Lecture 1: Introduction to the Transcriptome

BIOINF3005/7160: Transcriptomics Applications

Steve Pederson

University of Adelaide

March 2nd, 2020

THE UNIVERSITY of ADELAIDE

Welcome to the University of Adelaide

- The start of any academic year is an exciting time - but it can also be challenging for our students, for many reasons
- For some of our students, this year is especially challenging
- That's because many of our students who wanted to be here to study can't be here - because of COVID-19 and the travel restrictions to Australia
- This is not a situation that those students - or any of us in this room today - have any direct control over
- We may have some students watching this very lecture from overseas

Welcome to the University of Adelaide

- Before the lecture begins, there are some points worth making so that everyone understands what the current situation is here at the University of Adelaide
- We have a diverse, inclusive, and welcoming community at our University, which is something we're very proud of
- In times of difficulty, that community is more important than ever.
- This is a time when our community comes together as a whole so we can care for and support each other together
- It's particularly important for those who may be going through a more difficult time than otherwise.
- We must be respectful of other people - and that includes the many students who couldn't join us in the room today

COVID-19

- The impact of COVID-19 is being felt right around the world.
- The University's own response to COVID-19 is being shaped by the latest advice from the Government and from health authorities.
- So far, there have only been a small number of COVID-19 cases in Australia ... only three in South Australia ... and none at the University.
- There is no evidence of transmission within the general community in Australia.
- Health authorities are advising we take the same basic precautions we would when it's cold or flu season – wash your hands, cover your mouth, dispose of tissues.
- This is good advice for any time you have a cough or a cold.

COVID-19

- There are resources available for students who need someone to talk to:
  - Student Life, if you need some extra assistance or support
  - Our Safer Campus Community website has a lot of resources to help us address harassment, including ways to report and get help
  - Adelaide UniCare is our on-campus medical service, and you can make appointments online.
- The University also has an online FAQ page about COVID-19 that's being constantly updated - there's a banner on our main webpage that links to the FAQ.
- The University has made it very clear that the safety and well-being of staff and students is paramount, and we will communicate with you clearly if anything changes.
- Welcome again, and good luck with your studies.

Course Timetable

- Lectures:
  - Monday 9:10am - 10:00am, Room B18, Ingkarni Wardli
- Practicals:
  - Wednesday 9:10am - 11:00am, Room 111, Johnson Building
  - Friday 9:10am - 11:00am, Room 111, Johnson Building

