

COMP90016 – Computational Genomics *Introduction*

Department of Computing and
Information Systems
The University of Melbourne

Welcome

- To the University of Melbourne!
- To the the department/school of Computing and Information Systems.
- To Computational Genomics/Bioinformatics - a field of research with a bright future.

<http://researchaustralia.org/health-medical-research/medical-research-future-fund/>

General Information

- Lectures: Tuesdays 10-11am
Electrical Engineering Brown Theatre
Wednesdays 1-2pm
Chemical and Biomolecular
Engineering Theatre
- Labs: Tuesday 11am-12pm AH 236
Thursday 2:15-3:15pm AH 211
- Staff: Jan Schröder Ruining Dong
Office 8.11 DmD
Email jsch@unimelb.edu.au ruining@student.unimelb.edu.au



Walter+Eliza Hall
Institute of Medical Research



Who is sitting in this class?

Current Degree	Number
Masters of bioinformatics	
Computer Science	
Biomedical Engineering	
Software Engineering	

Prior Degree - Major (focus)	Number
Science - Bio	
Science - Computing	
Science - Maths/Stats	
Engineering	

Genomics!

- Genomics is the study of genomes:
 - Structural Genomics
 - Functional Genomics
 - Comparative Genomics
 - Computational Genomics
- Computational Genomics:
 - Computation for asking questions about genomics.
 - Developing new computational techniques as needed.

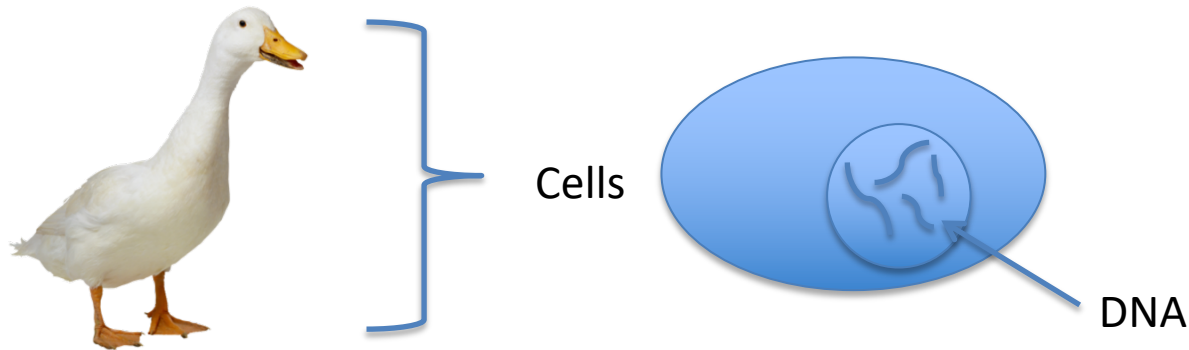


Related Disciplines

- Bioinformatics:
 - Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data. *
- Computational Genomics:
 - The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems. *
- Computational Biology
- ...

* Definitions by the NIH

The Genome



- “An adult human being has approximately 100,000 billion cells, all originating from a single cell, the fertilized egg cell.”*
- The collection of DNA molecules in each cell is (mostly) identical.
- The collection of DNA in cells is called the organism's genome.

*Nobel prize website

The Promise of Genomics

- Wouldn't it be nice if you could predict a patient's response to a drug? By looking at their genome?
- Wouldn't it be nice if you could assess a person's disease risk? And advise accordingly. By looking at their genome?
- State of the art today:
<http://www.youtube.com/watch?v=Lbb8OHQNeXo>

Personalized Medicine

- An important goal.
- The hope that information about a patient (proteins, genes, metabolites, etc.) could be used to tailor medical care to that *individual's* needs.

Genome Analysis in Individual Medicine: Breast Cancer

- The lifetime risk of breast cancer (women): 11-12%
 - For women with a BRCA1 mutation: 56-87% risk
 - For women with a BRCA2 mutation: 37-85% risk
- The lifetime risk of ovarian cancer is 1.4%
 - BRCA1 mutation: 39% risk
 - BRCA2: 11-17% risk
- Figures from:
<http://www.cancer.gov/cancertopics/factsheet/Risk/BRCA>

Genome Analysis in Individual Medicine: Breast Cancer (2)

- Genetic knowledge and testing:
 - Single site (known family mutation) \$300
 - Multi site 3 common BRCA mutations: \$400
 - Sequencing BRCA1 & BRCA2: \$2700
- Projected price decrease: Ozcelik et al. The Journal of Molecular Diagnostics, Volume 14, Issue 5 (September 2012)



