COSC 348:
Computing for Bioinformatics

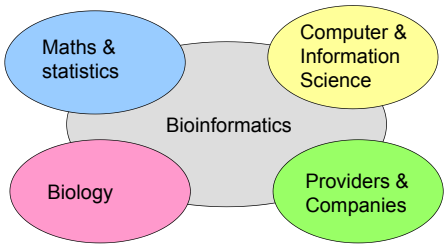
Lecture 1: Introduction

Lubica Benuskova

http://www.cs.otago.ac.nz/cosc348/

Bioinformatics

- Bioinformatics is the field of science, in which biology, mathematics, statistics, computer science, and information technology merge into a single discipline.



Why Computing for Bioinformatics?

- Bioinformatics is an *applied* area of computing.
- We encounter problems that are unlike any in studying compilers or operating systems, computer architecture, networks, graphics, etc.
- The kind of problems Bioinformatics presents us with are fundamentally *algorithmic* problems.
- We have to understand the limitations of our answers, especially how well they scale with problem size.
- We have to start by understanding what the biological data and problem actually are.

Computing for Bioinformatics

- This course is designed for "tool-developers" rather than tool users.
  - It is not about statistical analysis of data neither. This is the subject of STAT 435 at the Dept. of Math & Stats.
- Three major sub-disciplines in relation to computation:
  - Algorithms to assess relationships between molecules;
  - Algorithms to analyse various types of molecular data;
  - Tools to enable efficient access and management of different types of data and information (= database and IR tools).

Assessment

- Assessment consists of two programming assignments (worth respectively 14% and 14% of the final mark):
  - http://www.cs.otago.ac.nz/cosc348/assignments.php
- 12% for lab work (1% for each lab work starting at week 2):
  - http://www.cs.otago.ac.nz/cosc348/labs.php
- a comprehensive closed book final examination (worth 60%):
  - http://www.otago.ac.nz/library/exams/
- Lecture notes and other supplementary material for study:
  - http://www.cs.otago.ac.nz/cosc348/lectures.php

Bioinformatics jobs (quotes from real ads)

- Job description: Design, development, implementation and testing of bioinformatics tools.
- Typical requirements:
  - Experience with the use of bioinformatics applications and tools for genomic analysis;
  - Knowledge of necessary biological databases;
  - Knowledge and experience with scripting and programming languages;
  - Knowledge and well developed IT technical skills with emphasis on `Unix`, pipeline development and web based technologies.

Bioinformatics companies and research in NZ

- NZ Genomics (provides genomics technology and bioinformatics services to underpin research in a broad range of areas)
- AgResearch (analysis of bovine, sheep and plant genomes)
- PEBL (Pacific Edge Biotechnology), Dunedin (development of cancer diagnostic tests)
- University of Otago
  - Database of RNA sequences / motifs, analysis of various gene data, etc.
- University of Auckland / Institute of Bioinformatics
  - New pharmaceuticals, bacterial genomes analysis, etc.

Introduction: 3 lectures

- Basic concepts
  - Elements, atoms, molecules
  - Cell, nucleus, chromosomes
  - DNA, RNA, proteins
  - Genetic code, genes, "junk" DNA
  - Gene transcription regulation
- Big picture
  - Individual development during life
  - Genetic diseases
  - Evolution

Elements in nature

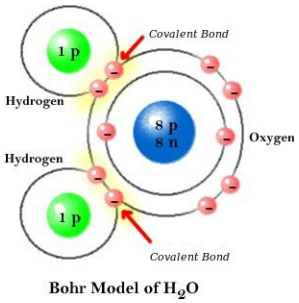
(Mendeleyev) Periodic Table

Chemical series of the periodic table

Alkali metals	Alkaline earth metals	Lanthanides	Actinides	Transition metals	
Poor metals	Metalloids	Nonmetals	Halogens	Noble gases	

Elements, atoms, molecules

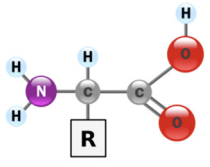
- Number assigned to each element is the number of protons/neutrons in the nucleus of its atom and the corresponding number of electrons in its orbit
- Atoms form molecules via chemical bonds
  - E.g. molecule of water H_2O, is made of two atoms of hydrogen and one atom of oxygen