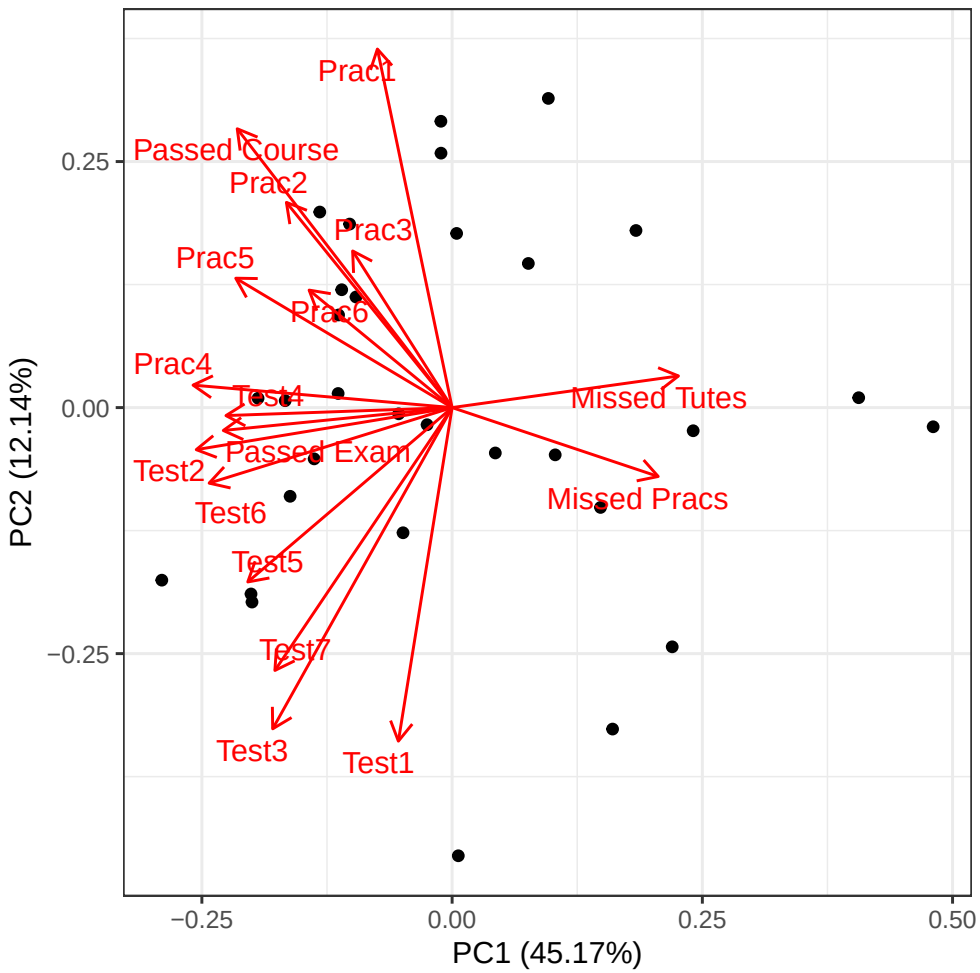
Course Outline

- Discussion of underlying biology
- Experimental approaches utilised in transcriptomic analysis
- Statistical and computational approaches utilised in transcriptomic analysis
- Praticals will be entirely computational
  - Focussed on working in `R` with some `bash`

Assessment

- Continuous Assessment (*i.e. no exam*)
- 6 assignments (+ Major Project for Bioinf7160)
- Assessment is strongly focussed on practical material
- Attendance at practicals is not compulsory, but is **strongly advised**

Results from 2019 (BIOTECH7005)



The Transcriptome

Definition

The transcriptome can be defined as the complete set of transcripts in a cell, or a population of cells, for a specific developmental stage or physiological condition[1]

[1] Z. Wang, M. Gerstein, and M. Snyder. "RNA-Seq: a revolutionary tool for transcriptomics". In: *Nat. Rev. Genet.* 10.1 (2009), pp. 57–63.

The Transcriptome

This can be summarised as the *RNA content of a cell*.

- Can include messenger RNA (*mRNA*), non-coding RNA (*ncRNA*), small RNA (*miRNA, piRNA etc.*)
- Can also include transfer RNA (*tRNA*) and ribosomal RNA (*rRNA*), but this is less common

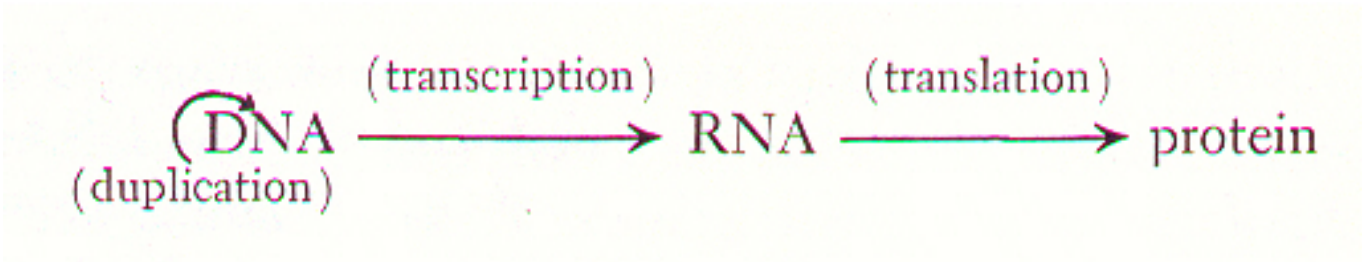
We are always dealing with a **snapshot of a dynamic process**

Why Study The Transcriptome?

