing mainly the relative ratios between 2 samples, control and disease.
    Briefly, IP relies on using specific antibody against one or all target proteins, with the drawbacks of potential low specificity and cross-reactivity. If the protein is ectopically expressed and there's no available antibody we use a TAP-Tag bound to the recombinant protein of interest consisting on a TEV cleavage site and an affinity tag for the purification using LC. Dual tag has a calmodulin binding peptide, making possible to use a double step affinity column using IgG beads first and calmodulin beads in the second step, with a divalent cation affinity column interacting with calcium ions.
    Labeling approaches are definitely the most used, either PAM or MAP.
    We can also use a label free approach, in which the sample and the control don't mix together at all and the quantification is based on spectral counting.
    Some complexes are too transient to survive the AP process, so we need to freeze the interactions first using cross linking before cell lysis and mixing. This way we can also compare the 2 methods to see which are the loosely attached partners. We rely on Formaldehyde, which is a short cross linker working on amino groups, or longer cross linker like succinimide esters, that works on lysine. Using a tag that is stable we can wash the transient interacting peptides using stringent washing conditions.
    Another possibility is using nanobodies or using proximity labeling of interacting complexes through a bioID. The protein of interest is tagged with the birA enzyme that can biotinylate all the proteins in close proximity with the proteins of interest. We can then recover all the interacting proteins using streptavidin. A better version of this technology is called APEX, which uses ascorbate peroxidase derivative to catalyze the same reaction but way faster then the birA e.Coli enzyme.

- PAM/MAP:
  - Purification After Mixing is a early stage SILAC based method that uses a label that doesn't affect the complex, since it's isotope based labeling. Sample and control are grown on different medium, labeling arg and lys with heavy or light isotope for identification, so ideally every peptide will incorporate at least one label. Cells are lysate and mixed, then we use AP before enzymatic digestion and LC-MS/MS. The two samples are easily distinguished in mass spectrometry because of the difference in molecular weight.
    MAP or Mixing After Purification instead utilizes the approach of mixing the 2 lysate only after the purification steps, so we preserve the integrity of our complexes before enzymatic digestion. MAP can be performed by both metabolic and chemical labeling. If we choose chemical labeling, we can use ICAT at the protein step, binding with the cysteine with a biotin tag, so we recover only the tagged peptides, reducing the complexity of the analysis. The problem is that not all the proteins can be modified, so we could use a ITRAQ approach at the level of peptides, before mixing.
    So one technical obstacle with PAM-SILAC approach is the perturbation of the thermodynamic environment, so all complexes are mixed. In this new environment, only the complexes with a high enough affinity are still bound together, the others can reorganize based on their concentrations. So to catch dynamic and loose interaction we can use a faster purification time or a time-controlled PAM-SILAC experiment. To catch the interactors with extremely high on/off rates we need to switch to a MAP approach, to preserve the PPi during purification. Working with both approach in any case let's us compare them, giving information on the dynamic characteristics of the complexes.

- Methods to define complex topology:
  - To understand the topology of a complex, or how are the proteins associated, we need a set of techniques associated to structural proteomics, because AP MS gives just the list and the quantity of the interactors. We can work at different levels: protein level, looking at the intact complex through CryoEM or native/non-denaturing MS; or we can work at the peptide level, using H/D exchange, covalent labelling or chemical cross-linking.
    HD exchange is an easy protocol and widely applicable but it does cost a lot.
    Covalent labeling is based on the introduction of an irreversible modification at reactive side chain of surface exposed residues in proteins and protein complexes. The labeling occurs through a radical based polymerization. We add small molecules that reacts on primary amines, usually epsilon amino groups on lysins on the surface, like anhydrides. We can also use an hydroxyl radical and we can control when the radical will be produced and react by switching the light.
    Cross linking freezes transient interactions, it's a covalent coupling that implies the use of a cross-linker, like formaldehyde.

- Describe the H/D exchange and what it's used for:
  - H/D exchange is one of the structural proteomics techniques used to study the topology of the complex in PPi. It's a peptide level techniques, based on the observation that solvated proteins have a tendencies of exchange H with the ones surrounding the protein itself. If H2O is replaced with something isotopically different like D2O, the resulting exchange will be visible in mass spectrometry, since D is about 1006Da. The protein complex is put in deuterium so the exchange can take place. To avoid reverting back when the protein is put back in water, we chemically disfavor the exchange using low pH and low temperature. At this point we can proceed with enzymatic digestion at low pH, for example with pepsin, and LC-MS/MS analysis still at low pH and low temp, obtaining the residues that are part of the accessible area of the complex.
    We can perform the opposite experiment as a control, so put both partners individually in deuterium, then mixed and placed back into water. At this point we have a double exchange in which only the interacting peptides will be bound to deuterium.