For more information see:

<http://www.cancer.gov/cancertopics/factsheet/Risk/BRCA>

<http://www.melbournebreastcancersurgery.com.au/brca-gene-testing.html>

NIH: Undiagnosed Diseases Program

- Nic Volker developed inflammatory bowel disease at less than 2 years old.
- “Nic Volker May Be Leading Edge of New Wave of Genetic Medicine” Milwaukee Wisconsin Journal Sentinel, 27 Feb 2011
- Evidence is emerging that the case of Nic Volker, a young undiagnosed patient in Wisconsin who was treated after exome sequencing revealed a specific genetic lesion, may be the leading edge of a wave moving across medicine. Doctors at the Medical College and Children's Hospital have half-a-dozen new cases moving through an evaluation process while other institutions are expanding the use of new genomic analysis tools.

NIH: Undiagnosed Diseases Program

- Amanda Young had a series of illnesses from a young age, and almost died several times.
- After 21 years of misery, extensive genomic testing eventually found a deficiency in IRAK-4. Appropriate management was put in place.
- <http://www.genome.gov/27026392>

The Requirement For Computation

- So lots of genetic information is available, but
 - *not* all helpful ...
 - ... often because analytic methods haven't caught up with the data generation.
- High throughput techniques and analysis are needed to look at low frequency signals.
- High throughput techniques and analysis require computation.
- High throughput techniques and analysis inspire new approaches to computation.

Computation is Needed

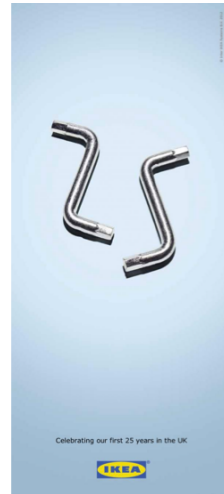
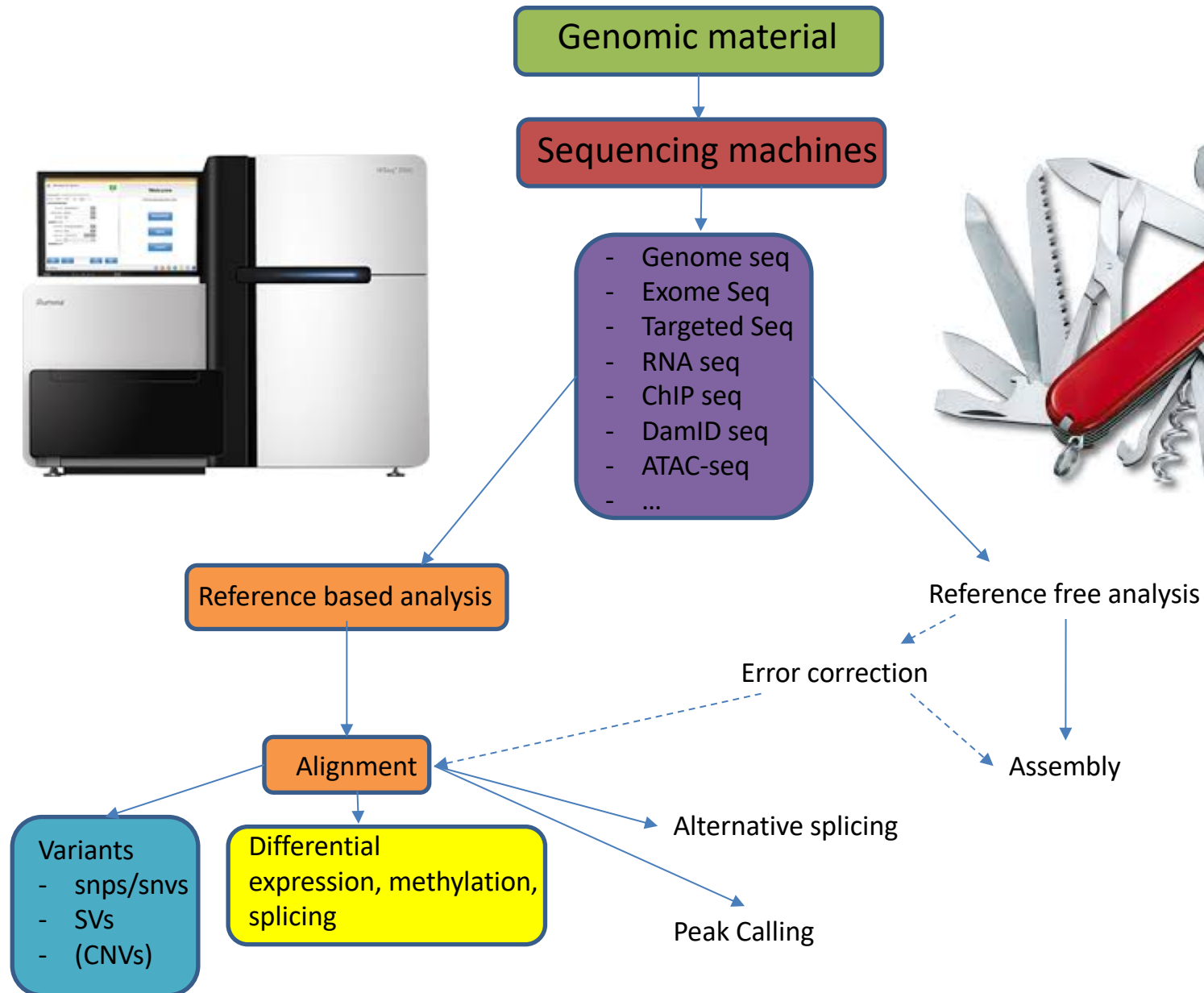
- Current genomics needs computation for:
 - Data gathering:
 - Wet lab: assembling sequences, preprocessing raw data, etc.,
 - Dry lab: simulation, modelling.
 - Data Analysis:
 - drawing meaningful information from data.
 - annotation, comparison, genetic linkage determination, etc.
 - Communicating results:
 - coordinating diverse data and databases, large databases, etc.
- And more...

