Bioinformatics 2

Introduction

Lecture 1

- Course Overview & Assessment
- Introduction to Bioinformatics Research
- Careers and PhD options
- Core topics in Bioinformatics
  - the central dogma of molecular biology

About us...

- Started in Biology (behaviour genetics)
- Got interested in databases (anatomy)
- Commercial and Academic Experience
- 'wet lab' and bioinformatics projects
- Office in forum, Lab in HRB
- Lysimachos Zografos – labs and tutorials

The class (2010)

- M.Sc. Classes:
- Quantitative Genetics and Genome Analysis (assignment 1 and term paper)
- Bioinformatics 2 (assignment 1 and exam)

What do I think you know?

- Variety of backgrounds and experience:
  - Biological Sciences
  - Computing Sciences
  - Mathematics, Statistics and Physics

Course Outcomes

- Know the core algorithms in bioinformatics
- Experience in using and/or implementing simple solutions
- Appreciate the current 'state of the art'
  - what has been solved?
  - what are the key limitations?
- Be familiar with the available resources

Course Design

- Lectures cover essential background
- Guest lectures present research level
- Self-study and assignments designed to cover practical implementation
- Tutorial / Lab support

Assessment (Bio2)

- Written assignment
  - Experimental design and data analysis mini project
  - Plagiarism will be refereed externally
    - Cite all sources!!!
  - Late submissions get 0 marks!

Bioinformatics?

- Introduce yourselves to each other.
- What is Bioinformatics?
- What does Bioinformatics do for CS?
- What does Bioinformatics do for Biology?
- What guest Bioinformatics lecture would you like?

- Discuss in groups for 10 min.

What is BioInformatics?

- Sequence analysis and genome building
- Molecular Structure prediction
- Evolution, phylogeny and linkage
- Automated data collection and analysis
- Simulations
- Biological databases and resources

BioInf and CS

- Provides CS with new challenges with clear medical significance.
- Complex and large datasets sometimes very noisy with hidden structures.
- Can biological solutions be used to inspire new computational tools and methods?

BioInf and Biology

- High-throughput biology:
  - around 1989, the sequence of a 1.8kb gene would be a PhD project
  - by 1993, the same project was an undergraduate project
  - in 2000 we generated 40kb sequence per week in a non-genomics lab
  - Illumina/Solexa systems Gigabases per expt.

BioInf and Biology

- High-throughput biology
- Data management and mining
- Modeling of Biological theories
- Analysis of complex systems
- Design and re-engineering of new biological entities

bioinformatics

- http://www.bbsrc.ac.uk/science/grants/index.html
  - Awarded grants database
- Bioinformatics.oxfordjournals.org
- www.biomedcentral.org/bmcbioinformatics
- www.nature.com/msb

Bioinformatics@ed

- Database integration
- Data provenance
- Evolutionary and genetic computation
- Gene expression databases
- High performance data structures for semi-structured data (Vectorised XML)

1/2

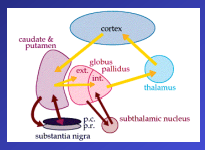
Bioinformatics@ed

- Machine learning
- Microarray data analysis
- Natural language and bio-text mining
- Neural computation, visualisation and simulation
- Protein complex modeling
- Systems Biology
- Synthetic Biology

bioInf activities @ ed

- Self organised reading groups
- http://www.bioinformatics.ed.ac.uk
- http://www.bioinformatics.ed.ac.uk/wiki/SysBioClub/
  - Regular seminars at 1200 thursdays
- Network science journal club (Lysimachos)
- Edinburgh Proteomics Group (Juri)

Neuroinformatics



information-processing in the nervous system

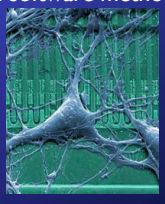
Computational Models

£10M

£4M

Neural Engineering

inspire new hardware and software methods



£12M

Software Systems

collect, analyze, archive, share, simulate and visualize data and models

Career Options

- Academic Routes
  - Get Ph.D, do Postdoctoral Research - lectureship and independent group
  - M.Sc. RA - becomes semi independent usually linked to one or more academic groups. Career structure is less defined but improving. RAs can do Ph.D. part-time.

