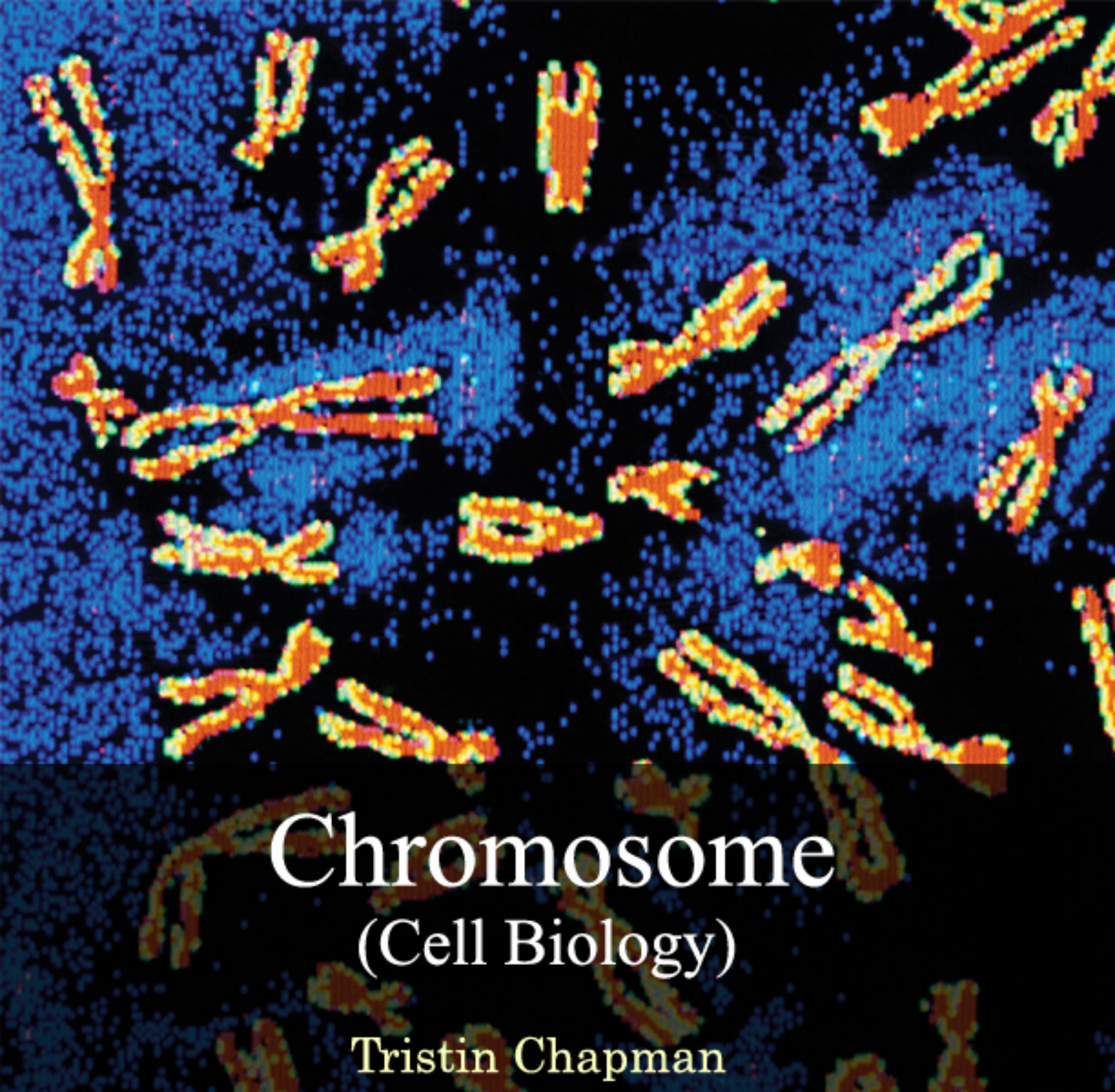


Chromosome
(Cell Biology)

Tristin Chapman

First Edition, 2012

ISBN 978-81-323-4103-1

Published by:
White Word Publications
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

Chromosome

A **chromosome** is an organized structure of DNA and protein that is found in cells. It is a single piece of coiled DNA containing many genes, regulatory elements and other nucleotide sequences. Chromosomes also contain DNA-bound proteins, which serve to package the DNA and control its functions.

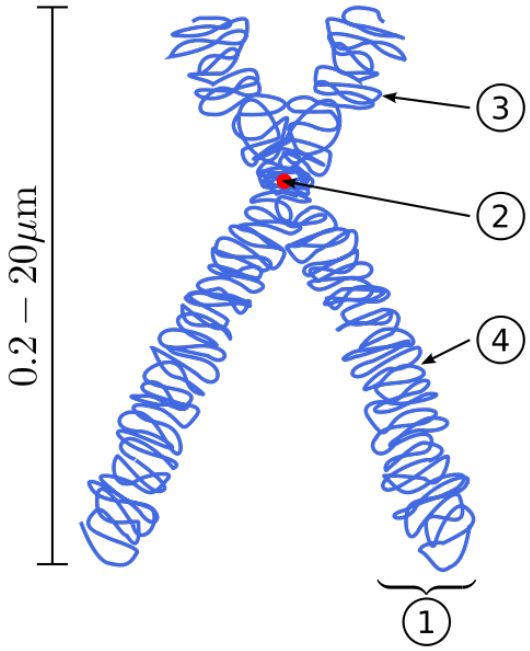


Diagram of a replicated and condensed metaphase eukaryotic chromosome. (1) Chromatid – one of the two identical parts of the chromosome after S phase. (2) Centromere – the point where the two chromatids touch, and where the microtubules attach. (3) Short arm. (4) Long foot.

Chromosomes vary widely between different organisms. The DNA molecule may be circular or linear, and can be composed of 10,000 to 1,000,000,000 nucleotides in a long chain. Typically, eukaryotic cells (cells with nuclei) have large linear chromosomes and

prokaryotic cells (cells without defined nuclei) have smaller circular chromosomes, although there are many exceptions to this rule. Also, cells may contain more than one type of chromosome; for example, mitochondria in most eukaryotes and chloroplasts in plants have their own small chromosomes.

In eukaryotes, nuclear chromosomes are packaged by proteins into a condensed structure called chromatin. This allows the very long DNA molecules to fit into the cell nucleus. The structure of chromosomes and chromatin varies through the cell cycle. Chromosomes are the essential unit for cellular division and must be replicated, divided, and passed successfully to their daughter cells so as to ensure the genetic diversity and survival of their progeny. Chromosomes may exist as either duplicated or unduplicated. Unduplicated chromosomes are single linear strands, whereas duplicated chromosomes (copied during synthesis phase) contain two copies joined by a centromere.

Compaction of the duplicated chromosomes during mitosis and meiosis results in the classic four-arm structure (pictured to the right). Chromosomal recombination plays a vital role in genetic diversity. If these structures are manipulated incorrectly, through processes known as chromosomal instability and translocation, the cell may undergo mitotic catastrophe and die, or it may unexpectedly evade apoptosis leading to the progression of cancer.

In practice "chromosome" is a rather loosely defined term. In prokaryotes and viruses, the term genophore is more appropriate when no chromatin is present. However, a large body of work uses the term chromosome regardless of chromatin content. In prokaryotes, DNA is usually arranged as a circle, which is tightly coiled in on itself, sometimes accompanied by one or more smaller, circular DNA molecules called plasmids. These small circular genomes are also found in mitochondria and chloroplasts, reflecting their bacterial origins. The simplest genophores are found in viruses: these DNA or RNA molecules are short linear or circular genophores that often lack structural proteins.

The word *chromosome* comes from the Greek χρῶμα (*chroma*, colour) and σῶμα (*soma*, body) due to their property of being very strongly stained by particular dyes.

History

In a series of experiments beginning in the mid-1880s, Theodor Boveri gave the definitive demonstration that chromosomes are the vectors of heredity. His two principles were the *continuity* of chromosomes and the *individuality* of chromosomes. It is the second of these principles that was so original. Wilhelm Roux suggested that each chromosome carries a different genetic load. Boveri was able to test and confirm this hypothesis. Aided by the rediscovery at the start of the 1900s of Gregor Mendel's earlier work, Boveri was able to point out the connection between the rules of inheritance and the behaviour of the chromosomes. Boveri influenced two generations of American cytologists: Edmund Beecher Wilson, Walter Sutton and Theophilus Painter were all influenced by Boveri (Wilson and Painter actually worked with him).

In his famous textbook *The Cell in Development and Heredity*, Wilson linked together the independent work of Boveri and Sutton (both around 1902) by naming the chromosome theory of inheritance the "Sutton-Boveri Theory" (the names are sometimes reversed). Ernst Mayr remarks that the theory was hotly contested by some famous geneticists: William Bateson, Wilhelm Johannsen, Richard Goldschmidt and T.H. Morgan, all of a rather dogmatic turn-of-mind. Eventually, complete proof came from chromosome maps in Morgan's own lab.

In eukaryotes

Eukaryotes (cells with nuclei such as those found in plants, yeast, and animals) possess multiple large linear chromosomes contained in the cell's nucleus. Each chromosome has one centromere, with one or two arms projecting from the centromere, although, under most circumstances, these arms are not visible as such. In addition, most eukaryotes have a small circular mitochondrial genome, and some eukaryotes may have additional small circular or linear cytoplasmic chromosomes.

In the nuclear chromosomes of eukaryotes, the uncondensed DNA exists in a semi-ordered structure, where it is wrapped around histones (structural proteins), forming a composite material called chromatin.

Chromatin

Chromatin is the complex of DNA and protein found in the eukaryotic nucleus, which packages chromosomes. The structure of chromatin varies significantly between different stages of the cell cycle, according to the requirements of the DNA.

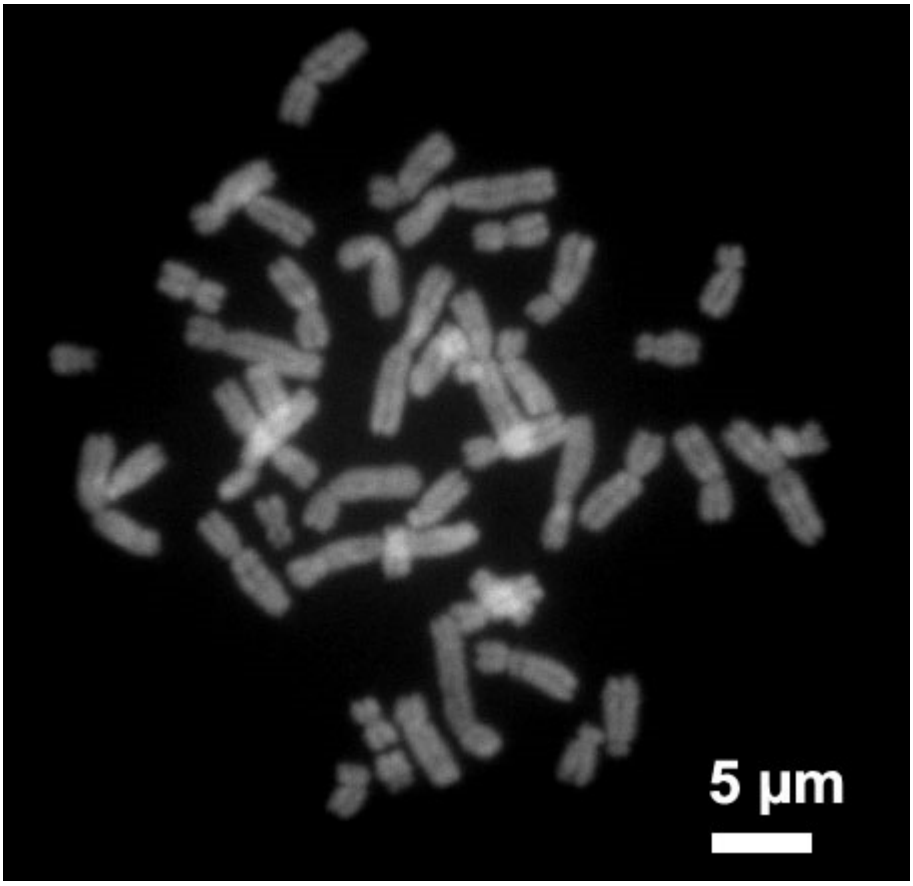
Interphase chromatin

During interphase (the period of the cell cycle where the cell is not dividing), two types of chromatin can be distinguished:

- Euchromatin, which consists of DNA that is active, e.g., being expressed as protein.
- Heterochromatin, which consists of mostly inactive DNA. It seems to serve structural purposes during the chromosomal stages. Heterochromatin can be further distinguished into two types:
    - *Constitutive heterochromatin*, which is never expressed. It is located around the centromere and usually contains repetitive sequences.
    - *Facultative heterochromatin*, which is sometimes expressed.

Individual chromosomes cannot be distinguished at this stage – they appear in the nucleus as a homogeneous tangled mix of DNA and protein.

Metaphase chromatin and division



Human chromosomes during metaphase.

In the early stages of mitosis or meiosis (cell division), the chromatin strands become more and more condensed. They cease to function as accessible genetic material (transcription stops) and become a compact transportable form. This compact form makes the individual chromosomes visible, and they form the classic four arm structure, a pair of sister chromatids attached to each other at the centromere. The shorter arms are called *p arms* (from the French *petit*, small) and the longer arms are called *q arms* (*q* follows *p* in the Latin alphabet). This is the only natural context in which individual chromosomes are visible with an optical microscope.

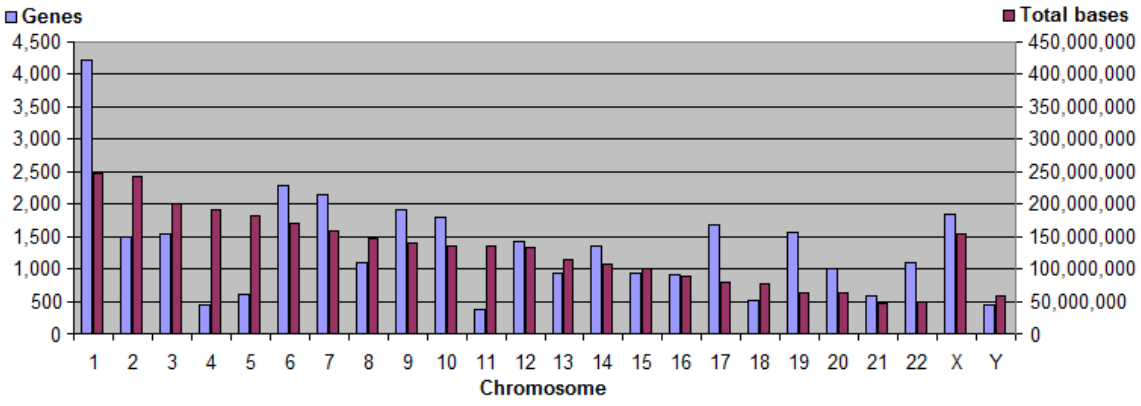
During divisions, long microtubules attach to the centromere and the two opposite ends of the cell. The microtubules then pull the chromatids apart, so that each daughter cell inherits one set of chromatids. Once the cells have divided, the chromatids are uncoiled and can function again as chromatin. In spite of their appearance, chromosomes are structurally highly condensed, which enables these giant DNA structures to be contained within a cell nucleus (Fig. 2).

The self-assembled microtubules form the spindle, which attaches to chromosomes at specialized structures called kinetochores, one of which is present on each sister

chromatid. A special DNA base sequence in the region of the kinetochores provides, along with special proteins, longer-lasting attachment in this region.

In humans

Chromosomes can be divided into two types—autosomes, and sex chromosomes. Certain genetic traits are linked to your sex, and are passed on through the sex chromosomes. The autosomes contain the rest of the genetic hereditary information. All act in the same way during cell division. Human cells have 23 pairs of large linear nuclear chromosomes, (22 pairs of autosomes and one pair of sex chromosomes) giving a total of 46 per cell. In addition to these, human cells have many hundreds of copies of the mitochondrial genome. Sequencing of the human genome has provided a great deal of information about each of the chromosomes. Below is a table compiling statistics for the chromosomes, based on the Sanger Institute's human genome information in the Vertebrate Genome Annotation (VEGA) database. Number of genes is an estimate as it is in part based on gene predictions. Total chromosome length is an estimate as well, based on the estimated size of unsequenced heterochromatin regions.



Chromosome	Genes	Total bases	Sequenced bases
1	4,220	247,199,719	224,999,719
2	1,491	242,751,149	237,712,649
3	1,550	199,446,827	194,704,827
4	446	191,263,063	187,297,063
5	609	180,837,866	177,702,766
6	2,281	170,896,993	167,273,993
7	2,135	158,821,424	154,952,424
8	1,106	146,274,826	142,612,826
9	1,920	140,442,298	120,312,298
10	1,793	135,374,737	131,624,737
11	379	134,452,384	131,130,853
12	1,430	132,289,534	130,303,534

13	924	114,127,980	95,559,980
14	1,347	106,360,585	88,290,585
15	921	100,338,915	81,341,915
16	909	88,822,254	78,884,754
17	1,672	78,654,742	77,800,220
18	519	76,117,153	74,656,155
19	1,555	63,806,651	55,785,651
20	1,008	62,435,965	59,505,254
21	578	46,944,323	34,171,998
22	1,092	49,528,953	34,893,953
X (sex chromosome)	1,846	154,913,754	151,058,754
Y (sex chromosome)	454	57,741,652	25,121,652
Total	**32,185**	**3,079,843,747**	**2,857,698,560**

In prokaryotes

The prokaryotes – bacteria and archaea – typically have a single circular chromosome, but many variations do exist. Most bacteria have a single circular chromosome that can range in size from only 160,000 base pairs in the endosymbiotic bacterium *Candidatus Carsonella ruddii*, to 12,200,000 base pairs in the soil-dwelling bacterium *Sorangium cellulosum*. Spirochaetes of the genus *Borrelia* are a notable exception to this arrangement, with bacteria such as *Borrelia burgdorferi*, the cause of Lyme disease, containing a single linear chromosome.

Structure in sequences

Prokaryotic chromosomes have less sequence-based structure than eukaryotes. Bacteria typically have a single point (the origin of replication) from which replication starts, whereas some archaea contain multiple replication origins. The genes in prokaryotes are often organized in operons, and do not usually contain introns, unlike eukaryotes.

DNA packaging

Prokaryotes do not possess nuclei. Instead, their DNA is organized into a structure called the nucleoid. The nucleoid is a distinct structure and occupies a defined region of the bacterial cell. This structure is, however, dynamic and is maintained and remodeled by the actions of a range of histone-like proteins, which associate with the bacterial chromosome. In archaea, the DNA in chromosomes is even more organized, with the DNA packaged within structures similar to eukaryotic nucleosomes.

Bacterial chromosomes tend to be tethered to the plasma membrane of the bacteria. In molecular biology application, this allows for its isolation from plasmid DNA by centrifugation of lysed bacteria and pelleting of the membranes (and the attached DNA).

Prokaryotic chromosomes and plasmids are, like eukaryotic DNA, generally supercoiled. The DNA must first be released into its relaxed state for access for transcription, regulation, and replication.

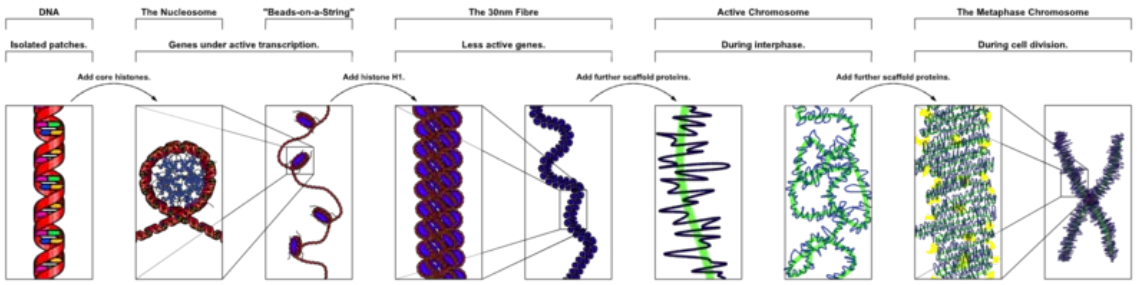


Fig. 2: The major structures in DNA compaction; DNA, the nucleosome, the 10nm "beads-on-a-string" fibre, the 30nm fibre and the metaphase chromosome.

Number of chromosomes in various organisms

Eukaryotes

These tables give the total number of chromosomes (including sex chromosomes) in a cell nucleus. For example, human cells are diploid and have 22 different types of autosome, each present as two copies, and two sex chromosomes. This gives 46 chromosomes in total. Other organisms have more than two copies of their chromosomes, such as bread wheat, which is *hexaploid* and has six copies of seven different chromosomes – 42 chromosomes in total.

Chromosome numbers in some plants

Plant Species	**#**
Arabidopsis thaliana (diploid)	10
Rye (diploid)	14
Maize (diploid or palaeotetraploid)	20
Einkorn wheat (diploid)	14

Durum wheat (tetraploid)	28
Bread wheat (hexaploid)	42
Cultivated tobacco (tetraploid)	48
Adder's Tongue Fern (diploid)	}

Chromosome numbers (2n) in some animals

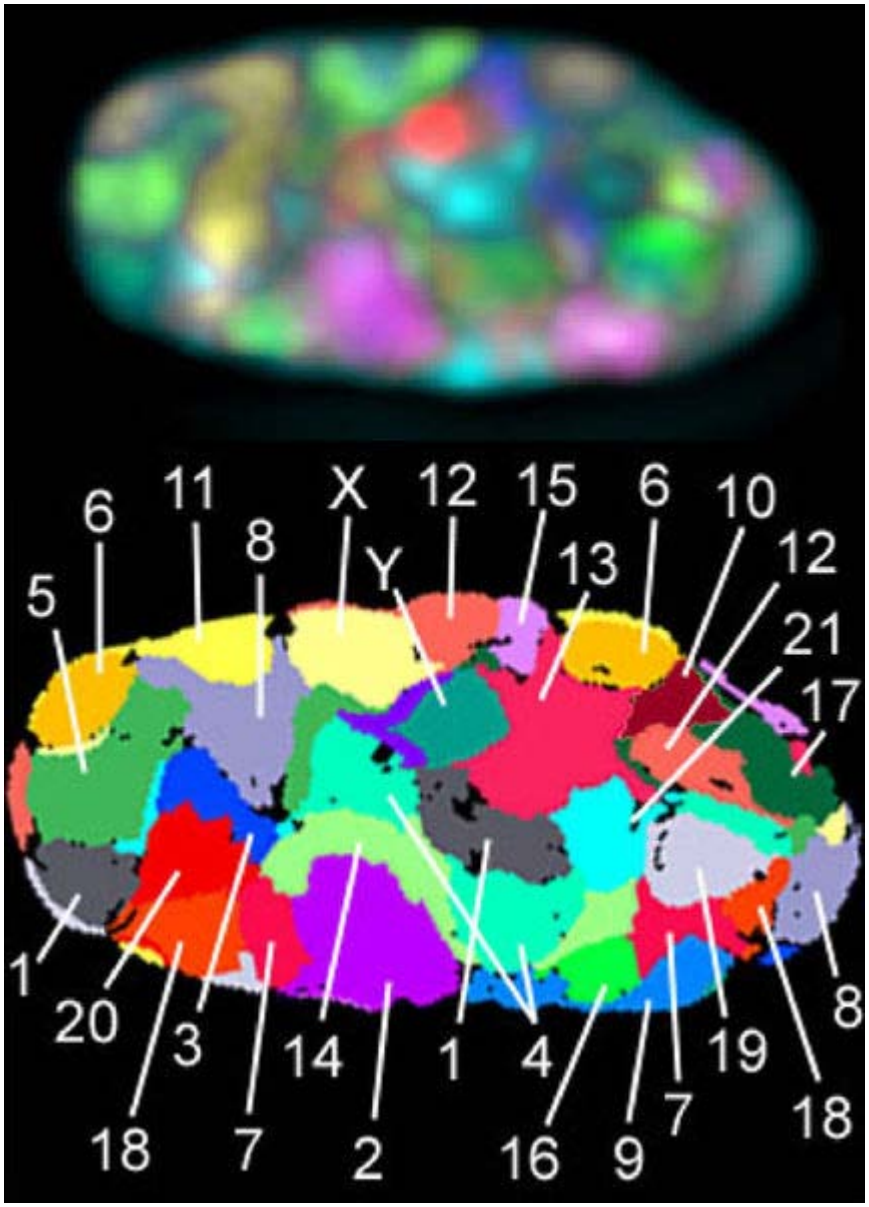
Species	**#**	**Species**	**#**
Common fruit fly	8	Guinea Pig	64
Guppy (*poecilia reticulata*)	46	Garden snail	54
Earthworm (*Octodrilus complanatus*)	36	Tibetan fox	36
Domestic cat	38	Domestic pig	38
Laboratory mouse	40	Laboratory rat	42
Rabbit (*Oryctolagus cuniculus*)	44	Syrian hamster	44
Hares	48	Human	46
Gorillas, Chimpanzees	48	Domestic sheep	54

Elephants	56	Cow	60
Donkey	62	Horse	64
Dog	78	Kingfisher	132
Goldfish	100-104	Silkworm	56

Chromosome numbers in other organisms

Species	Large Chromosomes	Intermediate Chromosomes	Microchromosomes
Trypanosoma brucei	11	6	~100
Domestic Pigeon (*Columba livia domestics*)	18	-	59-63
Chicken	8	2 sex chromosomes	60

Normal members of a particular eukaryotic species all have the same number of nuclear chromosomes. Other eukaryotic chromosomes, i.e., mitochondrial and plasmid-like small chromosomes, are much more variable in number, and there may be thousands of copies per cell.



The 23 human chromosome territories during prometaphase in fibroblast cells.

Asexually reproducing species have one set of chromosomes, which are the same in all body cells. However, asexual species can be either haploid or diploid.

