RNA Synthesis and Processing
Overview

Genome structure
Mendelian genetics
Linkage and recombination
Chromosome Painting
Chromosome structure

Short arm: p

long arm: q
A chromosome contains two (almost) identical copies of DNA molecules. Each copy is called a chromatid and two chromatids are joined at their centromeres.

Chromosomes and regions in a chromosome are named. For example by notation 6p21.3, we mean

- 6 : chromosome number
- p : short arm (from petit in French), q for long arm
- 21.3 : band 21, sub-band 3 (microscope)
Chromosomes in Cell Division
Chromosome Structure
Chromosome structure

- Chromosome of $1.5 \times 10^8$ nucleotide pairs, containing about 3000 genes.
- 0.5% of chromosome, containing 15 genes.
- One gene of $10^5$ nucleotide pairs.
- Regulatory DNA sequences.

DNA TRANSCRIPTION

Primary RNA transcript

RNA SPlicing

mRNA

From The Art of MRnC³ © 1995 Garland Publishing Inc
Mendel started genetics research before we knew about chromosome and gene.

*Phenotype* -- observable difference among members in a population

- For example: hair color, eye color, blood type

What controls a phenotype?

- This is the question that Mendel tried to answer
  - Is still the central question of modern genetics

He used pea, a simple organism, and quantitative method to study phenotypes.

- We call a quantitative study of biology *computational biology* now.
Gene linkage

Experiments: white eyed female flies cross with wild type males (red eye, dominant)

What should we expect:
  ▪ All red eyed flies in F1

No! We have 50% red eyed and 50% white eyed
  ▪ Why?
    ▪ A further study show that all red eyed are female and all white eyed are male
Gene Linkage:

Morgan's explanation:

\[
\frac{W}{W} \quad X \quad \frac{\pm}{Y}
\]

Female, red eye

Male, while eye
Recombination

F1 generation: white eyed, miniature wing female crossed to wild type male

F2 generation (ie. heterozygous female crossed to a wm male.

We know that white eye gene and the miniature gene are in the same chromosome

What should we expect:
- female: 50% wm, 50% normal
- male: 50% wm, 50% normal
Recombination

What we have: F2 generation (ie. heterozygous female crossed to a wm male):

- white eyes, normal wings 223
- red eyes, miniature wings 247
- red eyes, normal wings 395
- white eyes, miniature wings 382

We have some combination that does not appear in parents!
Recombination

chromtides exchange their DNA components.
Forward genetics: given a phenotype, how do we identify the genes that contribute to the phenotype?
  ▪ Cancer, Parkinson's Disease,…

Reverse genetics: given a gene, how could we know what are the phenotype that the gene might control?
  ▪ Diagnostics
How are Phenotypes Caused?

Through a biochemical pathways:

A \overset{W}{\rightarrow} B \overset{X}{\rightarrow} C \overset{Y}{\rightarrow} D \overset{Z}{\rightarrow} E

Genes, mRNA, and proteins
Chromosome structure

chromosome of $1.5 \times 10^8$ nucleotide pairs, containing about 3000 genes

0.5% of chromosome, containing 15 genes

one gene of $10^5$ nucleotide pairs

regulatory DNA sequences

intron

exon

DNA TRANSCRIPTION

primary RNA transcript

RNA SPLICING

mRNA

exon sequence

intron sequence
Chromosome Structure

Chromosomes are made up of a single continuous DNA molecule can’t be straight-- would be many times length of the cell

Chromatin: ordered aggregate of DNA and proteins visible in the cell cycle
heterochromatin: dense, compact structure during interphase generally near the centromere and **telomeres** (chromosome ends) composed of long tracks of fairly short base pair repeats few genes compared to euchromatin

euchromatin: less dense DNA that only becomes visible after condensing typically has genes being actively transcribed
Gene Expression

- genes are located at specific places on the chromosome (loci)
- regions corresponding to genes are transcribed
- sequences flanking the coding sequence control the start and stop of transcription
- not all genes are expressed at the same rates or at the same times
RNA Synthesis and Processing

• Transcription in Prokaryotes
• Eukaryotic RNA Polymerases and General Transcription Factors
• Regulation of Transcription in Eukaryotes
• RNA Processing and Turnover
Regulation of gene expression allows cells to adapt to environmental changes and is responsible for the distinct activities of the differentiated cell types that make up complex organisms.
RNA polymerase catalyzes polymerization of ribonucleoside 5′-triphosphates (NTPs) as directed by a DNA template, always in the 5′ to 3′ direction.

