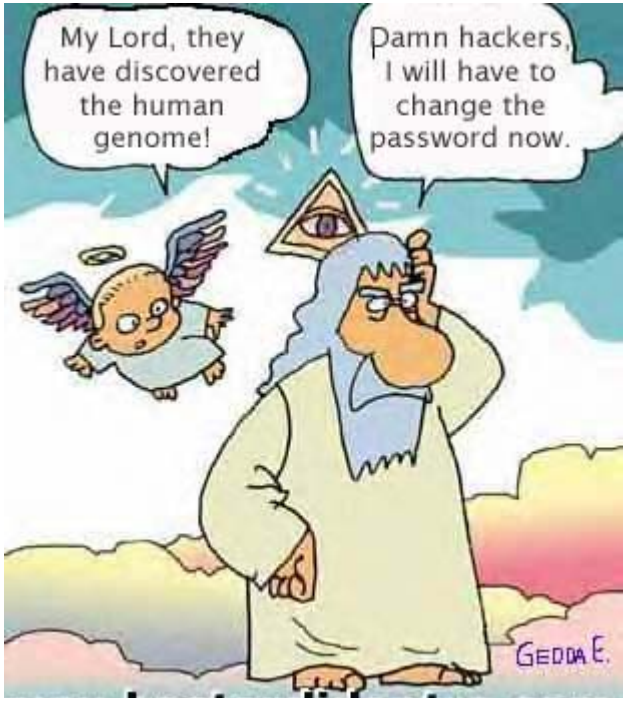
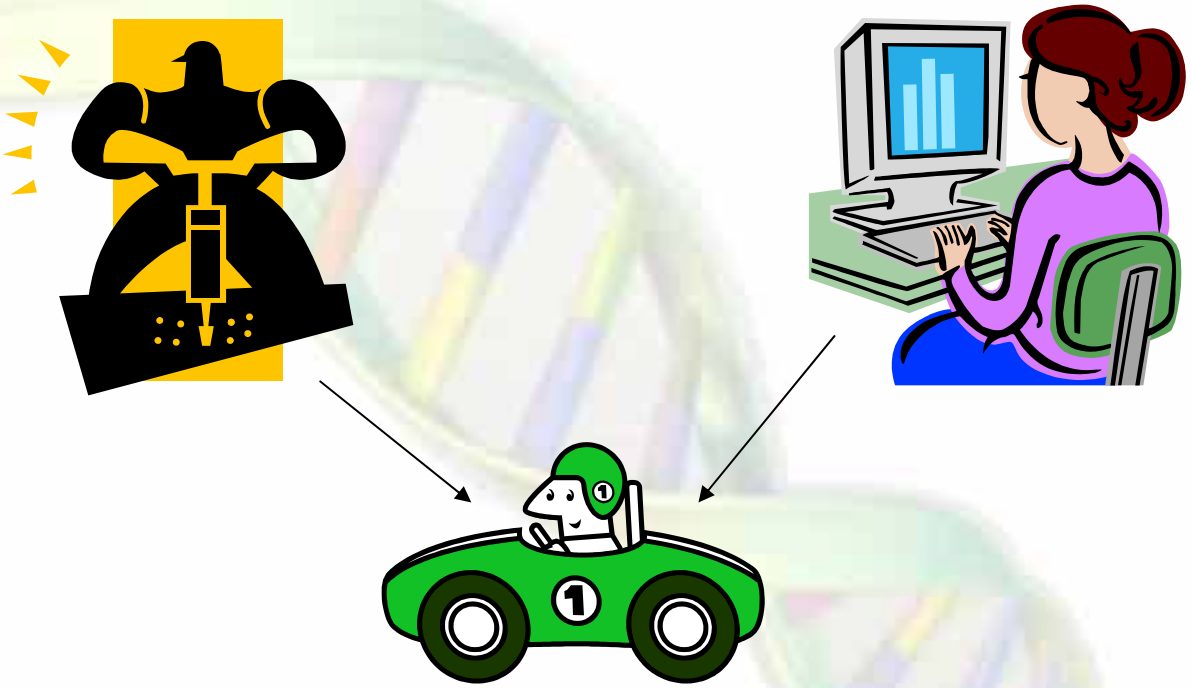


Gene expression in the immune system

This talk

- Chromosomes
- DNA
- Genes (and gene expression)
- Proteins
- Operons
- General database problems in bioinformatics
- Gene expression in the immune system?

