Bioanalytical Proteomics and Interactomics

Saul Pierotti

December 8, 2019

Introduction

- Most proteomic studies are based on mass spectrometry
- The program will be finished in December, in January we will only do exercises
- The examination is composed in a poster presentation and an oral part
  - A poster is a self-sufficient work
  - Usually it is prepared at the beginning of a scientist's career
  - We should base our poster on selected papers
  - The poster is not mandatory, but if we do it we will only be asked 1 question in the exam
- Proteomics is one of the main omics, but also interactomics and secretomics are becaming really interesting
- The data produced by omics are on the order of zetta and yottabytes
- Personalized medicine is important, and now we have the tools to implement it
- Proteomics can be structural, functional and differential (expression)
- Three main approaches: top-down, bottom-up and shotgun
- Many funding sources now require researchers to share raw data
- In silico-cloning: we need to know at least one software
- Other useful softwares are Pymol, Graphpad Prism, ImageJ
  - Graphpad Prism is a spreadsheet used to elaborate data and do statistics
- Proteomics is the large scale study of proteins, the proteome is the total amount of proteins coded in the genome
- Proteomics involves a spatial and temporal dimension, while the genome is static
  - It is really noisy (!)
  - Proteins are really diverse, and their processing is complex
  - They are present in a really broad concentration range
  - There is no PCR for proteins (!)
- Unnatural amino acids are really interesting, and there are many patens about them
  - They are used frequently in cosmetics (!)
- Proteases are everywhere (!)
- Aromatic amino acid tend to be in β conformations
- MALEK tend to be in α helices
- 1/3 of drugs target GPCR receptors
- Membrane proteins are hard to study
- An antibody is around 150 kDa, it is a big protein
- Is of huge clinical interest to find fluorescent proteins that emit in the red spectrum
  - Most tissues absorb really well green light, and red is the least absorbed part of the spectrum
- Mutagenesis can be performed rationally or randomly
  - Random mutagenesis is made with kits that make an imprecise PCR
  - We can substitute any possible aminoacid in a specific postion that we know to be important
- The top-down aprroach starts from the protein
- The bottom-up approach uses peptides deriving form digestion of the protein
- The shotgun approach uses different peptides deriving from many proteins

- The less complex the protein mixture, the better (!)
- Informations on chemical compounds can be retrieved in PubChem
- PubMed does not indicize articles outside of biology and medicine (!)
  - Better to use Scopus or Web of Science for chemistry and other fileds
  - For patens, Espacenet is the main database
- A volcano plot has the fold change in expression in the x axis and the significance in the y axis
  - It is used to rapidly identify differentially expressed proteins (it is also applicable to other fields)

Random notes

- Bioluminescence seems that appeared as a way to get rid of oxygen, when it firs appeared in geological times

Protein separation

- Proteins can be separed with different chromatographies or electrophoresis, microdialysis
- We can use depletion techniques to remove typical high-abbundance proteins
  - There are kits to remove albumin with spin columns
- Protein electrophoresis can be done in native conformation or by SDS-PAGE
- Immunological techniques are Western blot, ELISA, immunoprecipitation
  - Antibodies have affinity constants ranging from 10^9 to 10^12 M
- The upper limit for the number of resolvable proteins in a 2D-PAGE is 10000
- Spot identification in 2D-PAGE can be done by MS or by image analysis
  - Image analysis can be done by comparing our image with gel images of identified proteins present in databases
  - I usually work on master images created by combining 5-10 images of gels run in the same condition, to reduce noise
- Databases for 2D-PAGE, like ExPasy proteomic server, allow to search for images of specific tissues or cells, or conditions
  - I can then compare my spots with the ones already identified
- Differential PAGE (DIGE) runs 2 samples on the same gel, with different labels, so to be able to compare expression differences between them
  - It is a good practice to repeat the experiment by switching labels
  - Labels called cyainine derivatives (CyDies) are available with many different emission spectra
    - Labelling is done prior to IEF
    - We can use many samples at the same time (multiplexing)
    - They are really small and bind to the ϵ amino group of Lys in proteins
- The staining method for 2D-PAGE should be chosen considering the sensitivity required
- 2D-PAGE is generally speaking non-quantitative
- We can perform WB also on 2D-PAGE (!)
- HILC (Hydrophilic interaction liquid chromatography) uses a polar stationary phase and a polar mobile phase
  - It is really useful for glycosilated proteins
  - Elution is done in water gradient
- N-terminal microsequencing is based on Edman degradation and can be used for spot indentification, but it is obsolete
  - It has been replaced by MS
  - It uses phenyl-isothiocyanate that reacts with the N-termini forming thiazolinone derivatives
  - The derivative is then released and identified by cromatography or electrophoreys, before repeating the cycle
  - Amino acids are red at 269 nm