Sexually reproducing species have somatic cells (body cells), which are diploid [2n] having two sets of chromosomes, one from the mother and one from the father. Gametes, reproductive cells, are haploid [n]: They have one set of chromosomes. Gametes are produced by meiosis of a diploid germ line cell. During meiosis, the matching chromosomes of father and mother can exchange small parts of themselves (crossover), and thus create new chromosomes that are not inherited solely from either parent. When a male and a female gamete merge (fertilization), a new diploid organism is formed.

Some animal and plant species are polyploid [Xn]: They have more than two sets of homologous chromosomes. Plants important in agriculture such as tobacco or wheat are often polyploid, compared to their ancestral species. Wheat has a haploid number of seven chromosomes, still seen in some cultivars as well as the wild progenitors. The more-common pasta and bread wheats are polyploid, having 28 (tetraploid) and 42 (hexaploid) chromosomes, compared to the 14 (diploid) chromosomes in the wild wheat.

Prokaryotes

Prokaryote species generally have one copy of each major chromosome, but most cells can easily survive with multiple copies. For example, *Buchnera*, a symbiont of aphids has multiple copies of its chromosome, ranging from 10–400 copies per cell. However, in some large bacteria, such as *Epulopiscium fishelsoni* up to 100,000 copies of the chromosome can be present. Plasmids and plasmid-like small chromosomes are, as in eukaryotes, very variable in copy number. The number of plasmids in the cell is almost entirely determined by the rate of division of the plasmid – fast division causes high copy number, and vice versa.

Karyotype

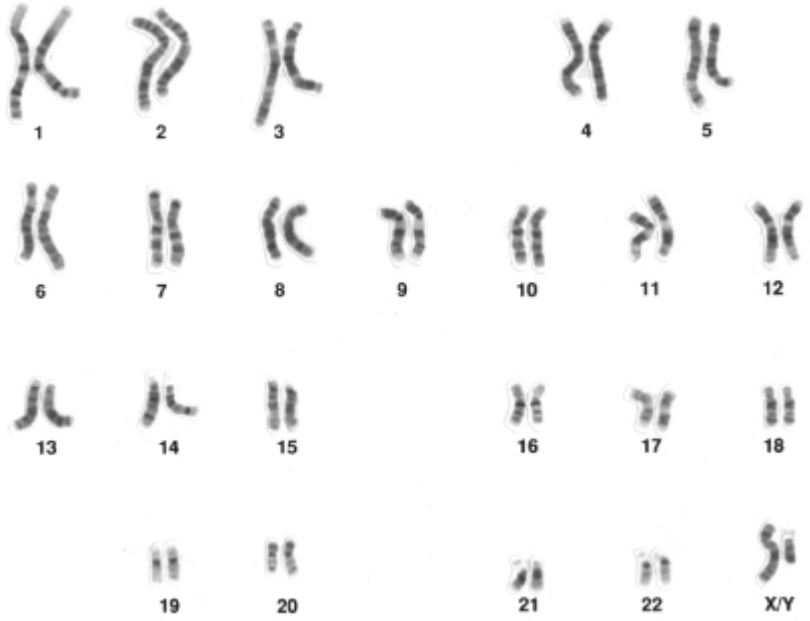


Figure 3: Karyogram of a human male

In general, the **karyotype** is the characteristic chromosome complement of a eukaryote species. The preparation and study of karyotypes is part of cytogenetics.

Although the replication and transcription of DNA is highly standardized in eukaryotes, *the same cannot be said for their karyotypes*, which are often highly variable. There may

be variation between species in chromosome number and in detailed organization. In some cases, there is significant variation within species. Often there is:

1. variation between the two sexes
2. variation between the germ-line and soma (between gametes and the rest of the body)
3. variation between members of a population, due to balanced genetic polymorphism
4. geographical variation between races
5. mosaics or otherwise abnormal individuals.

Also, variation in karyotype may occur during development from the fertilised egg.

The technique of determining the karyotype is usually called *karyotyping*. Cells can be locked part-way through division (in metaphase) in vitro (in a reaction vial) with colchicine. These cells are then stained, photographed, and arranged into a *karyogram*, with the set of chromosomes arranged, autosomes in order of length, and sex chromosomes (here X/Y) at the end: Fig. 3.

Like many sexually reproducing species, humans have special gonosomes (sex chromosomes, in contrast to autosomes). These are XX in females and XY in males.

Historical note

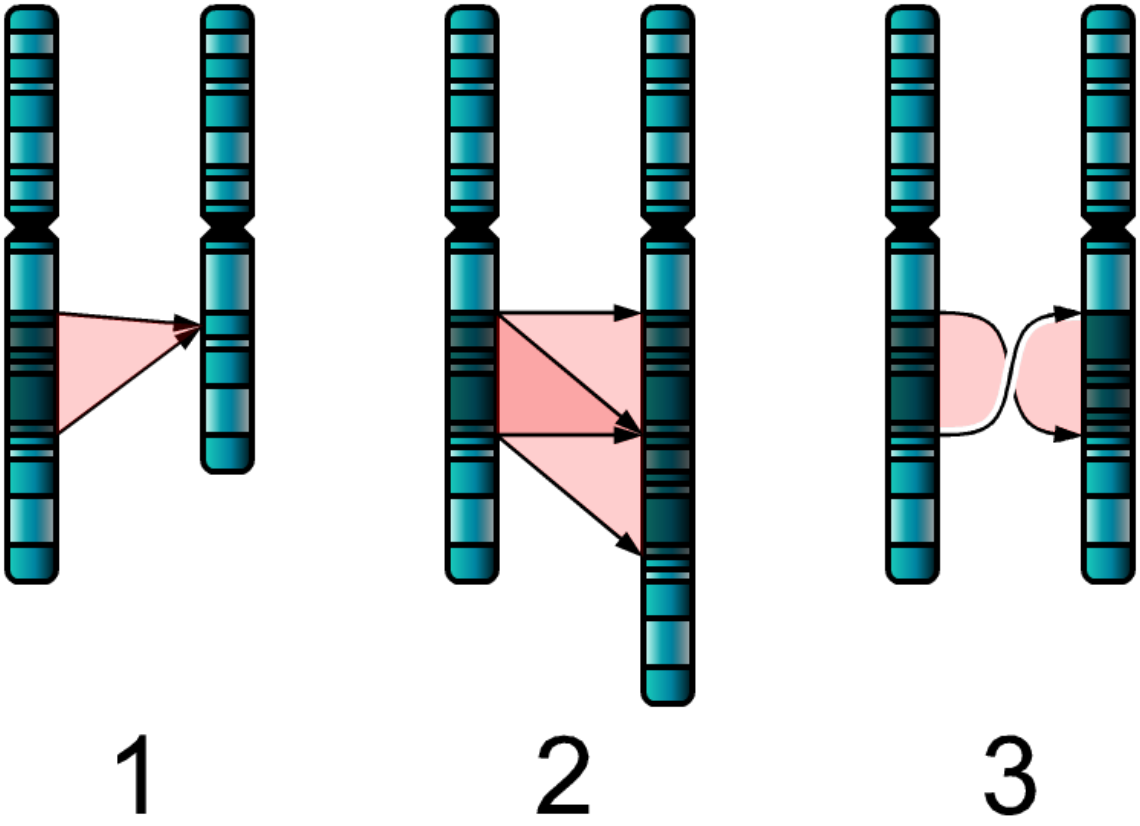
Investigation into the human karyotype took many years to settle the most basic question. How many chromosomes does a normal diploid human cell contain? In 1912, Hans von Winiwarter reported 47 chromosomes in spermatogonia and 48 in oogonia, concluding an XX/XO sex determination mechanism. Painter in 1922 was not certain whether the diploid number of man is 46 or 48, at first favouring 46. He revised his opinion later from 46 to 48, and he correctly insisted on humans having an XX/XY system.

New techniques were needed to definitively solve the problem:

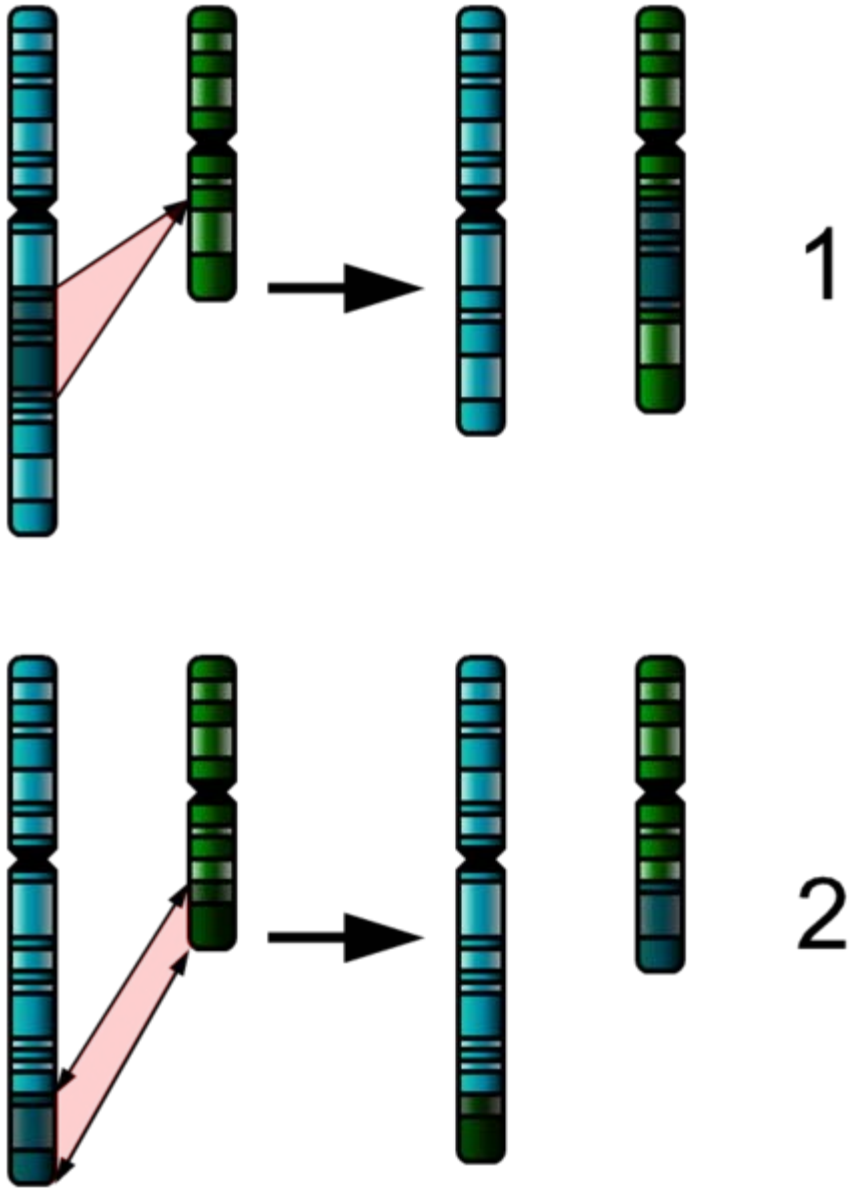
1. Using cells in culture
2. Pretreating cells in a hypotonic solution, which swells them and spreads the chromosomes
3. Arresting mitosis in metaphase by a solution of colchicine
4. Squashing the preparation on the slide forcing the chromosomes into a single plane
5. Cutting up a photomicrograph and arranging the result into an indisputable karyogram.

It took until the mid-1950s for it to become generally accepted that the human karyotype include only 46 chromosomes. Considering the techniques of Winiwarter and Painter, their results were quite remarkable. Chimpanzees (the closest living relatives to modern humans) have 48 chromosomes.

Aberrations



The three major single chromosome mutations; deletion (1), duplication (2) and inversion (3).



The two major two-chromosome mutations; insertion (1) and translocation (2).

In Down syndrome, there are three copies of chromosome 21

Chromosomal aberrations are disruptions in the normal chromosomal content of a cell and are a major cause of genetic conditions in humans, such as Down syndrome. Some chromosome abnormalities do not cause disease in carriers, such as translocations, or chromosomal inversions, although they may lead to a higher chance of birthing a child with a chromosome disorder. Abnormal numbers of chromosomes or chromosome sets, aneuploidy, may be lethal or give rise to genetic disorders. Genetic counseling is offered for families that may carry a chromosome rearrangement.

The gain or loss of DNA from chromosomes can lead to a variety of genetic disorders. Human examples include:

- Cri du chat, which is caused by the deletion of part of the short arm of chromosome 5. "Cri du chat" means "cry of the cat" in French, and the condition was so-named because affected babies make high-pitched cries that sound like those of a cat. Affected individuals have wide-set eyes, a small head and jaw, moderate to severe mental health issues, and are very short.
- Down syndrome, usually is caused by an extra copy of chromosome 21 (trisomy 21). Characteristics include decreased muscle tone, stockier build, asymmetrical skull, slanting eyes and mild to moderate developmental disability.
- Edwards syndrome, which is the second-most-common trisomy; Down syndrome is the most common. It is a trisomy of chromosome 18. Symptoms include motor retardation, developmental disability and numerous congenital anomalies causing serious health problems. Ninety percent die in infancy; however, those that live past their first birthday usually are quite healthy thereafter. They have a characteristic clenched hands and overlapping fingers.
- Idic15, abbreviation for Isodicentric 15 on chromosome 15; also called the following names due to various researches, but they all mean the same; IDIC(15), Inverted duplication 15, extra Marker, Inv dup 15, partial tetrasomy 15
- Jacobsen syndrome, also called the terminal 11q deletion disorder. This is a very rare disorder. Those affected have normal intelligence or mild developmental disability, with poor expressive language skills. Most have a bleeding disorder called Paris-Trousseau syndrome.
- Klinefelter's syndrome (XXY). Men with Klinefelter syndrome are usually sterile, and tend to have longer arms and legs and to be taller than their peers. Boys with the syndrome are often shy and quiet, and have a higher incidence of speech delay and dyslexia. During puberty, without testosterone treatment, some of them may develop gynecomastia.
- Patau Syndrome, also called D-Syndrome or trisomy-13. Symptoms are somewhat similar to those of trisomy-18, but they do not have the characteristic hand shape.
- Small supernumerary marker chromosome. This means there is an extra, abnormal chromosome. Features depend on the origin of the extra genetic material. Cat-eye syndrome and isodicentric chromosome 15 syndrome (or Idic15) are both caused by a supernumerary marker chromosome, as is Pallister-Killian syndrome.

- Triple-X syndrome (XXX). XXX girls tend to be tall and thin. They have a higher incidence of dyslexia.
- Turner syndrome (X instead of XX or XY). In Turner syndrome, female sexual characteristics are present but underdeveloped. People with Turner syndrome often have a short stature, low hairline, abnormal eye features and bone development and a "caved-in" appearance to the chest.
- XYY syndrome. XYY boys are usually taller than their siblings. Like XXY boys and XXX girls, they are somewhat more likely to have learning difficulties.
- Wolf-Hirschhorn syndrome, which is caused by partial deletion of the short arm of chromosome 4. It is characterized by severe growth retardation and severe to profound mental health issues.

Chromosomal mutations produce changes in whole chromosomes (more than one gene) or in the number of chromosomes present.

- Deletion – loss of part of a chromosome
- Duplication – extra copies of a part of a chromosome
- Inversion – reverse the direction of a part of a chromosome
- Translocation – part of a chromosome breaks off and attaches to another chromosome

Most mutations are neutral – have little or no effect. Chromosomal aberrations are the changes in the structure of chromosomes. It has a great role in evolution. A detailed graphical display of all human chromosomes and the diseases annotated at the correct spot may be found at the Oak Ridge National Laboratory.

Chapter 2

Chromatin

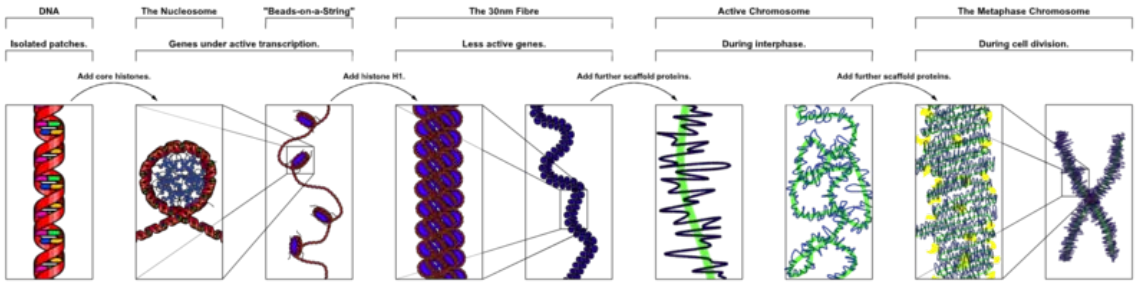


Fig. 1: The major structures in DNA compaction; DNA, the nucleosome, the 10nm "beads-on-a-string" fibre, the 30nm fibre and the metaphase chromosome.

Chromatin is the combination of DNA and other proteins that make up the contents of the nucleus. The primary protein component of chromatin is histones which act to compactly package the DNA. Chromatin is only found in eukaryotic cells, prokaryotic cells have a very different organisation of their DNA which is not normally referred to as chromatin. The primary functions of chromatin are to package DNA into a smaller volume to fit in the cell, to strengthen the DNA to allow mitosis and meiosis and prevent DNA damage, and to control gene expression and DNA replication.

The structure of chromatin depends on a combination of several factors; the overall structure depends on the stage of the cell cycle. During interphase the chromatin is structurally loose to allow access of RNA and DNA polymerases to transcribe and replicate the DNA. The local structure of chromatin during interphase depends on the genes present on the DNA, DNA coding actively transcribed genes are more loosely packaged found associated with RNA polymerases (referred to as euchromatin) while DNA coding inactive genes are found associated with structural proteins and are more tightly packaged (heterochromatin). Epigenetic chemical modification of the structural proteins in chromatin also alter the local chromatin structure, in particular chemical modifications of histone proteins by methylation and acetylation. As the cell prepares to divide, i.e. enters mitosis or meiosis, the chromatin packages more tightly to facilitate

segregation of the chromosomes during anaphase. During this stage of the cell cycle this makes the individual chromosomes in many cells visible by optical microscope.

In general terms, there are three levels of chromatin organization:

1. DNA wraps around histone proteins forming nucleosomes; the "beads on a string" structure (euchromatin).
2. Multiple histones wrap into a 30 nm fibre consisting of nucleosome arrays in their most compact form (heterochromatin).
3. Higher-level DNA packaging of the 30 nm fibre into the metaphase chromosome (during mitosis and meiosis).

There are, however, many of cells which do not follow this organisation. For example spermatozoa and avian red blood cells have more tightly packed chromatin than most eukaryotic cells and trypanosomatid protazoa do not condense their chromatin into visible chromosomes for mitosis.

During interphase

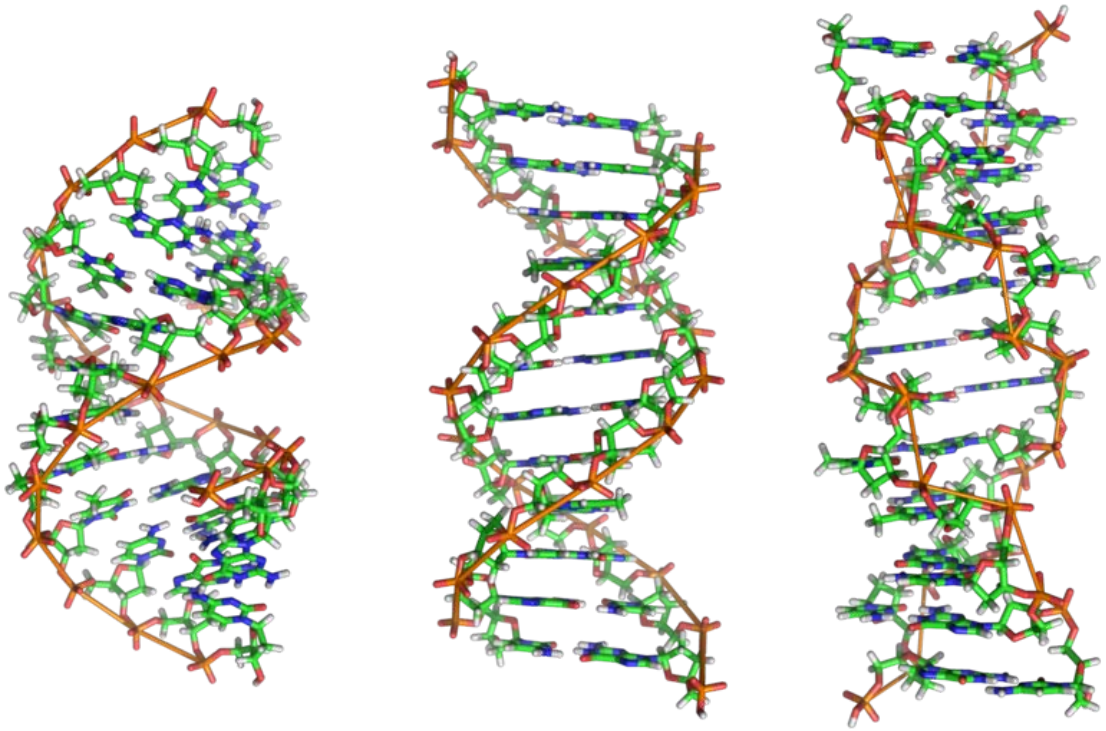
The structure of chromatin during interphase is optimised to allow easy access of transcription and DNA repair factors to the DNA while compacting the DNA into the nucleus. The structure varies depending on the access required to the DNA. Genes that require regular access by RNA polymerase require the looser structure provided by euchromatin.

Change in structure

Chromatin undergoes various forms of change in its structure. Histone proteins, the foundation blocks of chromatin, are modified by various post-translational modification to alter DNA packing. Acetylation results in the loosening of chromatin and lends itself to replication and transcription. When certain residues are methylated they hold DNA together strongly and restrict access to various enzymes. A recent study showed that there is a bivalent structure present in the chromatin: methylated lysine residues at location 4 and 27 on histone 3. It is thought that this may be involved in development; there is more methylation of lysine 27 in embryonic cells than in differentiated cells, whereas lysine 4 methylation positively regulates transcription by recruiting nucleosome remodeling enzymes and histone acetylases.

Polycomb-group proteins play a role in regulating genes through modulation of chromatin structure.

DNA structure

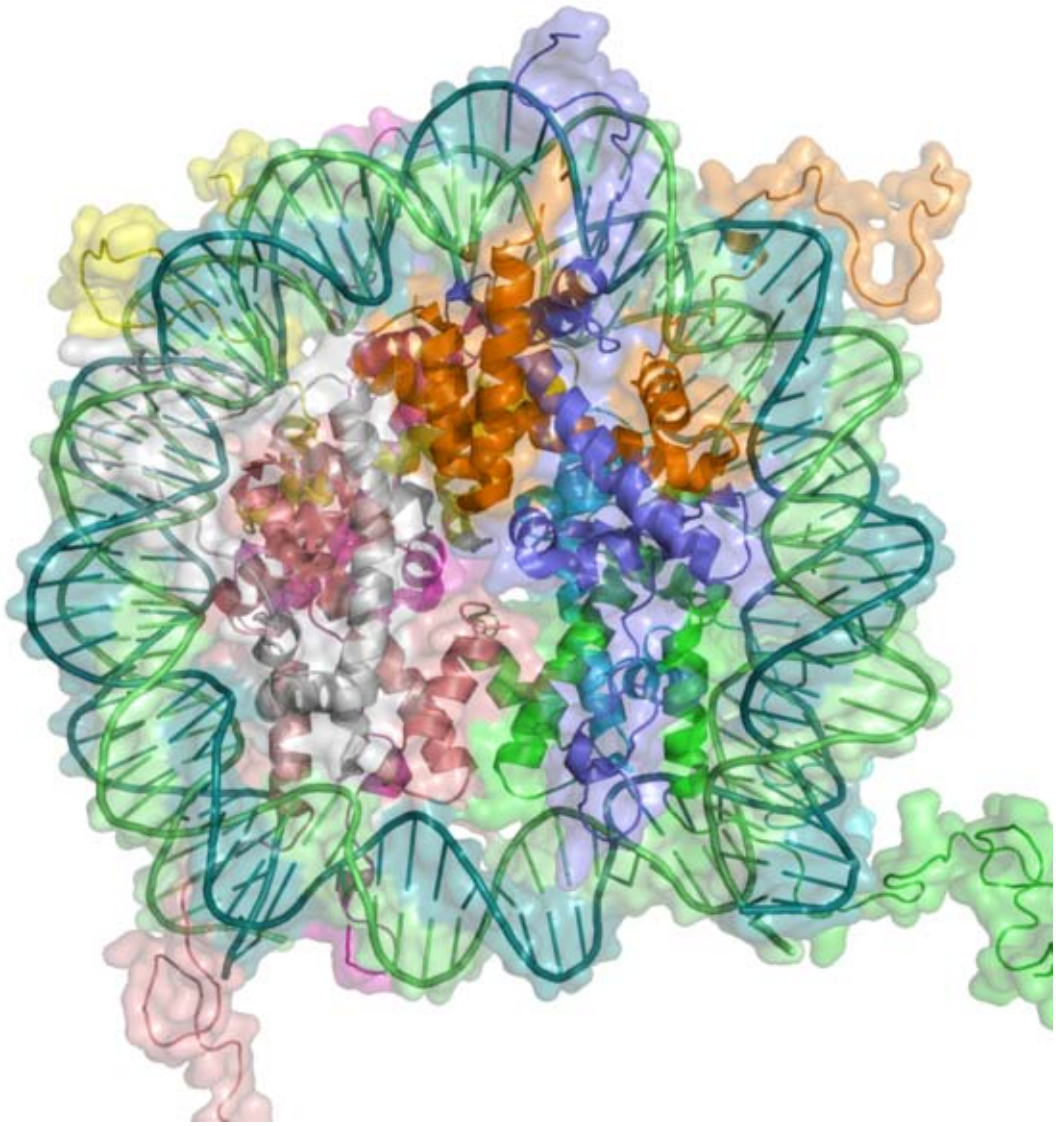


The structures of A-, B- and Z-DNA.