Our chromosomes

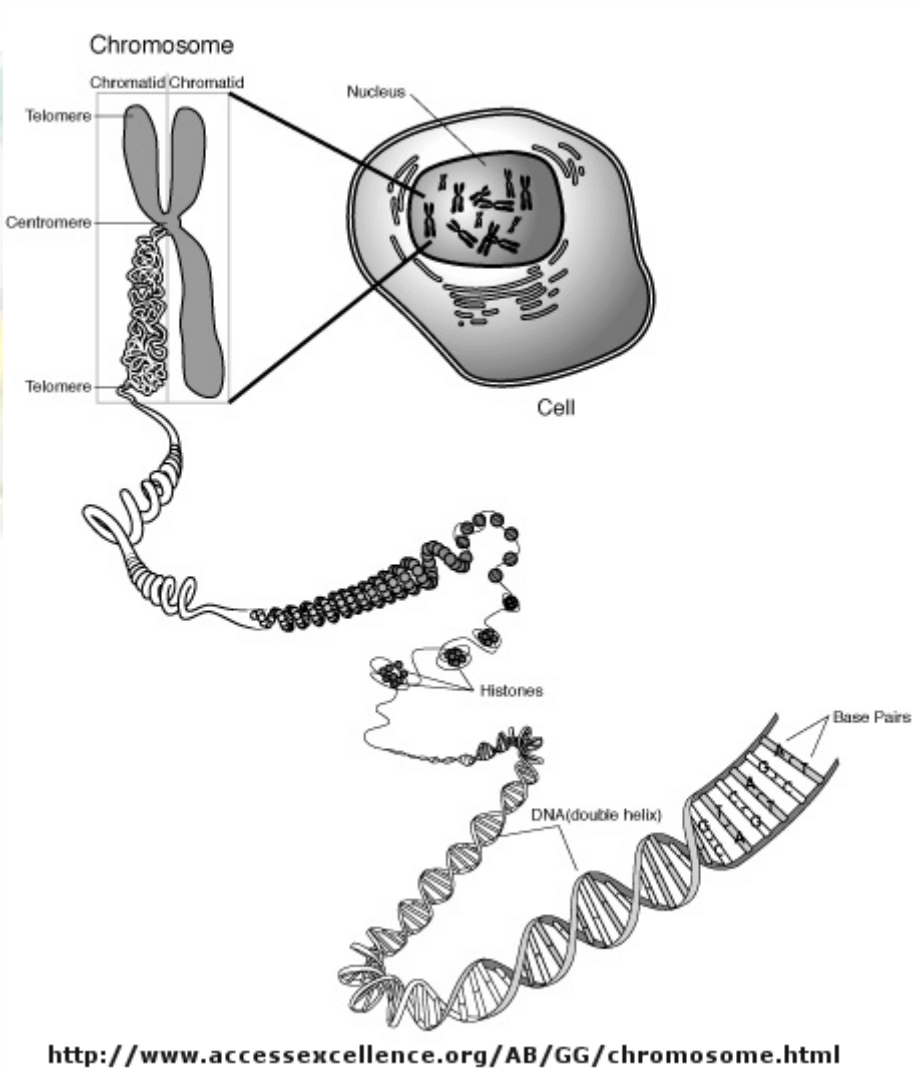


All in all 24 chromosomes

Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

cont…

- Packaging
  - the shortest human chromosome contains 4.6×10^7 bp of DNA
  - ~14,000 μm of extended DNA
  - ~2 μm long
  - Packing ratio = 7000 (14,000/2)