- Crosslinking:
  - It's a method used to freeze transient interactions, it's a covalent coupling that implies the use of a cross-linker. It's used to study the topology and the list of proteins involved in a complex. Using XL we obtain informations on the spatial proximity of the actual ligands, we can localize precisely the contact points between the two partners and we can also define the distance between the two. This is all done because when using affinity purification we may not recover all the proteins of the complex, because the affinity of some protein is lower than the stringency of our protocol. Primary amines (lysin) are the most targeted with a very high pK. Depending on the crosslinker that we use we have a space arm length that varies from 4.4 A to 53.4 A, this means that we can choose the most appropriate length depending on what we are studying, looking at the very near interactions or more distant ones.
    There are Homobifunctional crosslinkers that have identical reactive groups at either end and Heterobifunctional crosslinkers, that have an interactive group and a photocleavable one, which can be cleaved during MS analysis to create a marker ion for the low part of the mass spectrum in the non polluted area, something like we do with iTRAQ.
    Lastly, we have different type of interaction between a crosslinker and our target PPi. It can be a dead end cross link, so there's no link; a type one interaction in which the spacer is reacting with the same ligand; a type 2 which is the actual cross link interaction between 2 partners. This is why using a photocleavable crosslinker can help reduce the complexity of our analysis. For example, if we have a dead end crosslink we obtain a characteristic mass that reveals the fact that the 2 proteins are not linked together.

- Phospoproteome:
  - Phosphorylation is a key PTM in which and amino acid residue is phosphorylated by a protein kinase by the addiction of covalently bound phosphate group, causing the protein to become activated, inactivated or modifying it's function. The reverse reaction is done by a phosphatase that removes the phosphoric group. Recovering the phosphoproteome is key to understand the functional protein level.
    Target amino acid are Ser, Thr and Tyr with a ratio of 90%, 10% and 0,05%, which is not correlated with their abundance. The classical approach to recover the phosphoproteome is done by using specific stain like Pro-q diamond or by using a Phosphate tag to improve the sensibility composed by a chemical structure interacting with the phospho group forming a dinuclear metal alkoxide-bridge complex. The tag either has a biotin group to use with AC or an acrylamide group to use with a retarding gel.
    Gel free approach is done by using an immuno-enrichment pY column in conjunction with the biotin Pos-tag, the problem is that we don't have Ab truly specific for Ser and Thr. So we can use a MOAC, metal oxide affinity chromatography, that uses a metal like titanium oxide which can bond to the Pos-tag, the only problem is that it binds to every negative charge, like carboxylic ending, so we have to work in high acidic condition with trifluoroacetic acid as modulator.
    Another type of metallo oxide to use is Zirconium oxide, it's similar to TiO2 but we can prepare a sort of nanofiber to insert into pipette tips to selectively enrich the Phosphoproteome, both at protein and at peptide level. We actually need to use both systems because the enrichment have different performance, Zr is very suitable for monophosphorylated peptides, while TiO2 is working better for double and triple phosphorylation peptides.
    The best method is using a Ti4+ iMAC (immobilized metal affinity chromatography) that utilizes a chelator to immobilize a Titanium atom which has a very high affinity with the phos group that recognize either pSer/pThr/pTyr and an SH2 superbinder. The SH2-superbinder comes from the SH2 binding domain of the Tyrosine kinase, with a Kd value in the 0.1 - 10 μM range, so way lower that what we could achieve with a specific antibody. Combining those 2 methods we can map the whole phopshoproteome.
    The last method would be utilizing chemical derivatisation, so using phosphate to selectively perform reactions that either change the phos group through phosphoramidate chemistry or remove the phosphogroup via beta-elimination reactions. This way we can target analysis of subphosphoproteome for protein abundance, phosphorylation stoichiometry and relative changes in the phosphorylation state.