The vast majority of DNA within the cell is the normal DNA structure. However in nature DNA can form three structures, A-, B- and Z-DNA. A and B chromosomes are very similar, forming right-handed helices, while Z-DNA is a more unusual left-handed helix with a zig-zag phosphate backbone. Z-DNA is thought to play a specific role in chromatin structure and transcription because of the properties of the junction between B- and Z-DNA.

At the junction of B- and Z-DNA one pair of bases is flipped out from normal bonding. These play a dual role of a site of recognition by many proteins and as a sink for torsional stress from RNA polymerase or nucleosome binding.

The nucleosome and "beads-on-a-string"



A cartoon representation of the nucleosome structure.

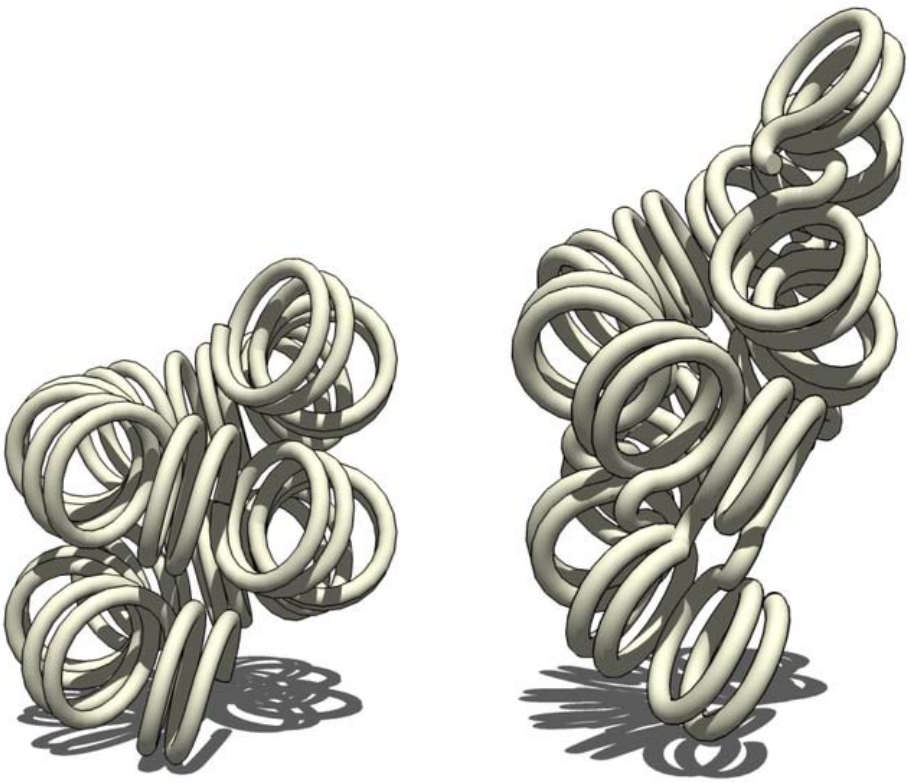
The basic repeat element of chromatin is the nucleosome, interconnected by sections of linker DNA, a far shorter arrangement than pure DNA in solution.

In addition to the core histones, there is the linker histone, H1, which contacts the exit/entry of the DNA strand on the nucleosome. The nucleosome core particle, together with histone H1, is known as a chromatosome. Nucleosomes, with about 20 to 60 base pairs of linker DNA, can form, under non-physiological conditions, an approximately 10 nm "beads-on-a-string" fibre. (Fig. 1-2). .

The nucleosomes bind DNA non-specifically, as required by their function in general DNA packaging. There are, however, large DNA sequence preferences that govern

nucleosome positioning. This is due primarily to the varying physical properties of different DNA sequences: For instance, adenosine and thymine are more favorably compressed into the inner minor grooves. This means nucleosomes can bind preferentially at one position approximately every 10 base pairs (the helical repeat of DNA)- where the DNA is rotated to maximise the number of A and T bases that will lie in the inner minor groove.

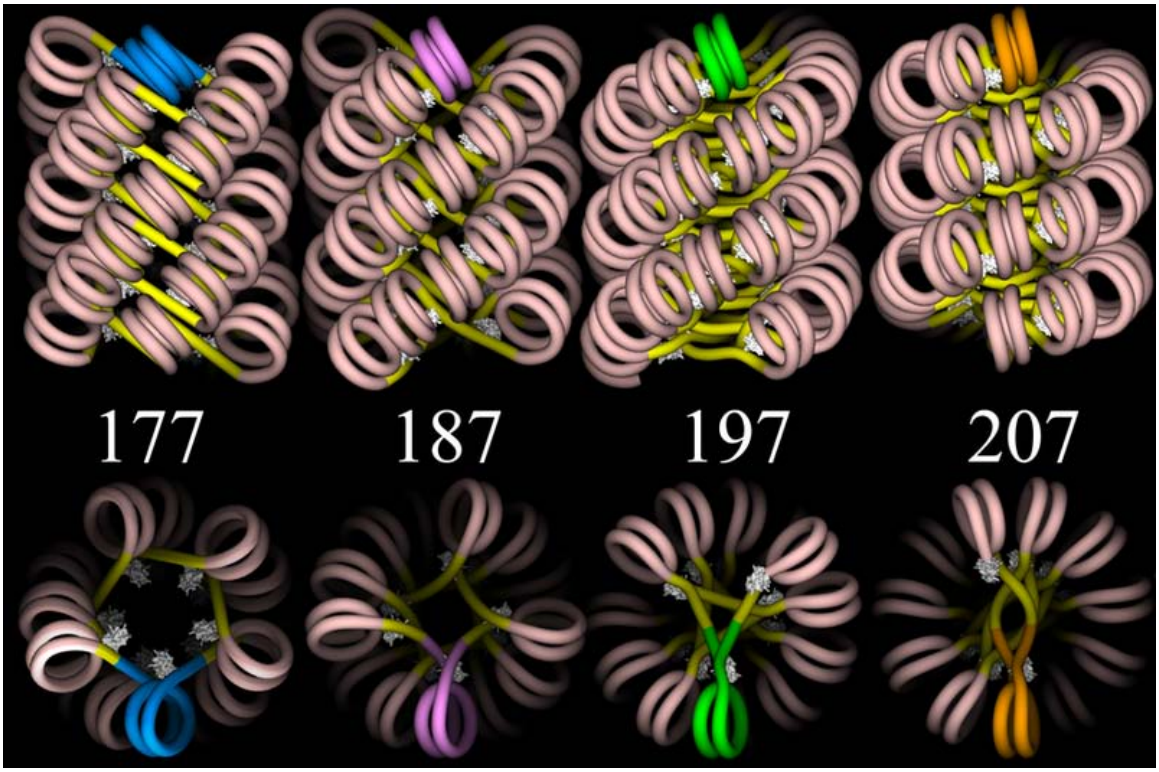
30 nm chromatin fibre



Two proposed structures of the 30nm chromatin filament.
Left: 1 start helix "solenoid" structure.
Right: 2 start loose helix structure.
Note: the histones are omitted in this diagram - only the DNA is shown.

With addition of H1, the "beads-on-a-string" structure in turn coils into a 30 nm diameter helical structure known as the 30 nm fibre or filament. The precise structure of the chromatin fibre in the cell is not known in detail, and there is still some debate over this.

This level of chromatin structure is thought to be the form of euchromatin, which contains actively transcribed genes. EM studies have demonstrated that the 30 nm fibre is highly dynamic such that it unfolds into a 10 nm fiber ("beads-on-a-string") structure when transversed by an RNA polymerase engaged in transcription.



Four proposed structures of the 30 nm chromatin filament for DNA repeat length per nucleosomes ranging from 177 to 207 bp. Linker DNA in yellow and nucleosomal DNA in pink.

The existing models commonly accept that the nucleosomes lie perpendicular to the axis of the fibre, with linker histones arranged internally. A stable 30 nm fibre relies on the regular positioning of nucleosomes along DNA. Linker DNA is relatively resistant to bending and rotation. This makes the length of linker DNA critical to the stability of the fibre, requiring nucleosomes to be separated by lengths that permit rotation and folding into the required orientation without excessive stress to the DNA. In this view, different length of the linker DNA should produce different folding topologies of the chromatin fiber. Recent theoretical work, based on electron-microscopy images of reconstituted fibers support this view.

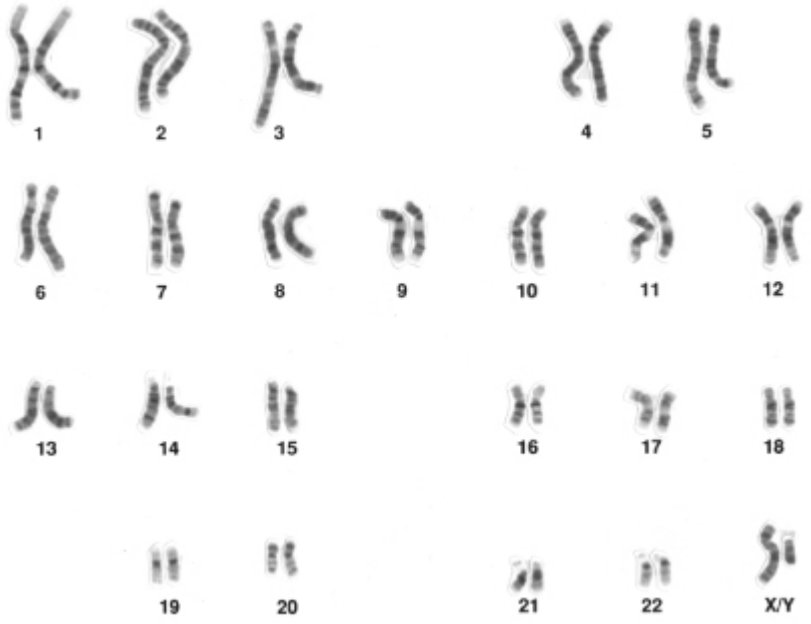
Spatial organization of chromatin in the cell nucleus

The layout of the genome within the nucleus is not random - specific regions of the genome have a tendency to be found in certain spaces. Specific regions of the chromatin are enriched at the nuclear membrane, while other regions are bound together by protein complexes. The layout of this is not, however, well characterised apart from the compaction of one of the two X chromosomes in mammalian females into the Barr body. This serves the role of permanently deactivating these genes, which prevents females getting a 'double dose' relative to males. The extent to which the inactive X is actually compacted is a matter of some controversy.

Chromatin and bursts of transcription

Fluctuations between open and closed chromatin may contribute discontinuity of transcription, or transcriptional bursting. Other factors are probably involved, such as the association and dissociation of transcription factor complexes with chromatin. The phenomenon, as opposed to simple probabilistic models of transcription, can account for the high variability in gene expression occurring between cells in isogenic populations.

Metaphase chromatin



Karyogram of human male using Giemsa staining, showing the classic metaphase chromatin structure.

The metaphase structure of chromatin differs vastly to that of interphase. It is optimised for physical strength and manageability, forming the classic chromosome structure seen in karyotypes. The structure of the condensed chromosome is thought to be loops of 30 nm fibre to a central scaffold of proteins. It is, however, not well characterised.

The physical strength of chromatin is vital for this stage of division to prevent shear damage to the DNA as the daughter chromosomes are separated. To maximise strength the composition of the chromatin changes as it approaches the centromere, primarily through alternative histone H1 anologues.

It should also be noted that, during mitosis, while most of the chromatin is tightly compacted, there are small regions that are not as tightly compacted. These regions often correspond to promoter regions of genes that were active in that cell type prior to entry into cromitosis. The lack of compaction of these regions is called bookmarking, which is an epigenetic mechanism believed to be important for transmitting to daughter cells the

"memory" of which genes were active prior to entry into mitosis. This bookmarking mechanism is needed to help transmit this memory because transcription ceases during mitosis.

Chromatin: alternative definitions

1. **Simple and concise definition:** Chromatin is DNA plus the proteins (and RNA) that package DNA within the cell nucleus.
2. **A biochemists' operational definition:** Chromatin is the DNA/protein/RNA complex extracted from eukaryotic lysed interphase nuclei. Just which of the multitudinous substances present in a nucleus will constitute a part of the extracted material will depend in part on the technique each researcher uses. Furthermore, the composition and properties of chromatin vary from one cell type to the another, during development of a specific cell type, and at different stages in the cell cycle.
3. **The *DNA* + *histone* = *chromatin* definition:** The DNA double helix in the cell nucleus is packaged by special proteins termed histones. The formed protein/DNA complex is called chromatin. The structural entity of chromatin is the nucleosome.

Alternative chromatin organisations

During metazoan spermiogenesis, the spermatid's chromatin is remodelled into a more spaced-packaged, widened, almost crystal-like structure. This process is associated with the cessation of transcription and involves nuclear protein exchange. The histones are mostly displaced, and replaced by protamines (small, arginine-rich proteins).

Nobel Prizes

The following scientists were recognized for their contributions to chromatin research with Nobel Prizes:

Year	Who	Award
1910	Albrecht Kossel (University of Heidelberg)	Nobel Prize in Physiology or Medicine "in recognition of the contributions to our knowledge of cell chemistry made through his work on proteins, including the nucleic substances"
1933	Thomas Hunt Morgan (California Institute of Technology)	Nobel Prize in Physiology or Medicine "for his discoveries concerning the role played by the chromosome in heredity"
1962	Francis Crick, James Watson and Maurice Wilkins (MRC Laboratory of Molecular Biology, Harvard	Nobel Prize in Physiology or Medicine "for their discoveries concerning the molecular structure of nucleic acids and its significance

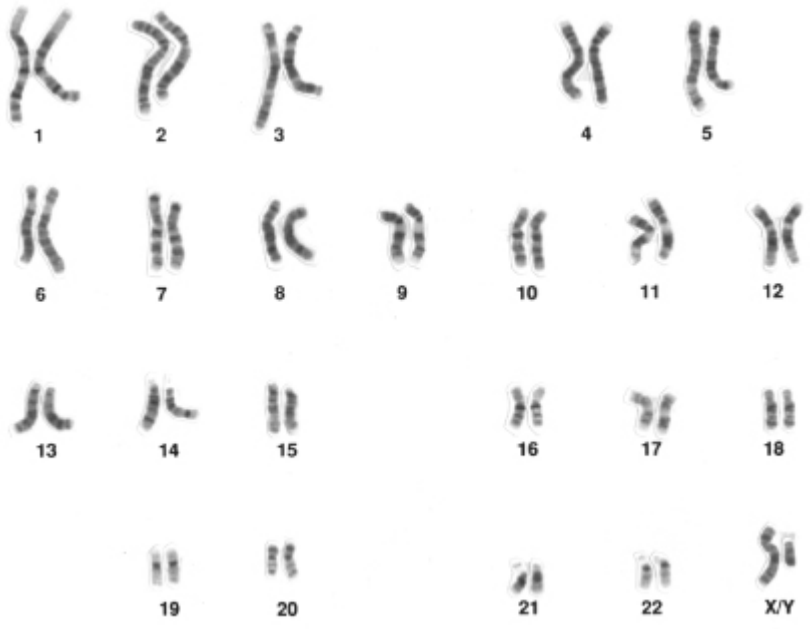
	University and London University respectively)	for information transfer in living material"
1982	Aaron Klug (MRC Laboratory of Molecular Biology)	Nobel Prize in Chemistry "for his development of crystallographic electron microscopy and his structural elucidation of biologically important nucleic acid-protein complexes"
1993	Roberts and Sharp	Nobel Prize in Physiology "for their independent discoveries of split genes"
2006	Roger Kornberg (Stanford University)	Nobel Prize in Chemistry "for his studies of the molecular basis of eukaryotic transcription"

Chapter 3

Karyotype

A **karyotype** is the number and appearance of chromosomes in the nucleus of a eukaryotic cell. The term is also used for the complete set of chromosomes in a species, or an individual organism.

Karyotypes describe the number of chromosomes, and what they look like under a light microscope. Attention is paid to their length, the position of the centromeres, banding pattern, any differences between the sex chromosomes, and any other physical characteristics. The preparation and study of karyotypes is part of cytogenetics.



Karyogram of human male using Giemsa staining.

The study of whole sets of chromosomes is sometimes known as *karyology*. The chromosomes are depicted (by rearranging a microphotograph) in a standard format

known as a *karyogram* or *idiogram*: in pairs, ordered by size and position of centromere for chromosomes of the same size.

The basic number of chromosomes in the somatic cells of an individual or a species is called the *somatic number* and is designated *2n*. Thus, in humans 2n = 46. In the germ-line (the sex cells) the chromosome number is *n* (humans: n = 23).

So, in normal diploid organisms, autosomal chromosomes are present in two copies. There may, or may not, be sex chromosomes. Polyploid cells have multiple copies of chromosomes and haploid cells have single copies.

Karyotypes can be used for many purposes; such as, to study chromosomal aberrations, cellular function, taxonomic relationships, and to gather information about past evolutionary events.

History of karyotype studies

Chromosomes were first observed in plant cells by Karl Wilhelm von Nägeli in 1842. Their behavior in animal (salamander) cells was described by Walther Flemming, the discoverer of mitosis, in 1882. The name was coined by another German anatomist, von Waldeyer in 1888.

The next stage took place after the development of genetics in the early 20th century, when it was appreciated that the set of chromosomes (the karyotype) was the carrier of the genes. Levitsky seems to have been the first to define the karyotype as the phenotypic appearance of the somatic chromosomes, in contrast to their genic contents. The subsequent history of the concept can be followed in the works of Darlington and White.

Investigation into the human karyotype took many years to settle the most basic question: how many chromosomes does a normal diploid human cell contain? In 1912, Hans von Winiwarter reported 47 chromosomes in spermatogonia and 48 in oogonia, concluding an XX/XO sex determination mechanism. Painter in 1922 was not certain whether the diploid number of humans was 46 or 48, at first favouring 46. He revised his opinion later from 46 to 48, and he correctly insisted on humans having an XX/XY system. Considering their techniques, these results were quite remarkable.

New techniques were needed to definitively solve the problem:

1. Using cells in culture
2. Pretreating cells in a hypotonic solution, which swells them and spreads the chromosomes
3. Arresting mitosis in metaphase by a solution of colchicine
4. Squashing the preparation on the slide forcing the chromosomes into a single plane
5. Cutting up a photomicrograph and arranging the result into an indisputable karyogram.

It took until the mid 1950s until it became generally accepted that the karyotype of humans included only 46 chromosomes. Rather interestingly, the great apes have 48 chromosomes. Human chromosome 2 was formed by a merger of ancestral chromosomes, reducing the number.

Observations on karyotypes

Staining

The study of karyotypes is made possible by staining. Usually, a suitable dye, such as Giemsa, is applied after cells have been arrested during cell division by a solution of colchicine. For humans, white blood cells are used most frequently because they are easily induced to divide and grow in tissue culture. Sometimes observations may be made on non-dividing (interphase) cells. The sex of an unborn fetus can be determined by observation of interphase cells.

Observations

Six different characteristics of karyotypes are usually observed and compared:

1. Differences in absolute sizes of chromosomes. Chromosomes can vary in absolute size by as much as twenty-fold between genera of the same family: *Lotus tenuis* and *Vicia faba* (legumes), both have six pairs of chromosomes (n=6) yet *V. faba* chromosomes are many times larger. This feature probably reflects different amounts of DNA duplication.
2. Differences in the position of centromeres. This is brought about by translocations.
3. Differences in relative size of chromosomes can only be caused by segmental interchange of unequal lengths.
4. Differences in basic number of chromosomes may occur due to successive unequal translocations which finally remove all the essential genetic material from a chromosome, permitting its loss without penalty to the organism (the dislocation hypothesis). Humans have one pair fewer chromosomes than the great apes, but the genes have been mostly translocated (added) to other chromosomes.
5. Differences in number and position of satellites, which (when they occur) are small bodies attached to a chromosome by a thin thread.
6. Differences in degree and distribution of heterochromatic regions. Heterochromatin stains darker than euchromatin, indicating tighter packing, and mainly consists of genetically inactive repetitive DNA sequences.

A full account of a karyotype may therefore include the number, type, shape and banding of the chromosomes, as well as other cytogenetic information.

Variation is often found:

1. Between the sexes

2. Between the germ-line and soma (between gametes and the rest of the body)
3. Between members of a population (chromosome polymorphism)
4. Geographical variation between races
5. Mosaics or otherwise abnormal individuals.

The human karyotype

Most (but not all) species have a standard karyotype. The normal human karyotypes contain 22 pairs of autosomal chromosomes and one pair of sex chromosomes. Normal karyotypes for females contain two X chromosomes and are denoted 46,XX; males have both an X and a Y chromosome denoted 46,XY. Any variation from the standard karyotype may lead to developmental abnormalities.

Diversity and evolution of karyotypes

Although the replication and transcription of DNA is highly standardized in eukaryotes, the same cannot be said for their karyotypes, which are highly variable. There is variation between species in chromosome number, and in detailed organization, despite their construction from the same macromolecules. This variation provides the basis for a range of studies in evolutionary cytology.

In some cases there is even significant variation within species. In a review, Godfrey and Masters conclude:

"In our view, it is unlikely that one process or the other can independently account for the wide range of karyotype structures that are observed... But, used in conjunction with other phylogenetic data, karyotypic fissioning may help to explain dramatic differences in diploid numbers between closely related species, which were previously inexplicable.

Although much is known about karyotypes at the descriptive level, and it is clear that changes in karyotype organization has had effects on the evolutionary course of many species, it is quite unclear what the general significance might be.

"We have a very poor understanding of the causes of karyotype evolution, despite many careful investigations... the general significance of karyotype evolution is obscure." Maynard Smith.

Changes during development

Instead of the usual gene repression, some organisms go in for large-scale elimination of heterochromatin, or other kinds of visible adjustment to the karyotype.

- Chromosome elimination. In some species, as in many sciarid flies, entire chromosomes are eliminated during development.

- Chromatin diminution (founding father: Theodor Boveri). In this process, found in some copepods and roundworms such as *Ascaris suum*, portions of the chromosomes are cast away in particular cells. This process is a carefully organised genome rearrangement where new telomeres are constructed and certain heterochromatin regions are lost. In *A. suum*, all the somatic cell precursors undergo chromatin diminution.

- X-inactivation. The inactivation of one X chromosome takes place during the early development of mammals. In placental mammals, the inactivation is random as between the two Xs; thus the mammalian female is a mosaic in respect of her X chromosomes. In marsupials it is always the paternal X which is inactivated. In human females some 15% of somatic cells escape inactivation.

Number of chromosomes in a set

A spectacular example of variability between closely related species is the muntjac, which was investigated by Kurt Benirschke and his colleague Doris Wurster. The diploid number of the Chinese muntjac, *Muntiacus reevesi*, was found to be 46, all telocentric. When they looked at the karyotype of the closely related Indian muntjac, *Muntiacus muntjak*, they were astonished to find it had female = 6, male = 7 chromosomes.

"They simply could not believe what they saw... They kept quiet for two or three years because they thought something was wrong with their tissue culture... But when they obtained a couple more specimens they confirmed [their findings]" Hsu p73-4

The number of chromosomes in the karyotype between (relatively) unrelated species is hugely variable. The low record is held by the nematode *Parascaris univalens*, where the haploid n = 1; the high record would be somewhere amongst the ferns, with the Adder's Tongue Fern *Ophioglossum* ahead with an average of 1262 chromosomes. Top score for animals might be the shortnose sturgeon *Acipenser brevirostrum* at a mere 372 chromosomes. The existence of supernumerary or B chromosomes means that chromosome number can vary even within one interbreeding population; and aneuploids are another example, though in this case they would not be regarded as normal members of the population.

Fundamental number

The fundamental number, *FN*, of a karyotype is the number of visible chromosomal arms per set of chromosomes. Thus, FN ≤ 2n, the difference depending on the number of chromosomes considered single-armed (acrocentric or telocentric) present. Humans have FN = 82, due to the presence of five acrocentric chromosome pairs (13, 14, 15, 21 and 22).

Ploidy

Ploidy is the number of complete sets of chromosomes in a cell.

- Polyploidy, where there are more than two sets of homologous chromosomes in the cells, occurs mainly in plants. It has been of major significance in plant evolution according to Stebbins. The proportion of flowering plants which are polyploid was estimated by Stebbins to be 30-35%, but in grasses the average is much higher, about 70%. Polyploidy in lower plants (ferns, horsetails and psilotales) is also common, and some species of ferns have reached levels of polyploidy far in excess of the highest levels known in flowering plants. Polyploidy in animals is much less common, but it has been significant in some groups.
  Polyploid series in related species which consist entirely of multiples of a single basic number are known as euploid.

- Haplo-diploidy, where one sex is diploid, and the other haploid. It is a common arrangement in the Hymenoptera, and in some other groups.

- Endopolyploidy occurs when in adult differentiated tissues the cells have ceased to divide by mitosis, but the nuclei contain more than the original somatic number of chromosomes. In the *endocycle* (endomitosis or endoreduplication) chromosomes in a 'resting' nucleus undergo reduplication, the daughter chromosomes separating from each other inside an *intact* nuclear membrane. In many instances, endopolyploid nuclei contain tens of thousands of chromosomes (which cannot be exactly counted). The cells do not always contain exact multiples (powers of two), which is why the simple definition 'an increase in the number of chromosome sets caused by replication without cell division' is not quite accurate.
  This process (especially studied in insects and some higher plants such as maize) may be a developmental strategy for increasing the productivity of tissues which are highly active in biosynthesis.
  The phenomenon occurs sporadically throughout the eukaryote kingdom from protozoa to man; it is diverse and complex, and serves differentiation and morphogenesis in many ways.

Aneuploidy

Aneuploidy is the condition in which the chromosome number in the cells is not the typical number for the species. This would give rise to a chromosome abnormality such as an extra chromosome or one or more chromosomes lost. Abnormalities in chromosome number usually cause a defect in development. Down syndrome and Turner syndrome are examples of this.