http://www.accessexcellence.org/AB/GG/chromosome.html

Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

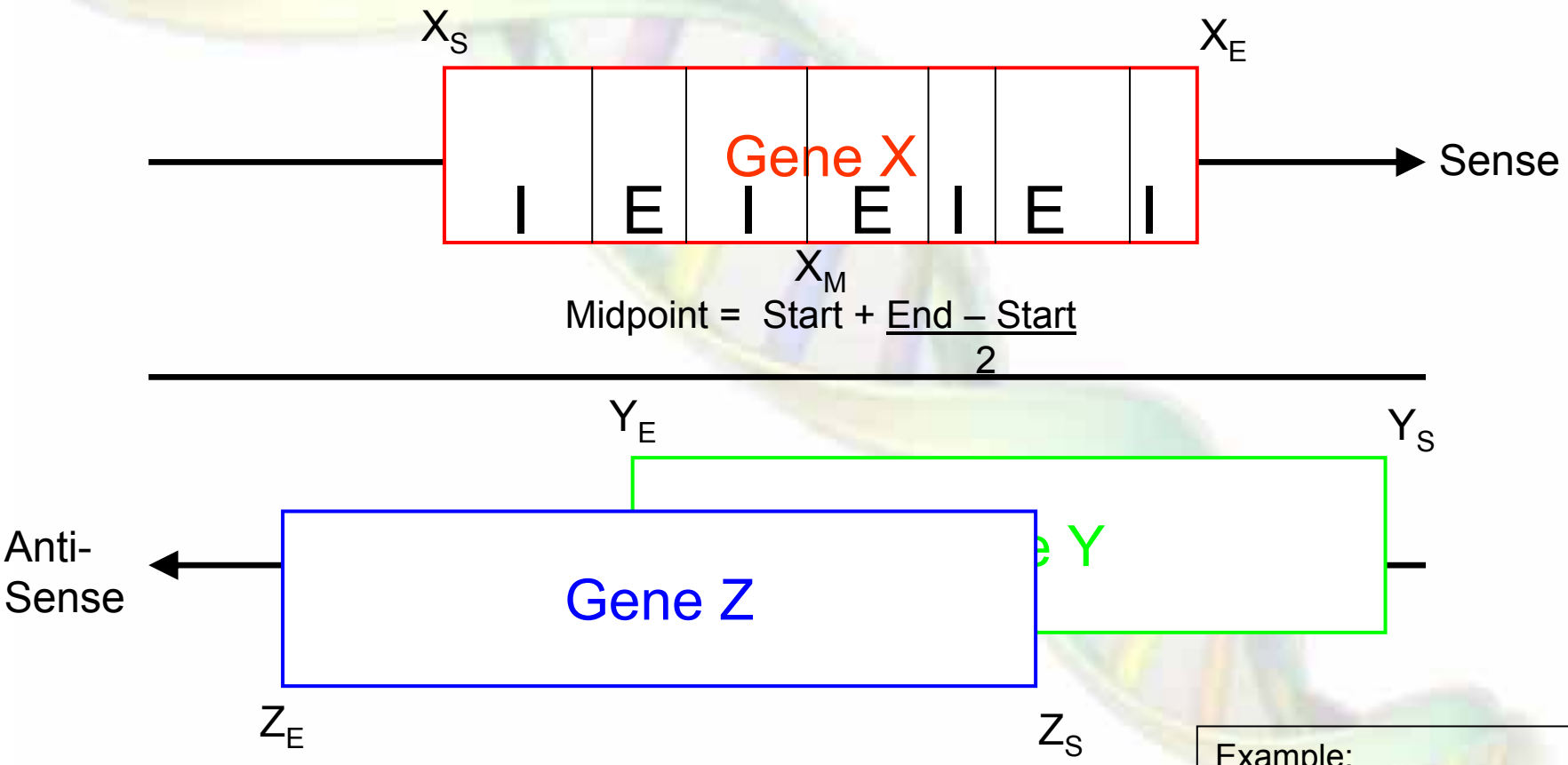
DNA

- ~ 3 billion base pairs in the human genome
- 4 repeating bases
  - A C T G
- These bases pair up
  - AT double-binding
  - CG triple-binding
- ~ 6 million base pairs difference between you and me
- Are we really so different from a mouse?