- *mRNA* is the intermediary step between the genome (DNA) and proteins
- small RNA and *ncRNA* play significant roles in gene regulation
- Make inference about the **biological processes driving our observations**
  - Targets for curing disease
  - Biomarkers for tumour detection
  - Understanding drought/salinity tolerance

The Central Dogma of Molecular Biology

A common (but simplistic) version of the Central Dogma:

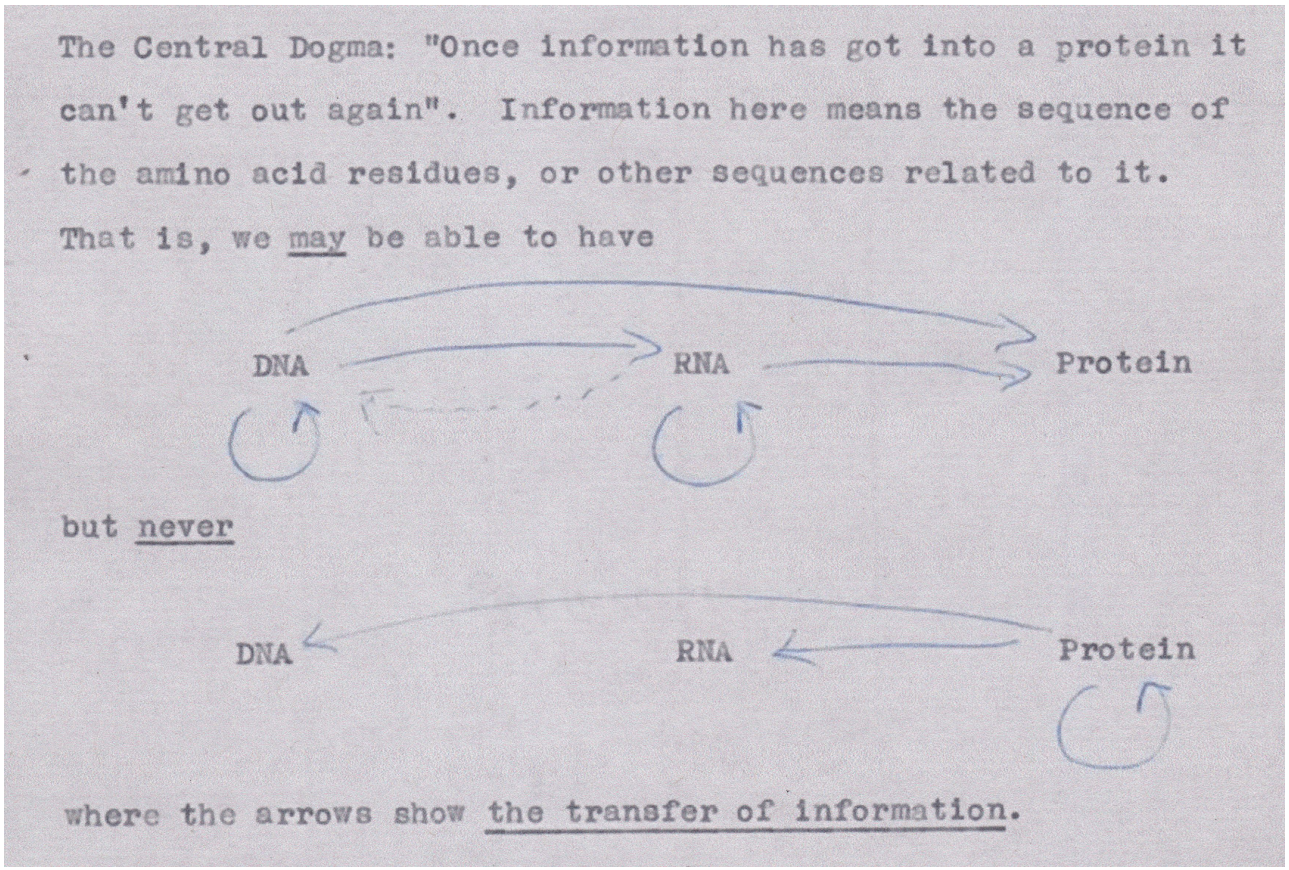


DNA makes RNA makes Protein. (Figure taken from *The Molecular Biology of the Gene*[2], p298)

[2] James Watson. *The Molecular Biology of the Gene*. W.A. Benjamin. Inc. new York, 1965.