Transcription initiates de novo (no preformed primer required) at specific sites—this is a major step at which regulation of transcription occurs.
Figure 7.5  Structure of bacterial RNA polymerase
Figure 7.1 E. coli RNA polymerase
Figure 7.2 Sequences of *E. coli* promoters
Figure 7.4 Transcription by *E. coli* RNA polymerase (Part 1)

- Polymerase bound nonspecifically to DNA
- Closed-promoter complex
- Open-promoter complex

Specific binding of σ to -35 and -10 promoter sequences
Unwinding of DNA around the initiation site
Initiation of transcription
Figure 7.4  Transcription by *E. coli* RNA polymerase (Part 2)

Release of $\sigma$

Elongation of RNA chain

*THE CELL 5e, Figure 7.4 (Part 2)*
Transcription of the GC-rich inverted repeat results in a segment of RNA that can form a stable stemloop structure. This disrupts its association with the DNA template and terminates transcription.
Figure 7.6 Transcription termination

Formation of stem loop

Dissociation of RNA from DNA template

THE CELL 5e, Figure 7.6
Figure 7.8 Negative control of the lac operon
Cis-acting control elements only affect the expression of linked genes on the same DNA molecule (e.g. the operator).

Trans-acting factors can affect expression of genes located on other chromosomes (e.g. the repressor).

The lac operon is an example of negative control—binding of the repressor blocks transcription.
An example of positive control in *E. coli*: Presence of glucose (the preferred energy source) represses expression of genes for enzymes that break down other sugars, such as the *lac* operon.
Low glucose levels activate adenylyl cyclase, which converts ATP to cAMP. cAMP then binds to catabolite activator protein (CAP).

CAP then binds to its target DNA sequences, 60 bases upstream of the transcription start site in the *lac* operon.
Figure 7.9 Positive control of the lac operon by glucose
Eukaryotic cells have three nuclear RNA polymerases that transcribe different classes of genes.

They are complex enzymes, consisting of 12 to 17 different subunits each.

They all have 9 conserved subunits, 5 of which are related to subunits of bacterial RNA polymerase.
### TABLE 7.1 Classes of Genes Transcribed by Eukaryotic RNA Polymerases

<table>
<thead>
<tr>
<th>Type of RNA synthesized</th>
<th>RNA polymerase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear genes</td>
<td></td>
</tr>
<tr>
<td>mRNA</td>
<td>II</td>
</tr>
<tr>
<td>miRNA</td>
<td>II</td>
</tr>
<tr>
<td>tRNA</td>
<td>III</td>
</tr>
<tr>
<td>rRNA</td>
<td></td>
</tr>
<tr>
<td>5.8S, 18S, 28S</td>
<td>I</td>
</tr>
<tr>
<td>5S</td>
<td>III</td>
</tr>
<tr>
<td>snRNA and scRNA</td>
<td>II and III&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mitochondrial genes</td>
<td>Mitochondrial&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Chloroplast genes</td>
<td>Chloroplast&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Some small nuclear (sn) and small cytoplasmic (sc) RNAs are transcribed by polymerase II and others by polymerase III.

<sup>b</sup>The mitochondrial and chloroplast RNA polymerases are similar to bacterial enzymes.
Figure 7.10 Structure of yeast RNA polymerase II
Promoters contain several different regulatory sequence elements.

Promoters of different genes contain different combinations of promoter elements, which appear to function together to bind general transcription factors.
Figure 7.13 RNA polymerase II/Mediator complexes and transcription initiation

Phosphorylation of CTD
Initiation of transcription

CTD

Elongation and processing factors

RNA

THE CELL 5e, Figure 7.13
Figure 7.14 The ribosomal RNA gene

- rDNA
- Promoter
- 18S
- 5.8S
- 28S

Transcription

45S pre-rRNA

Processing

rRNAs

18S
5.8S
28S
Some regulatory sequences are farther away—called **enhancers**.

They were first identified during studies of the promoter of another virus, SV40.
Figure 7.19 The SV40 enhancer
Activity of enhancers doesn’t depend on either their distance from, or orientation with respect to, the transcription initiation site.
Figure 7.20 Action of enhancers (Part 1)

(A) 

Basal transcription

P Gene

(B) 

Stimulated transcription

E Gene

(C) 

Stimulated transcription

E Gene
Chromatin immunoprecipitation identifies DNA regions that bind to transcription factors.

