Genes and Chromosomes

- Date: August 15, 2005 *
- Time: 10:00 am - 10:50 am *
- Room: G-202 Biomolecular Building
- Lecturer: Dale Ramsden
  32-044 Lineberger
dale_ramsden@med.unc.edu
  966-9839

- *Please consult the online schedule for this course for the definitive date and time for this lecture.

- Office Hours: by appointment

- Assigned Reading: This syllabus.
Important terms are bolded. Illustrative and supplementary information (rare) is italicised.

Basic Principles: For those of you with little background in molecular biology, a review of basic principles of nucleic acid structure and metabolism will be presented (GUTS, or Get Up to Speed session) immediately before this lecture, on August 15 between 9:00 and 9:50 PM

Overall objectives for “The genome, and genomic instability”
Genome Maintenance I: Genes and Chromosomes: August 15th
Genome Maintenance II: Genome Replication: August 17th
Genomic Instability I: DNA damage and repair: August 18th
Genomic Instability II: Recombination: August 19th
This section consists of four lectures: In the the first two you will learn how to define the genome, and describe how genomes are maintained. In the second two lectures, you will learn why genomes are unstable, the consequences of genomic instability, and how cells cells mitigate the risk of genomic instability.

Lecture objectives:
At the end of this lecture, you should have an idea
a) how your genome is different from your genes
b) how your genome is organized into chromosomes
c) how chromosomes are maintained
I. General considerations.

The genome is the genetic complement of an organism. All cells of all individuals of a given species have roughly the same genetic complement. There are some obvious and important exceptions (sometimes referred to as “genomic instability”) - two lectures worth! Before cells divide, they must therefore duplicate their genetic material (replication; see next lecture) so that each daughter cell also has a full genome. The amount of DNA that encodes genes is often profoundly less than the total genome size, and we will discuss some reasons for this (“ploidy”, repetitive DNA).

The genome is stored in DNA, in chromosomes, defined as a single molecule of DNA and its associated proteins. We will discuss several important mechanisms exist to facilitate the stable maintenance of chromosomes in cells (centromere, telomere, chromatin).

II. Organization of genomes.

A major division in the way in which genomes are organized: prokaryotes (bacteria) vs. eukaryotes (almost everything else).

A. The prokaryotic genome.

Prokaryotic genomes:
- small ($10^5$-$10^6$ base pairs)
- simple: Genes and proteins in prokaryotes are generally co-linear; the gene is simply the linear, triplet code required to make the protein.
- typically one chromosome, and one copy of that chromosome.
- The Chromosome is often circular.

B. The eukaryotic genome.

- very large genomes ($10^7$ [yeast] to $10^9$ [human] base pairs), in the nucleus
- large expanses of DNA with no obvious purpose (junk DNA?) between the genes (intervening sequences) and interrupting the gene (introns; see transcription lecture). Less than 5% of the human genome encodes proteins.
- Organisms that do sex have a duplicated genome after reproduction. A single copy of the genome, “haploid”, is found in sperm and oocytes. After fusion of sperm/oocyte, you have two copies of each chromosome – one from the sperm (paternal) and one from the oocyte (maternal). This is termed diploid. All cells except germ cells (sperm and oocyte) in most sexually reproducing species are diploid.
- Organized into multiple linear chromosomes (yeast have 16, humans have 24 different ones (Chromosomes 1 through 22, X, and Y)
The primary genome is found in the nucleus, but eukaryotes also have small, circular, prokaryotic-like chromosomes with a different set of genes in organelles (mitochondria, chloroplasts).

**Repetitive sequences; satellites and micro-satellites.**

Eukaryotes have large tracts of repetitive DNA that can span thousands of base pairs. Repetitive DNA is usually between genes, but not always. Some regions of repeated sequence have a special function; see telomeres and centromeres below. The majority of repetitive sequence has no known function.

The unit of repeated sequence can range from a single nucleotide to several 100 nucleotides. Regions of repetitive sequence are often called satellites (e.g. the centromeric alpha satellite); repeats of three nucleotides (triplet) or smaller are termed microsatellites. The number of repeats in some regions of repetitive sequence can be highly variable between individuals (these regions are sometimes termed Variable number tandem repeats, or VNTRs). This variation can be employed as a means of identifying individuals with much higher confidence than other techniques (say fingerprinting) (covered in greater detail by Dr. Lee).

**Microsatellite instability causes disease.**

Microsatellites are frequently not stable; the number of copies can increase or decrease after replication (lecture on mismatch repair will discuss why this happens) this is termed microsatellite instability. The expansion and contraction of microsatellites can result in disease. E.g. the triplet repeat diseases: Huntington’s Chorea, Freiderich’s ataxia, Myotonic Dystrophy, fragile X syndrome; there are many more. In some of these diseases, but not all, microsatellite instability is observed in the coding sequence of the disease causing gene.