Mass spectrometry - basis is not for exam

- A mass spectrometer contains an ion source, a mass filter and a detector
- Some ion sources can be interfaced with a separation technique, others must be operated manually
- Soft ionization method do not break chemical bonds, while hard methods do that
  - In proteomics we use soft ionization: MALDI or ESI
  - There is also a semi-soft technique: fast atom bombardment (FAB)
    - It is used for small proteins and peptides
    - It uses typically a glycerol matrix
  - In ionization techniques, the matrix is used in order to prevent sample fragmentation
- In ESI the sample is in solution and it is sprayed by a small tube in a really strong electric field in presence of a hot nitrogen flow
  - The ionization is performed at atmosferic pressure
  - The sample forms droplets that evaporate concentrating charge, until they undergo columbic explosion
  - The produce ions are multiple-charged, and this is useful for detecting big molecules
  - We can operate both in positive or negative ionization
    - In positive mode I need to add formic acid to the solvent, in negative mode ammonia
    - If I have acidic groups in my analite I want to operate in negative modality, while with basic groups I want to operate in positive modality
  - It is possible to form protonated ions, deprotonated ions or cation adducts
    - It is very sensitive to salts and detergents
  - Since I have multiple-charged ions, I have many peaks for each analite
    - With molecules under 1200 Da I tend to have single ions
    - The multiple peaks are normally distributed and can be deconvoluted via software to derive the MW of the ion
  - Native and complexed proteins tend to have non-normal distribution of ions
  - It is very sensitive, but I need really pure samples
- In MALDI the solid sample is heated by a laser while embedded in the matrix
  - The matrix becomes ionized and transfer charge to the analite, while exploding at supersonic speed
  - We can operate in negative or positive modality
  - It generates single-charged ions M+H and M-H
  - It is more tolerant to contaminats
  - It can directly analize cells (!)
  - MALDI resolution can be influenced by the differential time of ionization of analite molecules, and differential initial velocity of the ions before the TOF acceleration
- A mass spectrum is a plot with m/z on the x axis and relative abundance in the y axis
  - Peaks are typically really sharp
  - From each ion I get a carachteristic pattern of peaks, due to spontaneous fragmentation
  - I have spontaneous fragmentation only if my ionizing source is strongh enough
- The accuracy of the mass spectrometer is measured in ppm
- The resolution of a mass specrtrometer is obtained by dividing height of the peak by its width at half height
- A mass filter is a device that separates analite ions based on their m/z
- A magnetic sensor mass analyzer (MSA) deflect the ions by differen radii depending on their m/z
- The quadrupole mass filter allows a stable path only for a selected m/z, and we can get a spectrum by scanning all the target m/z values
  - It is used also in tandem MS
- Orbitrap (Makarov, 2000) is the newest and most sensitive analyzer
  - There is a spindle-like electrode around which the analite ions spin and a barrel-like electrode
  - The ions oscillate with a frequency related to their m/z, giving a signal
  - The signal is deconvoluted with the Fourier transform to yeld the fundamental armonics, which are related to m/z
- The time of flight analyzer (TOF) accelarates ions with an electric field in a way that is proportional to

m/z
  - The time that is required for the ions to reach the detector is proportional to m/z
- The reflectron is a variant of TOF that improve resolution by compensating for different initial velocities
- MS detectors are diods that generate secondary electrons when struck by an ion
  - Electron multipliers produce many secondary electron for each ion. and therefore are more sensitive, but require extensive maintenance
- Tandem MS combines 2 MS in the same instrument
  - I avoid fragmentation in the first step using a soft ionization of the molecular ion
  - A collision cell between the 2 MS filled with inert gas allows to fragment the analite