Career Options

- Commercial Sector
  - Big Pharma - Accept PhD and MSc entry. Normally assigned to projects and work within defined teams. Defined career structure (group leaders, project managers etc)
  - Spin-out/Small biotech - Accept PhD and MSc entry. More freedom and variety. A degree of 'maintenance' work is to be expected.

Career Options

- Hybrid Approaches
  - Commercial and Academic research groups are becoming much closer linked.
  - University academics encouraged to exploit their IPR (intellectual property rights).
  - Companies can get government support to collaborate with academic research groups.

Ph.D.

- Assuming a start date of September 2010
- 'prize' studentships advertised on jobs.ac.uk, Nature, Science etc starting NOW!
  - Many linked to nationality/residency (Check details carefully).
- UK 'quota' studentships vary with department but contact/apply early.

Ph.D.

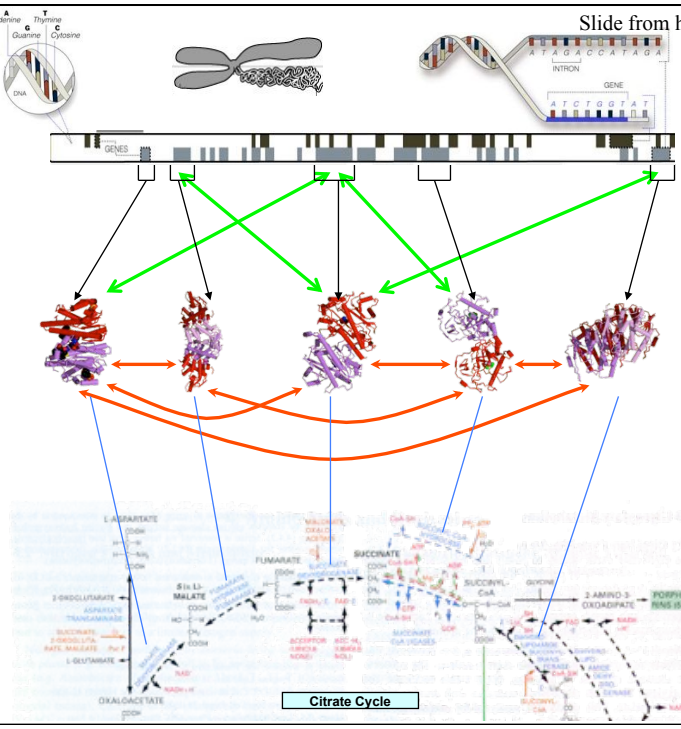
- US studentships take longer but are better paid and have extra training/coursework
  - require an entry exam
  - deadlines around summer for '11