The Central Dogma of Molecular Biology

According to Crick[3]



[3] Frances Crick. "Ideas on protein synthesis". In: *Unpublished Note*. Wellcome Library, 1956. URL: https://wellcomelibrary.org/item/b18174139.

What is a Gene?

Definition

The gene is the basic physical unit of inheritance. Genes are passed from parents to offspring and contain the information needed to specify traits. Genes are arranged, one after another, on structures called chromosomes. A chromosome contains a single, long DNA molecule, only a portion of which corresponds to a single gene. Humans have approximately 20,000 genes arranged on their chromosomes.[4]

[4] https://www.genome.gov/genetics-glossary/Gene

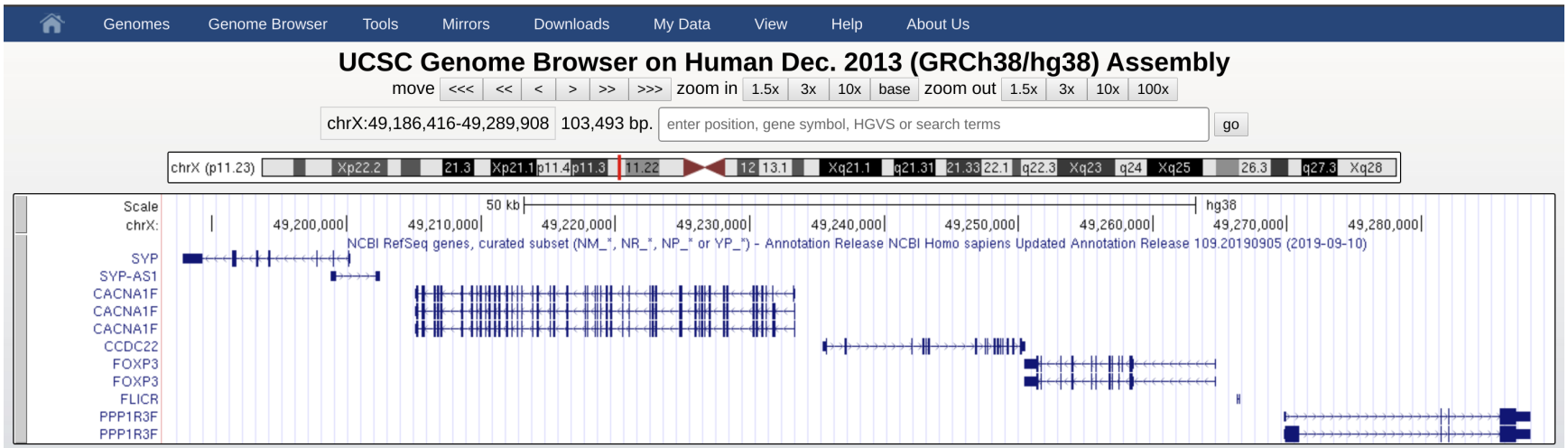
What is a Gene?

- Classically, a *gene* is a genomic locus that is **transcribed from DNA into RNA**[5]
- Some genes are protein coding, others are not
- A gene may have numerous *transcripts*, or *isoforms*
  - Some transcripts of a gene may be protein coding
  - Some transcripts from *the same gene* may not be protein coding
- Genes can range in size from dozens to millions of nucleotides

[5] NB: This definition does not include chimeric transcripts from more than one DNA locus

What is a Gene?

Here is an example region on the human X chromosome[6]