Genes

- The part of the DNA that encodes proteins are called genes
- Genes are only translated into proteins (by the ribosome) when the gene is expressed/on.
- Microarrays is a technology for measuring the LEVEL of gene expression in a given tissue at a specific time point.
  - Ex: diseased and sick tissue
- The level of gene expression is assumed to mirror the protein level

Gene expression



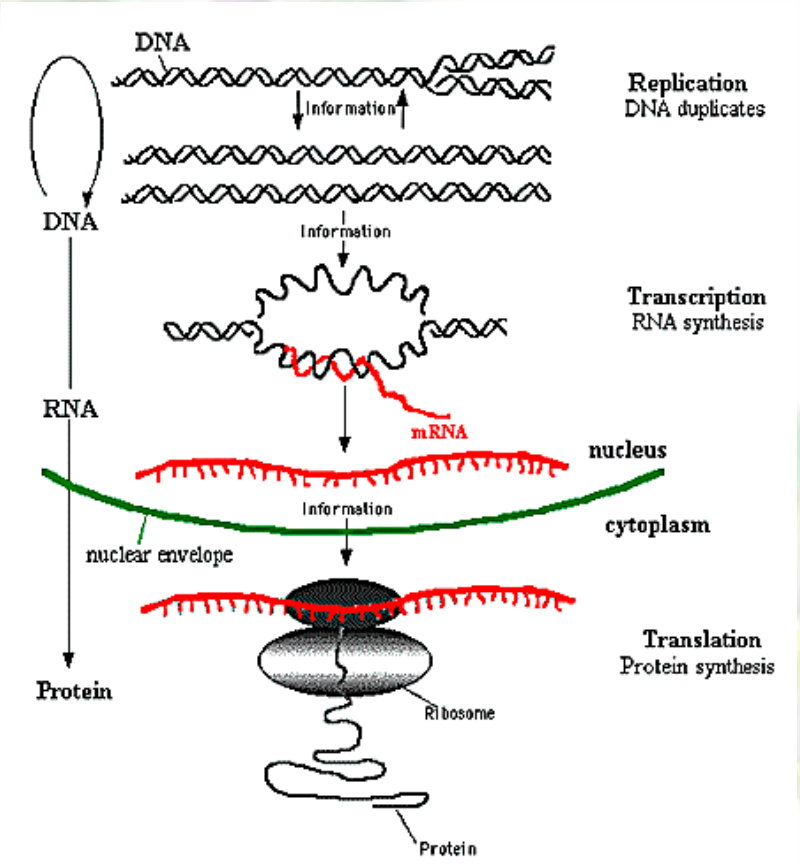
Example:
X: 100 + (200-100)/2 = 150
Y: 230 + (130-230)/2 = 180

Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

Proteins

- The molecules responsible for “everything”
  - Responding to external stimuli
  - Regulating the gene expression level
    - i.e. switching genes and/or cascades on/off
    - Catalysing/suppressing reactions (“enzymes”)
  - etc
- Dimensionality issues
  - One protein can be encoded by multiple genes
  - One gene can encode multiple proteins

The Central Dogma of Molecular Biology



The Central Dogma of Molecular Biology

Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

Operons

- In prokaryotes and nematodes
- A segment of DNA containing adjacent genes, including structural genes, an operator gene and a regulatory gene (promoter/suppressor)
  - OPERATOR = DNA that regulates the activity of the structural genes in an operon, by interacting with a specific repressor or activator
  - PROMOTER = “binding site for transcription enzyme”
- That is the ENTIRE region is switched on or off,

General database ”problems” in bioinformatics regarding this project

- Need information from multiple databases
  - => Have to map on set of identifiers (GenBank/RefSeq) against another set of identifiers (Ensembl)
    - An “allegedly” up-to-date mapping still contain incorrect information
    - The information in the GenBank database change rapidly
      - The researcher found an error in the sequence, the “gene” was realised to be a part of another etc etc
  - => What source to trust?
    - Why Ensembl?
  - => one db for human and one for mouse
    - If I start “mixing” the information, the problem is multiplied
    - Now I have “hacked scripts” for uploading a new version of the database...

Cont…

- A microarray contains information from other organism as well
- BLAST probes against Ensembl
  - BLAST goes against NCBI
  - Where to get the probe sequence?
  - What e-value and score should I use for automatic definition, manual definition etc etc

Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

- END OF BASIC BIOLOGY
  - **********
- START OF GENE EXPRESSION IN THE IMMUNE SYSTEM

Gene expression in the immune system

- What are we trying to do?
  - Why?
  - How?
- "Results"
  - "Conclusions"
- What next?

Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

What?