Peptide mass fingerprinting

- It is the main bottom-up approach
- The standard workflow starts with 2D-PAGE that allows to recover unique spots
- It is important to chose a staining method that is compatible with MS
- We can also use multidimensional HPLC as an alternative to 2D-PAGE
- Spots are then cut and destained
- Spot picking can be done manually or with a robot
  - Manual picking is susceptible to keratin contamination
  - There is a risk for gel deformation
- The protein is then digested by trypsin to yeld peptides
  - Trypsin cuts after K or R, but only if not followed by P
  - It is important to have a complete digestion to avoid missing cleavage sites
  - Better to use volatile buffers to eliminate them easily afterwards
  - Trypsin can also self-digest (!) and the deriving peptides are really useful as a standard for MS
    * It is an internal calibrator
- Peptides are then purified by reverse chromatography or Zip tips
  - RC uses apolar stationary phase on polar mobile phase
  - Zip tips are a miniature RC column (!)
- We perform MS on the peptides, getting a fingerprint of the protein
  - In MS peptides must be ionized in a gas phase
  - MS measures the m/z ratio of the peptide ions
- I can then identify the protein by searching for my fingerprint in databases
- Masses can be reported as monoisotopic or average
  - Entry level instrument cannot differentiate isotopes, therefore report only average masses
- To match a fingerprint with a database, I check how many hits on the same protein I have with my masses
  - I choose the protein with the maximum number of hits
- Main PMF databases are Mascot and ProFound

Peptide de novo sequencing

- Peptide de novo sequencing uses MS/MS spectra to determine the sequence of a protein without using any previous knowledge
- This is in contrast with database search, that identifies the peptide using databases
- The fragmentation process produces different kinds of ions
  - Give the low collision energy employed, most fragmentations involve peptide bonds
  - If the charge is retained in the N-terminal fragment, the ion is termed a, b or c
  - If the charge is retained in the C-terminal fragment, the ion is termed x, y or z
  - Fragmentation of the peptide bond produces y or b fragments
- I can recover the mass of the residues by analyzing the mass difference between ions
- Given a MS/MS spectrum, the sofware Peaks can give the protein sequence, with a confidence for each residue

Synthetic biology - Not for exam

- In order to provide proprieties that are not available in nature we can use non-natural amino acids
- We can also modify polymerases in order to use non-natural nucleotides that still pair among themselves in a specific way

Molecular cloning - Not for exam

- Molecular cloning is essential because it is our only way to obtain high amounts of proteins
- The general workflow is to isolate the cDNA of the protein, insert it in a plasmid and transfect bacteria with it
  - The cDNA is amplified with primers containing a 3' overhang that allows to introduce the appropriate restriction sites
- When I cut my plasmid, I want to use single cutters so to avoid that the plasmid could close on itself
  - Usually I can cut exactly where I want because plasmids are engineered to have many restriction sites
- Selection of recombinants is done with antibioitics for bacteria and mammalian cells, and with auxotrofs for yeast
- Cells are made competent with a $CaCl_2$ solution
- Commercially available competent cells are usually much more efficient, even though a little bit more expensive
- Transfection can be done by heat-shock or electroporation
- Plasmids tend to be toxic for bacteria, because of their metabolic burden
- DNA absorbs at 260 nm, proteins at 280 nm
- Protein biotinylation can be performed in vivo or chemically
  - Chemically, I usually bind biotin to Lys residues
  - In vivo I can fuse my protein with a biotin binding domain and clone it toghether with BirA, a biotinylating enzyme
- For in silico cloning, we can use Vector NTI Advance or the online platform Benchling
- When I create recombinant proteins, it is smart to add a linker
  - Usually Gly is used
- When I clone a protein into an expression vector, I need to optimize codon usage for that organism
  - An old approach was to use the most used codon
  - Now it is common to respect the codon frequencies of its proteins
  - When I do codon optimization, I can play a little bit so to avoid or introduce cut sites where needed
- Once I have designed my sequence, I can order it from companies
- Addgene is the most used plasmid repository
- You can also order a plasmid from your sequence
- There are also spin-off that do all the work for you and give you the plasmid to be transfected
  - It is expensive (some thousand €)
- Bacterial cultures and eukariotic cell lines can be purchased from ATCC (mostly for cell lines) or from DSMZ (Liebniz, for bacteria)
  - Cell lines from ATCC are expensive
  - Bacterial strains from DSMZ are cheap (20€)
- Site directed mutagenesis can be used for changing the proprieties of proteins
  - We can change emission wavelenght, increase stability