- Glycoproteome:
  - Glycosylation is another one of the most important PTM, it's predicted to affect 20% of all the proteins. Glycoproteins are normally found on the cell surface and they are correlated to recognition/adhesion. Glycosilation is not a single addition of a particular sugar but it appears like a number of modifications that comes with many different configurations. If the glycosilation pattern is not correct, this is a marker for disease or disease progression. These modifications produce different classes of protein glycosylation, the 2 most relevant are N-linked on the Asparagine and O-linked on Ser/Thr/Tyr or hydroxylysine. The glycomoyeties can be further modified in the ER or in Golgi, giving different glycoforms.
    We need to enrich to study the glycosylation pattern and study it in LC-MS/MS, but the glycoproteome is difficult to decode from the mass, so we need to split our sample treating half of it with deglycosylation enzymes. Quantification is either label free or label based.
    For enrichment, the older methods relies on the utilization of lectins to bind effectively the sugars, using either single AC or serial lectin chromatography. Problem is the affinity in the mM range, it's a rather weak interactions and also there are lots of competitors.
    Newer methods uses HILIC (Hydrophilic Interaction Liquid Chromatography) that works using a stationary phase interacting with the glycopeptides surrounded by a water layer. Using progressively more apolar eluants like acetonitrile, methanol and isopropanol we can separate the nonglycosylated peptydes from the glycosylated.
    Another method is relying on TiO_2 to enrich in sialylated glycopeptides or boronic acid, which is able to form a reversible covalent coupling with the glycol moieties. In this case we use a magnet to recover the ferromagnetic beads, coated with BoH_2, to recover what is interacting, eliminating the flow through, changing the pH to break the reversible covalent interaction, then the peptides are analyzed in LC MS/MS.
    Last method revolves around chemical derivatization so we can rely on chemistry to modify the sugars. An early enrichment strategy was Hydrazide capture technology, modifying the glycol moiety with mild oxidation to produce an aldehyde/cheto group, then adding proteins containing a PTM. Using a PNGase F enzyme we can cleave selectively the amine bond to the asparagine chain, releasing only peptides that are modified on N (asparagine). Another method is using an unnatural monosaccharide analogue like we do on SILAC that has a chemical reporter we can use to chemically tag with an enrichment probe.
    A possibility to study the functional aspect of the glycoprotein is using a ligand base receptor capture technology, using a TRICEPS probe, which crosslink to oxidized glycans on our target protein, has a ligand-attachment site and a biotin end for AC.

- How many proteins are there in serum, order of magnitude of dynamic range and why 6-mer ligand is the best for CPLL:
  - Proteins in serum are a huge number, around 7000 mg/100mL of plasma and the dynamic concentration is in the range of 10^8. Plasma proteins like albumin are very stable and always over-represented, then we have Ig and long distance receptor ligands. The lowest abundant proteins, so the one we're really interested in studying, are tissue leakage proteins, products that don't belong in the blood stream so they could be a marker of disease, aberrant secretions for example coming from a tumor, and foreign proteins coming from viruses or bacteria.
    So Alb is in the range of 35-50mg/mL while something interesting like IL6 is present ranging from 0-5pg/mL. To study the proteome then it's needed to decrease the dynamic range of concentration by either removing the HAPs (highly abundant proteins) or equalizing the representation of the proteins.
    Early techniques are based on solvent precipitation or immuno-depletion, like using consumer ready affinity column like the Seppro IgY14 columns, that is pre-packed to remove the 14 most HAPs. If we pass the eluate in another column we can also remove the MAPs (medium abundant proteins) to get only the low abundant one.
    Like always, the problem with immuno affinity is the high amount of cross-reactivity.
    So the best approach is to use something like a Protein equalizer technology or CPLL (combinatorial peptide ligand library). It compromises of a diverse library of peptide ligands coupled with spherical beads. The peptide ligand libraries are made of different mixtures of amino acids with different lengths, from 4 to 6 AA, in a concentration of 50pmol/mL. This amount to millions of copies of a single ligand for each bead and considering we use all of the 20 amino acids if we go for 6 reaction cycles we will have 64 million different ligands.
    Performance study going from a 0-mer (so just the support beads) to the 6-mer have estabilished which is the best approach. We can clearly see that the number of peptides captured lessen as we increase the peptide bait length, with the highest increase going from 0-mer to 1-mer. At 6-mer length we reach a sort of plateau so there's no point in using a longer peptide as bait. Interestingly enough, smaller bait is able to capture large size protein (>40 kDa) while 6-mer capture smaller size (10-50 kDa).

- Bioinformatic:
  - Usual proteomic workflow gives us a list of masses, which we need to compare to a in-silico process data. The in-silico proteins are “digested” and the masses analyzed, so we can compare the masses and identify our proteins.
    To perform this search we initially used a software called Mascot, it enabled us to discover proteins trough mass fingerprint analysis, calculating the probability of association between the experimental data and the theoretical mass data, scoring the association. Mascott used a probability method to identify a protein in a database and calculate that the probability of association between the experimental and the theoretical mass data is a random event. So it calculates a p-value and a score for the protein identification, which is the sum of the score of all peptides composing the protein.
    So for this reason, in all proteomics analysis it’s key to have a database, a repository of proteins, otherwise we cannot draw any conclusion. ProteomeXchange started as a repo and it became a consortium that encourages optimal data dissemination.
    Today UniProt is one of the biggest protein knowledge databank and it’s also curated, it gives information about the identity of the proteins, the organism, the protein coding gene, if there’s an annotation score, if there’s info about the function of the protein, about PTM, the literature, domain, structural features. So a huge list.
    But if we want to study the function we need to move from the structural information to something that can attribute and explain the function of the protein, so we need Gene Ontology. It’s a major bioinformatic initiative to unify the representation of gene and gene product attributes across all species, to maintain a vocabulary, annotate genes and gene products and to enable functional interpretation of experimental data. It’s a representation of the body of knowledge within a given domain. Ontologies consists of a set of classes with relations that operate between them, divided in 3 domains, so cellular component, molecular function and biological process.

