**Group project plan**

Software specification

Specifically, the web application should:

1. Provide the user with background information about all known human protein kinases:
   * their name and gene symbol,
   * Accession number
   * the families they belong to,
   * where in the cell they are typically found, and
   * the sites they are known to phosphorylate.
   * taxonomy(uniprot or <https://cansarblack.icr.ac.uk/protein-family/RAF%20subfamily>)
2. Allow the user to browse phosphosites in terms of:

* their genomic location,
* neighbouring sequence, etc.

3. Provide the user with background information about all known inhibitors of human protein kinases, e.g.

* their name
* chemical structure, and
* which kinases they are known to inhibit.

4. Allow the user to upload quantitative phosphoproteomics data of the format provided in example file az20.tsv (see QMPlus for details). The web application should then summarise this data 2 graphically and provide an estimate of the relative activity of every human kinase in the sample represented by the data.

How to begin the project

Successful completion of this project requires a combination of technical skill, good organisation, logical thinking, web-based research and possibly a visit to the library. Your first task is to work together to do the following:

1. Ensure that you understand the software specification and sketch out some kind of architecture for the software (i.e. what components are needed and how they should interact).

2. Determine which data and technologies you need to produce the software.

3. Find out enough about the data and technologies so that you can approximate how long the different parts of project will take, and who in your team is best suited to complete them.

4. Agree on the optimal way of working together to complete the project.

5. Identify any specific new skills that need to be learned by team members.

6. Produce a development plan for the duration of the project, detailing the various tasks and who will be responsible for them

Retrieving info

* Kinase data from Uniprot - name and gene symbol, the families they belong to, where in the cell they are typically found, and the sites they are known to phosphorylate, mass, binary interactions (could quantify this for interactive visualisation using JQuery or Bokeh)
* Phosphosite data from uniprot - genomic location, neighbouring sequence from ensembl
* PKIDB for kinase inhibitors (<http://www.icoa.fr/pkidb/>) cross reference the site info with our human kinase list and filter the targets - target, FDA approved, chemical structure, similar compounds using ChEMBL API, drug description, weight, chemical formula, associated conditions and references using DrugBank API
* By end of week 1

Developing the website server

* Decide how things will be displayed: Welcome page with introduction, site name, links to external sites, possibly include uniprot ID that can take you to uniprot, filter, advanced search, API service, Filter search bar so that you choose to look for either inhibitors or kinases of phosphosites
* Try out a dummy website
* Potentially add a phylogeny tree of the human kinases(kinase.com)
* Add blast or sequence alignment

uploading the quantitative phosphoproteomics data

* What information are we going to summarize from a statistical standpoint
* What statistical test they ran to obtain the p value?
* How to correct p-value
* Volcano plot?

Front end

* Flask – in terminal ‘pip install flask’
* CSS to design
* bootstrap

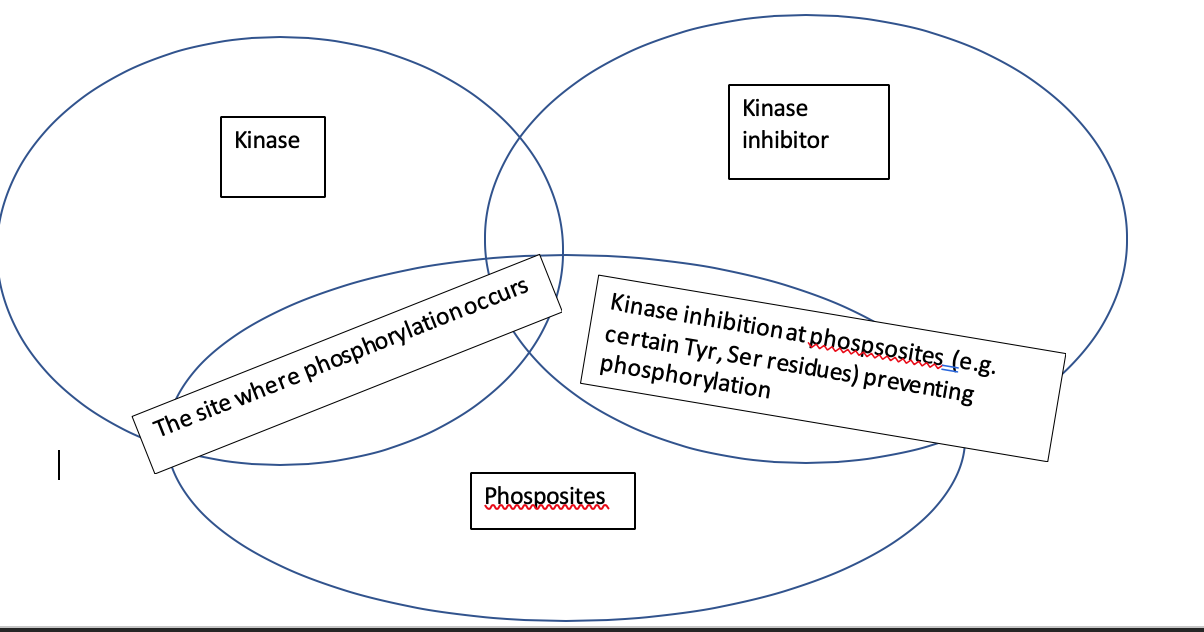
Back end

* SQLite to link the databases; 3 to start with including the Kinase data, Phosphosite data and kinase inhibitor data
* Dbdiagram.io?

New skills required

* SQL
* HTML
* CSS
* FLASK

Relational database – Venn diagram showing the link for the three databases we are going to build using SQL



Touch ups

* Decide on group name
* Choose/make an icon

Questions for thursday

* What do you mean by neighbouring sequence? (5 amino acids to the right and 5 to the left for example?)
* The family ‘other’? When checking these kinases in uniprot we can see they belong to the ser/thr family, should we change this or leave it as it is? (ubiquitin-like?), tyr family is not written as ‘other’.
* Sites the kinases are known to phosphorylate? Cant find a site that shows this