Cells are treated with formaldehyde to cross-link transcription factors to the DNA sequences to which they were bound.

Chromatin is extracted and fragmented. Fragments of DNA linked to a transcription factor can then be isolated by immunoprecipitation.
Figure 7.25 Chromatin immunoprecipitation (Part 1)

1. Treat cells with formaldehyde
2. Sonicate to produce fragments of chromatin
3. Fragments of chromatin with transcription factors cross-linked to DNA
4. Immunoprecipitate with antibody against transcription factor of interest
Antibody binds specific transcription factor

Collect chromatin-antibody complex

Reverse cross-links
Purify DNA

DNA fragment containing specific transcription factor binding site
One of the first transcription factors to be isolated was Sp1, in studies of virus SV40 DNA, by Tjian and colleagues.

Sp1 binds to GC boxes in the SV40 promoter. This established the action of Sp1 and also suggested a method for purification of transcription factors.
Key Experiment 7.1 Isolation of a Eukaryotic Transcription Factor: Purification of Sp1

THE CELL 5e, Key Experiment 7.1
Figure 7.28  Examples of DNA-binding domains (Part 2)

(C) **Leucine zipper**

- Leucine side chain
- Leucine zipper region
- DNA-binding helix

(D) **Helix-loop-helix**

- Helix
- Loop
- DNA-binding helix
The most common is the **zinc finger domain**, which binds zinc ions and folds into loops ("fingers") that bind DNA.

**Steroid hormone receptors** contain zinc fingers; they regulate gene transcription in response to hormones such as estrogen and testosterone.
**Helix-turn-helix** domain: one helix makes most of the contacts with DNA, the other helices lie across the complex to stabilize the interaction.

They include **homeodomain** proteins, important in the regulation of gene expression during embryonic development.
Homeodomain proteins were first discovered as developmental mutants in *Drosophila*. They result in development of flies in which one body part is transformed into another.

In *Antennapedia*, legs rather than antennae grow from the head.
Figure 7.29 The *Antennapedia* mutation
Leucine zipper and helix-loop-helix proteins contain DNA-binding domains formed by dimerization of two polypeptide chains.

Different members of each family can dimerize with one another—combinations can form an expanded array of factors.
Regulation of Transcription in Eukaryotes

The activation domains of transcription factors are not as well characterized as their DNA-binding domains.

Activation domains stimulate transcription by two mechanisms:

• Interact with Mediator proteins and general transcription factors
• Interact with coactivators to modify chromatin structure.
Figure 7.30  Action of transcriptional activators

DNA-binding domain

Activation domain

Interactions with Mediator and general transcription factors

Modification of chromatin structure
Gene expression in eukaryotic cells is also regulated by repressors which bind to specific DNA sequences and inhibit transcription.

In some cases, they simply interfere with binding of other transcription factors.

Other repressors compete with activators for binding to specific regulatory sequences.
Figure 7.33  Decondensed chromosome regions in *Drosophila*
Histone modification provides a mechanism for epigenetic inheritance—transmission of information that is not in the DNA sequence.

Modified histones are transferred to both progeny chromosomes where they direct similar modification of new histones—maintaining characteristic patterns of histone modification.
Figure 7.36 Epigenetic inheritance of histone modifications (Part 1)

**DNA replication**

- Parental chromatin
- Incorporation of new nucleosomes

*THE CELL 5e, Figure 7.36 (Part 1)*
Figure 7.36  Epigenetic inheritance of histone modifications (Part 2)
Chromatin remodeling factors are protein complexes that alter contacts between DNA and histones. They can reposition nucleosomes, change the conformation of nucleosomes, or eject nucleosomes from the DNA.
Figure 7.37 Chromatin remodeling factors

1. Chromatin remodeling factor
2. Binding of transcriptional activator and chromatin remodeling factor
3. Nucleosome displacement
4. Binding of general transcription factors and RNA polymerase
Transcription can also be regulated by noncoding RNA molecules, including small-interfering RNAs (siRNAs) and microRNAs (miRNAs).

