Introduction to PPIs

Monday 3rd September 2012

EMBO Practical Course:

Computational Analysis of Protein-Protein Interactions
for Bench Biologists

September 2nd - 8th 2012

MDC Berlin

Aidan Budd
EMBL Heidelberg, Germany

How we Think About PPIs

We asked you to prepare answers to several questions before coming to the course

A. Imagine you have been asked to introduce a new student working in your lab to the concept of protein-protein interactions.

Write down some of the key ideas you would want them to understand before starting to work on their own, and associated advice

For example:

3D structures of PPIs are extremely helpful for building hypotheses about the biology and biochemistry of an interaction, e.g. which residues are likely to be most important for binding strength/affinity

Thus, when working in detail on an interaction, it's often good to check whether there is already a 3D structure for it in e.g. PDBe

How we Think About PPIs

We asked you to prepare answers to several questions before coming to the course

A. Imagine you have been asked to introduce a new student working in your lab to the concept of protein-protein interactions.

Write down some of the key ideas you would want them to understand before starting to work on their own, and associated advice

Take some minutes to explain your ideas to your neighbour

Trainers volunteer some answers to the question

Let us know if you came up with other ideas not mentioned by the trainers, or any questions about their answers

Introduction to Bioinformatics

Monday 3rd September 2012

EMBO Practical Course:

Computational Analysis of Protein-Protein Interactions for Bench Biologists

September 2nd - 8th 2012

MDC Berlin

Aidan Budd
EMBL Heidelberg, Germany

Niall Haslam
UCD Dublin
Conway Institute, Ireland

Thinking Like a Bioinformatician

You all want to "get better at bioinformatics"

That's part of why you're here

But... general goals like this can be a bit overwhelming, it can be hard to know where to begin...

A more specific way of describing this goal could be...
...you want to:

- be able to work and think more like "expert" bioinformaticians
- overcome bioinformatics "problems" more easily/more quickly

Thinking Like a Bioinformatician

To help with this, we will begin by:

- talking together about your current bioinformatics expertise
- presenting some of the ideas/knowledge "expert" bioinformaticians regularly use in their work

Your Bioinformatics Expertise

It can be hard to answer the question:

"What do I already know about bioinformatics"

(again, the question is too general)

So we try to help explore this with you by asking:

What have you learnt about bioinformatics that you wish you'd known at the start of your research career?

How have these ideas been useful for you?

For example...

Bioinformatics Expertise: Example

Knowledge/experience:

Most databases identify each record with a unique identifying "number" (the "Accession Number")

How does this help?

Knowing the accession number for my record of interest allows me to unambiguously identify this record within a data resource

Example - query SwissProt to find human Src record:

with accession number (P12931) - match 1 record

with "human src" - match 594 records

Bioinformatics Expertise: Example

Knowledge/experience:

Many data resources contain similar but different data i.e. have partial but not complete overlap of their content.

How does this help?

If I can't find the data I expect in one resource, I know I may find it by looking in a similar resource

Example: UCSC and Ensembl human genome browsers

I don't get so stressed/panic when I don't find what I want in a resource

Anything that keeps my stress levels down is great!

Your Bioinformatics Expertise

Talk with your neighbour to find some examples of your own

Think about bioinformatics "problems" you've solved, and how you solved them...

Your Bioinformatics Expertise

Knowledge/experience:

There are **lots** of different bioinformatics tools

How does this help?

When I have a new "question", encourages me to spend time searching for an existing tool that might help with it

Helps me feel more comfortable not knowing everything about how any given tool works

I'm less stressed by the thought "I don't understand how this works" as I'm used to it

Your Bioinformatics Expertise

Knowledge/experience:

Most bioinformatics tools address one or both of these questions:

- what experimental data has been reported concerning my entity (e.g. protein/PPI) of interest? [e.g. much of the data in UniProt]
- what predictions can I make about the structure/function of my entity (e.g. protein/PPI) of interest? [e.g. BLAST, IUPRED]

How does this help?

Makes it easier for me to understand the purpose/function of a "new" tool - important given how many tools there are!

Mediates my expectations of what I can "get" from an analysis, i.e. I frame my question in these terms

Your Bioinformatics Expertise

Knowledge/experience:

Comparing (and identifying identical) strings (i.e. lists of characters e.g. "asd815") is used to link/connect information about the "same"/similar entities in many bioinformatics tools

How does this help?

I know my analyses will have fewer errors when I choose and use:

- consistent naming schemes within (and even between) projects (so identical entities have the same "name")
- names/strings without "white space", using just A-Z, 0-9
- consistent ways of structuring my data files e.g. separating columns using the "tab" character

Your Bioinformatics Expertise

Knowledge/experience:

Bioinformatics tools and data resources may change with time, both their interface and the underlying data

How does this help?

Knowing this, where possible, I save/keep local versions of data obtained from elsewhere, to help understand/protect about future changes

e.g. working with a predicted protein sequence from a newly-sequence genome, I record both accession numbers and sequences in case it is missing from later releases

Knowing these changes are inevitable (and important) helps reduce my frustration when it happens!