Aneuploidy may also occcur within a group of closely related species. Classic examples in plants are the genus *Crepis*, where the gametic (= haploid) numbers form the series x = 3, 4, 5, 6, and 7; and *Crocus*, where every number from x = 3 to x = 15 is represented by at least one species. Evidence of various kinds shows that that trends of evolution have gone in different directions in different groups. Closer to home, the great apes have 24x2

chromosomes whereas humans have 23x2. Human chromosome 2 was formed by a merger of ancestral chromosomes, reducing the number.

Chromosomal polymorphism

Some animal species are polymorphic for chromosome fusions or dissociations. When this happens, the chromosome number is variable from one individual to another. Well-researched examples are the ladybird beetle *Chilocorus stigma*, some mantids of the genus *Ameles*, the European shrew *Sorex araneus*. There is some evidence from the case of the mollusc *Thais lapillus* (the dog whelk) on the Brittany coast, that the two chromosome morphs are adapted to different habitats.

Species trees

The detailed study of chromosome banding in insects with polytene chromosomes can reveal relationships between closely related species: the classic example is the study of chromosome banding in Hawaiian drosophilids by Hampton Carson.

In about 6,500 sq mi (17,000 km^2), the Hawaiian Islands have the most diverse collection of drosophilid flies in the world, living from rainforests to subalpine meadows. These roughly 800 Hawaiian drosophilid species are usually assigned to two genera, *Drosophila* and *Scaptomyza*, in the family Drosophilidae.

The polytene banding of the 'picture wing' group, the best-studied group of Hawaiian drosophilids, enabled Carson to work out the evolutionary tree long before genome analysis was practicable. In a sense, gene arrangements are visible in the banding patterns of each chromosome. Chromosome rearrangements, especially inversions, make it possible to see which species are closely related.

The results are clear. The inversions, when plotted in tree form (and independent of all other information), show a clear "flow" of species from older to newer islands. There are also cases of colonization back to older islands, and skipping of islands, but these are much less frequent. Using K-Ar dating, the present islands date from 0.4 million years ago (mya) (Mauna Kea) to 10mya (Necker). The oldest member of the Hawaiian archipelago still above the sea is Kure Atoll, which can be dated to 30 mya. The archipelago itself (produced by the Pacific plate moving over a hot spot) has existed for far longer, at least into the Cretaceous. Previous islands now beneath the sea (guyots) form the Emperor Seamount Chain.

All of the native *Drosophila* and *Scaptomyza* species in Hawai‘i have apparently descended from a single ancestral species that colonized the islands, probably 20 million years ago. The subsequent adaptive radiation was spurred by a lack of competition and a wide variety of niches. Although it would be possible for a single gravid female to colonise an island, it is more likely to have been a group from the same species.

There are other animals and plants on the Hawaiian archipelago which have undergone similar, if less spectacular, adaptive radiations.

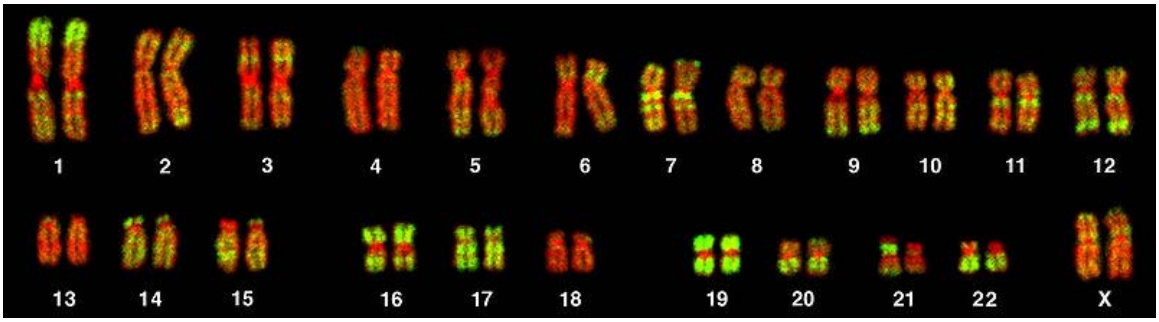
Depiction of karyotypes

Types of banding

Cytogenetics employs several techniques to visualize different aspects of chromosomes:

- G-banding is obtained with Giemsa stain following digestion of chromosomes with trypsin. It yields a series of lightly and darkly stained bands - the dark regions tend to be heterochromatic, late-replicating and AT rich. The light regions tend to be euchromatic, early-replicating and GC rich. This method will normally produce 300-400 bands in a normal, human genome.
- R-banding is the reverse of G-banding (the R stands for "reverse"). The dark regions are euchromatic (guanine-cytosine rich regions) and the bright regions are heterochromatic (thymine-adenine rich regions).
- C-banding: Giemsa binds to constitutive heterochromatin, so it stains centromeres.
- Q-banding is a fluorescent pattern obtained using quinacrine for staining. The pattern of bands is very similar to that seen in G-banding.
- T-banding: visualize telomeres.
- Silver staining: Silver nitrate stains the nucleolar organization region-associated protein. This yields a dark region where the silver is deposited, denoting the activity of rRNA genes within the NOR.

Classic karyotype cytogenetics

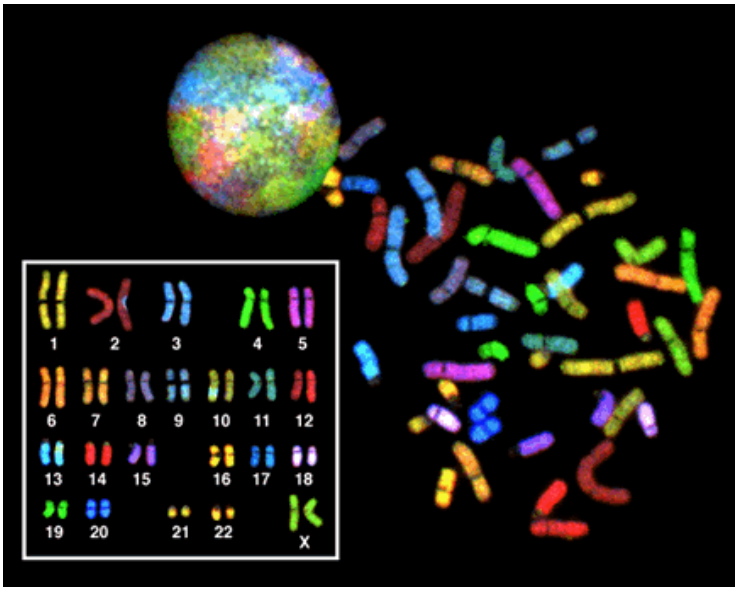


Karyogram from a human female lymphocyte probed for the Alu sequence using FISH.

In the "classic" (depicted) karyotype, a dye, often Giemsa *(G-banding)*, less frequently Quinacrine, is used to stain bands on the chromosomes. Giemsa is specific for the phosphate groups of DNA. Quinacrine binds to the adenine-thymine-rich regions. Each chromosome has a characteristic banding pattern that helps to identify them; both chromosomes in a pair will have the same banding pattern.

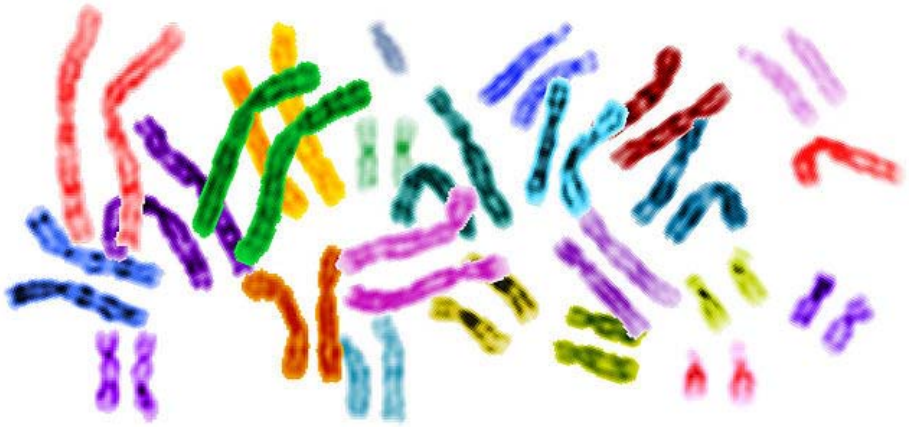
Karyotypes are arranged with the short arm of the chromosome on top, and the long arm on the bottom. Some karyotypes call the short and long arms *p* and *q*, respectively. In addition, the differently stained regions and sub-regions are given numerical designations from proximal to distal on the chromosome arms. For example, Cri du chat syndrome involves a deletion on the short arm of chromosome 5.

Spectral karyotype (SKY technique)



Spectral karyogram of a human female

Spectral karyotyping is a molecular cytogenetic technique used to simultaneously visualize all the pairs of chromosomes in an organism in different colors. Fluorescently labeled probes for each chromosome are made by labeling chromosome-specific DNA with different fluorophores. Because there are a limited number of spectrally-distinct fluorophores, a combinatorial labeling method is used to generate many different colors. Spectral differences generated by combinatorial labeling are captured and analyzed by using an interferometer attached to a fluorescence microscope. Image processing software then assigns a pseudo color to each spectrally different combination, allowing the visualization of the individually colored chromosomes.



Spectral human karyotype

This technique is used to identify structural chromosome aberrations in cancer cells and other disease conditions when Giemsa banding or other techniques are not accurate enough.

Digital karyotyping

Digital karyotyping is a technique used to quantify the DNA copy number on a genomic scale. Short sequences of DNA from specific loci all over the genome are isolated and enumerated. This is method is also known as virtual karyotyping.

Chromosome abnormalities

Chromosome abnormalities can be numerical, as in the presence of extra or missing chromosomes, or structural, as in derivative chromosome, translocations, inversions, large-scale deletions or duplications. Numerical abnormalities, also known as aneuploidy, often occur as a result of nondisjunction during meiosis in the formation of a gamete; trisomies, in which three copies of a chromosome are present instead of the usual two, are common numerical abnormalities. Structural abnormalities often arise from errors in homologous recombination. Both types of abnormalities can occur in gametes and therefore will be present in all cells of an affected person's body, or they can occur during mitosis and give rise to a genetic mosaic individual who has some normal and some abnormal cells.

Chromosomal abnormalities that lead to disease in humans include

- Turner syndrome results from a single X chromosome (45, X or 45, X0).
- Klinefelter syndrome, the most common male chromosomal disease, otherwise known as 47, XXY is caused by an extra **X** chromosome.

- Edwards syndrome is caused by trisomy (three copies) of chromosome 18.
- Down syndrome, a common chromosomal disease, is caused by trisomy of chromosome 21.
- Patau syndrome is caused by trisomy of chromosome 13.
- Also documented are trisomy 8, trisomy 9 and trisomy 16, although they generally do not survive to birth.

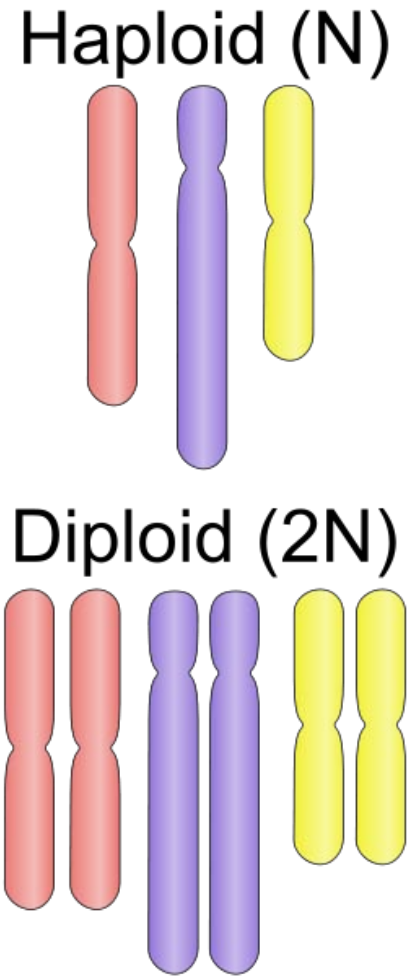
Some disorders arise from loss of just a piece of one chromosome, including

- Cri du chat (cry of the cat), from a truncated short arm on chromosome 5. The name comes from the babies' distinctive cry, caused by abnormal formation of the larynx.
- 1p36 Deletion syndrome, from the loss of part of the short arm of chromosome 1.
- Angelman syndrome – 50% of cases have a segment of the long arm of chromosome 15 missing; a deletion of the maternal genes, example of imprinting disorder.
- Prader-Willi syndrome – 50% of cases have a segment of the long arm of chromosome 15 missing; a deletion of the paternal genes, example of imprinting disorder.

Chromosomal abnormalities can also occur in cancerous cells of an otherwise genetically normal individual; one well-documented example is the Philadelphia chromosome, a translocation mutation commonly associated with chronic myelogenous leukemia and less often with acute lymphoblastic leukemia.

Chapter 4

Ploidy



Diploid cells have two homologous copies of each chromosome.

Ploidy is the number of sets of chromosomes in a biological cell.

Human sex cells (sperm and egg) have one complete set of chromosomes from the male or female parent. Sex cells, also called gametes, combine to produce somatic cells. Somatic cells therefore have twice as many chromosomes. The **haploid number** (n) is the number of chromosomes in a gamete. A somatic cell has twice that many chromosomes (2n).

Humans are **diploid**. A human somatic cell contains 46 chromosomes: 2 complete haploid sets, which make up 23 homologous chromosome pairs. However, many organisms have more than two sets of homologous chromosomes and are called polyploid.

The number of chromosomes in a single (non-homologous) set is called the **monoploid number** (x), and is different from the haploid number (n). Both numbers n, and x, apply to every cell of a given organism. For humans, x = n = 23, which is also written as 2n = 2x = 46. Bread wheat is an organism where x and n differ. It has six sets of chromosomes, two sets from each of three different diploid species that are its distant ancestors. The somatic cells are **hexaploid**, with six sets of chromosomes, 2n = 6x = 42. The gametes are both haploid and **triploid**, with three sets of chromosomes. The monoploid number x = 7, and the haploid number n = 21.

Tetraploidy (four sets of chromosomes, 2n = 4x) is common in plants, and also occurs in amphibians, reptiles, and insects.

The Australian bulldog ant, *Myrmecia pilosula*, a haplodiploid species, has n = x = 1, the lowest chromosome number theoretically possible. Haploid individuals of this species have a single chromosome, and diploid individuals have two chromosomes.

Euploidy is the state of a cell or organism having an integral multiple of the monoploid number, possibly excluding the sex-determining chromosomes. For example, a human cell has 46 chromosomes, which is an integer multiple of the monoploid number, 23. A human with abnormal, but integral, multiples of this full set (e.g. 69 chromosomes) would also be considered as euploid. **Aneuploidy** is the state of not having euploidy. In humans, examples include having a single extra chromosome (such as Down syndrome), or missing a chromosome (such as Turner syndrome). Aneuploid karyotypes are given names with the suffix *-somy* (rather than *-ploidy*, used for euploid karyotypes), such as trisomy and monosomy.

Etymology

The term *ploidy* is a back-formation from *haploid* and *diploid*. These two terms are from Greek ἁπλόος *haplóos* "single" and διπλόος *diplóos* "double" combined with εἶδος *eîdos* "form" (compare *idol* from Latin *īdōlum*, that from Greek εἴδωλον *eídōlon* derived from εἶδος *eîdos*). The two *haploid* and *diploid* terms were borrowed from German through William Henry Lang's 1908 translation of a 1894 textbook by Eduard Strasburger and colleagues. Strasburger used diploid to refer to an organism with twice the number of chromosomes of a haploid organism, hence "double" and "single".

Haploid and monoploid

As stated above, the **haploid number** (n) is the number of chromosomes in a gamete of an individual, and this is distinct from the monoploid number (x) which is the number of unique chromosomes in a single complete set. Gametes (sperm, and ova) are haploid cells. The haploid gametes produced by (most) diploid organisms are monoploid, and these can combine to form a diploid zygote. For example, most animals are diploid and produce monoploid gametes.

During meiosis, sex cell precursors have their number of chromosomes halved by randomly "choosing" one homologue, resulting in haploid gametes. Because homologous chromosomes usually differ genetically, gametes usually differ genetically from one another.

All plants and many fungi and algae switch between a haploid and a diploid state (which may be polyploid), with one of the stages emphasized over the other. This is called alternation of generations. Most fungi and algae are haploid during the principal stage of their life cycle.

Male bees, wasps, and ants are haploid organisms because of the way they develop from unfertilized, haploid egg cells.

In humans, the monoploid number (x) equals the haploid number (n), x = n = 23, but in some species (especially plants), these numbers differ. Common wheat has six sets of chromosomes in the somatic cells, derived from its three different ancestral species. The gametes of common wheat are considered as haploid since they contain half the genetic information of somatic cells, but are not monoploid as they still contain three complete sets of chromosomes (n = 3x).

Diploid

Diploid (indicated by $2n = 2x$) cells have two homologous copies of each chromosome, usually one from the mother and one from the father. Nearly all mammals are diploid organisms (the viscacha rats *Pipanacoctomys aureus* and *Tympanoctomys barrerae* are the only known exceptions as of 2004), although all individuals have some small fraction of cells that display polyploidy. Human diploid cells have 46 chromosomes and human haploid gametes (egg and sperm) have 23 chromosomes.

Retroviruses that contain two copies of their RNA genome in each viral particle are also said to be diploid. Examples include human foamy virus, human T-lymphotropic virus, and HIV.

Haploidisation

Haploidisation (haploidization) is the process of creating a haploid cell (usually from a diploid cell).

A laboratory procedure called haploidisation forces a normal cell to expel half of its chromosomal complement. In mammals this renders this cell chromosomally equal to sperm or egg. This was one of the procedures used by Japanese researchers to produce Kaguya, a fatherless mouse.

Haploidisation sometimes occurs in plants when meiotically reduced cells (usually egg cells) develop by parthenogenesis.

A rare genetic disorder that has occurred in a total of 7 recorded cases is Detrimental Haploidy Syndrome where the somatic cells of the human body are haploid after the first division of cells from fertilisation. As a result of this a human with this syndrome is unfortunately prone to other diseases and unable to reproduce.

Polyploidy

Polyploidy is the state where all cells have multiple sets of chromosomes beyond the basic set, for example, in triploids $2n = 3x$, in tetraploids $2n = 4x$. The chromosome sets may be from the same species or from closely related species. In the latter case these are known as allopolyploids (or amphidiploids, which are allopolyploids that behave as if they were normal diploids). Allopolyploids are formed from the hybridization of two separate species. In plants, this probably most often occurs from the pairing of meiotically unreduced gametes, and not by diploid–diploid hybridization followed by chromosome doubling. The so-called Brassica triangle is an example of allopolyploidy, where three different parent species have hybridized in all possible pair combinations to produce three new species.

Polyploidy occurs commonly in plants, but rarely in animals. Even in diploid organisms many somatic cells are polyploid due to a process called endoreduplication where duplication of the genome occurs without mitosis (cell division).

The extreme in polyploidy occurs in the fern-ally genus *Ophioglossum*, the adder's-tongues, in which polyploidy results in chromosome counts in the hundreds, or in at least one case, well over one thousand. Interestingly, these plants seem to have simplified structures in their phenotype.

Variable or indefinite ploidy

Depending on growth conditions, prokaryotes such as bacteria may have a chromosome copy number of 1 to 4, and that number is commonly fractional, counting portions of the chromosome partly replicated at a given time. This is because under exponential growth conditions the cells are able to replicate their DNA faster than they can divide.

Mixoploidy

Mixoploidy refers to the presence of two cell lines, one diploid and one polyploid. Though polyploidy in humans is not viable, mixoploidy has been found in live adults and

children. There are two types: diploid-triploid mixoploidy, in which some cells have 46 chromosomes and some have 69, and diploid-tetraploid mixoploidy, in which some cells have 46 and some have 92 chromosomes.

Dihaploidy and Polyhaploidy

Dihaploid and polyhaploid cells are formed by haploidisation of polyploids, i.e., by halving the chromosome constitution.

Dihaploids (which are diploid) are important for selective breeding of tetraploid crop plants (notably potatoes), because selection is faster with diploids than with tetraploids. Tetraploids can be reconstituted from the diploids, for example by somatic fusion.

The term “dihaploid” was coined by Bender to combine in one word the number of genome copies (diploid) and their origin (haploid). The term is well established in this original sense, but it has also been used for doubled monoploids or doubled haploids, which are homozygous and used for genetic research.

Chapter 5

Microchromosome and Chromosomal Inversion

Microchromosome

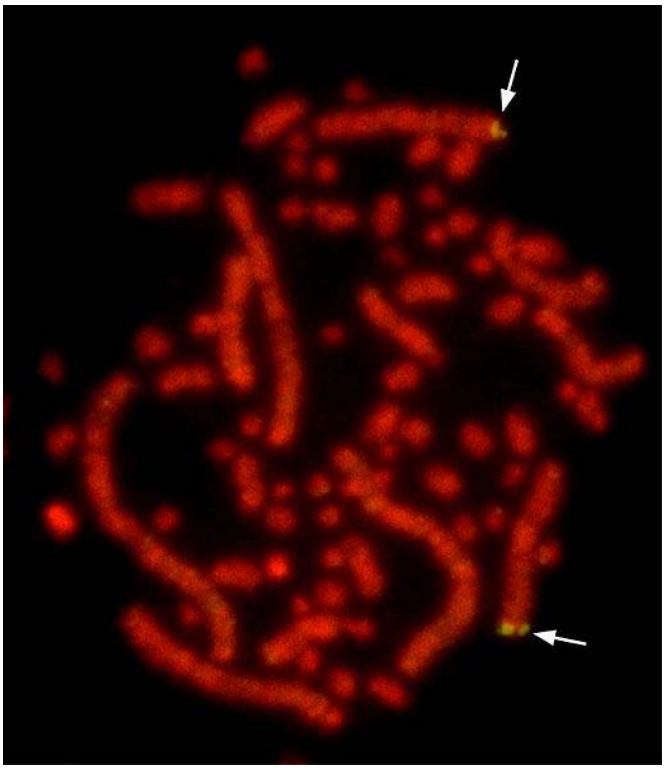


Image of chicken chromosomes featuring the many microchromosomes (appearing as dots). The arrows indicate a stained gene locus on homologous macrochromosomes.

A **microchromosome** is a type of very small chromosome which is a typical component of the karyotype of birds, some reptiles, fish, and amphibians; they tend to be absent in

mammals. They are less than 20 Mb in size; chromosomes which are greater than 40 Mb in size are known as **macrochromosomes**, while those between 20 and 40 Mb are classified as **intermediate chromosomes**. Microchromosomes are characteristically very small and often cytogenetically indistinguishable in a karyotype. While originally thought to be insignificant fragments of chromosomes, in species where they have been studied they have been found to be rich in genes. In chickens, microchromosomes have been estimated to contain between 50 and 75% of all genes. The presence of microchromosomes makes ordering and identifying chromosomes into a coherent karyotype particularly difficult. During metaphase, they appear merely as 0.5-1.5 μm long specks. Their small size and poor condensation into heterochromatin means they generally lack the diagnostic banding patterns and distinct centromere locations used for chromosome identification.

In avians

Birds (except falconiformes) usually have karyotypes of approximately 80 chromosomes (*2n = 80*), with only a few being distinguishable macrochromosomes and an average of 60 being microchromosomes. They are more abundant in avians than any other group of animals. Chickens are an important model organism for studying microchromosomes. Examination of microchromosomes in birds has led to the hypotheses that they may have originated as conserved fragments of ancestral macrochromosomes, and conversely that macrochromosomes could have arisen as aggregates of microchromosomes. Comparative genomic analysis shows that microchromosomes contain genetic information which has been conserved across multiple classes of chromosomes. This indicates that at least ten chicken microchromosomes arose from fission of larger chromosomes and that the typical bird karyotype arose 100-250 mya.

Chickens

Chickens have a diploid number of 78 ($2n = 78$) chromosomes, and as is usual in birds, the majority are microchromosomes. Classification of chicken chromosomes varies by author. Some classify them as 8 pairs of macrochromosomes, one pair of sex chromosomes, with the remaining 32 pairs being intermediate or microchromosomes. Other arrangements such as that used by the International Chicken Genome Sequencing Consortium include five pairs of macrochromosomes, five pairs of intermediate chromosomes, and twenty-eight pairs of microchromosomes. Microchromosomses represent approximately one third of the total genome size, and have been found to have a much higher gene density than macrochromosomes. Because of this, it is estimated that the majority of genes are located on microchromosomes, though due to the difficulty in physically identifying microchromosomes and the lack of microsatellite markers, it has been difficult to place genes on specific microchromosomes.

Replication timing and recombination rates have been found to differ between microchromosomes and macrochromosomes in chickens. Microchromomes replicate earlier in the S phase of interphase than macrochromosomes. Recombination rates have also been found to be higher on microchromosomes. Possibly due to the high

recombination rates, chicken chromosome 16 (a microchromosome) has been found to contain the most genetic diversity of any chromosome in certain chicken breeds. This is likely due to the presence on this chromosome of the major histocompatibility complex (MHC).

For the many small linkage groups in the chicken genome which have not been placed on chromosomes, the assumption has been made that they are located on the microchromosomes. Interestingly, groups of these correspond almost exactly with large sections of certain human chromosomes. For example linkage groups E29C09W09, E21E31C25W12, E48C28W13W27, E41W17, E54 and E49C20W21 correspond with chromosome 7.