Quantitative proteomics

- Quantitative proteomics gives us an understanding of the state of the cell
- In MS, I can do quantitative analyses with labeled or label-free approaches
  - Labeled approaches are SILAC, ICAT, iTRAQ, O18/O16 enxymatic labeling

  - In labeled approaches I can run all my samples toghether, while in label-free approaches I can only run one sample at a time

SILAC

- The SILAC approach uses a labeled aminoacid in a culture medium deprived of that aminoacid
- The aminoacid has to be essential (!)
- It is frequent to use deuterated Leu (Leud3)
- It can be used only for cell cultures and not for samples like urine, blood, ecc
- The 2 samples are mixed early in the sample preparation process and analysed together by LC-MS
- The LC treats in the same way the samples
- In MS, I can easily distinguish the peaks from the 2 samples, and compare protein expression levels
- I can also perform SILAC in whole organisms by using appropriate food (!)
  - This has to be done for more than 1 generation (!)

ICAT

- The ICAT approach uses a reactive group that can label an aminoacid side chain
- I can use biotin linked with a thiol-specific reactive group (iodoacetamide) through an etilene glycole polimer linker
- The linker can be deuterated/normal
- Biotin is used for purifing tagged peptides in both normal and deuterated samples, so to minimize error
- It can be used only for proteins with accessible Cys, which is much rarer than Leu (SILAC) (!)

iTRAQ

- In the iTRAQ appraoch links an isobaric tag to amine groups
  - The tag is composed of a balance group and a reporter group
  - There are 4 different couples of reporter-balance groups, but the couples have all the same mass
  - The tag is made so to fragment by CID in MS/MS analises between peptide and balance group, and between balance and reporter group
  - In the first MS the different tags behave in an identical way
  - After fragmentation, the tags are quantified to yeld the relative and absolute abundances of proteins

Label-free approaches

- In label-free approaches I usually compare LC-MS or LC peaks, and then identify the protein by MS/MS (LC-MS/MS or LC/LC-MS/MS)
  - The main problem of this approach is the experimental error between different runs
  - The large volume of data requires automation
  - Peak intensity is not-quantitative, but in a specific setting with ESI it was found to be roughly
- Spectral count is another label-free approach where I count the number of peptides identified from my sample
  - The most represented the protein, the more peptides will be identified
  - It is not so precise (!)

MALDI imaging

- In 2D MALDI tissue is sliced with a cryostat and a MALDI matrix is applied
- We can do a comparative study for different samples (normal vs patological) or study a single sample in order to reconstruct the protein distribution in the tissue
- In 3D MALDI I use consecutive sections of tissue slice to get a 3D model of protein distribution in a tissue
- It is similar to immunohistochemistry, but without the need for ABs (!)