- Basically describe the way genes are ordered on the chromosomes

Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

Why?

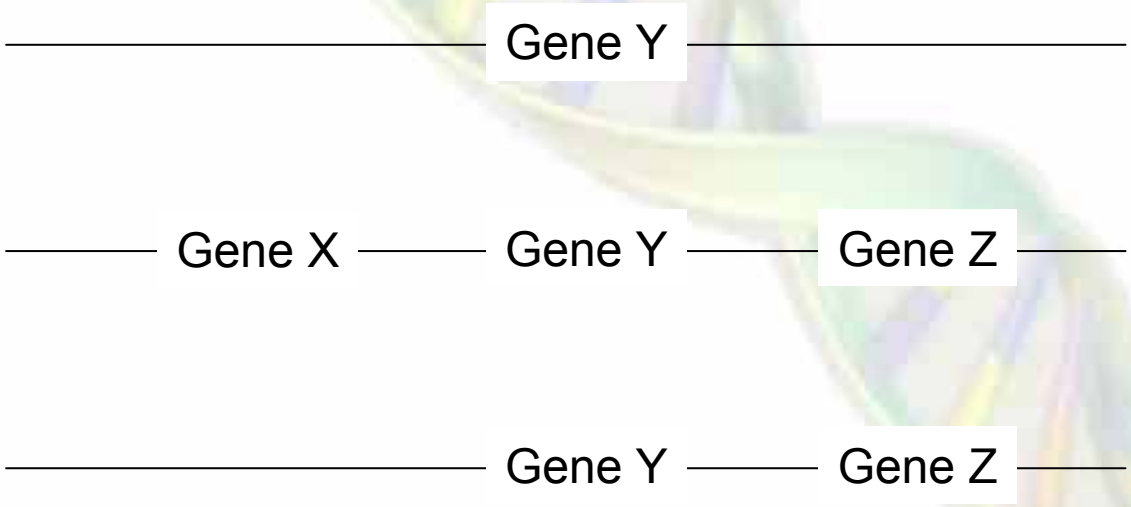
- The chromosomes are densely packed and tightly organised, why would not the genes be treated in the same way?

RIDGEs

- Hypothesised that they exist in eukaryotes
  - Same “idea” as operons; a group off adjacent genes are transcribed at the same time.
  - The reason is that since it is expensive to unwound the DNA going through all that trouble just to express ONE gene does not seem likely
  - Therefore, it is assumed that RIDGEs would consists of housekeeping genes
    - HK genes are generally expressed in all tissues at all times

