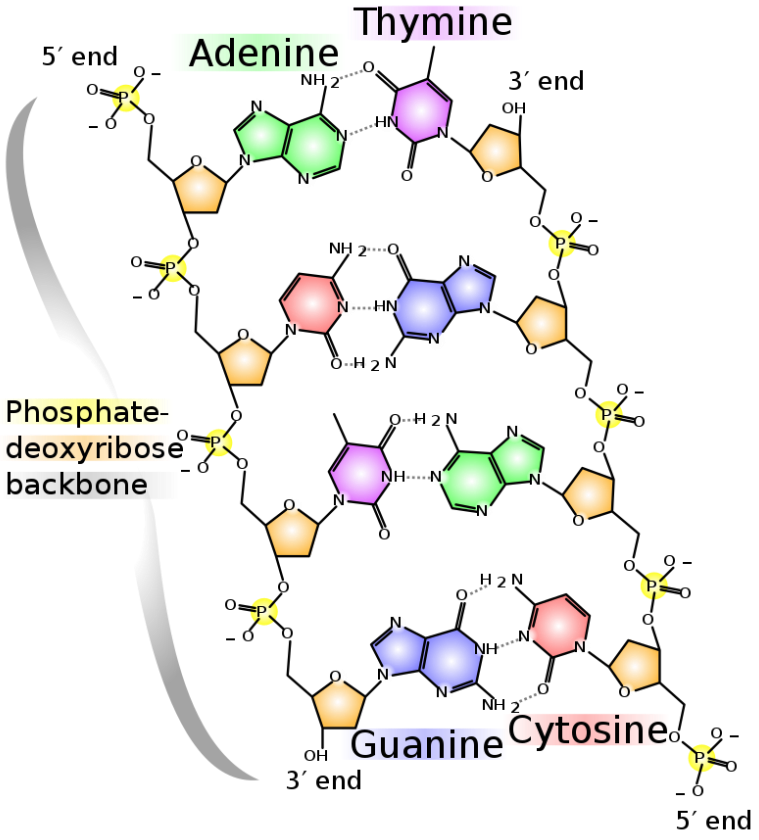
Gene Expression

SML201: Introduction to Data Science, Spring 2019

Michael Guerzhoy

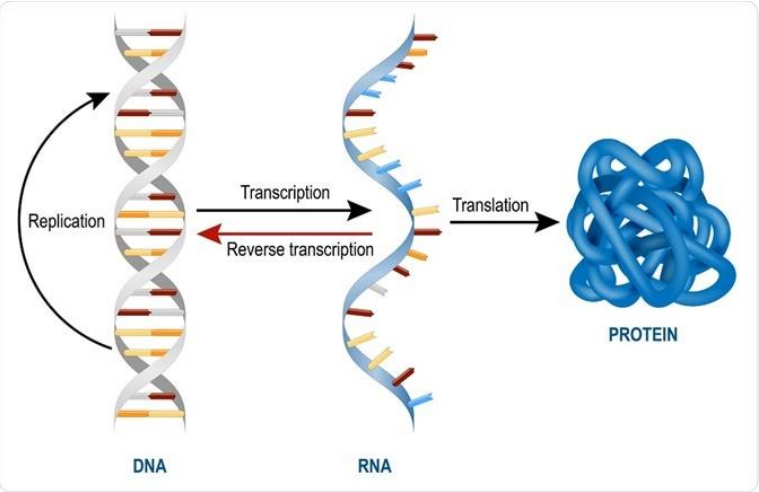
DNA

- A DNA molecule consists of two chains of nucleobases
- The same (give or take) in every cell in the body
- The human genome consists of over 3 billion base pairs (less than 1 gigabyte of data)
  - DNA differs between individuals (except for identical twins), but not by very much



RNA

- An RNA molecule is essentially a copy of a part of the DNA
- The ribosome synthesizes protein molecules using information encoded in the RNA



news-medical.net

Genes

- A gene is a sequence of nucleobases in DNA (or RNA) that produces a molecule with a particular function
  - E.g., the MC1R gene (i.e., the sequence in a particular location in DNA; MC1R is the name of the location) is mainly responsible for red hair
  - Things are complicated

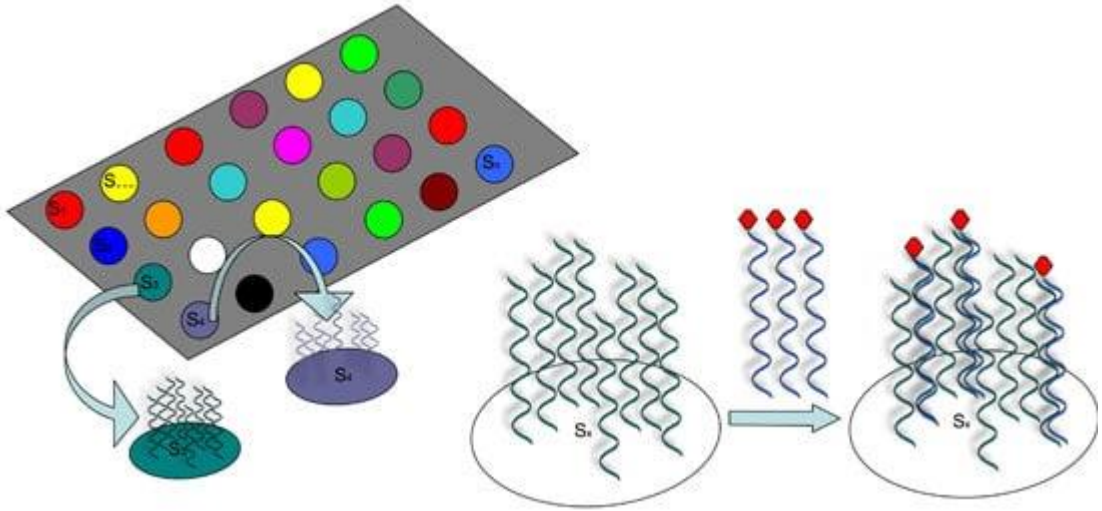
Gene expression

- Different cells in the organism do completely different things
- All genes are in the DNA in every cell, but not all of them result in RNA that then makes protein
- Gene expression is the amount of RNA molecules produced by a particular gene
  - More RNA molecules => the gene is doing something
- Gene expression regulation is hella complicated

Cancer

- Diseases involving abnormal cell growth
- Cell reproducing and/or producing proteins they're not suppose to produce
- Can try to measure gene expression to tell whether a cell is cancerous and/or what kind of cancer it is

Microarrays



bitesizebio.com/

A tool that measures the gene expressions for thousands of different genes simultaneously by detecting RNA molecules