- It can be used not only for proteins, but also other molecules
- I can study many different molecules at the same time
- The spatial resolution is not as good as with IHC

SELDI

- Surface enhanced LDI (SELDI) is a variation of MALDI that instead of mixing the sample with the matrix, uses a surface that has biochemical affinity to the target molecule
- The surface is functionalised so to be able to bind the target
  - Common functionalizations are CM10 (weakly positive), C6-C12 (hydrophobic), IMAC30 (metal binding), Q10 (anion exchanger), antibodies, DNA, proteins
- After sample application, we can remove contaminant through washing steps and then apply a matrix on it
- Some surfaces select for a range of protein with certain proprieties, others for a specific target

Many different -omics

- Cellomics is the monitoring of the whole content of single cells
  - Micro-arrayed donut shaped chambers (DSCs) of different volumes can be used for studying single molecules, cells or multicellular arrangements
  - Each chamber acts as a separate reaction vessel
  - Single cells can be placed in pL and nL chambers, while fL chambers are for molecular studies
  - In this way single cells can be studied in a non invasive and time resolved manner using a microscope
  - Raman spectroscopy can be used for doing label-free cellomics studies
- The secretome is the ensemble of molecules secreted by a cell type
  - The secretome of mesenchimal stem cells (MSCs) can be useful for cartilage regeneration
  - This terapeutic effect can be achieved by transplanting MSCs, extracellular matrix or secreted molecules
- The regulome is the set of regulatory components in a cell type
  - SCRAT is a software that can analyze single cell regulome data
  - It is useful for discriminating cell populations and subpopulations

Molecular interactions

- Protein protein interactions (and interaction of proteins with other molecules) are at the heart of any biological process
- We can charachterize an interaction by studying who the interactors are, and which residues partecipate in the interaction
- Interactions in a cell type for the interactome
- The interactome can be studied with computational or experimental tools
  - Computational tools can be based on structural studies, co-mutation studies
  - Experimental approaches involve X-ray and NMR spectroscopy, yeast two hybrid, co-immunoprecipitation and other methods based on affinity
- String is a database of experimental and predicted interactions
- Some experimental method can be used in vivo, others in vitro
- Physical interactions are experimentally determined with a bait-prey approach
- Frequently we want to identify the partner of a specific protein
  - In this case we have 1 bait and a library of preys

Yeast two-hybrid

- The yeast two-hybrid system uses a trascription factor as a reporter
- The TF is composed of a DNA binding domain and an activating domain

- The 2 domains are separately linked to bait and prey
  - The DB domain is fused to the bait
  - The AD is fused to the prey
- A reporter gene is expressed if the 2 proteins interact
- We build bait and prey plasmids with the recombinant proteins, which are then transfected in yeast
- The DB domains used are those of Gal4 or LexA
- The AD domains used are those of Gal4, VP16 or B42
- The reporter gene can be a survival gene, or beta galactosidase (blue colonies in X-gal)
- This system is useful for a firts screening of large cDNA libraries

Mammalian two-hybrid

- The mammalian two-hybrid uses 3 plasmids
- A plasmid harbours the BD, one the AD, and a third comprises a minimal TATA box and the reporter gene
- It is used for bait and prey proteins that are not appropiately modified post-translationally by yeast
- It is useful also for proteins that reside in a specific organelle
- Another advantage is that it is faster: the reporter gene typically used give a visible effect in 48h, while with yeast we have to wait for the colonies to grow 3-4 days
- The most used reporter genes are SEAP, luciferase, and β-galactosidase
- SEAP (secreted alkaline phosphatese) is a secreted protein that can withstand 65°C
  - We can degrade the endogenous alkalin phosphatases of the cells by heating them to 65°C so to remove background noise
  - We can detect alkaline phosphatase activity with fluorescence or chemioluminescence methods using the culture medium
  - It is advantageous because it is secreted, so we do not need to lyse our cells (!)
- Luciferase from Photinus pyralis (north american firefly) is an enzyme that emits in the yellow-green spectrum (560 nm)
  - It converts the carboxylic acid luciferin into luciferin-adenylate
  - The adenylate is reactive and reacts in the luciferase active site with molecular oxygen producing an excited product
  - This product rapidly returns to the ground state emitting a photon
- Bacteria galactosidase is a really stable enzyme
  - Its activity can be tested with ONPG, which is hydrolized to onityrophenol
  - The product can be detected with fluorescent, chemioluminescent and colorimetric methods
- This system is usually used for a refinment of the first screening done with the yeast two-hybrid
- It is not suited for the screening of large libraries