Aims of This Subject

- What you will come out with:
 - Understanding of flow of genetic information.
 - Understanding of role of computational approaches to analysis of biological information.
 - Familiarity with mainstream thought.
 - Awareness of open problems in the discipline.
- To be prepared for:
 - Further study, or
 - Work in the discipline.

Objectives

- On completion of this subject, you should be able to:
 - Explain key concepts in bioinformatics.
 - Locate and use existing bioinformatics resources.
 - Compare and contrast different approaches to bioinformatics problems.
 - Read and critique research articles in the bioinformatics literature.
 - Describe major research issues in genomic era and post-genomic bioinformatics.
 - Build simple bioinformatics computer applications.



Provisional Lecture Schedule

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Intro/Genome and DNA	Sequencing	Genome Organisation and Content	Alignment	SNPs	Paper Discussion/pysam
	Group Assignment start (no marks)		Assignment 1 start		Assignment 1 due/ Assignment 2 start
Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
HMMs	SV 1	SV 2	SV 3	Hierarchical clustering/ Sequencing Assays	Recap
		Assignment 2 due/ assignment 3 start			Assignment 3 due

Subject Materials

- LMS: <http://www.lms.unimelb.edu.au>
- NCBI: National Center for Biotechnology Information (USA): major web resource for computational genomics:
<http://www.ncbi.nlm.nih.gov>
- Ensembl: Ensembl is a joint project between EMBL - EBI and the Wellcome Trust Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes:
<http://www.ensembl.org/>
- UCSC Genome Browser: <https://genome.ucsc.edu/>
- Journals
 - Bioinformatics, available through library
 - PLoS Biology/PLoS One, freely available online
- Books:
 - *An Introduction to Bioinformatics Algorithms* by Neil C. Jones and Pavel A. Pevzner (available in library).
 - *Genomes* (<https://www.ncbi.nlm.nih.gov/books/NBK21128/>) by Terrence Brown.

Assessment

- 60% final examination
- 40% practical work
 - Spread over 3 assignments.
- Hurdles:
 - Examination: 30/60
 - Practical work: 20/40 and every assignment must be submitted.
- Unless otherwise noted, assignments are individual work.
- Late assignments will not be marked unless prior arrangements have been made with the lecturer.

Programming environment

- We will be using Python and a bit of R to work with and analyse genomics data.
- We are going to test assignment submissions on the MSE server digitalis (digitalis.eng.unimelb.edu.au).

Make sure that you have access to the server, or otherwise lodge a ticket with IT services to get an account (you should have access automatically if you logged in to the MSE lab PCs before).

Lab Workshops

- There are weekly labs.
- Labs start this week with an introduction to Python programming.
- Each week the workshops are deepening some concepts from the lectures.

Subject Communication

- My office hours
9am each Tuesday at Doug McDonnell 8.11
- My email
jsch@unimelb.edu.au
- Discussion forums
LMS discussion board
Other preferences: Slack, Yammer?

See you tomorrow.

THANK YOU