Huntington’s Chorea is one of several neurological diseases caused by expansion of a poly-glutamine encoding stretch in a gene. Shown below is the first 100 nucleotides or so of coding sequence of the huntingtin gene from a normal person; DNA sequence on top, translated amino acid sequence in bold/italics below. Note 21 copies of the CAG microsatellite, which encodes glutamine (single letter aa code=Q).

```
  atg gcc acc ctg gaa aag ctg atg aag gcc ttc gag tcc ctc aag tcc ttc cag cag cag cag
  M   A   T   L   E   K   L   M   K   A   F   E   S   L   K   S   F   Q   Q   Q   Q
  cag cag cag cag cag cag cag cag cag cag cag cag cag cag cag cag cag caa cag ccg cca……
  Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   P   P…. 
```

People with Huntington’s Chorea have 35 or greater CAG repeats. Disease severity/age of onset often reflects the number of repeats. 500 repeats: very severe, early onset. <40 repeats: often subclinical, late onset.

Diseases of microsatellite instability show genetic anticipation. A person with a slightly higher than normal number of copies (say, 30 copies) will likely not get huntingtons disease. However, there is a much greater chance than normal that this microsatellite will expand to a disease causing level in the next generation.
II. Centromeres and Telomeres

There are three minimal requirements for proper maintenance of chromosomes in cells.

1) A centromere, to mediate segregation during cell division.

2) A telomere, to protect chromosome ends.

3) Origin (s) of replication.

Centromeres and Telomeres are both repetitive DNA elements that clearly can’t be called “junk” DNA. Origins of replication will be discussed in the next lecture.

Figure 1. Chromosome immediately before mitosis
A. Centromeres

Function.
Immediately before cell division, replication (see replication lecture) generates two copies of each chromosome (this makes 4 copies total; e.g. for a given chromosome, say chromosome 2, you would have two copies of your paternal chromosome 2, 2 copies of your maternal chromosome 2).

Each copy is termed a **chromatid** until the copies are segregated into different cells. When the cell divides into two daughter cells, each daughter cell ends up having one, and only one, of each of the copies; once again, one maternal set and one paternal set of chromosomes. This is controlled by a highly organized segregation of chromatids during the **mitosis** phase of cell division. Chromatids are joined together at a specific region along the length of the chromosome termed the **centromere**.

The chromosome condenses….and then **mitosis** happens…which includes a lot of stuff that you learn about in detail in Dr. Burridge’s lecture called “mitosis and cell cycle”. For the purposes of this course, understand that during cell division centromeres are a region of the chromosome where the **kinetochore** attaches. The kinetochore is a protein machine that mediates the following events:

1) The duplicated chromosomes are aligned along a plane in the middle of the cell  
2) Chromosomes are then split into their individual chromatid copies  
3) One of each copy is dragged to the two poles of the cell. This is mediated by the kinetochore attached to the centromere, so the centromere is the basis for segregation, and the rest of the chromosome just follows the lead of the centromere.

A diagram of a mitotic cell can be found in the Cell biology Mitosis lecture; a microscope picture of the analogous cell is shown below.

![Centromeres](image)

**Figure 2.** Cells in onion root tip during cell division. Note condensed state of chromosomes.

Then the cells divide (cytokinesis; see Burridge lecture). If segregation isn’t done properly, you might wind up with the wrong number of chromosomes in each daughter cell. For example, if
segregation of chromosome 3 screws up, in one cell you could have two paternal copies of chromosome 3 and one maternal copy of chromosome 3 (3 copies overall, and 47 chromosomes total); in the other daughter cell you would have only one maternal copy, and NO paternal copies of chromosome 3 (45 chromosomes total). Cells with the wrong number of chromosomes are termed **aneuploid**; aneuploidy is commonly seen in cancers.

**Structure.**
The placement of centromeres relative to the length of the chromosome is different for each chromosome. Centromeres can be in the center of the chromosome (as the name would suggest), or towards an end. Centromeres consist of some defined sequence elements, but are mostly less well defined repetitive sequence. A minimal centromere in yeast can be defined that is 120 base pairs long. However, in multicellular eukaryotes, centromeres typically occupy several hundred thousand base pairs, and include the alpha satellite repetitive sequence.

Although the exact DNA sequence of a centromere is fairly variable from one chromosome to another, chromosome centromeres nevertheless are functionally equivalent. They are readily interchangeable between chromosomes; e.g. the X chromosome centromere could be swapped with the Y chromosome centromere with no ill effect.