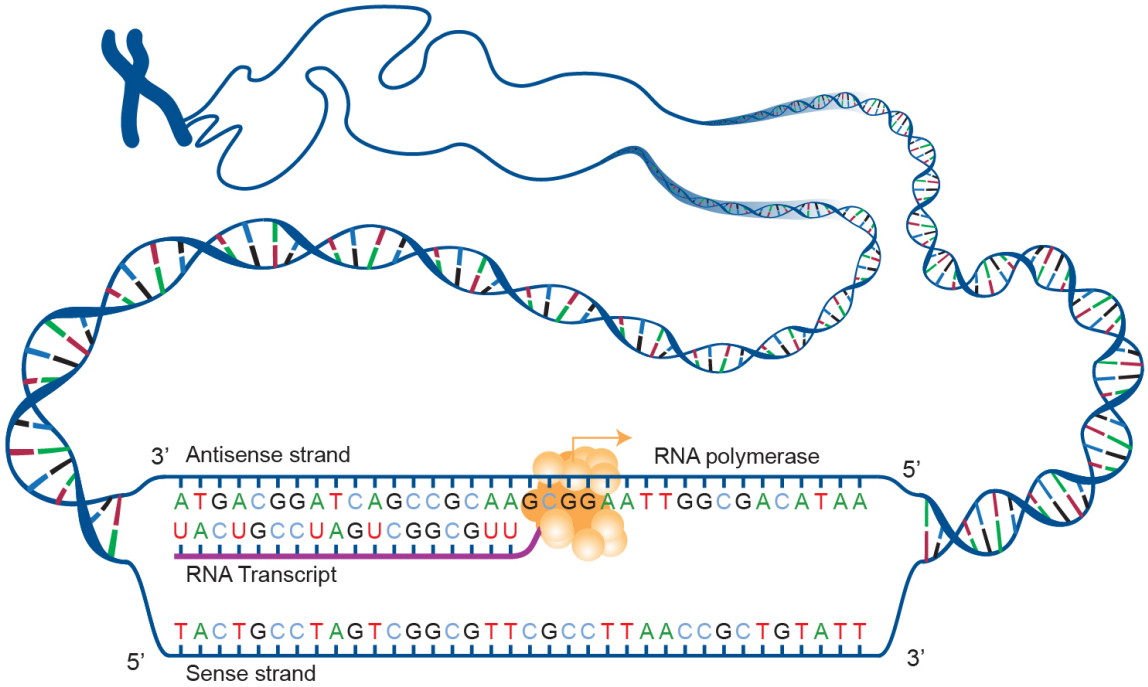


[6] UCSC Genome Browser

What is Transcription?

Transcription

Transcription is the process of making an RNA copy of a gene sequence[7]



[7] https://www.genome.gov/genetics-glossary/Transcription?id=197

The Steps of Transcription

1. RNA polymerase, together with one or more transcription factors, **binds to the promoter**
2. RNA polymerase creates a transcription bubble, which **separates the two strands of the DNA helix**, breaking the hydrogen bonds between complementary DNA nucleotides.
3. RNA polymerase **adds RNA nucleotides** complementary to the nucleotides of one DNA strand.
4. RNA sugar-phosphate backbone forms to create **single-stranded RNA**.
5. Hydrogen bonds of the RNA–DNA complex break, **freeing the newly synthesized RNA strand**.

The Steps of Transcription

If the cell has a nucleus (i.e. a *eukaryotic cell*):

6. The RNA may be further processed. This may include **polyadenylation**, **capping**, and **splicing**.
7. The RNA may remain in the nucleus or **exit to the cytoplasm** through the nuclear pore complex.

Before transcription even starts, the relevant section of DNA must be unpacked from any histones

What is a Promoter?

Definition

A promoter is a region of DNA that leads to initiation of transcription of a particular gene. Promoters are located near the transcription start sites of genes, upstream on the DNA (towards the 5' region of the sense strand). [8]

[8] https://en.wikipedia.org/wiki/Promoter_(genetics)

Key Elements of a Eukaryotic Promoter

The **core promoter** is the minimal portion of the promoter required to properly initiate transcription.

- Includes the transcription start site (TSS) and elements directly upstream
- A binding site for RNA polymerase
- General transcription factor binding sites, e.g. TATA box, B recognition element.
- Many other elements/motifs may be present. There is no such thing as a set of "universal elements" found in every core promoter.

Key Elements of a Eukaryotic Promoter

The **Proximal promoter** is the proximal sequence upstream of the gene that tends to contain primary regulatory elements

- Approximately 250 base pairs upstream of the start site
- Specific transcription factor binding sites

The **Distal promoter** is the distal sequence upstream of the gene often with a weaker influence than the proximal promoter

- Additional (weaker) regulatory elements,
- Anything further upstream (but not an enhancer or other regulatory region whose influence is positional/orientation independent)
- Specific transcription factor binding sites