Gene regulation

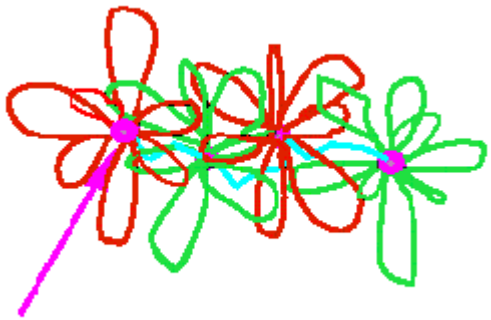
PROTEIN A

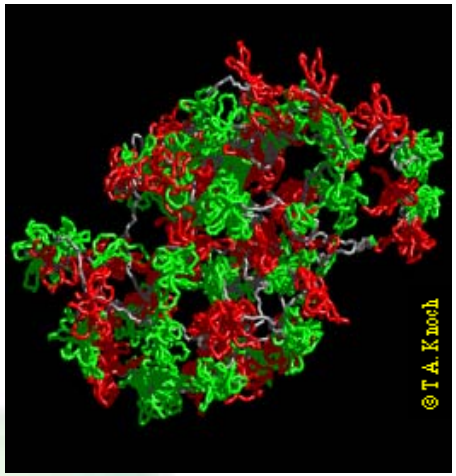


Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

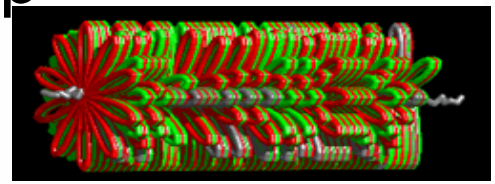
Sloppy gene regulation

- Rosette



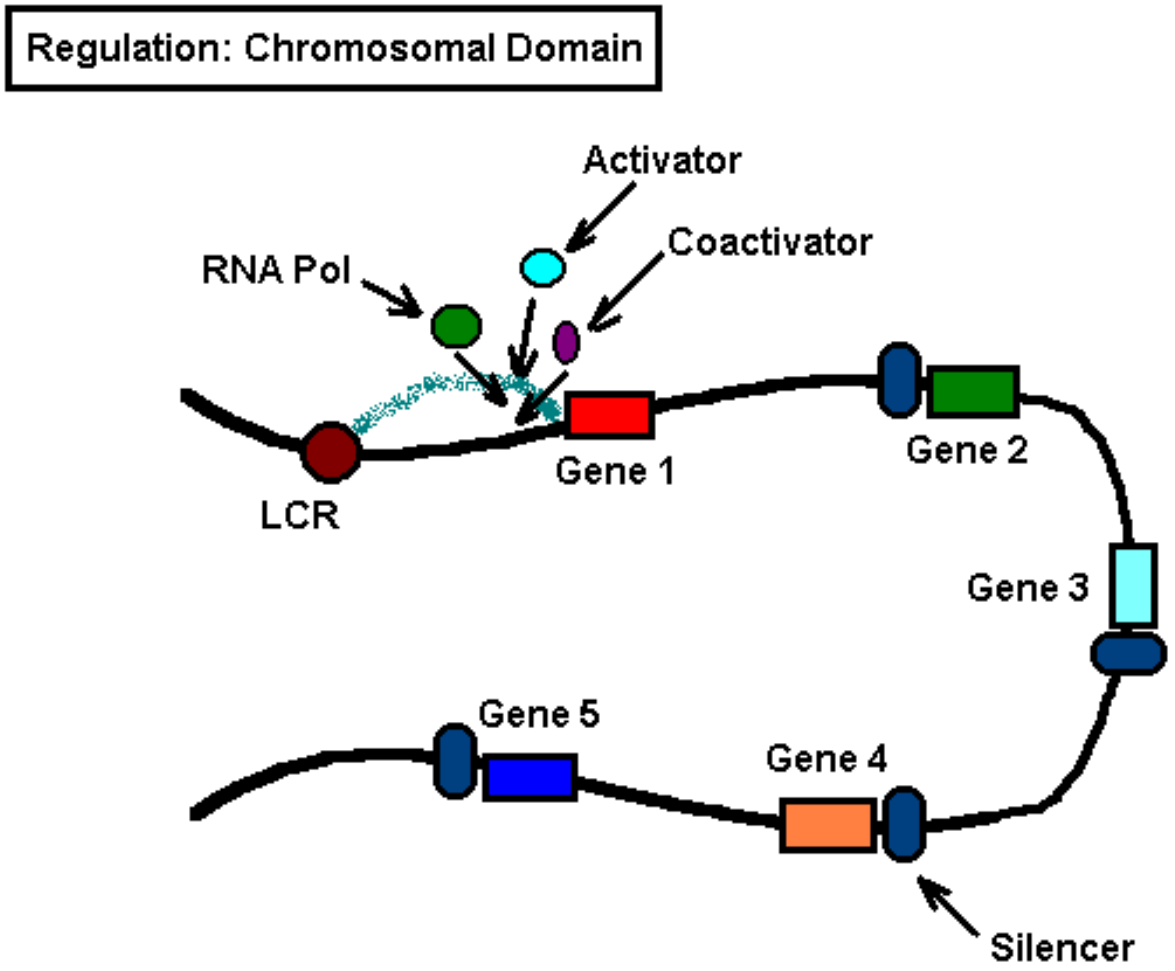


- One loop ~126 kbp



Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

RIDGE – Region of IncreaseD Gene Expression



Definition

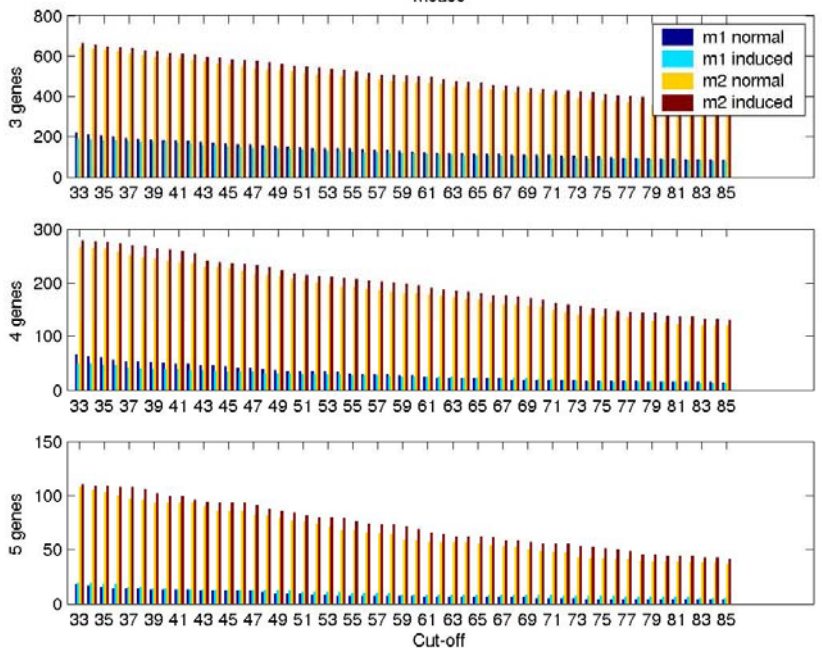