Bioinformatics 2

Introduction

Armstrong, 2008 Bio2 lecture 1

Slide from http://www.nd.edu/~networks/



GENOME

protein-gene interactions

PROTEOME

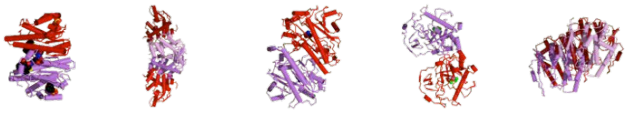
protein-protein interactions

METABOLISM

Bio-chemical reactions

Slide from http://www.nd.edu/~networks/

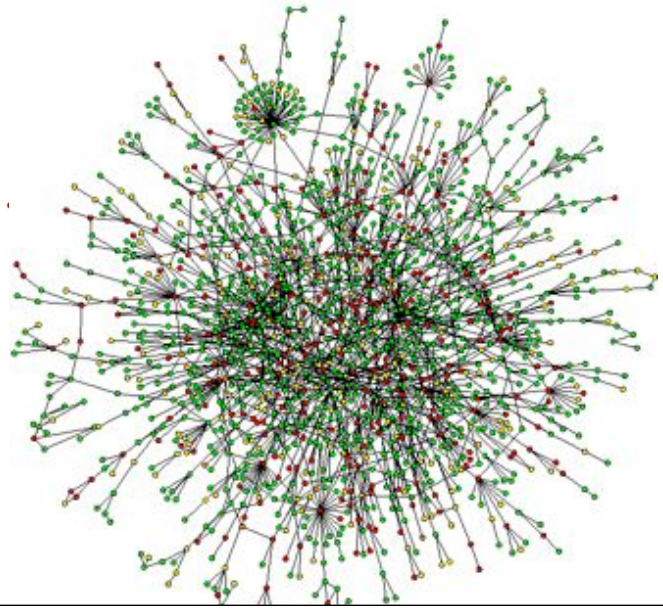
Lecture 2



PROTEOME

Slide from http://www.nd.edu/~networks/

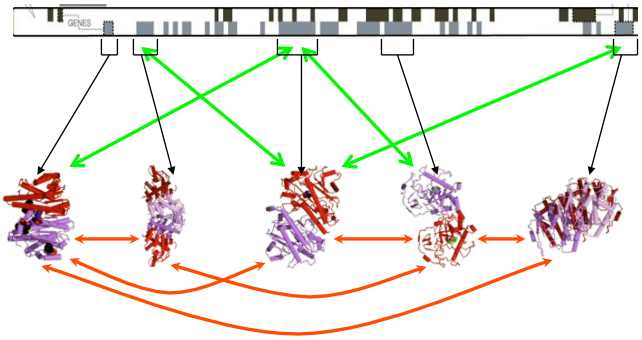
Lecture 2 and later (interaction networks)