Pull-down assays

- The Pull-down assay uses a recombinant bait-tag complex expressed in E. coli
- We can immobilize the bait using antibodies anti-tag, or molecules that bind the tag
- After incubation of the immobilised bait with the prey, I perform washing steps
  - The incubation can be done with a cell protein extract
- The bait-prey complexes then can be eluted using SDS loading buffer, or a competitive ligand
- The resulting eluate is analysed by SDS-PAGE
- Interaction can be detected in the gel by doing a WB with Ab against both bait and prey
- Another detection approach is to use a radioactive bait with S^{35}
- Glutathion-S-transferase is a commonly used tag, pulled down with reduced glutathion
  - The formation of a fusion bait with GST has also the advantage of making the chimera more soluble
- Once the complexes have been detected, the interactors can be identified by MS

Co-immunoprecipitation

- The co-immunoprecipitation assay exploits the specificity of an Ab for the bait
- The first step is to obtain a cell extract in non-denaturing conditions
- The prey bound to the bait co-precipitates
- The Ab-bait complexes are immobilised on a sepharose gel
- After a washing step, we can elute and then identify our interactors
- The main advantage is that I operate with a cell extract, so I do not have alterations due to expression systems and the environment is as similar as possible to the in vivo situation
- The identification is similar to the other pull-down assays: I detach the complexes with SDS sample buffer and then identify them with WB
- There is the underlying assumption that proteins that co-precipitate with the prey interact with it in a biologically significant way: this has to be proven

Protein arrays and chips

- Arrays are differentiated from chips because in arrays we have molecules immobilized in high density, while chips exploit microfluidic technologies
- Protein arrays can be used to determine interacting partners
- An array can be configured in 3 main ways
  - With an immobilised capturing agent and the target in solution
  - With an immobilised target and a series of putative interactors in solution
  - With an immobilised complex sample (reverse phase)
- Protein microarrays allow an high throughput analysis of proteins immobilised on a surface
  - We can use protein, nucleic acids, lipids, small molecules as probes
- The protein microarray substrate can be glass, polyacrilamide or agarose gels, nanowells
- The immobilisation of proteins can be achieved in different ways
  - By diffusion into the surface
  - By adsorption
  - Covalent attachment thorugh specific side chains
  - By affinity ligands
- Immobilization of nucleic acids on a chip is easy, but with proteins it is more tricky because you can loose functionality
  - The orientation with which the protein binds the surface is the main problem
  - This can be overcome by using affinity binding, that forces the protein to orient in a specific and modulable way
  - Covalent binding is usually achieved with Lys residues
- Protein array printing is done by robots, with contanct or non contact methods
  - Actually not only proteins, but any biomolecule can be deposited in this way onto an array
  - Non contact methods are similar to a traditional inkjet printer
    - Small droplets of probe solution are expelled and dry onto the glass surface of the array
    - It allows to achieve a slightly higher spot density
    - It avoids to alter the surface by touching it
  - In contact printing, each print pin applies the probe solution onto the surface
    - The printing pin is a needle with a capillary channel inside it
    - After touching the surface, a small aliquote of solution adheres to the surface
  - The main danger of these methods is cross-contamination between the spots
  - A single spot typically consists of a few nL and covers 100-150 μm
- Expression arrays are used to identify expressed proteins
- Interaction arrays are used for studying the interaction parteners of a target
- Differential profiling is done by mixing a disease and an healthy sample marked with different Cy dyes
  - I can incubate an Ab array with the mixture, and get information on the different proteins expressed in the different states
  - Labelling complex samples with Cy dyes can be difficult