- Protein networks:
  - One of the goal of functional investigation and of system biology is to explain the relationships between structure, function and regulation of molecular networks by combining experimental and theoretical approaches, with the final goal of mapping all the molecular interactions happening inside the cell.
    So graph theory enable us to create a visual way to annotate mapping relationships using node and edges, to draw interactions and understand PPi. Points and nodes represents our proteins/genes while lines and edges represent the various interactions between them.
    We can go further by adding direction and weights to our lines, attributing a function.
    Shortcomings of this approach are false positive and negatives, meaning that we could overestimate or underestimate the number of PPi, and we also don’t have any spatial and temporal informations.
    We gather the data coming from AP-QMS and we use bioinformatics to enrich and clean all the data coming from experiments, using known interactors coming from the literature or using statistical calculations to identify the interactors that could be obtained by chance. Then we use the high confidence interaction data to create a network topology by using a number of PPi databases to create an interaction network. We can also enrich our network by using databases with relevant clinical data to map the disease related proteins.
    At this point usually the data was clustered, grouping protein interactors into functional modules based on user-defined criteria, from the looser interactors to the stronger interactors. Besides that, it’s very helpful to use a functional enrichment analysis to screen our results against GO data, to place our network in the right biological process and to see which protein interactors are most likely to be biologically significant, to establish which have to be further investigated. Then we have a multiple sources to enrich my initial network, to add all the information presents in the literature, using web-inference, literature meta-analysis and from genomics/transcriptomics/proteomics data. We can use physical interactions databases or functional interactions databases, related to the pathways, like Reactome, which is one of the most cited by the literature. I have multiple ways to place my protein in a greater landscape.
    At last another platform we can use is String 9.1, is a database of known and predicted protein interactions, that include direct and indirect associations, derived from genomic context, high through-output experiments, conserved regions or previous knowledge. String enable us to enrich our network drawing directly on it a lot of informations.

- Degradome, what is and how to study it:
  - Degradomics is the identification, study and quantification of the proteases, their inhibitors, the availability, the specificity of the cleavage sites and their substrates.
    Degradation is always occurring so we're interested in the relationship between a physiological baseline and the perturbation caused by external factors or a pathology.
    Protein degradation at cellular level is a complex mechanism and if there's a mismatch in the degradation pathway it can cause disease like tumors, Alzheimer's and so on. We want to know which proteases are available and present in the cell at a certain moment, how much there are, and at the same time we're interested in the specificity of the cleavage sites, to understand how protease works. We want to have information about which substrate and which proteins are cleaved via identification and quantification and what are the interactors of this process. This way we can get a global picture of the protein network of the degradation, which can lead to the highlighting or the discovery of potential biomarkers.
    We use a 3 step approach, so discovery degradomics, verification of target proteomics then the validation with classical immunoassays.
    The idea is to have a proteome and a control proteome with a disease sample, something like a dysregulation of proteolytic enzymes or chronic inflammation that is disturbing the equilibrium between protein synthesis and protein degradation.
    Cleavage by a protease of interest generates fragments that are either neo-N-terminal or neo-C-terminal and to study those we can't use a typical proteomics workflow, since the terminal peptides will be lost in the sea of digested peptides.
    We have then 3 main procedures to enrich in protein terminal peptides prior to MS analysis, TAILS and COFRADIC study the N-termi while we can use C-TAILS as the only method to enrich for neo-C-termini peptides.
    COFRADIC methods uses a negative selection, meaning we modify all the mature and neo-N-termini peptides and fractionate them with RP-HPLC, containing both N-termini and internal peptides. A second modification and fractionation step let us discard the internal peptides while eluting only the N-termini peptides.
    TAILS is a versatile and simpler negative selection method that utilizes a ICAT or iTRAQ label. After denaturation, reduction and alkylation primary amines are labeled, we then digest with trypsin and using a high molecular polymer specific for the newly created free amino groups at the N-terminus we can remove them. We can then proceed with the LC-MS/MS analysis.
    Last method is called C-TAILS and it's the only method that can enrich for both N-terminal and C-terminal peptides, while simultaneously introducing a modification to unambiguously distinct the C-terminal. N and C termini are first labeled on the alfa and epsilon amine, then the carboxyl groups are protected with ethanolamine. After digestion we protect the newly created N-termini and we remove the internal N-termini using a polymer. The polymer with covalently bound internal and N-terminal peptides is removed by filtration, enriching only in the C-terminal peptides for LC-MS/MS analysis.