PROTEOME

protein-protein interactions

Slide from http://www.nd.edu/~networks/



GENOME

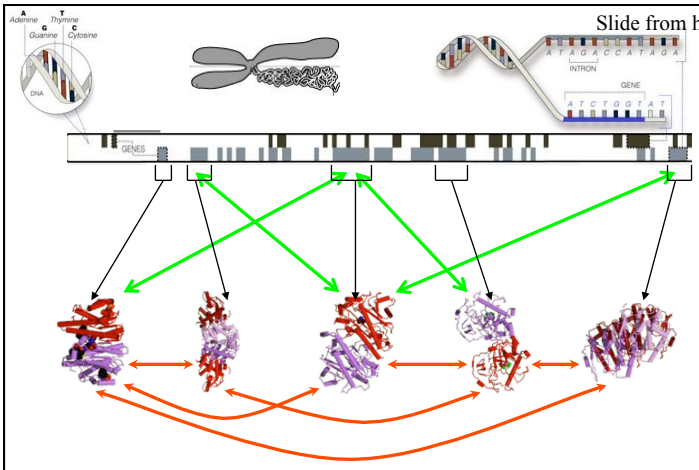
protein-gene interactions

PROTEOME

protein-protein interactions

Lectures 3,4
Functional genomics, microarrays and protein-gene interactions

Slide from http://www.nd.edu/~networks/



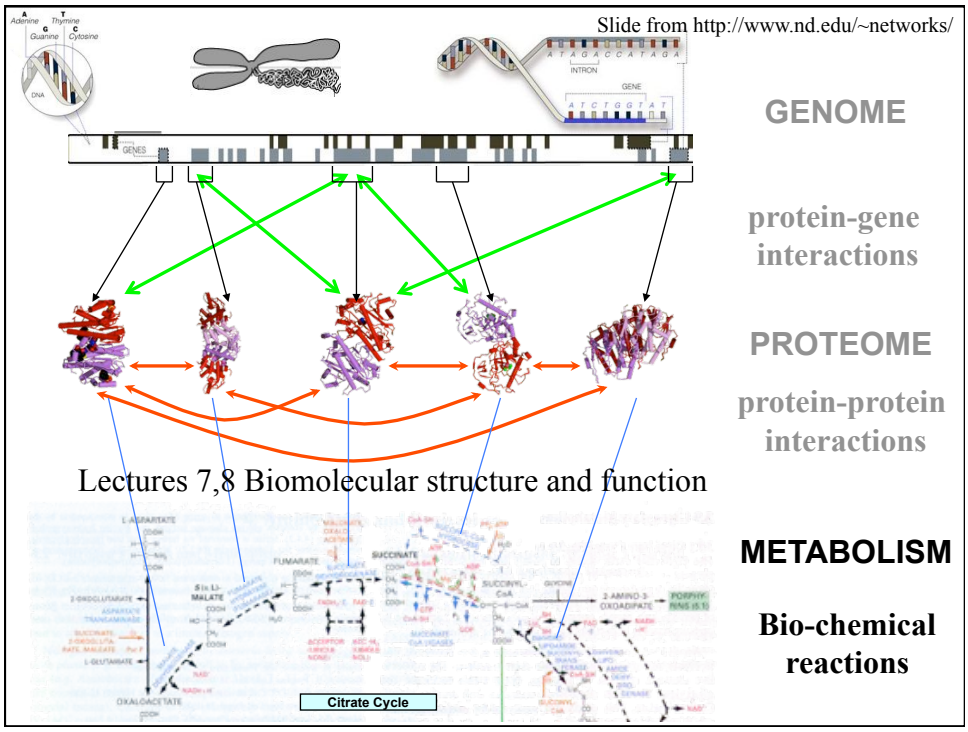
GENOME

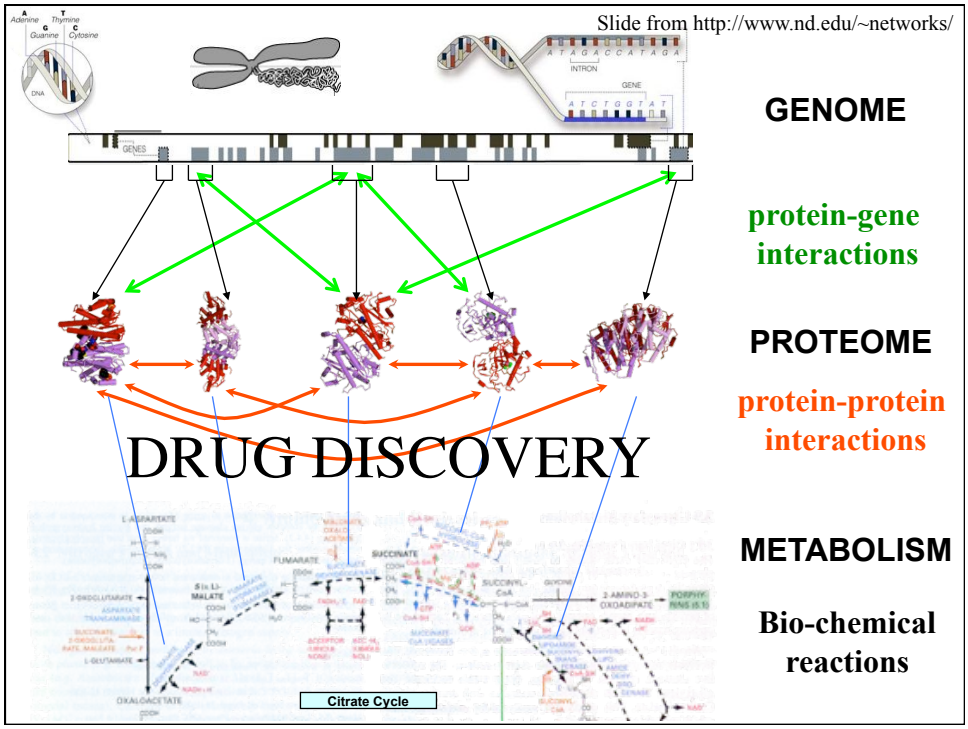
protein-gene interactions

PROTEOME

protein-protein interactions

Lectures 5,6 genetics and sequence level bioinformatics


Slide from http://www.nd.edu/~networks/
GENOME
protein-gene interactions
PROTEOME
protein-protein interactions
Lectures 7,8 Biomolecular structure and function
METABOLISM
Bio-chemical reactions
Citrate Cycle


Slide from http://www.nd.edu/~networks/
GENOME
protein-gene interactions
PROTEOME
protein-protein interactions
DRUG DISCOVERY
METABOLISM
Bio-chemical reactions
Citrate Cycle

Guest lectures and topics (subject to change)

Yuri Rappsilber (Biological Sciences) Proteomics

Donald Dunbar (Centre for Inflamation Research)
Microarray technologies

Malcolm Walkinshaw (Biological Sciences)
Structure based in silico drug design

Chris Larminie (GalaxoSmithKline)
Bioinformatics in the Pharmaceutical Sector

Preparation work for lecture 2

Understand the 'central dogma of molecular biology'
What is a protein?
What is an amino acid?
How are amino acids joined together to make proteins?
What are the key characteristics of amino acids and proteins
(pH, charge, acidity, hydrophobicity)
What are peptidases? (enzymes that cut protein bonds)

Given a protein sequence how would you predict the weight of a protein?
Understand that protein sequences vary between species and are stored/queried from databases.