- A RIDGE is a number of consecutive genes (according to the midpoint concept) that have a correlated expression level
  - Should gaps be allowed?
- A RIDGE can also be **any** number of genes (starting from 1) that stretches a genetic distance about 126 kbp
  - Should gaps be allowed?

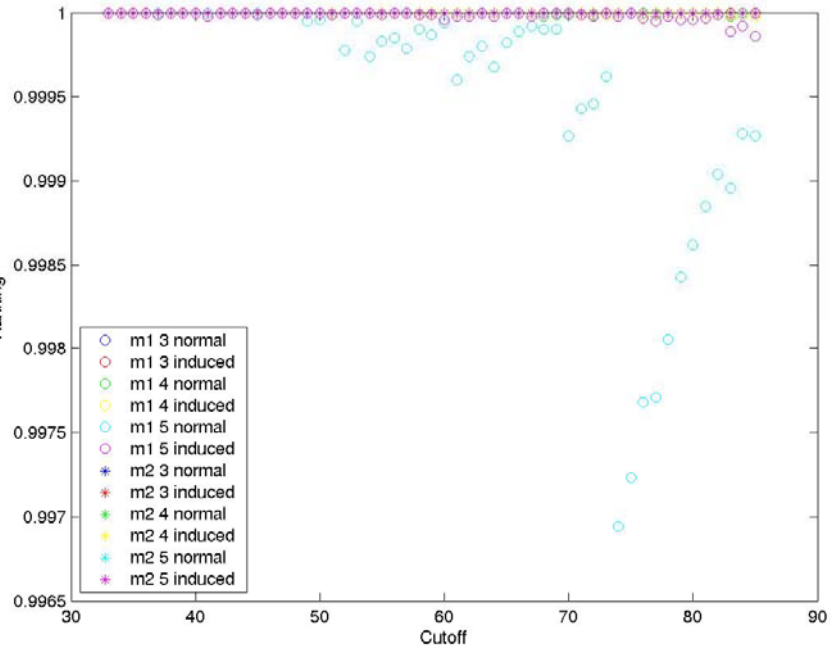
How?

- Determining whether or not RIDGEs exist in the genome
  - And if they do, what does that entail about protein expression levels
  - What changes in assumptions (if any) about gene and protein cascades do we have to make