Aptamers

- Aptamers are a syntetic peptides or oligonucleotides that can bind a target in a really specific way because of their 3D structure
- They can be used as drugs, delivery mechanisms, in imaging, analytical applications
- They are engineered with the SELEX strategy
  - I choose an initial library of nucleic acids/peptides
  - I incubate the library with the desired target
  - I perform washing steps so that only aptamers that bind the target are retained
  - I charachterise the bound aptamers and I amplify them
    - This can be done by PCR for DNA
    - For proteins I sequence them and then prepare oligonucleotides with the right coding sequence
  - In this way I get a new library, and I can repeat the process again
  - Typically 10-15 rounds are performed
- This is a targeted evolution process
- I am selecting for aptamers that interact with the tag in the experimental conditions, but I do not know if they will bind to it also in vivo (!)
- Peptide nucleic acids (PNAs) are molecules with an aminoacidic backbone linked by peptide bonds, but have the traditional 4 nucleotides as side chains
  - THe nucleotides are linked not to the Cα
  - They bind DNA and RNA in a really specific and strong manner, because they lack charge at the level of the backbone and because of their flexibility
- I have to test also for selectivity
- It is a really expensive system, and to date only the aptamer for thrombin is really effective

Phage display

- Phage display is an in vitro selection technique that uses a target protein fused with the coat protein of a phage
- I can easily build a library of putative interactors to my protein of interest
- The selection is done with a standard affinity techniques using the protein/molecule of interest
- High affinity phages bind the support, and then can be eluted with free bait
- The phages can be amplified by infecting E. coli, and then selected again
- Through repetitive selection and infection cycles, we can select for strong binders to the protein of interest
- I can use error-prone PCR to mutate my protein and select advantageous mutation for binding to the substrate
- Once I am satisfied with the result, I can just sequence the phage and get the corresponding protein sequence
- It is used in drug discovery for selecting targets and for selecting high specificity antibodies
- It is well-suited for screening large libraries
- As always, are we sure that our findings will translate in vivo?
- If the first screening is washed too arshly, I can loose some variability
- I can only use small peptides, big proteins disrupt the fold of the coat protein

FRET

- It uses 2 fluorescent molecules, one emitting on the absorption wavelenght of the other
  - There has to be spectral overlap between emission of the donor and absorption of the acceptor
- The energy transfer is non-radiative: the relative orientation of the dipole moment of the molecules must be approximately parallel
- The resonace efficiency depends on the inverse 6th power of the distance (generally 1-10 nm is the range)
  - It is really strongly dependent on the distance between the 2 molecules, so it can detect interactions
- RET signal is rationmetric: it does not depend on the number of cells, volume and other variables

- The signal is calculated as a ratio of the emissions of acceptor and donor, minus a correction factor that accounts for the spectral overlapping of donor and acceptor
  - A signal of 1 means that the emission of the acceptor is equal to that of the donor
- FRET can be used to show in vivo interactions
- BRET is like FRET but uses luciferase as donor molecule
- In BRET luminescence is a consequence of a chemical reaction: we do not need to excite
  - There can not be autofluorescent background, and no photobleaching
- BRET is a naturally occurring phenomenon in Acquorea victoria, where GFP is the acceptor
- Molecules used in FRET are Cy dyes, fluorescent proteins, FITC, AlexaFluor, Texas Red, Europium compounds
- For fluorescent proteins I can just clone a recombinant protein (in vivo!), while for organic molecules I need click chemistry (in vitro!)
- FRET can be used for detecting not only interaction, but also conformational changes (!)
  - This is useful for detecting enzimatic activity, intramolecular interactions, ligand binding
  - I can engeneer a sensor that emits FRET when there is kinase or protease activity in the medium
- The SRET (sequential BRET-FRET) uses a 3-step energy trasfer
  - Its efficiency is quite low (!)
  - The first step is the transfer of energy from a luciferase to an acceptor protein (GFP), which act as a donor for a final acceptor (YFP)
  - It has been used for studying GPCRs
- Since we use fusion proteins, we can alterate the structure of our protein (!)
- When we use BRET for monitoring homodimerization, I have to consider that I will have many dimers with the same tag (so no fluorescence (!))
- we can also use quantum dots as acceptors
  - A quantum dot is a semiconductor nanocrystal of 2-8 nm
  - Their abosrption spectrum is really wide, it can be excited by almost any fluorescent biomolecule
  - Its emission can be finely tuned by particle size, and its emission spectrum is really narrow
  - They have a very large Stokes shift (absorption/emission spectral separation)
- QD-BRET uses a quantum dot bioconjugate with Renilla Luciferase that is able of self-excitation
  - It is advantageous because we can make many differemt QD with really specific spectra, thus we can do multiplexing
- Inteins are peptides capable of excising themselves from proteins, in a process known as protein splicing
- QD-BRET was pioneerly used to detect metalloprotease activity in serum
  - The conjugation of QD to RLuc was mediated by an intein phused to RLuc, that catalysed the splicing of RLuc itslef with a QD containing a reactive NH group
  - Proteases act on the linker between QD and RLuc, impeding BRET to occur
- Modified Inteins can be used for tag-less protein purification
  - A mutated version of an intein is unable to splice itself out at slightly acidic pH
  - I can link insert the intein between the protein of interest and a tag
  - After purufucation, I can change the pH and release the native, tagless protein