Turkey

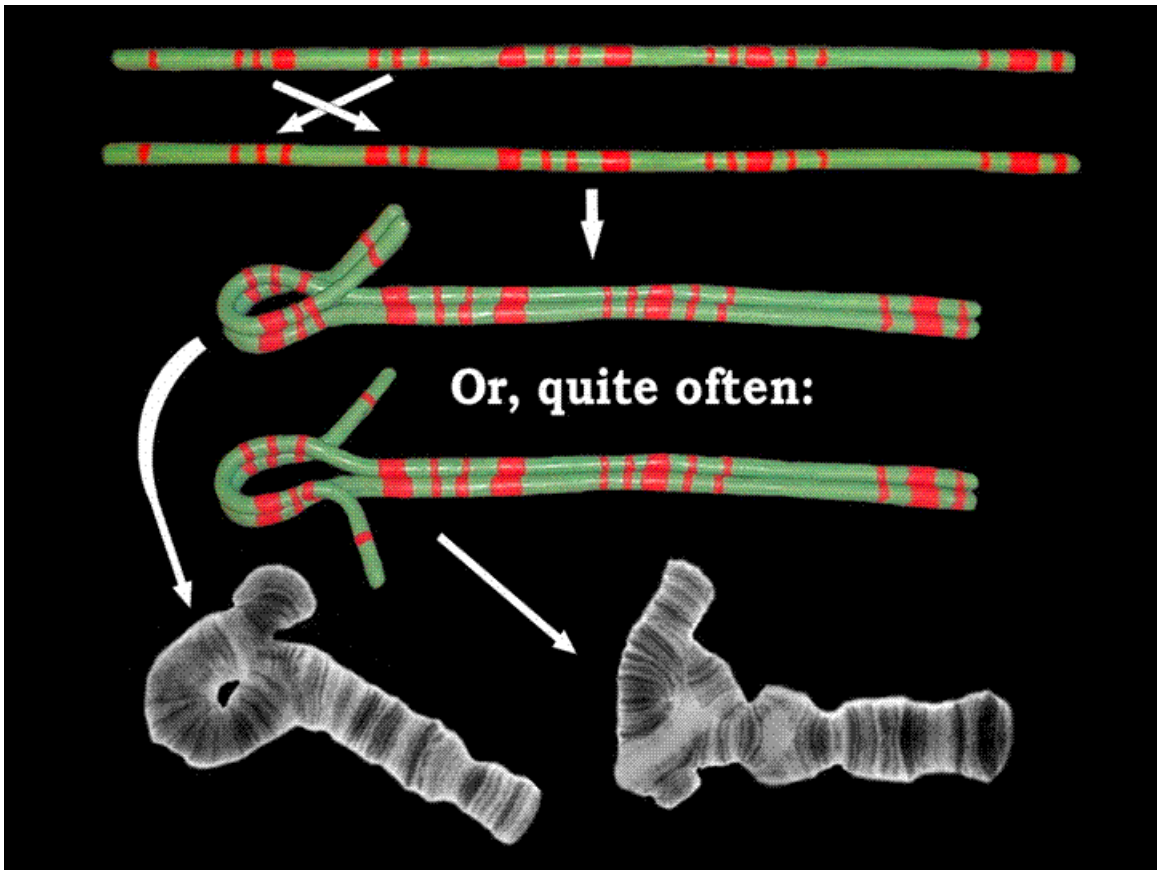
The turkey has a diploid number of 80 ($2n = 80$) chromosomes. The karyotype contains an additional chromosomal pair relative to the chicken due to the presence of at least two fission/fusion differences (GGA2 = MGA3 and MGA6 and GGA4 = MGA4 and MGA9). Given these differences involving the macrochromosomes, an additional fission/fusion must also exist between the species involving the microchromosomes if the diploid numbers are valid. Other rearrangements have been identified through comparative genetic maps, physical maps and whole genome sequencing.

In humans and other animals

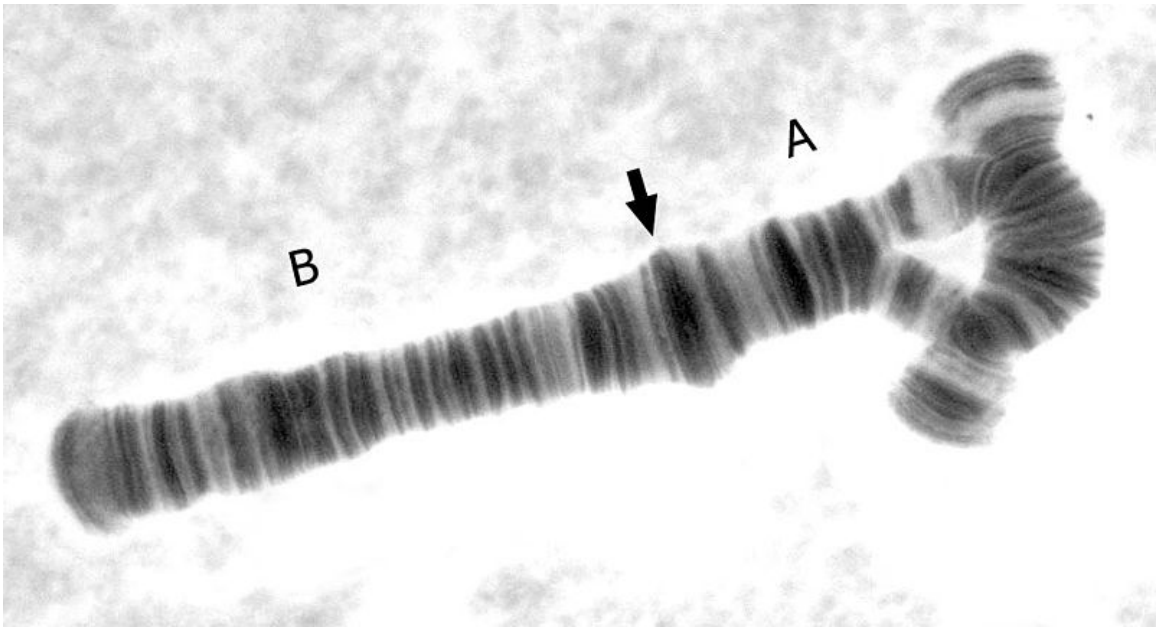
Microchromosomes are absent from the karyotypes of mammals, crocodilians, and frogs.

In rare cases, microchromosomes have been observed in the karotypes of individual humans. A link has been suggested between microchromosome presence and certain genetic disorders like Down syndrome and fragile X syndrome. The smallest chromosome in humans is normally chromosome 21, which is 47 Mb.

Chromosomal inversion



A clay model showing why heterozygous inversion loops are visible in polytene chromosome preparations



An inversion loop in the A arm of a chromosome from an *Axarus* species midge

An **inversion** is a chromosome rearrangement in which a segment of a chromosome is reversed end to end. An inversion occurs when a single chromosome undergoes breakage and rearrangement within itself. Inversions are of two types: **paracentric** and **pericentric**.

Paracentric inversions do not include the centromere and both breaks occur in one arm of the chromosome. Pericentric inversions include the centromere and there is a break point in each arm.

Cytogenetic techniques may be able to detect inversions, or inversions may be inferred from genetic analysis. Nevertheless, in most species small inversions go undetected. In insects with polytene chromosomes, for example Drosophila, preparations of larval salivary gland chromosomes allow inversions to be seen when they are heterozygous. This useful characteristic of polytene chromosomes was first advertised by Theophilus Shickel Painter in 1933.

Inversions usually do not cause any abnormalities in carriers as long as the rearrangement is balanced with no extra or missing genetic information. However, in individuals which are heterozygous for an inversion, there is an increased production of abnormal chromatids (this occurs when crossing-over occurs within the span of the inversion). This leads to lowered fertility due to production of unbalanced gametes.

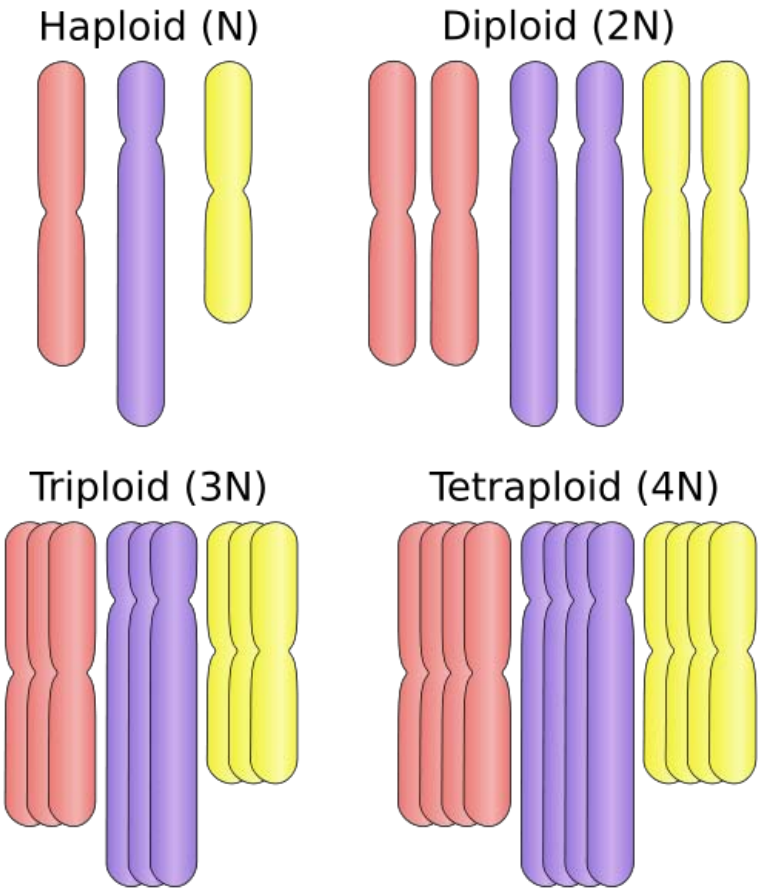
The most common inversion seen in humans is on chromosome 9, at inv(9)(p11q12). This inversion is generally considered to have no deleterious or harmful effects, but there is some evidence it leads to an increased risk for miscarriage for about 30% of affected couples.

An Inversion does not involve a loss of genetic information, but simply rearrange the linear gene sequence.

Families that may be carriers of inversions may be offered genetic counseling and genetic testing.

Chapter 6

Polyploid



This image shows haploid (single), diploid (double), triploid (triple), and tetraploid (quadruple) sets of chromosomes. Triploid and tetraploid chromosomes are examples of polyploidy.

Polyploid is a term used to describe cells and organisms containing more than two paired (homologous) sets of chromosomes. Most species are diploid, meaning they have two sets of chromosomes — one set inherited from each parent. However **polyploidy** is found in some organisms and is especially common in plants. In addition, polyploidy also

occurs in some tissues of animals who are otherwise diploid, such as human muscle tissues. This is known as **endopolyploidy**. (Monoploid organisms also occur; a monoploid has only one set of chromosomes.)

Polyploidy refers to a numerical change in a whole set of chromosomes. Organisms in which a particular chromosome, or chromosome segment, is under- or overrepresented are said to be **aneuploid** (from the Greek words meaning "not," "good," and "fold"). Therefore the distinction between aneuploidy and polyploidy is that aneuploidy refers to a numerical change in part of the chromosome set, whereas polyploidy refers to a numerical change in the whole set of chromosomes.

Polyploidy may occur due to abnormal cell division, either during mitosis, or commonly during metaphase I in meiosis.

Polyploidy occurs in some animals, such as goldfish, salmon, and salamanders, but is especially common among ferns and flowering plants, including both wild and cultivated species. Wheat, for example, after millennia of hybridization and modification by humans, has strains that are **diploid** (two sets of chromosomes), **tetraploid** (four sets of chromosomes) with the common name of durum or macaroni wheat, and **hexaploid** (six sets of chromosomes) with the common name of bread wheat. Many agriculturally important plants of the genus *Brassica* are also tetraploids. Polyploidization is a mechanism of sympatric speciation because polyploids are usually unable to interbreed with their diploid ancestors.

Polyploidy can be induced in plants and cell cultures by some chemicals: the best known is colchicine, which can result in chromosome doubling, though its use may have other less obvious consequences as well. Oryzalin also will double the existing chromosome content.

Polyploid types

Polyploid types are labeled according to the number of chromosome sets in the nucleus:

- **triploid** (three sets; 3x), for example seedless watermelons, common in the phylum Tardigrada
- **tetraploid** (four sets; 4x), for example Salmonidae fish
- **pentaploid** (five sets; 5x), for example Kenai Birch (*Betula papyrifera* var. *kenaica*)
- **hexaploid** (six sets; 6x), for example wheat, kiwifruit
- **octaploid** (eight sets; 8x), for example *Acipenser* (genus of sturgeon fish)
- **decaploid** (ten sets; 10x), for example certain strawberries
- **dodecaploid** (twelve sets; 12x), for example the plant *Celosia argentea* and the amphibian *Xenopus ruwenzoriensis*

Polyploidy in animals (non-human)

Examples in animals are more common in the 'lower' forms such as flatworms, leeches, and brine shrimp. Polyploid animals are often sterile, so they often reproduce by parthenogenesis. Polyploid lizards are also quite common and parthenogenetic. Polyploid mole salamanders (mostly triploids) are all female and reproduce by kleptogenesis, "stealing" spermatophores from diploid males of related species to trigger egg development but not incorporating the males' DNA into the offspring. While mammalian liver cells are polyploid, rare instances of polyploid mammals are known, but most often result in prenatal death.

One of the few known exceptions to this 'rule' is an octodontid rodent of Argentina's harsh desert regions, known as the Plains Viscacha-Rat (*Tympanoctomys barrerae*). This rodent is not a rat, but kin to guinea pigs and chinchillas. Its "new" diploid [2n] number is 102 and so its cells are roughly twice normal size. Its closest living relation is *Octomys mimax*, the Andean Viscacha-Rat of the same family, whose 2n = 56. It is surmised that an *Octomys*-like ancestor produced tetraploid (i.e., 4n = 112) offspring that were, by virtue of their doubled chromosomes, reproductively isolated from their parents; but that these likely survived the ordinarily catastrophic effects of polyploidy in mammals by shedding (via translocation or some similar mechanism) the "extra" set of sex chromosomes gained at this doubling. (The closely related Golden Vizcacha Rat, 2n = 96, is thought to have arisen via roughly the same process).

Polyploidy in humans

True polyploidy rarely occurs in humans, although it occurs in some tissues (especially in the liver). Aneuploidy is more common.

Polyploidy occurs in humans in the form of triploidy, with 69 chromosomes (sometimes called 69,XXX), and tetraploidy with 92 chromosomes (sometimes called 92,XXXX). Triploidy, usually due to polyspermy, occurs in about 2–3% of all human pregnancies and ~15% of miscarriages. The vast majority of triploid conceptions end as miscarriage and those that do survive to term typically die shortly after birth. In some cases survival past birth may occur longer if there is mixoploidy with both a diploid and a triploid cell population present.

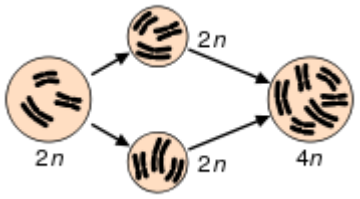
Triploidy may be the result of either digyny (the extra haploid set is from the mother) or diandry (the extra haploid set is from the father). Diandry is mostly caused by reduplication of the paternal haploid set from a single sperm, but may also be the consequence of dispermic (two sperm) fertilization of the egg. Digyny is most commonly caused by either failure of one meiotic division during oogenesis leading to a diploid oocyte or failure to extrude one polar body from the oocyte. Diandry appears to predominate among early miscarriages while digyny predominates among triploidy that survives into the fetal period. However, among early miscarriages, digyny is also more common in those cases <8.5 weeks gestational age or those in which an embryo is present. There are also two distinct phenotypes in triploid placentas and fetuses that are

dependent on the origin of the extra haploid set. In digyny there is typically an asymmetric poorly grown fetus, with marked adrenal hypoplasia and a very small placenta. In diandry, a partial hydatidiform mole develops. These parent-of-origin effects reflect the effects of genomic imprinting.

Complete tetraploidy is more rarely diagnosed than triploidy, but is observed in 1–2% of early miscarriages. However, some tetraploid cells are commonly found in chromosome analysis at prenatal diagnosis and these are generally considered 'harmless'. It is not clear whether these tetraploid cells simply tend to arise during *in vitro* cell culture or whether they are also present in placental cells *in vivo*. There are, at any rate, very few clinical reports of fetuses/infants diagnosed with tetraploidy mosaicism.

Mixoploidy is quite commonly observed in human preimplantation embryos and includes haploid/diploid as well as diploid/tetraploid mixed cell populations. It is unknown whether these embryos fail to implant and are therefore rarely detected in ongoing pregnancies or if there is simply a selective process favoring the diploid cells.

Polyploidy in plants



Speciation via polyploidy: A diploid cell undergoes failed meiosis, producing diploid gametes, which self-fertilize to produce a tetraploid zygote.

Polyploidy is pervasive in plants and some estimates suggest that 30–80% of living plant species are polyploid, and many lineages show evidence of ancient polyploidy (paleopolyploidy) in their genomes. Huge explosions in angiosperm species diversity appear to have coincided with the timing of ancient genome duplications shared by many species. It has been established that 15% of angiosperm and 31% of fern speciation events are accompanied by ploidy increase. Polyploid plants can arise spontaneously in nature by several mechanisms, including meiotic or mitotic failures, and fusion of unreduced (2n) gametes. Both autopolyploids (e.g. potato) and allopolyploids (e.g. canola, wheat, cotton) can be found among both wild and domesticated plant species. Most polyploids display heterosis relative to their parental species, and may display novel variation or morphologies that may contribute to the processes of speciation and eco-niche exploitation. The mechanisms leading to novel variation in newly formed allopolyploids may include gene dosage effects (resulting from more numerous copies of genome content), the reunion of divergent gene regulatory hierarchies, chromosomal rearrangements, and epigenetic remodeling, all of which affect gene content and/or expression levels. Many of these rapid changes may contribute to reproductive isolation and speciation.

Lomatia tasmanica is an extremely rare Tasmanian shrub which is triploid and sterile, and reproduction is entirely vegetative with all plants having the same genetic structure.

There are few naturally occurring polyploid conifers. One example is the giant tree *Sequoia sempervirens* or Coast Redwood which is a hexaploid (6x) with 66 chromosomes (2n = 6x = 66), although the origin is unclear.

Polyploid crops

Polyploid plants tend to be larger and better at flourishing in early succession habitats such as farm fields. In the breeding of crops, the tallest and best thriving plants are selected for. Thus, many crops (and agricultural weeds) may have unintentionally been bred to a higher level of ploidy.

The induction of polyploidy is a common technique to overcome the sterility of a hybrid species during plant breeding. For example, Triticale is the hybrid of wheat (*Triticum turgidum*) and rye (*Secale cereale*). It combines sought-after characteristics of the parents, but the initial hybrids are sterile. After polyploidization, the hybrid becomes fertile and can thus be further propagated to become triticale.

In some situations polyploid crops are preferred because they are sterile. For example many seedless fruit varieties are seedless as a result of polyploidy. Such crops are propagated using asexual techniques such as grafting.

Polyploidy in crop plants is most commonly induced by treating seeds with the chemical colchicine.

Examples of polyploid crops

- Triploid crops: apple, banana, citrus, ginger, watermelon
- Tetraploid crops: apple, durum or macaroni wheat, cotton, potato, cabbage, leek, tobacco, peanut, kinnow, Pelargonium
- Hexaploid crops: chrysanthemum, bread wheat, triticale, oat, kiwifruit
- Octaploid crops: strawberry, dahlia, pansies, sugar cane

Some crops are found in a variety of ploidies: tulips and lilies are commonly found as both diploid and as triploid; daylilies (Hemerocallis cultivars) are available as either diploid or tetraploid; apples and kinnows can be diploid, triploid, or tetraploid.

Terminology

Autopolyploidy

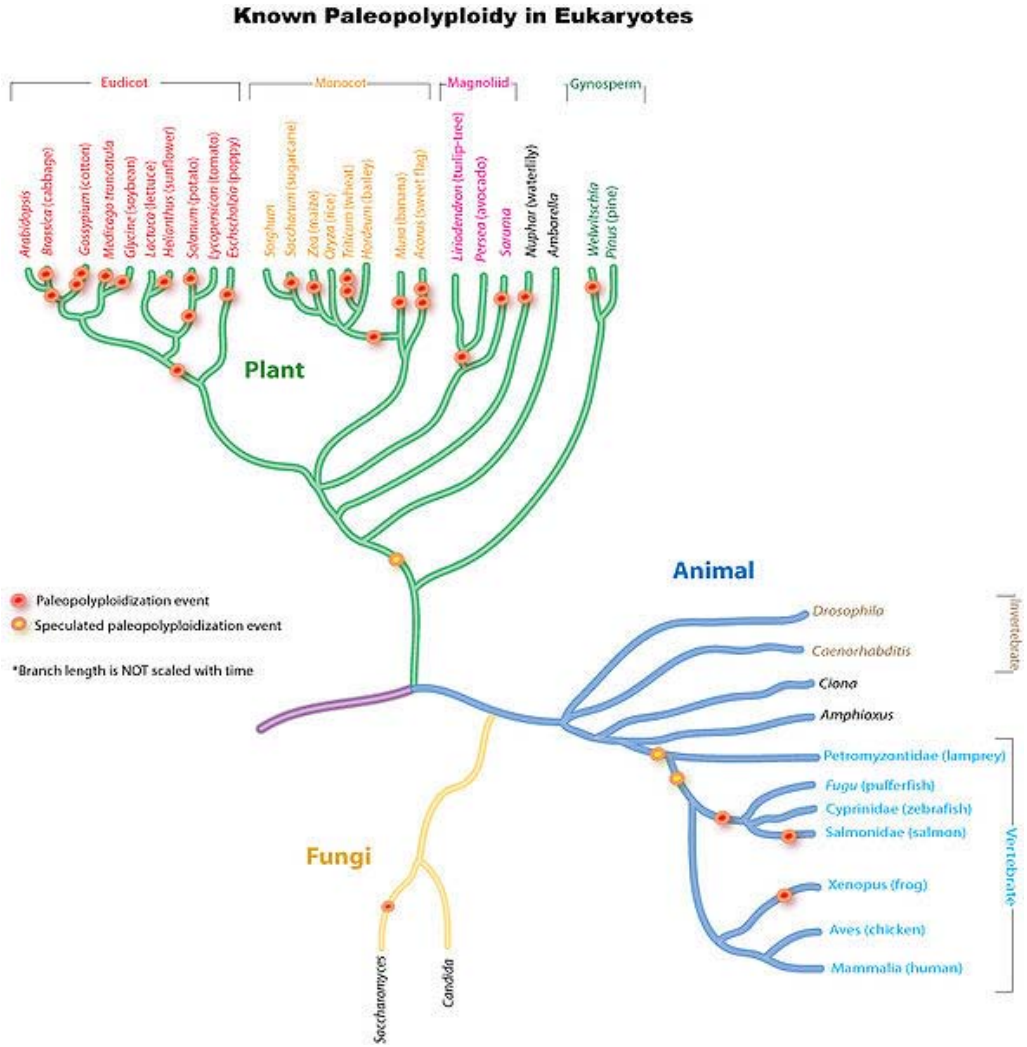
Autopolyploids are polyploids with multiple chromosome sets derived from a single species. Autopolyploids can arise from a spontaneous, naturally occurring genome doubling, like the potato. Others might form following fusion of 2n gametes (unreduced

gametes). Bananas and apples can be found as autotriploids. Autopolyploid plants typically display polysomic inheritance, and are therefore often infertile and propagated clonally perfect.

Allopolyploidy

Allopolyploids are polyploids with chromosomes derived from different species. Precisely it is the result of doubling of chromosome number in an F1 hybrid. *Triticale* is an example of an allopolyploid, having six chromosome sets, allohexaploid, four from wheat (*Triticum turgidum*) and two from rye (*Secale cereale*). *Amphidiploid* is another word for an allopolyploid. Some of the best examples of allopolyploids come from the Brassicas, and the Triangle of U describes the relationships among the three common diploid Brassicas (*B. oleracea, B. rapa,* and *B. nigra*) and three allotetraploids (*B. napus, B. juncea,* and *B. carinata*) derived from hybridization among the diploids.

Paleopolyploidy



This phylogenetic tree shows the relationship between the best-documented instances of paleopolyploidy in eukaryotes.

Ancient genome duplications probably occurred in the evolutionary history of all life. Duplication events that occurred long ago in the history of various evolutionary lineages can be difficult to detect because of subsequent diploidization (such that a polyploid starts to behave cytogenetically as a diploid over time) as mutations and gene translations gradually make one copy of each chromosome unlike its other copy.

In many cases, these events can be inferred only through comparing sequenced genomes. Examples of unexpected but recently confirmed ancient genome duplications include baker's yeast (*Saccharomyces cerevisiae*), mustard weed/thale cress (*Arabidopsis thaliana*), rice (*Oryza sativa*), and an early evolutionary ancestor of the vertebrates (which includes the human lineage) and another near the origin of the teleost fishes. Angiosperms (flowering plants) have paleopolyploidy in their ancestry. All eukaryotes probably have experienced a polyploidy event at some point in their evolutionary history.

Karyotype

A karyotype is the characteristic chromosome complement of a eukaryote species. The preparation and study of karyotypes is part of cytology and, more specifically, cytogenetics.

Although the replication and transcription of DNA is highly standardized in eukaryotes, the same cannot be said for their karotypes, which are highly variable between species in chromosome number and in detailed organization despite being constructed out of the same macromolecules. In some cases there is even significant variation within species. This variation provides the basis for a range of studies in what might be called evolutionary cytology.

Paralogous

The term is used to describe the relationship among duplicated genes or portions of chromosomes that derived from a common ancestral DNA. Paralogous segments of DNA may arise spontaneously by errors during DNA replication, copy and paste transposons, or whole genome duplications.

Homologous

The term is used to describe the relationship of similar chromosomes that pair at mitosis and meiosis. In a diploid, one homolog is derived from the male parent (sperm) and one is derived from the female parent (egg). During meiosis and gametogenesis, homologous chromosomes pair and exchange genetic material by recombination, leading to the production of sperm or eggs with chromosome haplotypes containing novel genetic variation.