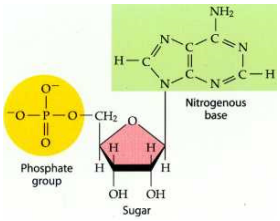
Bohr Model of H_2O

Biological molecules: COHN

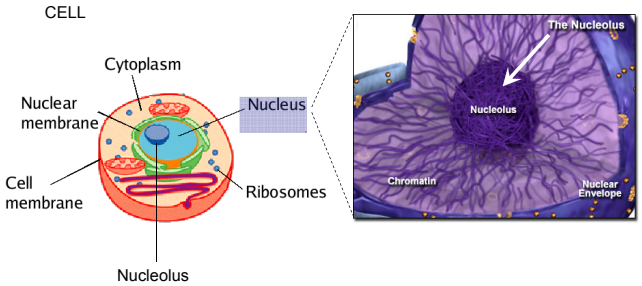
- Amino acids: building blocks of proteins
  - We are built by 20 amino acids
  - Each has a different residue group **R**



- Nucleotides: building blocks of DNA & RNA
  - 4 nucleotides
  - Each has a different base group

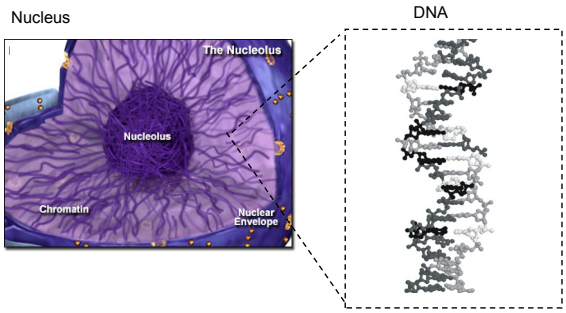


Cell and nucleus



Note: A cell is the basic organizational unit of all living organisms. A cell is comprised of organelles, which are comprised of molecules, which are comprised of atoms.

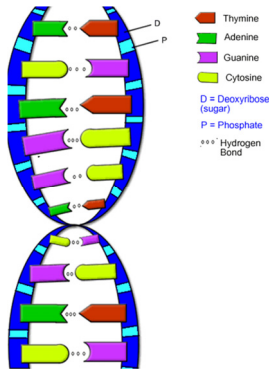
Nucleus and DNA



Note: Eukaryotes are organisms whose cells have a nucleus (animals, plants, fungi, protists). Bacteria and archaea are prokaryotes, i.e. are without a nucleus.

DNA - **D**eoxyribo**N**ucleic **A**cid

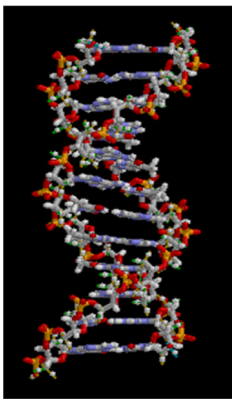
- Double-helix structure (discovered in 1953)
- 4 bases: Adenine (A), Thymine (T), Guanine (G), Cytosine (C)
- Bases pair in a base-pair or bp-rule: T-A and C-G
- The order (sequence) of the bases (**A, T, G, C**) determines the information similar to the way, in which letters of the alphabet form words and sentences.



Note: DNA was discovered by Swiss biologist Friedrich Miescher in 19th cent.