Your Bioinformatics Expertise

Knowledge/experience:

Often it is more effective to use different tools, or the same tool in different ways, to address different questions

How does this help?

How should I use UniProt?

it depends on what you want to use it for

How should I change parameters to improve my BLAST search?

it depends on what you want to use it for

Which MSA tools should I use to align my sequences?

it depends on what you want to use it for

etc.

Thus, having a clear understanding of precisely which question we want to address helps us use tools more effectively

Example Bioinformatics Analysis

Aidan Budd, EMBL Heidelberg

Example Bioinformatics Analysis

Give here an example bioinformatics analysis that illustrates many of the points made in the previous slides

Exercises using tools need to be done in the context of a particular problem/ question - because, as already pointed out, the way to use a tool effectively depends always on the question it is being used to address.

Scenario:

A friend working in a zebrafish lab has done a forward genetic analysis, using a phenotype, and has identified the mutated gene

They want to try and understand how/why the gene contributes to the phenotype, in particular by identifying or predicting proteins that physically interact with the gene

Perhaps knocking these out/silencing them will have a similar pheontype?

They tell us it's called ENSDARG00000046048

Example Bioinformatics Analysis

Would you like to:

1. Try this yourselves with no more information/ideas from me?
2. Try this yourselves with some hints on resources you might like to try?
3. I demo it to you first, then you have a go yourself? - in which case you'll get a short written description of how I did it to try and follow yourselves

If you try first, then do it in pairs, keep going until you get stuck - then get help! - we'll try and notice when several people are stuck, and then we'll move on. Think about, in this case, what contributed to you getting stuck

Example Bioinformatics Analysis

Hints on how I would/did do it... ENSDARG00000046048

- Find protein sequence of the ENSEMBL record
- Get the UniProt record - two ways of doing it, database cross-linking and BLAST (note that I try first swissprot and it's not there, but it is in uniprot)
- Read about the gene
- Look for interaction partners described in the record - via STRING maybe, but not very strong evidence
- Look for related PDB structures in complex? Yes, we find one by BLAST at NCBI
- Get the structure from PDB and look at it in PyMOL
- Look for a protein related to the interacting protein in ZF
- Get this info PDBsum - I find it easier to get the info on which sequences are in there compared to PDB
- Read about the interaction - is there a model for describing the interaction modules in the two proteins? Yes, it's the FFAT motif described by the ELM resource
- Is the pattern conserved in the interacting protein? Yes.
- Other proteins possessing this module in ZF might also interact with the query protein

Another example

Scenario:

A friend is working on the parasite Giardia. They want to study the role of nucleoporin proteins in the biology of the parasite, expecting this might be important understanding gene regulation there etc.

They want to find the sequences of these proteins to help with their cloning etc. They ask you for help

Discuss with your neighbours some of the ways you could begin to try and find the sequences of these proteins

Another example

Scenario:

A friend is working on the parasite Giardia. They want to study the role of nucleoporin proteins in the biology of the parasite, expecting this might be important understanding gene regulation there etc.

They want to find the sequences of these proteins to help with their cloning etc. They ask you for help

- Search for a Giardia genome resource - try a text search for nucleoporins
- Check which ones this is matching using BLAST
- Google for "nucleoporins"
- Choose one of them
- Try a BLAST at the NCBI
- If it doesn't work, what modules do SMART/PFAM predict in the sequence
- Try using these tools to identify similar proteins in the organism

Another example

Scenario:

A friend is working on the parasite Giardia. They want to study the role of nucleoporin proteins in the biology of the parasite, expecting this might be important understanding gene regulation there etc.

They want to find the sequences of these proteins to help with their cloning etc. They ask you for help

Try, together with your neighbour, to find some other Giardia nucleoporin sequences

Describing PPIs

Describing PPIs

We asked you to prepare answers to several questions before coming to the course

B. Imagine that one (or more) PPI(s) you are interested in studying in the lab are described in a bioinformatics database.

Write down, for example for the interaction between human CDK4 and a D-type cyclin (as described here http://www.pnas.org/content/106/11/4166.full) or some other interaction you are interested in, in a "structured" form, the information you would want/hope to find about this interaction in the database. For example:

Interactor 1

- UniProt Accession Number: P24385
- Organism: Homo sapiens
- Publication: PubMed ID: 19237565

Interactor 2

- UniProt etc...

Describing PPIs

We asked you to prepare answers to several questions before coming to the course

B. Imagine that one (or more) PPI(s) you are interested in studying in the lab are described in a bioinformatics database.

Write down, for example for the interaction between human CDK4 and a D-type cyclin (as described here http://www.pnas.org/content/106/11/4166.full) or some other interaction you are interested in, in a "structured" form, the information you would want/hope to find about this interaction in the database. For example:

Compare your answers with your neighbours

Try and put together a consensus of the most important information

Niall will present his ideas for this