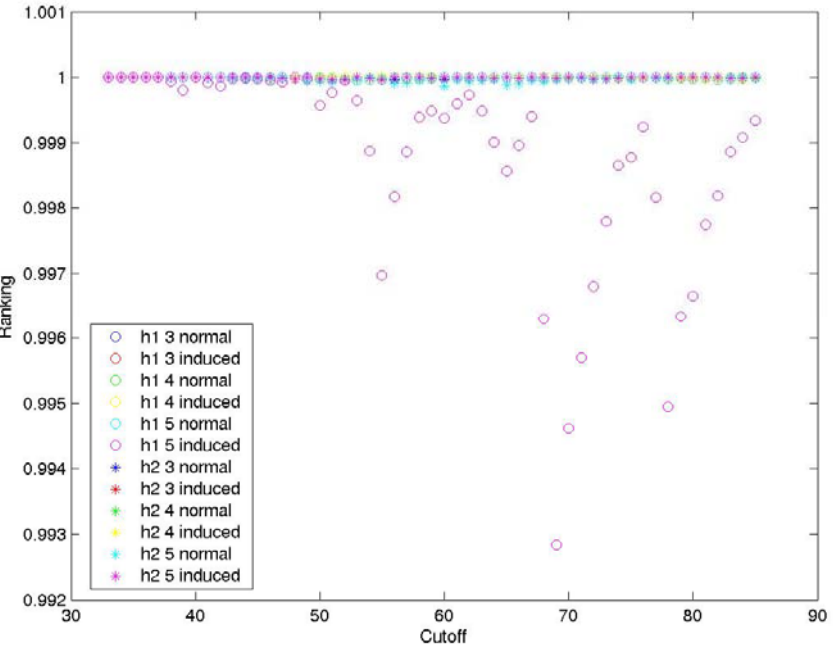
Results so far

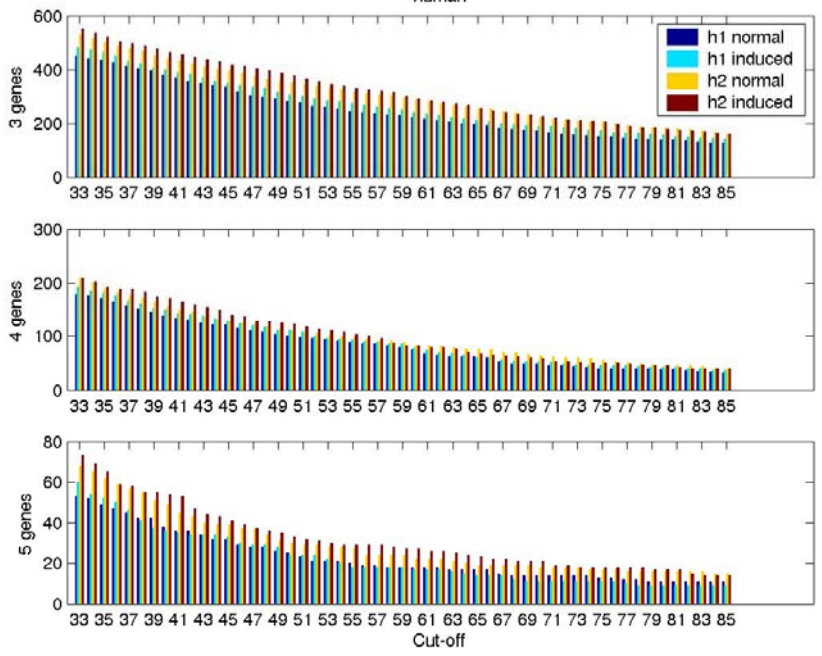
- Definition of a RIDGE
- I have tested the definition against two human and two mouse sets
  - The initial results seem to support the RIDGE hypothesis for mouse
  - Technically it supports it for human as well
  - Looking at the number of RIDGEs there is a difference, looking at the procentage of genes in a RIDGE, there is not
  - Housekeeping genes?





Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics





Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

Conclusions

- Can't make any
  - All we can say is that we can not refute RIDGEs with the current analysis
  - There might be a difference in the result corresponding to some parameters connected to the organism
    - It might be really interesting to find this difference
      - To do this, we have to run the analysis on multiple datasets and multiple species with different things in common, like genome size, like closely related, and not so closely related etc

What next?

- The specific case
  - “proof of concept”
- General case
  - ~ 1000 test cases

Lena Hansson: Lena.Hansson@ed.ac.uk
The School of Informatics @ The University of Edinburgh
Scottish Centre for Genomic Technology and Informatics

Acknowledgement

- The School of Informatics @ The University of Edinburgh
  - Douglas Armstrong
- The Scottish Centre for Genomic Technology and Informatics (GTI)
  - Peter Ghazal
  - The entire bioinformatics team