They can induce histone modifications that lead to chromatin condensation and formation of heterochromatin.
In the yeast *S. pombe*, siRNAs direct formation of heterochromatin at centromeres, by associating with the RNA-induced transcriptional silencing (RITS) complex.

RITS includes proteins that induce chromatin condensation and methylation of histone H3 lysine-9.
Figure 7.38 Regulation of transcription by siRNAs

siRNA
Association with RITS Unwinding of siRNA
RITS
Pairing with mRNA transcript at target gene
mRNA
siRNA
RITS
Methylation of H3 lysine-9 Heterochromatin formation
Transcription repressed
**DNA methylation** is another general mechanism that controls transcription in eukaryotes.

Methyl groups are added at the 5-carbon position of cytosines (C) that precede guanines (G) (CpG dinucleotides).
Figure 7.40  DNA methylation

Cytosine

\[ \text{Methylation} \]

5-Methylcytosine

THE CELL 5e, Figure 7.40
Methylation is common in transposable elements, it plays a key role in suppressing the movement of transposons.

DNA methylation is associated with transcriptional repression of some genes, and also has a role in X chromosome inactivation.
DNA methylation is a mechanism of epigenetic inheritance.

Following DNA replication, an enzyme methylates CpG sequences of a daughter strand that is hydrogen-bonded to a methylated parental strand.
Figure 7.41 Maintenance of methylation patterns

Methylated parental DNA

DNA replication

Methylation

Methylated daughter DNAs
DNA methylation also plays a role in **genomic imprinting**: expression depends on whether they are inherited from the mother or from the father.

Example: gene $H19$ is transcribed only from the maternal copy. It is specifically methylated during the development of male, but not female, germ cells.
Most newly-synthesized RNAs must be modified, except bacterial RNAs which are used immediately for protein synthesis while still being transcribed.

rRNAs and tRNAs must be processed in both prokaryotic and eukaryotic cells.

Regulation of processing steps provides another level of control of gene expression.
The ribosomal RNAs of both prokaryotes and eukaryotes are derived from a single long pre-rRNA molecule.

5S rRNA in eukaryotes is transcribed from a separate gene.
Figure 7.43 Processing of ribosomal RNAs

**Prokaryotes**

Pre-rRNA (5.5 kb) → 16S (1.5 kb) → Mature rRNAs

16S 23S 5S

16S (5.5 kb) → 23S (2.9 kb) → 5S (0.12 kb)

**Eukaryotes**

Pre-rRNA (13 kb) → 18S (1.9 kb) → Mature rRNAs

18S 5.8S 28S

18S (1.9 kb) → 5.8S (0.16 kb) → 28S (5 kb)
tRNAs also start as long precursor molecules (pre-tRNAs), in both prokaryotes and eukaryotes.

Processing of the 5’ end of pre-tRNAs involves cleavage by the enzyme RNase P.

RNase P is a ribozyme—an enzyme in which RNA rather than protein is responsible for catalytic activity.
RNA editing is processing (other than splicing) that alters the protein-coding sequences of some mRNAs.

It involves single base modification reactions such as deamination of cytosine to uridine and adenosine to inosine.
Ultimately, RNAs are degraded in the cytoplasm.

Intracellular levels of any RNA are determined by a balance between synthesis and degradation.

Rate of degradation can thus control gene expression.
Bacterial mRNAs are rapidly degraded, most have half-lives of 2 to 3 minutes.

Rapid turnover allows the cell to respond quickly to changes in its environment, such as nutrient availability.
In eukaryotic cells, mRNA half-lives vary; less than 30 minutes to 20 hours in mammalian cells.

Short-lived mRNAs code for regulatory proteins, levels of which can vary rapidly in response to environmental stimuli.

mRNAs encoding structural proteins or central metabolic enzymes have long half-lives.
Degradation of eukaryote mRNAs is initiated by shortening of the poly-A tails.

Rapidly degraded mRNAs often contain specific AU-rich sequences near the 3’ ends which are binding sites for proteins that can either stabilize them or target them for degradation.
Figure 7.56  mRNA degradation

5′ m^7G  A A A A  3′

5′ UTR  3′ UTR

Deadenylation

Decapping

3′ → 5′ degradation

5′ → 3′ degradation

THE CELL 5e, Figure 7.56
These RNA-binding proteins are regulated by extracellular signals, such as growth factors and hormones.

Degradation of some mRNAs is regulated by both siRNAs and miRNA.