Centromeres maintain a condensed, heterochromatin (see chromatin, below) structure throughout the life of the cell. Nucleosomes in centromeres possess the histone H3 variant CENP-A, which helps direct kinetochore attachment.

**B. Telomeres.**

**Structure and Function**
A Linear chromosome means it has ends, which presents two major problems.

1) A short amount of sequence at the end of the chromosome is left un-replicated after every round of replication (see lagging strand synthesis, replication lecture). If this proceeded indefinitely, genes near the ends of the chromosome would get deleted.

2) DNA ends are a problem for cells; they are sensitive to degradation, and are seen by the cell as a form of DNA damage. If seen as damage, chromosome ends could either trigger a “repair” reaction, where the ends of two chromosomes are fused together (see below) or the chromosome end could trigger cell death by “checkpoint pathways” (see DNA damage and repair lecture). Fusion of chromosomes will mean that they can’t segregate properly during mitosis.

![Figure 3. Chromosome fusion. Spots (pink in color picture) are telomeres.](image)
These problems are solved by **telomeres**.

1) The telomere is a short repeated sequence; the number and sequence of repeats varies between organisms (there are 500 to 2000 repeats of the TTAGGG sequence in each human telomere). Early in development, in actively dividing cells, several telomeric repeats are added every round of replication by a specific enzyme termed **telomerase**; this opposes the loss in telomere sequence due to the end replication problem. In other words, eukaryotic cells balance the loss of telomere sequence due to the inability to complete replication to the end with addition of new telomere repeats through telomerase activity.

![Figure 4. Telomerase adds TTAGGG repeats to end of chromosome.](image)

Telomerase is like standard DNA polymerases (see replication lecture) in that it adds new DNA to the 3' end of a DNA chain. However, unlike standard DNA polymerases, telomerase does not use a DNA template to direct synthesis; it uses an intrinsic RNA component of the enzyme to act as the template strand instead. The RNA component has, as part of its sequence, the complementary sequence to 5'TTAGGG3' (5'CCCUAA3'), so it can direct synthesis of TTAGGG repeats only. (see figure 4 on next page)

2) Telomeres form loops. A single stranded tail at the end of the telomere loops back and inserts itself into a region of telomere double helix. This is termed the “T-loop”, and was discovered in 1999 by Dr. Jack Griffith (UNC) and Titia de Lange. Formation of the loop, and protection of the branch point, is mediated by the TRF-2 protein (Telomere Repeat binding Factor-2). This loop essentially hides the chromosome end, and means the end of the chromosome is resistant to degradation or “repair”.

**Telomeres, aging, and cancer.**

There is a link between the potential of cells to regularly divide (proliferative capacity) and the length of their telomeres. Cells with short telomeres have limited proliferative capacity, cells with long telomeres have much greater proliferative capacity.

This leads to the idea that telomeres are a “clock” that helps define proliferative capacity. Early in development (when? Depends on cell type) cells stop expressing...
telomerase. These cells continue to divide, but because they are no longer expressing telomerase, their telomeres shorten. After a limited number of divisions (the number varies according to cell type) the telomeres have shortened to a point where a signal is sent that stops the cells from dividing further (senescence). Cellular senescence is a significant component of aging.

Cancer-causing mutations can force such a differentiated cell to start dividing again; however, the telomeres continue to shorten. These transformed cells will eventually reach “crisis”, a point at which the telomere becomes too short to permit cell survival. Although most cells are killed at this point, some survive to cause cancer because they have “learned” to re-express telomerase, and their telomeres no longer continue to shorten.

III. Packaging DNA into chromosomes: chromatin

The linear length of the human genome is about two meters, and needs to fit in a space that is 6 microns in diameter (this is like having a 100 kilometer long spaghetti noodle on your dinner plate). There is an obvious “compaction” problem. Moreover, an unsupported DNA molecule the length of a chromosome is certain to incur frequent breakage and entanglement unless organized into compact, discrete, structurally supported packets.

A. The nucleosome.

The basic unit of eukaryotic chromatin structure is the nucleosome. This consists of a core octamer containing two copies each of four Histone proteins: two each of H2A, H2B, H3 and H4. All histones are compact globular proteins with an extended tail rich in basic amino acids/positive charges.
The positive charge of the histone interacts with the negative charge of the DNA phosphate backbone. DNA is wrapped around the octamer slightly more than twice, about 70 bp/turn, so about 145 bp of DNA wraps around each octamer. The nucleosome core particle (octamer+DNA) is about 10 nm wide.

**B. Higher order chromatin structure.**

Higher order organization (spacing and packing of adjacent nucleosomes) varies from gene to gene, and over the life of the cell.