Homoeologous

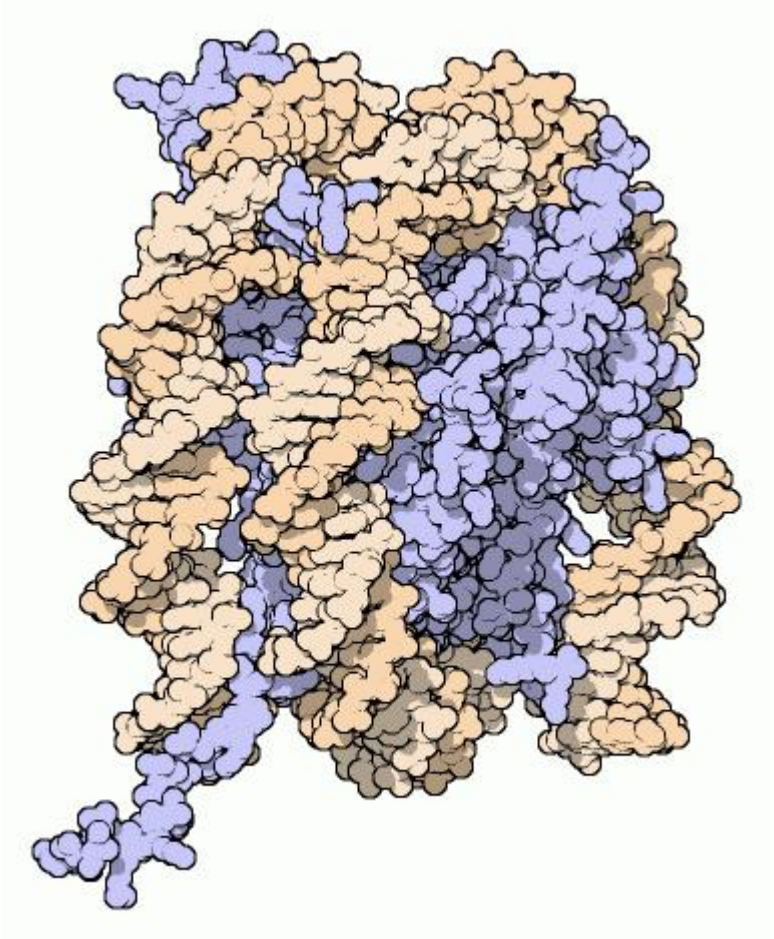
The term *homoeologous*, also spelled *homeologous*, is used to describe the relationship of similar chromosomes or parts of chromosomes brought together following inter-species hybridization and allopolyploidization, and whose relationship was completely homologous in an ancestral species. In allopolyploids, the homologous chromosomes within each parental sub-genome should pair faithfully during meiosis, leading to disomic inheritance; however in some allopolyploids, the homoeologous chromosomes of the parental genomes may be nearly as similar to one another as the homologous chromosomes, leading to tetrasomic inheritance (four chromosomes pairing at meiosis), intergenomic recombination, and reduced fertility.

Example of homoeologous chromosomes

Durum wheat is the result of the inter-species hybridization of two diploid grass species *Triticum urartu* and *Aegilops speltoides*. Both the diploid ancestors had two sets of 7 chromosomes, which were similar in terms of size and genes contained on them. Durum wheat contains two sets of chromosomes derived from *Triticum urartu* and two sets of chromosomes derived from *Aegilops speltoides*. Each chromosome pair derived from the *Triticum urartu* parent is **homoeologous** to the opposite chromosome pair derived from the *Aegilops speltoides* parent, though each chromosome pair unto itself is **homologous**.

Chapter 7

Nucleosome



Nucleosome (DNA in orange and histones in blue)

Nucleosomes are the basic unit of DNA packaging in eukaryotes, consisting of a segment of DNA wound around a histone protein core. This structure is often compared to thread wrapped around a spool.

Nucleosomes form the fundamental repeating units of eukaryotic chromatin, which is used to pack the large eukaryotic genomes into the nucleus while still ensuring

appropriate access to it (in mammalian cells approximately 2 m of linear DNA have to be packed into a nucleus of roughly 10 μm diameter). Nucleosomes are folded through a series of successively higher order structures to eventually form a chromosome; this both compacts DNA and creates an added layer of regulatory control which ensures correct gene expression. Nucleosomes are thought to carry epigenetically inherited information in the form of covalent modifications of their core histones. The nucleosome hypothesis was proposed by Don and Ada Olins in 1974 and Roger Kornberg.

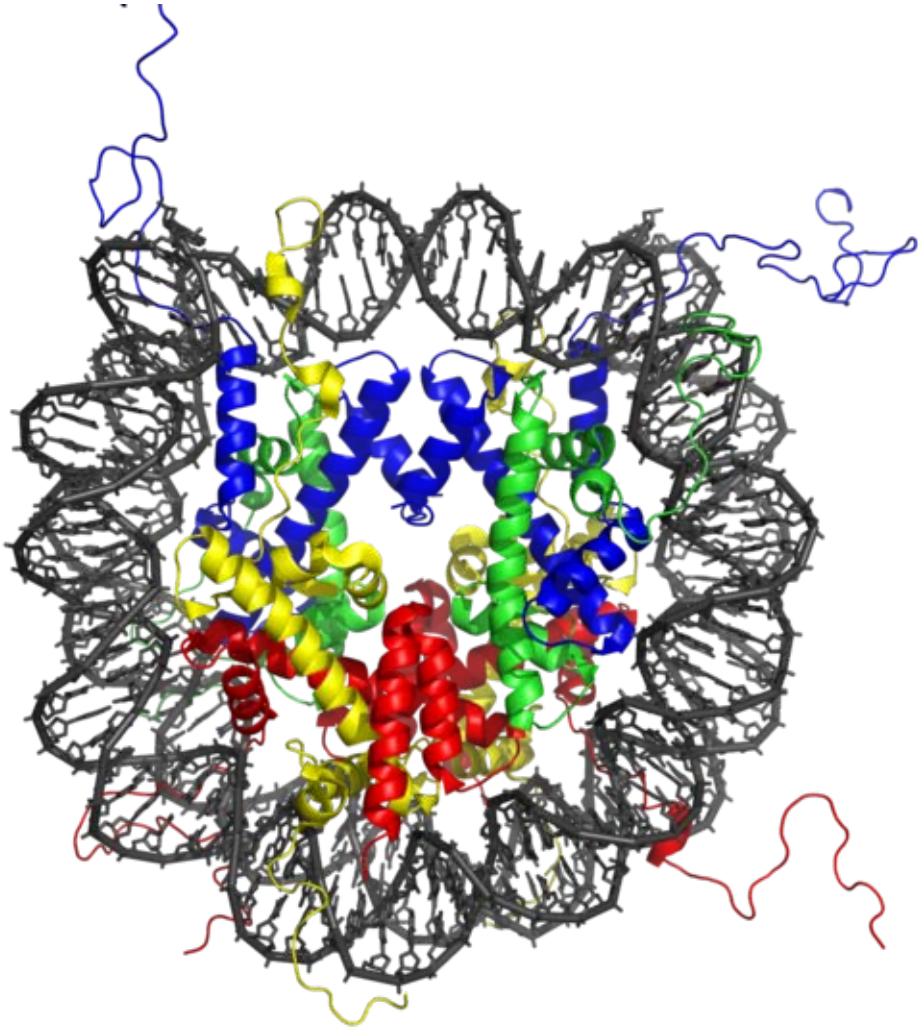
The nucleosome core particle consists of approximately 147 base pairs of DNA wrapped in 1.67 left-handed superhelical turns around a histone octamer consisting of 2 copies each of the core histones H2A, H2B, H3, and H4. Core particles are connected by stretches of "linker DNA", which can be up to about 80 bp long. Technically, a nucleosome is defined as the core particle plus one of these linker regions; however the word is often synonymous with the core particle.

Linker histones such as H1 and its isoforms are involved in chromatin compaction and sit at the base of the nucleosome near the DNA entry and exit binding to the linker region of the DNA. Non-condensed nucleosomes without the linker histone resemble "beads on a string of DNA" under an electron microscope.

In contrast to most eukaryotic cells, mature sperm cells largely use protamines to package their genomic DNA, most likely to achieve an even higher packaging ratio. Histone equivalents and a simplified chromatin structure have also been found in Archea, proving that eukaryotes are not the only organisms that use nucleosomes.

Structure

Structure of the core particle



The crystal structure of the nucleosome core particle consisting of H2A , H2B , H3 and H4 core histones, and DNA. The view is from the top through the superhelical axis.

Overview

Early structural studies provided evidence that an octamer of histone proteins wraps DNA around itself in about two turns of a left-handed superhelix. In 1997 the first near atomic resolution crystal structure of the nucleosome was solved by the Richmond group showing some of the most important details of the particle. It should be noted however, that the human alpha-satellite palindromic DNA critical to achieving the 1997 nucleosome crystal structure was actually developed at Oak Ridge National Laboratory,

Oak Ridge, Tennessee, by the Bunick group . The Bunick group pioneered the development of the human alpha-satellite palindromic DNA and this was a critical advancement enabling high-resolution structural determination of nucleosome core particles . The structures of over 20 different nucleosome core particles have been solved to date, including those containing histone variants and histones from different species. The structure of the nucleosome core particle is remarkably conserved, and even a change of over 100 residues between frog and yeast histones results in electron density maps with an overall root mean square deviation of only 1.6Å.

The nucleosome core particle

The nucleosome core particle (shown in the figure) consists of about 146 bp of DNA wrapped in 1.67 left-handed superhelical turns around the histone octamer, consisting of 2 copies each of the core histones H2A, H2B, H3, and H4. Adjacent nucleosomes are joined by a stretch of free DNA termed "linker DNA" (which varies from 10 - 80 bp in length depending on species and tissue type).

Protein interactions within the nucleosome

The core histone proteins contain a characteristic structural motif termed the "histone fold" which consists of three alpha-helices (α1-3) separated by two loops (L1-2). In solution the histones form H2A-H2B heterodimers and H3-H4 heterotetramers. Histones dimerise about their long α2 helices in an anti-parallel orientation, and in the case of H3 and H4, two such dimers form a 4-helix bundle stabilised by extensive H3-H3' interaction. The H2A/H2B dimer binds onto the H3/H4 tetramer due to interactions between H4 and H2B which include the formation of a hydrophobic cluster. The histone octamer is formed by a central H3/H4 tetramer sandwiched between two H2A/H2B dimers. Due to the highly basic charge of all four core histones, the histone octamer is only stable in the presence of DNA or very high salt concentrations.

Histone - DNA interactions

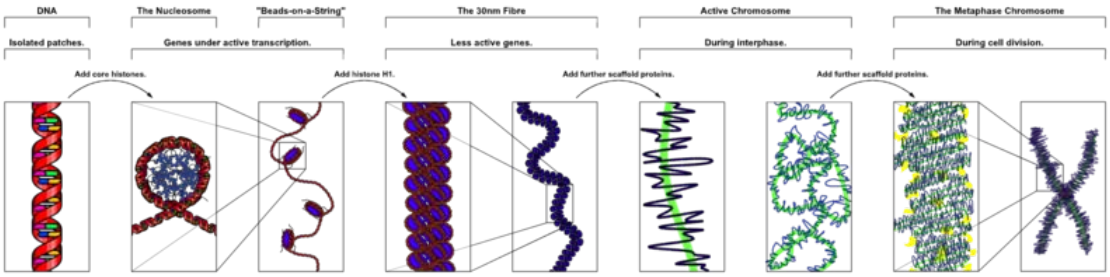
The nucleosome contains over 120 direct protein-DNA interactions and several hundred water mediated ones. Direct protein - DNA interactions are not spread evenly about the octamer surface but rather located at discrete sites. These are due to the formation of two types of DNA binding sites within the octamer; the α1α1 site which uses the α1 helix from two adjacent histones and the L1L2 site formed by the L1 and L2 loops. Salt links and hydrogen bonding between both side chain basic and hydroxyl groups and main chain amides with the DNA backbone phosphates form the bulk of interactions with the DNA. This is important given that the ubiquitous distribution of nucleosomes along genomes requires it to be a non-sequence-specific DNA-binding factor. Although nucleosomes tend to prefer some DNA sequences over others, they are capable of binding practically to any sequence, which is thought to be due to the flexibility in the formation of these water-mediated interactions. In addition, non-polar interactions are made between protein side chains and the deoxyribose groups, and an arginine side chain intercalates into the DNA minor groove at all 14 sites it faces the octamer surface. The

distribution and strength of DNA binding sites about the octamer surface distorts the DNA within the nucleosome core. The DNA is non-uniformly bent and also contains twist defects. The twist of free B-form DNA in solution is 10.5 bp per turn, however, the overall twist of nucleosomal DNA is only 10.2 bp per turn, varying from a value of 9.4 to 10.9 bp per turn.

Histone tail domains

The histone tail extensions constitute up to 30% by mass of histones, but are not visible in the crystal structures of nucleosomes due to their high intrinsic flexibility and have been thought to be largely unstructured. The N-terminal tails of histones H3 and H2B pass through a channel formed by the minor grooves of the two DNA strands, protruding from the DNA every 20 bp. The N-terminal tail of histone H4 on the other hand has a region of highly basic amino acids (16-25) which, in the crystal structure, forms an interaction with the highly acidic surface region of a H2A-H2B dimer of another nucleosome, being potentially relevant for the higher-order structure of nucleosomes. This interaction is thought to occur also under physiological conditions and suggests that acetylation of the H4 tail distorts the higher order structure of chromatin.

Higher order structure



The current chromatin compaction model.

The organization of the DNA that is achieved by the nucleosome can not fully explain the packaging of DNA observed in the cell nucleus. Further compaction of chromatin into the cell nucleus is necessary, but is not yet well understood. The current understanding is that repeating nucleosomes with intervening "linker" DNA form a *10-nm-fiber*, known descriptively as "beads on a string", and have a packing ratio of about five to ten. A chain of nucleosomes can be arranged in a *30 nm fiber*, a compacted structure with a packing ratio of ~50 and whose formation is dependent on the presence of the H1 histone.

A crystal structure of a tetranucleosome has been presented and used to build up a proposed structure of the 30 nm fiber as a two-start helix. There is still a certain amount of contention regarding this model as it is incompatible with recent electron microscopy data. Beyond this, the structure of chromatin is poorly understood, but it is classically suggested that the 30 nm fiber is arranged into loops along a central protein scaffold to form transcriptionally active euchromatin. Further compaction leads to transcriptionally inactive heterochromatin.

Nucleosome dynamics

Although the nucleosome is a very stable protein-DNA complex, it is not static and has been shown to undergo a number of different structural re-arrangements including nucleosome sliding and DNA site exposure. Depending on the context, nucleosomes can inhibit or facilitate transcription factor binding. Nucleosome positions are controlled by three major contributions: First, the intrinsic binding affinity of the histone octamer depends on the DNA sequence. Second, the nucleosome can be displaced or recruited by the competitive or cooperative binding of other protein factors. Third, the nucleosome may be actively translocated by ATP-dependent remodeling complexes.

Nucleosome sliding

Work performed in the Bradbury laboratory showed that nucleosomes reconstituted onto the 5S DNA positioning sequence were able to reposition themselves translationally onto adjacent sequences when incubated thermally. Later work showed that this repositioning did not require disruption of the histone octamer but was consistent with nucleosomes being able to "slide" along the DNA *in cis*. In 2008, It was further revealed that CTCF binding sites act as nucleosome positioning anchors so that, when used to align various genomic signals, multiple flanking nucleosomes can be readily identified. Although nucleosomes are intrinsically mobile, eukaryotes have evolved a large family of ATP-dependent chromatin remodelling enzymes to alter chromatin structure, many of which do so via nucleosome sliding.

DNA site exposure

Work from the Widom laboratory has shown that nucleosomal DNA is in equilibrium between a wrapped and unwrapped state. Measurements of these rates using time resolved FRET revealed that DNA within the nucleosome remains fully wrapped for only 250ms before it is unwrapped for 10-50ms and then rapidly rewrapped . This implies that DNA does not need to be actively dissociated from the nucleosome but that there is a significant fraction of time during which it is fully accessible. Indeed, this can be extended to the observation that introducing a DNA binding sequence within the nucleosome increases the accessibility of adjacent regions of DNA when bound . This propensity for DNA within the nucleosome to "breathe" is predicted to have important functional consequences for all DNA binding proteins that operate in a chromatin environment.

Modulating nucleosome structure

Eukaryotic genomes are ubiquitously associated into chromatin; however, cells need to spatially and temporally regulate specific loci independently of bulk chromatin. In order to achieve the high level of control required to co-ordinate nuclear processes such as DNA replication, repair and transcription, cells have developed a variety of means to locally and specifically modulate chromatin structure and function. This can involve

covalent modification of histones, the incorporation of histone variants and non-covalent remodelling by ATP-dependent remodelling enzymes.

Histone post-translational modifications

Since they were discovered in the mid 1960's histone modifications have been predicted to affect transcription. The fact that most of the early post-translational modifications found were concentrated within the tail extensions that protrude from the nucleosome core lead to two main theories regarding the mechanism of histone modification. The first of the theories suggested that they may affect electrostatic interactions between the histone tails and DNA to "loosen" chromatin structure. Later it was proposed that combinations of these modifications may create binding epitopes with which to recruit other proteins. Recently, given that more modifications have been found in the structured regions of histones it has been put forward that these modifications may affect histone-DNA and histone-histone interactions within the nucleosome core. Some modifications have been shown to be correlated with gene silencing, others seem to be correlated with gene activation. Common modifications include acetylation, methylation or ubiquitination of lysine; methylation of arginine and phosphorylation of serine. The information stored in this way is considered epigenetic since it is not encoded in the DNA but is still inherited to daughter cells. The maintenance of a repressed or activated status of a gene is often necessary for cellular differentiation.

Histone variants

Whilst histones are remarkably conserved throughout evolution, several variant forms have been identified. Interestingly, this diversification of histone function is restricted to H2A and H3, with H2B and H4 being mostly invariant. H2A can be replaced by H2AZ (which leads to reduced nucleosome stability) or H2AX (which is associated with DNA repair and T cell differentiation) whereas the inactive X chromosomes in mammals are enriched in macroH2A. H3 can be replaced by H3.3 (which correlates with activate genes) and in centromeres H3 is replaced by CENPA.

ATP-dependent nucleosome remodelling

A number of distinct reactions are associated with the term ATP-dependent chromatin remodelling. Remodelling enzymes have been shown to slide nucleosomes along DNA,disrupt histone-DNA contacts to the extent of destabilising the H2A/H2B dimer and to generate negative superhelical torsion in DNA and chromatin. Recently, the Swr1 remodelling enzyme has been shown to introduce the variant histone H2A.Z into nucleosomes. At present, it is not clear if all of these represent distinct reactions or merely alternative outcomes of a common mechanism. What is shared between all, and indeed the hallmark of ATP-dependent chromatin remodelling, is that they all result in altered DNA accessibility. Studies looking at gene activation *in vivo* and, more astonishingly, remodelling *in vitro* have revealed that chromatin remodelling events and transcription-factor binding are cyclical and periodic in nature. While the consequences

of this for the reaction mechanism of chromatin remodelling are not known, the dynamic nature of the system may allow it to respond faster to external stimuli.

Dynamic nucleosome remodelling across the Yeast genome

Studies in 2007 have catalogued nucleosome positions in yeast and shown that nucleosomes are depleted in promoter regions and origins of replication . About 80% of the yeast genome appears to be covered by nucleosomes and the pattern of nucleosome positioning clearly relates to DNA regions that regulate transcription, regions that are transcribed and regions that initiate DNA replication . Most recently, a new study examined ‘’dynamic changes’’ in nucleosome repositioning during a global transcriptional reprogramming event to elucidate the effects on nucleosome displacement during genome-wide transcriptional changes in yeast (*Saccharomyces cerevisiae*) . The results suggested that nucleosomes that were localized to promoter regions are displaced in response to stress (like heat shock). In addition, the removal of nucleosomes usually corresponded to transcriptional activation and the replacement of nucleosomes usually corresponded to transcriptional repression, presumably because transcription factor binding sites became more or less accessible, respectively. In general, only one or two nucleosomes were repositioned at the promoter to effect these transcriptional changes. However, even in chromosomal regions that were not associated with transcriptional changes, nucleosome repositioning was observed, suggesting that the covering and uncovering of transcriptional DNA does not necessarily produce a transcriptional event.

Nucleosome assembly in vitro

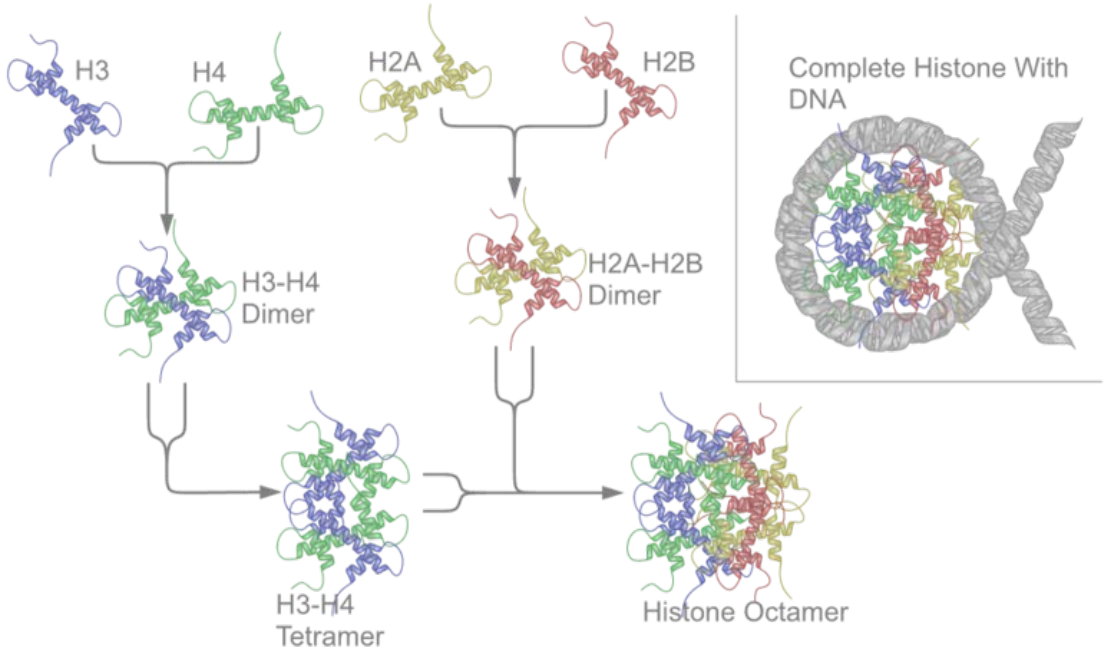
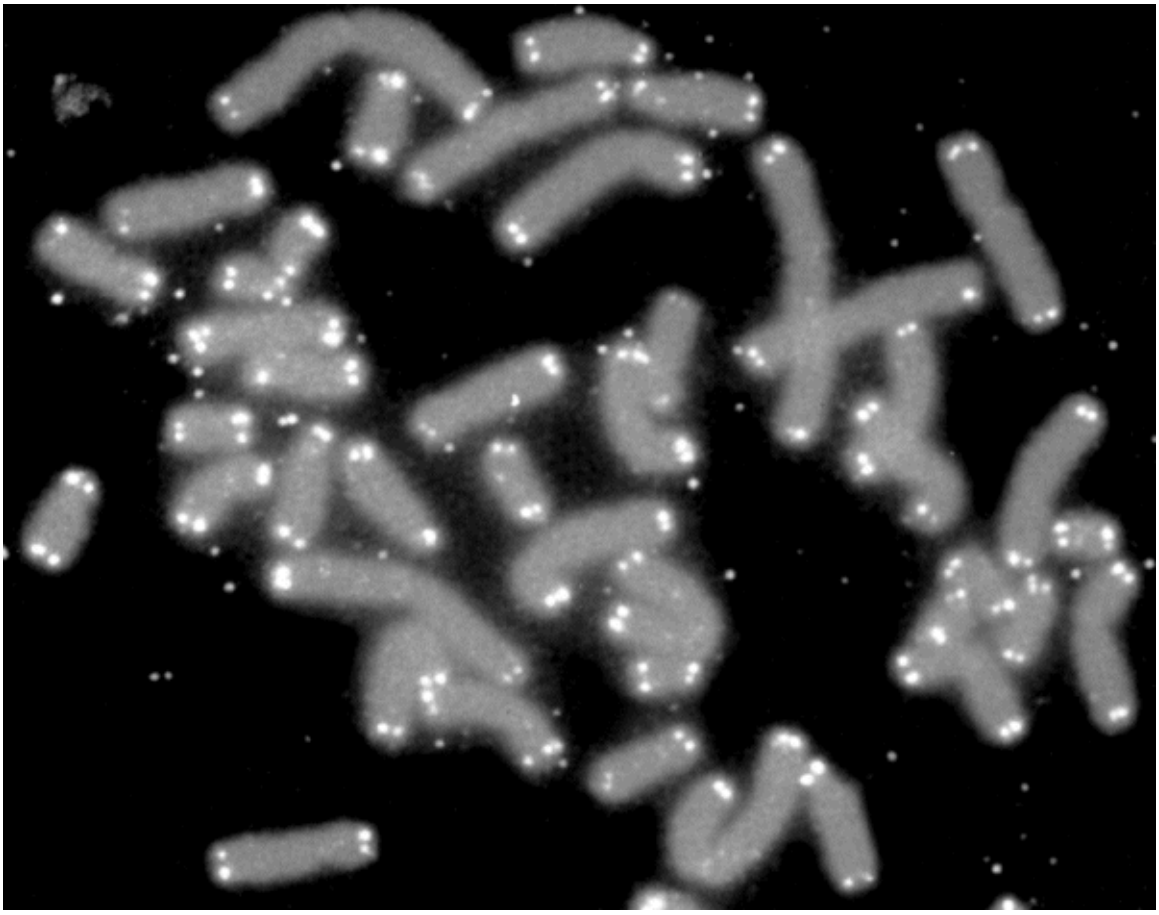


Diagram of nucleosome assembly.

Nucleosomes can be assembled *in vitro* by either using purified native or recombinant histones. One standard technique of loading the DNA around the histones involves the use of salt dialysis. A reaction consisting of the histone octamers and a naked DNA template can be incubated together at a salt concentration of 2 M. By steadily decreasing the salt concentration, the DNA will equilibrate to a position where it is wrapped around the histone octamers, forming nucleosomes. In appropriate conditions, this reconstitution process allows for the nucleosome positioning affinity of a given sequence to be mapped experimentally.