Protein complementation assay

- Protein complementation assay (PCA) is based on the expression of a luciferase as separated domains bound to proteins that we think could interact
- If the proteins interact, the luciferase assembles and we can detect luminescence
- PCA can also be used to detect intramolecular interactions
- The major drawback is that the 2 halves of the luciferase have high affinity for each other, so they could interact even without interaction of the target proteins (!)
- The binding of the 2 protein fragments of Luc is really cooperative, therefore it goes from completely undetactable to maximum signal in a really narrow range
  - This is in contrast to FRET, which has a really low dinamyc range and requires extensive optimization
- NanoBiT is a next-generation PCA approach that was develloped by targeted evolution

  - The 2 halves have low affinity for each other, so they do not give an aspecific signal
  - They have low steric hindrance, so they do not disturb interactions among the proteins of interest
- We can combine PCA and BRET/FRET to assay non-binary interactions
  - The donor or acceptor can be made of 2 halves that are functional only if assembled

In vivo imaging

- We can use these tools also for in vivo imaging
- In vivo imaging is in complaiance with the 3R rule
  - Reduction of the number of animals required
  - Refinement of the experiment so to minimize harm
  - Replace, if possible, the use of animals
- There is no need to kill the animal, so we can get imaging at different timepoints, minimizing cost and ethical concerns(!)
- We can use these tools to monitor the occurrence of specific biological events in live animals
- The general method is to overlap a light image of the animal with the signal from the imaging method used
  - Detection is done in a dark box
  - With the overlapping, we have information on both signal intensity and location
- Bioluminescent imaging (BLI) methods are superior to fluorescent ones because they are not affected by background fluorescence
- Some applications of BLI
  - We can engineer pathogenic organisms to emit bioluminescence so to study their localization in vivo, or the activation of specific genes
  - They can also be used to monitor drug efficacy
  - We can use a similar strategy to engineer bioluminescent tumor cells
  - We can detect the survival and replication of stem cells grafts
- The nude mouse is a really useful model because it lacks the thymus and it is nude
  - It is easier to use for imaging, and it readly accepts xenografts
- Traditionally BLI methods employ Photinus pyralis (North american firefly) luciferase, which emits in the green spectrum
- Green light is absorbed strongly by mammalian tissues, especially by hemoglobin
- It is better to use red light emission spectra, because they penetrate more easily through tissues
- NIR-II imaging uses a wavelenght in the second NIR, that minimizes scattering and has an higher tissue penetration
  - NIR-II fluorophores are single walled carbon nanotubes (SWCN), QDs and polymethine dyes
- For QD and SWCN, we have no information about the clinical safety of these devices

Function prediction and databases

- GO slim is a simplified subset of GO terms used to facilitate browsing
- Annotation trasfer is mainly done by homology (sequence similarity)
  - If sequence similarity is higher than 80% we usually have the same function
  - Sometimes this criterion is misleading so tools tend to integrate other informations on the protein
  - Most of the newly discovered proteins do not have high similarity with any protein of known function
- Interpro provides a functional charachterization of a sequence
  - It is a meta-database that gets information by many other specialized databases
  - Some databases are strong in a specific case, others in other cases, and InterPro gets the best of all of them
- The SKL tag targets proteins to peroxisomes and it is recognised by the PTS1 receptor
- When chosing a tool, we should be careful to check if it is still active and maintained