Packing of adjacent nucleosomes is usually mediated at least in part by histone H1 (linker histone).

Adjacent nucleosomes are often packed into higher order 30 nm wide fibers, a helical arrangement of stacked adjacent nucleosomes in a structure of approximately 6 nucleosomes/turn.

30 nm fibers can be further organized into 20-100 kilobase pair loops; the base of these loops are attached to protein scaffolds. These loops are packed together into a regular, tightly condensed structure immediately prior to cell division (e.g. see figure 1).

**C. Chromatin and DNA metabolism.**

The organization of DNA into nucleosomes and higher order chromatin solves the compaction problem, but raises another issue. The DNA is less accessible, and thus presents a barrier to replication and transcription. This is a problem - you must be able to loosen chromatin compaction somewhat to permit replication and transcription to proceed, but also an opportunity
the amount a gene is transcribed can be regulated by making the gene more or less accessible to
the transcription machinery. Cells therefore dynamically regulate chromatin state, condensing
chromosomes for easier segregation during cell division, then stretching certain parts of the
chromosome back out after cell division for transcription and replication.

**Heterochromatin and Euchromatin.**
Immediately prior to cell division chromosomes are condensed into tightly packed rod-like
structures, visible by standard microscopy. After cell division is completed chromosomes are
de-condensed, with regions that have actively transcribed genes reverting to a state with very
loosely organized chromatin (“euchromatin”). The extreme example of this is the heavily
transcribed ribosomal RNA genes in dividing cells, which are likely almost free of nucleosomes.
“Silent” regions, or heterochromatin, largely retain a highly condensed structure throughout the
life of the cell. Heterochromatin is found in areas of the chromosome that aren’t
transcriptionally active. This often includes repetitive sequences (and in particular, centromeres
and telomeres) regions near the ends of the chromosome (“sub-telomeric regions”), as well as
genes not necessary for a particular cell type.

**Chromatin remodeling.**
A large part of regulation of chromatin state occurs at the level of the nucleosome, by altering
how tightly the octamer associates with DNA, as well as how tightly adjacent nucleosomes
associate with each other. The ability to adjust nucleosome positioning in chromatin is termed
“remodeling”. Since nucleosomes are typically very stable, this requires a multi-protein
complex designed for this purpose. Remodeling complexes typically use the energy from ATP
hydrolysis to perform remodeling. Remodeling complexes are also important in assembly of
nucleosomes de novo on recently replicated DNA.

Chromatin state is altered, and remodeling facilitated or inhibited, by covalent modification of
histones. These modifications may act partly by loosening association of DNA with
nucleosomes, because the modifications often neutralize positive charges in histone tails so the
negatively charged DNA backbone no longer sticks to nucleosomes quite as tightly. Thus
remodeling complexes often have histone modifying enzymes: histone acetylases and de-
acetylases, histone methylases, and histone kinases.

Acetylation and methylation of histones are critical events both for assembly of new
nucleosomes after DNA replication, as well as regulating transcription (See also Gene regulation,
in Marzluff lectures). Phosphorylation is important for chromosome condensation prior to cell
division. Recently, histone phosphorylation has also been observed localized to regions of DNA
damage.

**Histone variants and chromatin structure**
The histone H3 variant CENP-A helps mediate attachment of the kinetochore to centromeres (see
above).
Regions of nucleosomes where normal H2A is substituted with the H2A variant H2AZ is
typically stably heterochromatic.
Figure 1 reproduced with permission from *Genes VII*, by B. Lewin, Oxford University Press, 2000, p.554.
Figure 2 reproduced with permission from *Molecular biology of the Cell*, Alberts et al., Macmillan inc., 1983, p. 610.
Figure 3 PNAS., Vol. 96, Issue 26, 14899-14904, December 21, 1999 DNA double-strand break repair proteins are required to cap the ends of mammalian chromosomes. Susan M. Bailey, Julianne Meyne, David J. Chen, Akihiro Kurimasa, Gloria C. Li, Bruce E. Lehner, and Edwin H. Goodwin,
Figure 5 reproduced with permission from *Genes VII*, by B. Lewin, Oxford University Press, 2000, p.564.
Figure 6 reproduced with permission from *Genes VII*, by B. Lewin, Oxford University Press, 2000, p.569.
Figure 8 reproduced with permission from *Genes VII*, by B. Lewin, Oxford University Press, 2000, p.579.
Figure 13 reproduced with permission from *Genes VII*, by B. Lewin, Oxford University Press, 2000, p.435.
Figure 14 reproduced with permission from *Genes VII*, by B. Lewin, Oxford University Press, 2000, p.436.

supercoiling