Chapter 8

Telomere



Human chromosomes (grey) capped by telomeres (white)

A **telomere** is a region of repetitive DNA at the end of a chromosome, which protects the end of the chromosome from deterioration. Its name is derived from the Greek nouns telos (*τέλος*) "end" and meros (*μέρος*, root: *μερ*-) "part". The telomere regions deter the

degradation of genes near the ends of chromosomes by allowing for the shortening of chromosome ends, which necessarily occurs during chromosome replication.

In the early 1970s, Russian theorist Alexei Olovnikov first recognized the problem of how chromosomes could replicate right to the tip, as such was impossible with replication in a 5' to 3' direction and the need for a primer sequence. To solve this and to accommodate Leonard Hayflick's idea of limited somatic cell division, Olovnikov suggested that DNA sequences would be lost in every replicative phase until they reached a critical level, at which point cell division would stop.

During cell division, enzymes that duplicate DNA cannot continue their duplication all the way to the end of the chromosome. If cells divided without telomeres, they would lose the ends of their chromosomes, and the necessary information they contain. The telomeres are disposable buffers blocking the ends of the chromosomes and are consumed during cell division and replenished by an enzyme, the telomerase reverse transcriptase.

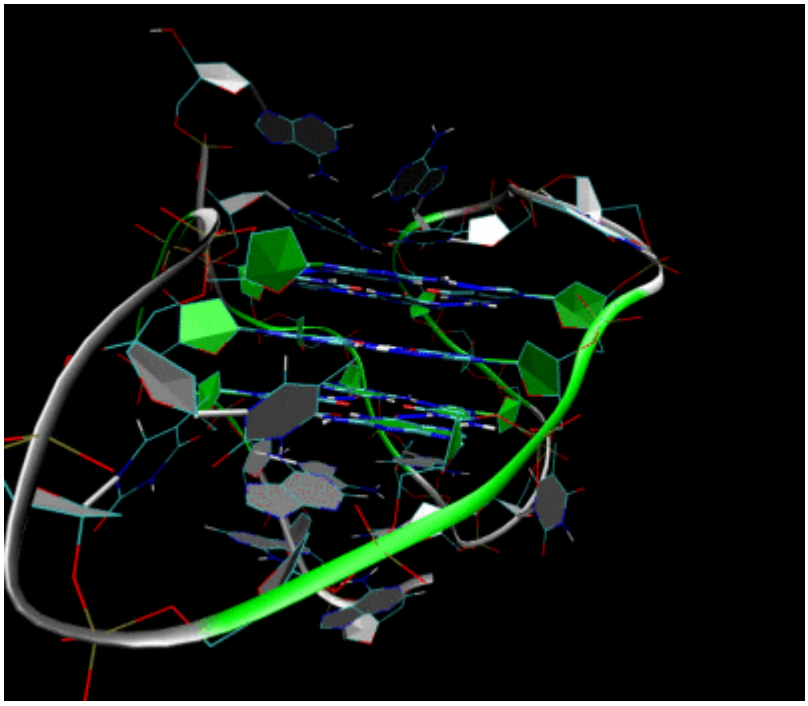
In 1975–1977, Elizabeth Blackburn, working as a postdoctoral fellow at Yale University with Joseph Gall, discovered the unusual nature of telomeres, with their simple repeated DNA sequences composing chromosome ends. Their work was published in 1978. The telomere shortening mechanism normally limits cells to a fixed number of divisions, and animal studies suggest that this is responsible for aging on the cellular level and sets a limit on lifespans. Telomeres protect a cell's chromosomes from fusing with each other or rearranging — abnormalities that can lead to cancer — and so cells are destroyed when their telomeres are consumed. Most cancers are the result of "immortal" cells that have ways of evading this programmed destruction.

Elizabeth Blackburn, Carol Greider, and Jack Szostak were awarded the 2009 Nobel Prize in Physiology or Medicine for the discovery of how chromosomes are protected by telomeres and the enzyme telomerase.

Nature and function of telomeres

Structure, function and evolutionary biology

Telomeres are repetitive DNA sequences located at the termini of linear chromosomes of most eukaryotic organisms, and a few prokaryotes. Telomeres compensate for incomplete semi-conservative DNA replication at chromosomal ends. The protection against homologous recombination (HR) and non-homologous end joining (NHEJ) constitutes the essential “capping” role of telomeres that distinguishes them from DNA double-strand breaks (DSBs).



Three-dimensional molecular structure of a telomere (G-quadruplex).

In most prokaryotes, chromosomes are circular and, thus, do not have ends to suffer premature replication termination. A small fraction of bacterial chromosomes (such as those in Streptomyces and Borrelia) are linear and possess telomeres, which are very different from those of the eukaryotic chromosomes in structure and functions. The known structures of bacterial telomeres take the form of proteins bound to the ends of linear chromosomes, or hairpin loops of single-stranded DNA at the ends of the linear chromosomes.

In most multicellular eukaryotic organisms, telomerase is active only in germ cells, stem cells and certain white blood cells. There are theories that claim that the steady shortening of telomeres with each replication in somatic (body) cells may have a role in senescence and in the prevention of cancer. This is because the telomeres act as a sort of time-delay "fuse", eventually running out after a certain number of cell divisions and resulting in the eventual loss of vital genetic information from the cell's chromosome with future divisions.

Telomere length varies greatly between species, from approximately 300 to 600 base pairs in yeast to many kilobases in humans, and usually is composed of arrays of guanine-rich, six- to eight-base-pair-long repeats. Eukaryotic telomeres normally terminate with 3′ single-stranded-DNA overhang, which is essential for telomere maintenance and capping. Multiple proteins binding single- and double-stranded telomere DNA have been identified. These function in both telomere maintenance and capping. Telomeres form large loop structures called telomere loops, or T-loops. Here, the single-stranded DNA curls around in a long circle stabilized by telomere-binding proteins. At the very end of the T-loop, the single-stranded telomere DNA is held onto a region of

double-stranded DNA by the telomere strand disrupting the double-helical DNA and base pairing to one of the two strands. This triple-stranded structure is called a displacement loop or D-loop.

Telomere shortening in humans can induce replicative senescence, which blocks cell division. This mechanism appears to prevent genomic instability and development of cancer in human aged cells by limiting the number of cell divisions. However, shortened telomeres impair immune function which might also increase cancer susceptibility. Malignant cells that bypass this arrest become immortalized by telomere extension due mostly to the activation of telomerase, the reverse transcriptase enzyme responsible for synthesis of telomeres. However, 5–10% of human cancers activate the Alternative Lengthening of Telomeres (ALT) pathway, which relies on recombination-mediated elongation.

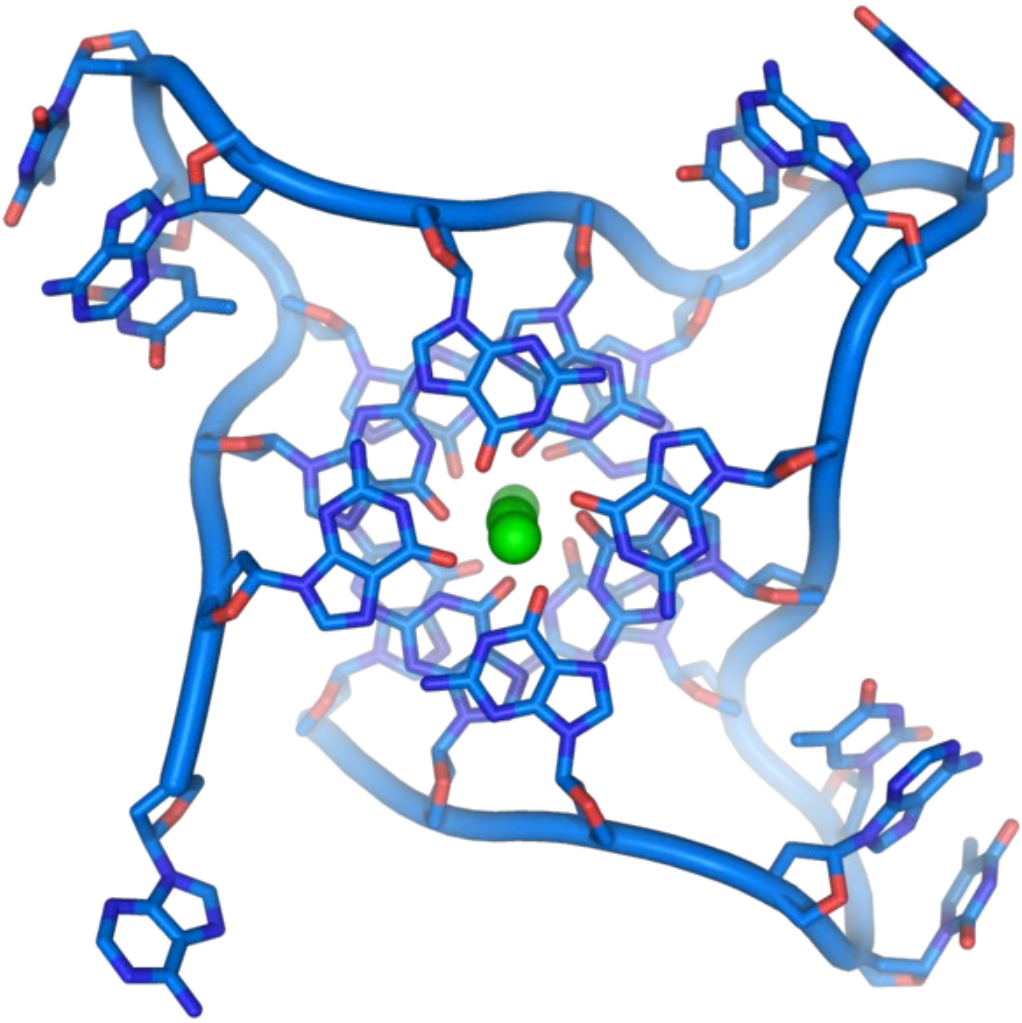
Since shorter telomeres are thought to be a cause of poorer health and aging, this raises the question of why longer telomeres are not selected for to ameliorate these effects. A prominent explanation suggests that inheriting longer telomeres would cause increased cancer rates (e.g. Weinstein and Ciszek, 2002). However, a recent literature review and analysis suggests this is unlikely, because shorter telomeres and telomerase inactivation is more often associated with increased cancer rates, and the mortality from cancer occurs late in life when the force of natural selection is very low. An alternative explanation to the hypothesis that long telomeres are selected against due to their cancer promoting effects is the "thrifty telomere" hypothesis which suggests that the cellular proliferation effects of longer telomeres causes increased energy expenditures. In environments of energetic limitation, shorter telomeres might be an energy sparing mechanism.

Human telomeres, cancer and ALT (Alternative lengthening of telomeres)

Human somatic cells lacking telomerase gradually lose telomeric sequences as a result of incomplete replication (Counter *et al.*, 1992). As human telomeres grow shorter, eventually cells reach the limit of their replicative capacity and progress into senescence. Senescence involves p53 and pRb pathways and leads to the arrest of cell proliferation (Campisi, 2005). It is thought that senescence plays an important role in suppression of emergence of cancer, although inheriting shorter telomeres probably does not protect against cancer (Eisenberg, 2011). With critically shortened telomeres, further cell proliferation can be achieved by inactivation of p53 and pRb pathways. Cells entering proliferation after inactivation of p53 and pRb pathways undergo crisis. Crisis is characterized by gross chromosomal rearrangements and genome instability, and almost all cells die. Rare cells emerge from crisis immortalized through telomere elongation by either activated telomerase or ALT (Colgina and Reddel, 1999; Reddel and Bryan, 2003). The first description of an ALT cell line demonstrated that the telomeres are highly heterogeneous in length and predicted a mechanism involving recombination (Murnane et al., 1994). Subsequent studies have confirmed a role for recombination in telomere maintenance by ALT (Dunham et al., 2000), however the exact mechanism of this pathway is yet to be determined. ALT cells produce abundant t-circles, possible products

of intratelomeric recombination and t-loop resolution (Tomaska *et al.*, 2000; 2009; Cesare and Griffith, 2004; Wang *et al.*, 2004).

Telomerase is a "ribonucleoprotein complex" composed of a protein component and an RNA primer sequence that acts to protect the terminal ends of chromosomes. The actions of telomerase are necessary because, during replication, DNA polymerase can synthesize DNA in only a 5' to 3' direction and can do so only by adding polynucleotides to an RNA primer that has already been placed at various points along the length of the DNA. These RNA strands must later be replaced with DNA. This replacement of the RNA primers is not a problem at origins of replication within the chromosome because DNA polymerase can use a previous stretch of DNA 5' to the RNA template as a template to backfill the sequence where the RNA primer was; at the terminal end of the chromosome, however, DNA polymerase cannot replace the RNA primer because there is no position 5' of the RNA primer where another primer can be placed, nor is there DNA upstream that can be used as a primer so that DNA polymerase can replace the RNA primer. Without telomeres at the end of DNA, this genetic sequence at the end of the chromosome would be deleted and the chromosome would grow shorter and shorter in subsequent replications. The telomere prevents this problem by employing a different mechanism to synthesize DNA at this point, thereby preserving the sequence at the terminal of the chromosome. This prevents chromosomal fraying and prevents the ends of the chromosome from being processed as a double-strand DNA break, which could lead to chromosome-to-chromosome telomere fusions. Telomeres are extended by telomerases, part of a protein subgroup of specialized reverse transcriptase enzymes known as TERT (**TE**lomerase **R**everse **T**ranscriptases) that are involved in synthesis of telomeres in humans and many other, but not all, organisms. However, because of DNA replication mechanisms, oxidative stress, and, because TERT expression is very low in many types of human cells, the telomeres of these cells shrink a little bit every time a cell divides, although, in other cellular compartments that require extensive cell division, such as stem cells and certain white blood cells, TERT is expressed at higher levels and telomere shortening is partially or fully prevented.



Structure of parallel quadruplexes that can be formed by human telomeric DNA. Image created from NDB UD0017.

In addition to its TERT protein component, telomerase also contains a piece of template RNA known as the TERC (**TE**lomerase **R**NA **C**omponent) or TR (**T**elomerase **R**NA). In humans, this TERC telomere sequence is a repeating string of TTAGGG, between 3 and 20 kilobases in length. There are an additional 100-300 kilobases of telomere-associated repeats between the telomere and the rest of the chromosome. Telomere sequences vary from species to species, but, in general, one strand is rich in G with fewer Cs. These G-rich sequences can form four-stranded structures (G-quadruplexes), with sets of four bases held in plane and then stacked on top of each other with either a sodium or a potassium ion between the planar quadruplexes.

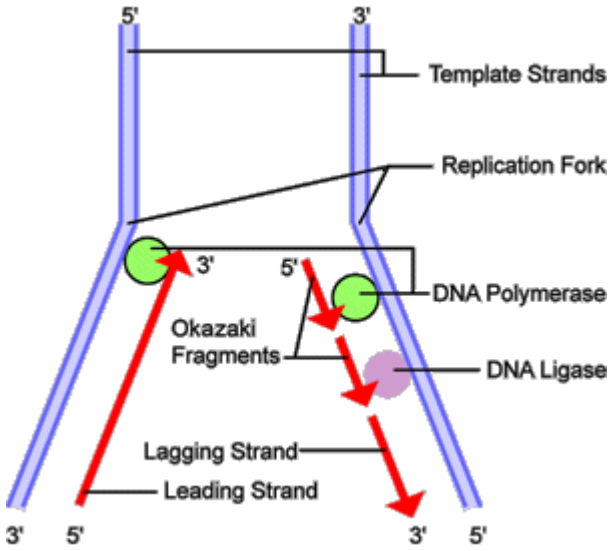
If telomeres become too short, they have the potential to unfold from their presumed closed structure. It is thought that the cell detects this uncapping as DNA damage and then enters cellular senescence, growth arrest, or apoptosis, depending on the cell's genetic background (p53 status). Uncapped telomeres also result in chromosomal fusions.

Since this damage cannot be repaired in normal somatic cells, the cell may even go into apoptosis. Many aging-related diseases are linked to shortened telomeres. Organs deteriorate as more and more of their cells die off or enter cellular senescence.

At the very distal end of the telomere is a 300 bp single-stranded portion, which forms the T-Loop. This loop is analogous to a *knot*, which stabilizes the telomere, preventing the telomere ends from being recognized as break points by the DNA repair machinery. Should non-homologous end joining occur at the telomeric ends, chromosomal fusion will result. The T-loop is held together by seven known proteins, the most notable ones being TRF1, TRF2, POT1, TIN1, and TIN2, collectively referred to as the shelterin complex.

A study published in the May 3, 2005 issue of the American Heart Association journal *Circulation* found that weight gain and increased insulin resistance are correlated with greater telomere shortening over time.

Telomere shortening



Lagging strand during DNA replication

Telomeres shorten in part because of the *end replication problem* that is exhibited during DNA replication in eukaryotes only. Because DNA replication does not begin at either end of the DNA strand, but starts in the center, and considering that all known DNA polymerases move in the 5' to 3' direction, one finds a leading and a lagging strand on the DNA molecule being replicated.

On the leading strand, DNA polymerase can make a complementary DNA strand without any difficulty because it goes from 5' to 3'. However, there is a problem going in the other direction on the lagging strand. To counter this, short sequences of RNA acting as primers attach to the lagging strand a short distance ahead of where the initiation site was. The DNA polymerase can start replication at that point and go to the end of the initiation

site. This causes the formation of Okazaki fragments. More RNA primers attach further on the DNA strand and DNA polymerase comes along and continues to make a new DNA strand.

Eventually, the last RNA primer attaches, and DNA polymerase, RNA nuclease, and DNA ligase come along to convert the RNA (of the primers) to DNA and to seal the gaps in between the Okazaki fragments. But, in order to change RNA to DNA, there must be another DNA strand in front of the RNA primer. This happens at all the sites of the lagging strand, but it does not happen at the end where the last RNA primer is attached. Ultimately, that RNA is destroyed by enzymes that degrade any RNA left on the DNA. Thus, a section of the telomere is lost during each cycle of replication at the 5' end of the lagging strand.

However, in vitro studies (von Zglinicki et al. 1995, 2000) have shown that telomeres are highly susceptible to oxidative stress. Telomere shortening due to free radicals explains the difference between the estimated loss per division because of the end-replication problem (ca. 20 bp) and actual telomere shortening rates (50-100 bp), and has a greater absolute impact on telomere length than shortening caused by the end-replication problem.

Lengthening telomeres

The phenomenon of limited cellular division was first observed by Leonard Hayflick, and is now referred to as the Hayflick Limit. Significant discoveries were made by the team led by Professor Elizabeth Blackburn at the University of California, San Francisco (UCSF).

Advocates of human life extension promote the idea of lengthening the telomeres in certain cells through temporary activation of telomerase (by drugs), or possibly permanently by gene therapy. They reason that this would extend human life because it would extend the Hayflick Limit. So far these ideas have not been proven in humans, but it has been demonstrated that telomere extension has successfully reversed some signs of aging in laboratory mice and the nematode worm species *Caenorhabditis elegans*. However, it has been hypothesized that longer telomeres and especially telomerase activation might cause increased cancer (e.g. Weinstein and Ciszek, 2002). Paradoxically, longer telomeres might also protect against cancer, because short telomeres are associated with cancer. It has also been suggested that longer telomeres might cause increased energy consumption.

Techniques to extend telomeres could be useful for tissue engineering, because they might permit healthy, noncancerous mammalian cells to be cultured in amounts large enough to be engineering materials for biomedical repairs.

However, there are several issues that still need to be cleared up. First, it is not even certain whether the relationship between telomeres and aging is causal. Changing

telomere lengths are usually associated with changing speed of senescence. This telomere shortening, however, might be a consequence of, and not a reason for, aging.

That the role of telomeres is far from being understood is demonstrated by two recent studies on long-lived seabirds. In 2003, scientists observed that the telomeres of Leach's Storm-petrel (*Oceanodroma leucorhoa*) seem to lengthen with chronological age, the first observed instance of such behaviour of telomeres. In 2006, Juola *et al.* reported that in another unrelated, long-lived seabird species, the Great Frigatebird (*Fregata minor*), telomere length did decrease until at least c.40 years of age (i.e. probably over the entire lifespan), but the speed of decrease slowed down massively with increasing ages, and that rates of telomere length decrease varied strongly between individual birds. They concluded that in this species (and probably in frigatebirds and their relatives in general), telomere length could not be used to determine a bird's age sufficiently well. Thus, it seems that there is much more variation in the behavior of telomere length than initially believed.

The telomere length varies in cloned animals. Sometimes the clones end up with shorter telomeres since the DNA has already divided countless times. Occasionally, the telomeres in a clone's DNA are longer because they get "reprogrammed".

Telomere sequences

Some known telomere sequences

Group	Organism	Telomeric repeat (5' to 3' toward the end)
Vertebrates	Human, mouse, *Xenopus*	TTAGGG
Filamentous fungi	*Neurospora crassa*	TTAGGG
Slime moulds	*Physarum*, *Didymium*	TTAGGG
	Dictyostelium	AG(1-8)
Kinetoplastid protozoa	*Trypanosoma*, *Crithidia*	TTAGGG
Ciliate protozoa	*Tetrahymena*, *Glaucoma*	TTGGGG
	Paramecium	TTGGG(T/G)
	Oxytricha, *Stylonychia*, *Euplotes*	TTTTGGGG
Apicomplexan protozoa	*Plasmodium*	TTAGGG(T/C)
Higher plants	*Arabidopsis thaliana*	TTTAGGG
Green algae	*Chlamydomonas*	TTTTAGGG
Insects	*Bombyx mori*	TTAGG
Roundworms	*Ascaris lumbricoides*	TTAGGC
Fission yeasts	*Schizosaccharomyces pombe*	TTAC(A)(C)G(1-8)

Budding yeasts	*Saccharomyces cerevisiae*	TGTGGGTGTGGTG (from RNA template) or G(2-3)(TG)(1-6)T (consensus)
	Saccharomyces castellii	TCTGGGTG
	Candida glabrata	GGGGTCTGGGTGCTG
	Candida albicans	GGTGTACGGATGTCTAACTTCTT
	Candida tropicalis	GGTGTA[C/A]GGATGTCACGATCATT
	Candida maltosa	GGTGTACGGATGCAGACTCGCTT
	Candida guillermondii	GGTGTAC
	Candida pseudotropicalis	GGTGTACGGATTTGATTAGTTATGT
	Kluyveromyces lactis	GGTGTACGGATTTGATTAGGTATGT

Systemic telomere length and aging

Peripheral blood leukocyte telomere length is the generally preferred measure of systemic telomere length. Systemic telomere length has been proposed as a marker of biological aging. A subject's systemic telomere length is predominantly genetically determined, but has several other known determinants: age (shorter telomeres in older people), paternal age at birth (longer telomeres in subjects with older fathers at their birth), and sex (shorter telomeres in men, probably due to a faster telomere attrition). Evidence suggests that elevated levels of oxidative stress and inflammation further increase the telomere attrition rate.

Vitamin D may have an effect on peripheral blood leukocyte telomere length. *Richards and coworkers* examined whether vitamin D concentrations would slow the rate of shortening of leukocyte telomeres. The authors stated that vitamin D is a potent inhibitor of the proinflammatory response and slows the turnover of leukocytes. Leukocyte telomere length (LTL) predicts the development of aging-related disease, and the length of these telomeres decreases with each cell division and with increased inflammation. Researchers measured serum vitamin D concentrations in 2160 women, aged 18–79 years (mean age: 49.4), from a large population-based cohort of twins. This study divided the group into thirds, based on vitamin D levels, and found that increased age was significantly associated with shorter LTL ($r = -0.40$, $P < 0.0001$). Higher serum vitamin D concentrations were significantly associated with longer LTL ($r = 0.07$, $P = 0.0010$) and this finding persisted even after adjustment for age ($r = 0.09$, $P < 0.0001$) and other variables that independently could affect LTL (age, season of vitamin D measurement, menopausal status, use of hormone replacement therapy, and physical activity). The difference in LTL between the highest and lowest tertiles of vitamin D was highly significant ($P = 0.0009$) and the authors stated that this was equivalent to 5.0 years of aging. The authors concluded that higher vitamin D levels, easily modifiable through nutritional supplementation, were associated with longer LTL. This underscores the potentially beneficial effects of vitamin D on aging and age-related diseases. Also, peer reviewed clinical studies indicate a relationship between regular exercise and the

minimizing of telomere erosion in both mice and humans, at least in the leukocytes and blood vessel walls .

Telomeres and cancer

As a cell begins to become cancerous, it divides more often and its telomeres become very short. If its telomeres get too short, the cell may die. It can escape this fate by up-regulating an enzyme called telomerase, which can prevent telomeres from getting shorter and even elongate them.

Studies have found shortened telomeres in many cancers, including pancreatic, bone, prostate, bladder, lung, kidney, and head and neck. In addition, people with many types of cancer have been found to possess shorter leukocyte telomeres than healthy controls.

Cancer cells require a mechanism to maintain their telomeric DNA in order to continue dividing indefinitely (immortalization). A mechanism for telomere elongation or maintenance is one of the key steps in cellular immortalization and can be used as a diagnostic marker in the clinic. Telomerase, the enzyme complex responsible for elongating telomeres, is activated in approximately 90% of tumors. However, a sizeable fraction of cancerous cells employ alternative lengthening of telomeres (ALT), a non-conservative telomere lengthening pathway involving the transfer of telomere tandem repeats between sister-chromatids.

Telomerase is the natural enzyme that promotes telomere repair. It is active in stem cells, germ cells, hair follicles, and 90 percent of cancer cells, but its expression is low or absent in somatic cells. Telomerase functions by adding bases to the ends of the telomeres. Cells with sufficient telomerase activity are considered immortal in the sense that they can divide past the Hayflick limit without entering senescence or apoptosis. For this reason, telomerase is viewed as a potential target for anti-cancer drugs.

Studies using knockout mice have demonstrated that the role of telomeres in cancer can both be limiting to tumor growth, as well as promote tumorigenesis, depending on the cell type and genomic context.