DNA double-helix discovery: Nobel Prize 1962

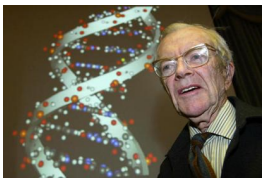
James Watson



Francis Crick

Maurice Wilkins: the 3rd Nobel laureate for DNA

- a New Zealand-born British molecular biologist known for his work at King's College London on the structure of DNA
- born in Pongaroa, north Wairarapa, east of Wellington, New Zealand

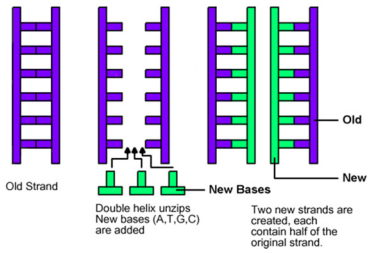


Basic facts about DNA

- "the blueprint of life" because it contains instructions for building and maintaining the organism
- DNA carries life's *hereditary* information, it's inherited from parents
- every cell in a person's body has the same DNA
- Human DNA consists of about 3 billion bases
- >99% of those bases are the same in all people on earth

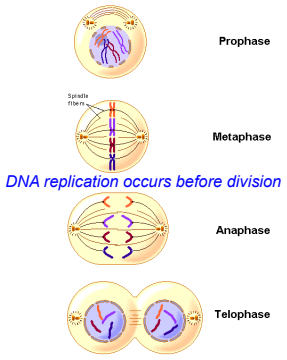
DNA replication

- **Replication**: DNA makes a copy of itself. DNA replicates right before a cell divides. Every new cell needs a copy of DNA.
- DNA replication is **semi-conservative**: when DNA makes its copy, one strand of the old DNA is always kept in the new DNA.



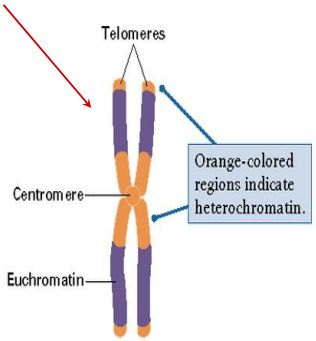
Cell division and development

- We all come from a single cell – a fertilized egg, which divides and divides and divides...
- Cells divide:
  - during the development and growth of organism
  - when the tissue is damaged and needs repair
  - all the time because cells need replacement when they age
- All unicellular organisms produce offspring by division.



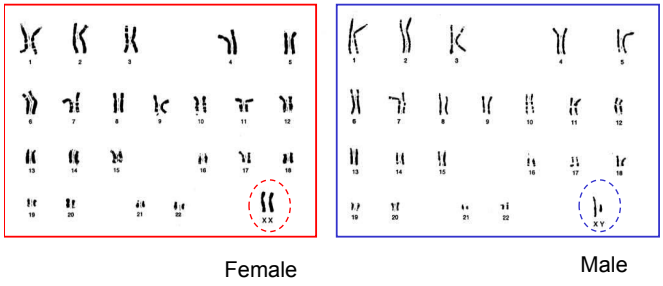
Chromosome

- DNA self-organizes in chromosomes during mitosis and meiosis (cell cycles in which cells divide)
- *Euchromatin* consists of DNA that is *expressed* as protein
- *Heterochromatin* consists of DNA that is not expressed as proteins but has a different function



Genome

is the whole hereditary information encoded in the DNA



Female Male

In humans, there are 22 pairs of autosomes (non-sex chromosomes) and 2 sex chromosomes: XX in females and XY in males

Big scientific questions

- Can we prolong our lives? Can we stop the aging process and be young forever? How?
- We all come from a single cell – a fertilized egg. As the embryo becomes multicellular, its cells specialize, and all this is governed by instructions from DNA and environment - how ?
- How are we to understand development from a single cell to a conscious being? How the genes shape our brains?
- Can we become smarter? And how?

Big scientific questions

- How evolution works on a molecular level?
- What happens if there are errors in DNA? Book Genes and Disease: http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gnd
- How can we treat it? How can we prevent this from happening? Design of new drugs.
- Developing the so-called personalised medicine of the future based on genetic profiling.
- Can we make an artificial life? I.e. to assemble a living thing from scratch?

Ethical / moral issues

"Your problem is in the gene that makes antibodies, but since the Biophase Corp. now has a patent on that gene, I can't do anything for you."

"IF WE CHANGE ONE MOLECULE OF THIS PAIN-KILLER, WE'LL GET AN EXCELLENT HAIR SPRAY. THE SALES PEOPLE ARE DECIDING WHICH WAY TO GO."