Telomeres and cardiovascular aging

Shorter (systemic) telomere length has been suggested as an independent risk factor for cardiovascular disease. The origin of this association is unclear, and several models have been proposed, particularly attributing the biomarker value to a genetic prediposition in subjects with shorter telomeres, to an effect of inflammation and oxidative stress or to a combination of both.

Measurement of telomere length in the laboratory

Several techniques are currently employed to assess average telomere length in eukaryotic cells. The most widely used method is the Terminal Restriction Fragment

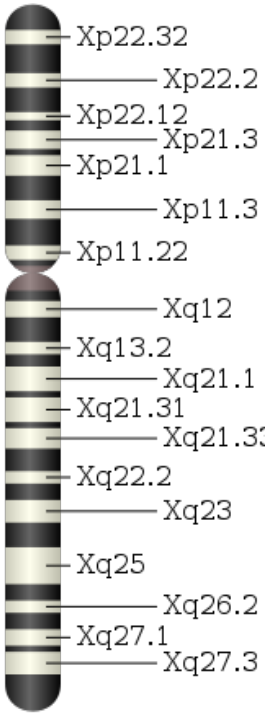
(TRF) southern blot, which involves hybridization of a radioactive 32P-(TTAGGG)n oligonucleotide probe to Hinf / Rsa I digested genomic DNA embedded on a nylon membrane and subsequently exposed to autoradiographic film or phosphoimager screen. Another histochemical method, termed Q-FISH, involves fluorescent in situ hybridization (FISH). Q-FISH, however, requires significant amounts of genomic DNA (2-20 micrograms) and labor that renders its use limited in large epidemiological studies. Some of these impediments have been overcome with a Real-Time PCR assay for telomere length and Flow-FISH. RT-PCR assay involves determining the Telomere-to-Single Copy Gene (T/S)ratio, which is demonstrated to be proportional to the average telomere length in a cell.

Another technique, referred to as single telomere elongation length analysis (STELA), was developed in 2003 by Duncan Baird. This technique allows investigations can target specific telomere ends, which is not possible with TRF analysis. However, due to this technique's being PCR-based, telomeres larger than 25Kb cannot be amplified and there is a bias towards shorter telomeres.

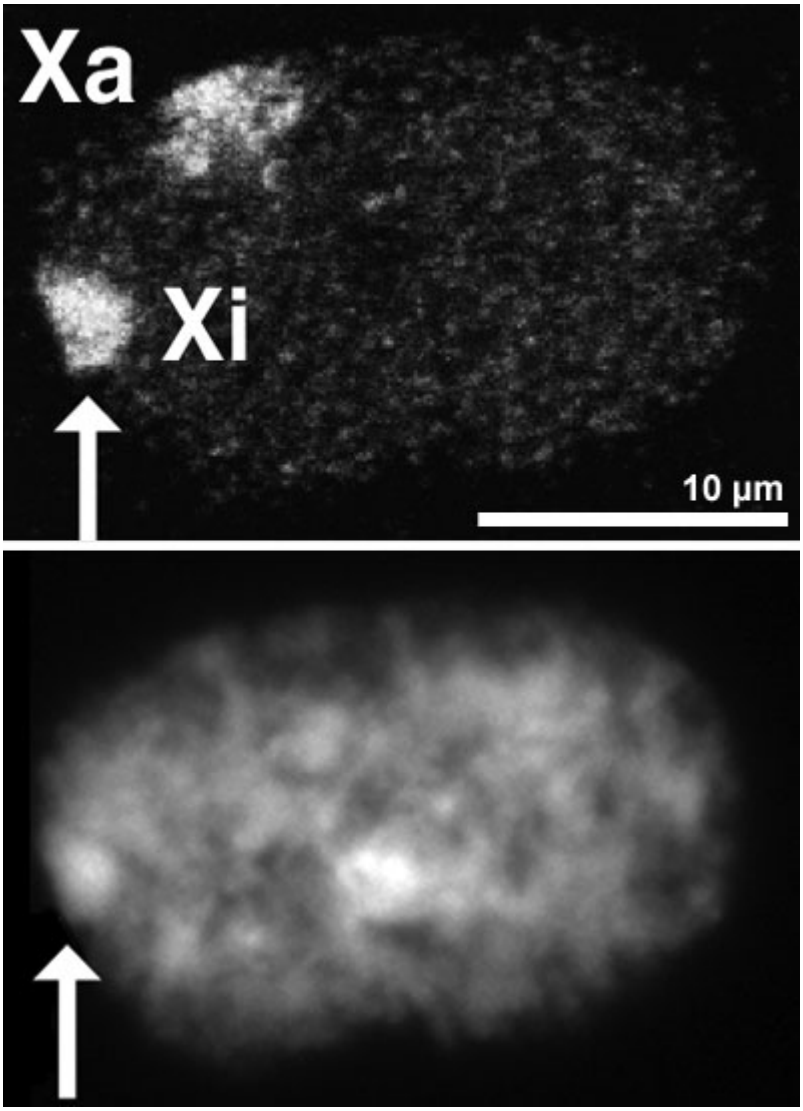
Chapter 9

X Chromosome and Y Chromosome

X chromosome



Scheme of the X chromatid



Nucleus of a female amniotic fluid cell. Top: Both X-chromosome territories are detected by FISH. Shown is a single optical section made with a confocal microscope. Bottom: Same nucleus stained with DAPI and recorded with a CCD camera. The Barr body is indicated by the arrow, it identifies the inactive X (Xi).

The **X chromosome** is one of the two sex-determining chromosomes in many animal species, including mammals (the other is the Y chromosome). It is a part of the XY sex-determination system and X0 sex-determination system. The **X** chromosome was named for its unique properties by early researchers, which resulted in the naming of its counterpart **Y** chromosome, for the next letter in the alphabet, after it was discovered later.

In humans

Function

The sex chromosomes X X are one of the 23 homologous pairs of chromosomes in a female. The X chromosome spans more than 153 million base pairs (the building material of DNA) and represents about 5% of the total DNA in women's cells, 2.5% in men's.

Each person normally has one pair of sex chromosomes in each cell. Females have two X chromosomes, whereas males have one X and one Y chromosome. Both males and females retain one of their mother's X chromosomes, and females retain their second X chromosome from their father. Since the father retains his X chromosome from his mother, a human female has one X chromosome from her paternal grandmother (father's side), and one X chromosome from her mother.

Identifying genes on each chromosome is an active area of genetic research. Because researchers use different approaches to predict the number of genes on each chromosome, the estimated number of genes varies. The X chromosome contains about 2000 genes compared to the Y chromosome containing 78 genes, out of the estimated 20,000 to 25,000 total genes in the human genome. Genetic disorders that are due to mutations in genes on the X chromosome are described as **X linked**.

The X chromosome carries a couple of thousand genes but few, if any, of these have anything to do directly with sex determination. Early in embryonic development in females, one of the two X chromosomes is randomly and permanently inactivated in nearly all somatic cells (cells other than egg and sperm cells). This phenomenon is called X-inactivation or Lyonization, and creates a Barr body. X-inactivation ensures that females, like males, have one functional copy of the X chromosome in each body cell. It was previously assumed that only one copy is actively used. However, recent research suggests that the Barr body may be more biologically active than was previously supposed.

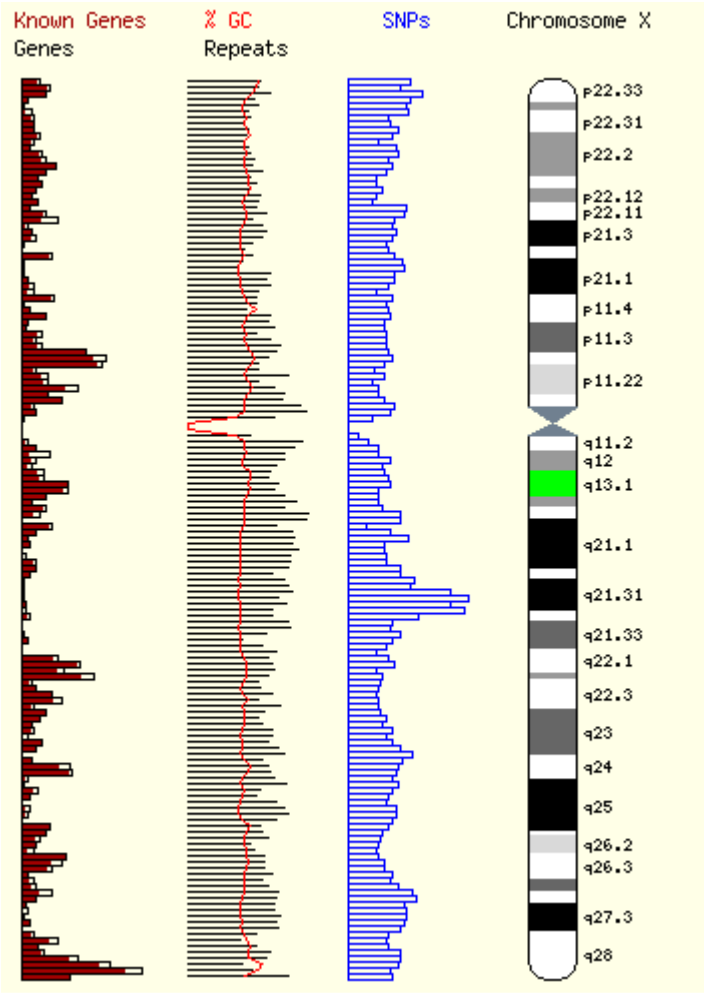
Structure

It is theorized by Ross et al. 2005 and Ohno 1967 that the X-chromosome is at least partially derived from the autosomal (non-sex-related) genome of other mammals evidenced from interspecies genomic sequence alignments.

The X-chromosome is notably larger and has a more active euchromatin region than its Y-chromosome counterpart. Further comparison of the X and Y reveal regions of homology between the two. However, the corresponding region in the Y appears far shorter and lacks regions that are conserved in the X throughout primate species, implying a genetic degeneration for Y in that region. Because males have only one X-chromosome, they are more likely to have an X chromosome-related disease.

It is estimated that about 10% of the genes encoded by the X-chromosome are associated with a family of "CT" genes, so named because they encode for markers found in both tumor cells (in **C**ancer patients) as well as in the human testis (in healthy patients).

Role in diseases



Numerical abnormalities

Klinefelter's syndrome:

- Klinefelter's syndrome is caused by the presence of one or more extra copies of the X chromosome in a male's cells. Extra genetic material from the X chromosome interferes with male sexual development, preventing the testicles from functioning normally and reducing the levels of testosterone.
- Males with Klinefelter's syndrome typically have one extra copy of the X chromosome in each cell, for a total of two X chromosomes and one Y chromosome (47,XXY). It is less common for affected males to have two or three extra X chromosomes (48,XXXY or 49,XXXXY) or extra copies of both the X and Y chromosomes (48,XXYY) in each cell. The extra genetic material may lead

to tall stature, learning and reading disabilities, and other medical problems. The average IQ in Klinefelter syndrome is in the normal range. When additional X and/or Y chromosomes are present in 48,XXXY, 48,XXYY, or 49,XXXXY, developmental delays and cognitive difficulties can be more severe and mild intellectual disability may be present.
- Klinefelter's syndrome can also result from an extra X chromosome in only some of the body's cells. These cases are called mosaic 46,XY/47,XXY.

Triple X syndrome (also called 47,XXX or trisomy X):

- This syndrome results from an extra copy of the X chromosome in each of a female's cells. Females with trisomy X have three X chromosomes, for a total of 47 chromosomes per cell. The average IQ of females with this syndrome is 90, while the average IQ of their normal siblings is 100 . Their stature on average is taller than normal females. They are fertile and their children do not inherit the condition.

- Females with more than one extra copy of the X chromosome (48, XXXX syndrome or 49, XXXXX syndrome) have been identified, but these conditions are rare.

Turner syndrome:

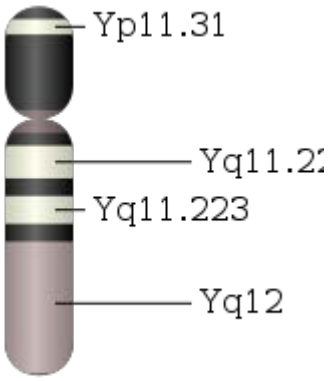
- This results when each of a female's cells has one normal X chromosome and the other sex chromosome is missing or altered. The missing genetic material affects development and causes the features of the condition, including short stature and infertility.
- About half of individuals with Turner syndrome have monosomy X (45,X), which means each cell in a woman's body has only one copy of the X chromosome instead of the usual two copies. Turner syndrome can also occur if one of the sex chromosomes is partially missing or rearranged rather than completely missing. Some women with Turner syndrome have a chromosomal change in only some of their cells. These cases are called Turner syndrome mosaics (45,X/46,XX).

Other disorders

XX male syndrome is a rare disorder, where the SRY region of the Y chromosome has recombined to be located on one of the X chromosomes. As a result, the XX combination after fertilization has the same effect as a XY combination, resulting in a male. However, the other genes of the X chromosome cause feminization as well.

X-linked endothelial corneal dystrophy is an extremely rare disease of cornea associated with Xq25 region. Lisch epithelial corneal dystrophy is associated with Xp22.3.

Y chromosome



Human Y-chromatid

The **Y-chromosome** is one of the two sex-determining chromosomes in most mammals, including humans. In mammals, it contains the gene SRY, which triggers testis development if present. The human Y-chromosome is composed of about 60 million base pairs. DNA in the Y-chromosome is passed from father to son, and Y-DNA analysis may thus be used in genealogy research.

Overview

Most mammals have one pair of sex chromosomes in each cell. Males have one Y-chromosome and one X-chromosome, while females have two X-chromosomes. In mammals, the Y-chromosome contains a gene, SRY, which triggers embryonic development as a male. The Y-chromosomes of humans and other mammals also contain other genes needed for normal sperm production.

There are exceptions, however. For example, the platypus relies on an XY sex-determination system based on five pairs of chromosomes. Platypus sex chromosomes in fact appear to bear a much stronger homology (similarity) with the avian Z chromosome, and the SRY gene so central to sex-determination in most other mammals is apparently not involved in platypus sex-determination. Among humans, some men have two Xs and a Y, or one X and two Ys, and some women have three Xs or a single X instead of a double X. There are other exceptions in which SRY is damaged (leading to an XY female), or copied to the X (leading to an XX male).

Origins and evolution

Before Y-chromosome

Many ectothermic vertebrates have no sex chromosomes. If they have different sexes, sex is determined environmentally rather than genetically. For some of them, especially reptiles, sex depends on the incubation temperature; others are hermaphroditic (meaning they contain both male and female gametes in the same individual).

Origin

The X and Y-chromosomes are thought to have evolved from a pair of identical chromosomes, termed autosomes, when an ancestral mammal developed an allelic variation, a so-called 'sex locus' - simply possessing this allele caused the organism to be male. The chromosome with this allele became the Y-chromosome, while the other member of the pair became the X-chromosome. Over time, genes which were beneficial for males and harmful to (or had no effect on) females either developed on the Y-chromosome, or were acquired through the process of translocation.

Until recently, the X and Y-chromosomes were thought to have diverged around 300 million years ago. However, recent research, particularly that stemming from the sequencing of the platypus genome, has suggested that the XY sex-determination system wouldn't have been present more than 166 million years ago, at the split of the monotremes from other mammals. This reestimation of the age of the therian XY system is based on the finding that sequences that are on the X-chromosomes of marsupials and eutherian mammals are present on the autosomes of platypus and birds. The older estimate was based on erroneous reports that the platypus X-chromosomes contained these sequences.

Recombination inhibition

Recombination between the X and Y-chromosomes proved harmful - it resulted in males without necessary genes formerly found on the Y-chromosome, and females with unnecessary or even harmful genes previously only found on the Y-chromosome. As a result, genes beneficial to males accumulated near the sex-determining genes, and recombination in this region was suppressed in order to preserve this male specific region. Over time, the Y-chromosome changed in such a way as to inhibit the areas around the sex determining genes from recombining at all with the X-chromosome. As a result of this process 95% of the human Y-chromosome is unable to recombine.

Shrinking

The human Y-chromosome has lost 1,393 of its 1,438 original genes over the course of its existence. With a rate of genetic loss of 4.6 genes per million years, the Y-chromosome may potentially lose complete function within the next 10 million years. Comparative genomic analysis, however, reveals that many mammalian species are

experiencing a similar loss of function in their heterozygous sex chromosome. Degeneration may simply be the fate of all nonrecombining sex chromosomes due to three common evolutionary forces: high mutation rate, inefficient selection and genetic drift. On the other hand, recent comparisons of the human and chimpanzee Y-chromosomes show that the human Y-chromosome has not lost any genes since the divergence of humans and chimpanzees between 6-7 million years ago, providing direct evidence that the linear extrapolation model may be flawed.

High mutation rate

The human Y-chromosome is particularly exposed to high mutation rates due to the environment in which it is housed. The Y-chromosome is passed exclusively through sperm, which undergo multiple cell divisions during gametogenesis. Each cellular division provides further opportunity to accumulate base pair mutations. Additionally, sperm are stored in the highly oxidative environment of the testis, which encourages further mutation. These two conditions combined put the Y-chromosome at a risk of mutation 4.8 times greater than the rest of the genome.

Inefficient selection

Without the ability to recombine during meiosis, the Y-chromosome is unable to expose individual alleles to natural selection. Deleterious alleles are allowed to "hitchhike" with beneficial neighbors, thus propagating maladapted alleles in to the next generation. Conversely, advantageous alleles may be selected against if they are surrounded by harmful alleles (background selection). Due to this inability to sort through its gene content, the Y-chromosome is particularly prone to the accumulation of "junk" DNA. Massive accumulations of retrotransposable elements are scattered throughout the Y. The random insertion of DNA segments often disrupts encoded gene sequences and renders them nonfunctional. However, the Y-chromosome has no way of weeding out these "jumping genes". Without the ability to isolate alleles, selection cannot effectively act upon them.

A clear, quantitative indication of this inefficiency is the entropy rate of the Y-chromosome. Whereas all other chromosomes in the human genome have entropy rates of 1.5-1.9 bits per nucleotide (compared to the theoretical maximum of exactly 2 for no redundancy), the Y-chromosome's entropy rate is only 0.84. This means the Y-chromosome has a much lower information content relative to its overall length; it is more redundant.

Genetic drift

Even if a well adapted Y-chromosome manages to maintain genetic activity by avoiding mutation accumulation, there is no guarantee it will be passed down to the next generation. The population size of the Y-chromosome is inherently limited to 1/4 that of autosomes: diploid organisms contain two copies of autosomal chromosomes while only half the population contains 1 Y-chromosome. Thus, genetic drift is an exceptionally

strong force acting upon the Y-chromosome. Through sheer random assortment, an adult male may never pass on his Y-chromosome if he only has female offspring. Thus, although a male may have a well adapted Y-chromosome free of excessive mutation, it may never make it in to the next gene pool. The repeat random loss of well-adapted Y-chromosomes, coupled with the tendency of the Y-chromosome to evolve to have more deleterious mutations rather than less for reasons described above, contributes to the species-wide degeneration of Y-chromosomes through Muller's ratchet.

Gene conversion

In 2003, researchers from MIT discovered a process which may slow down the process of degradation. They found that human Y-chromosome is able to "recombine" with itself, using palindrome base pair sequences. Such a "recombination" is called gene conversion or *recombinational loss of heterozygosity* (RecLOH).

In the case of the Y-chromosomes, the palindromes are not noncoding DNA; these strings of bases contain functioning genes important for male fertility. Most of the sequence pairs are greater than 99.97% identical. The extensive use of gene conversion may play a role in the ability of the Y-chromosome to edit out genetic mistakes and maintain the integrity of the relatively few genes it carries. In other words, since the Y-chromosome is single, it has duplicates of its genes on itself instead of having a second, homologous, chromosome. When errors occur, it can use other parts of itself as a template to correct them.

Findings were confirmed by comparing similar regions of the Y-chromosome in humans to the Y-chromosomes of chimpanzees, bonobos and gorillas. The comparison demonstrated that the same phenomenon of gene conversion appeared to be at work more than 5 million years ago, when humans and the non-human primates diverged from each other.

Future evolution

In the terminal stages of the degeneration of the Y-chromosome, other chromosomes increasingly take over genes and functions formerly associated with it. Finally, the Y-chromosome disappears entirely, and a new sex-determining system arises. Several species of rodent in the sister families Muridae and Cricetidae have reached these stages, in the following ways:

- The Transcaucasian mole vole, *Ellobius lutescens*, the Zaisan mole vole, *Ellobius tancrei*, and the Japanese spinous country rats *Tokudaia osimensis* and *Tokudaia muenninki*, have lost the Y-chromosome and SRY entirely. *Tokudaia* spp. have relocated some other genes ancestrally present on the Y-chromosome to the X-chromosome. Both genders of *Tokudaia* spp. and *Ellobius lutescens* have an XO genotype, whereas all *Ellobius tancrei* possess an XX genotype. The new sex-determining system for these rodents remains unclear.

- The wood lemming *Myopus schisticolor*, the arctic lemming, *Dicrostonyx torquatus*, and multiple species in the grass mouse genus *Akodon* have evolved fertile females who possess the genotype generally coding for males, XY, in addition to the ancestral XX female, through a variety of modifications to the X and Y-chromosomes.
- In the creeping vole, *Microtus oregoni*, the females, with just one X-chromosome each, produce X gametes only, and the males, XY, produce Y gametes, or gametes devoid of any sex chromosome, through nondisjunction.

Outside of the rodent family, the black muntjac, *Muntiacus crinifrons*, evolved new X and Y-chromosomes through fusions of the ancestral sex chromosomes and autosomes. Primate Y-chromosomes, including in humans, have degenerated so much that primates will also evolve new sex determination systems relatively soon, in about 14 million years in humans.

Human Y-chromosome

In humans, the Y-chromosome spans about 58 million base pairs (the building blocks of DNA) and represents approximately 2% of the total DNA in a male cell. The human Y-chromosome contains 86 genes, which code for only 23 distinct proteins. Traits that are inherited via the Y-chromosome are called holandric traits.

The human Y-chromosome is unable to recombine with the X-chromosome, except for small pieces of pseudoautosomal regions at the telomeres (which comprise about 5% of the chromosome's length). These regions are relics of ancient homology between the X and Y-chromosomes. The bulk of the Y-chromosome which does not recombine is called the "NRY" or non-recombining region of the Y-chromosome. It is the SNPs in this region which are used for tracing direct paternal ancestral lines.

Genes

Not including pseudoautosomal genes, genes include:

- NRY, with corresponding gene on X-chromosome
  - AMELY/AMELX (amelogenin)
  - RPS4Y1/RPS4Y2/RPS4X (Ribosomal protein S4)

- NRY, other
  - AZF1 (azoospermia factor 1)
  - BPY2 (basic protein on the Y-chromosome)
  - DAZ1 (deleted in azoospermia)
  - DAZ2
  - PRKY (protein kinase, Y-linked)
  - RBMY1A1
  - SRY (sex-determining region)
  - TSPY (testis-specific protein)

- USP9Y
- UTY (ubiquitously transcribed TPR gene on Y-chromosome)
- ZFY (zinc finger protein)

Y-chromosome-linked diseases

Y-chromosome-linked diseases can be of more common types, or very rare ones. Yet, the rare ones still have importance in understanding the function of the Y-chromosome in the normal case.

More common

No vital genes reside only on the Y-chromosome, since roughly half of humans (females) do not have Y-chromosomes. The only well-defined human disease linked to a defect on the Y-chromosome is defective testicular development (due to deletion or deleterious mutation of *SRY*). However, having two X-chromosomes and one Y-chromosome has similar effects. On the other hand, having Y-chromosome polysomy has other effects than masculinization.

Defective Y-chromosome

This results in the person presenting a female phenotype even though that person possesses an XY karyotype (i.e., is born with female-like genitalia). The lack of the second X results in infertility. In other words, viewed from opposite direction, the person goes through defeminization but fails to complete masculinization.

The cause can be seen as an incomplete Y-chromosome: the usual karyotype in these cases is 44X, plus a fragment of Y. This usually results in defective testicular development, such that the infant may or may not have fully formed male genitalia internally or externally. The full range of ambiguity of structure may occur, especially if mosaicism is present. When the Y fragment is minimal and nonfunctional, the child usually is a girl with the features of Turner syndrome or mixed gonadal dysgenesis.

XXY

Klinefelter's syndrome (47, XXY) is not an aneuploidy of the Y-chromosome, but a condition of having an extra X-chromosome, which usually results in defective postnatal testicular function. The mechanism is not fully understood; the extra X does not seem to be due to direct interference with expression of Y genes.

XYY

47,XYY syndrome is caused by the presence of a single extra copy of the Y-chromosome in each of a male's cells. 47, XYY males have one X-chromosome and two Y-chromosomes, for a total of 47 chromosomes per cell. Researchers have found that an extra copy of the Y-chromosome is associated with increased stature and an increased

incidence of learning problems in some boys and men, but the effects are variable, often minimal, and the vast majority do not know their karyotype. When chromosome surveys were done in the mid-1960s in British secure hospitals for the developmentally disabled, a higher than expected number of patients were found to have an extra Y-chromosome. The patients were mischaracterized as aggressive and criminal, so that for a while an extra Y-chromosome was believed to predispose a boy to antisocial behavior (and was dubbed the "criminal karyotype"). Subsequently, in 1968 in Scotland the only ever comprehensive nationwide chromosome survey of prisons found no overrepresentation of 47,XYY men, and later studies found 47,XYY boys and men had the same rate of criminal convictions as 46,XY boys and men of equal intelligence. Thus, the "criminal karyotype" concept is inaccurate and obsolete.

Rare

The following Y-Chromosome-linked diseases are rare, but notable because of their elucidating of the nature of the Y-chromosome.

More than two Y-chromosomes

Greater degrees of Y-chromosome polysomy (having more than one extra copy of the Y-chromosome in every cell, e.g., XYYYY) are rare. The extra genetic material in these cases can lead to skeletal abnormalities, decreased IQ, and delayed development, but the severity features of these conditions are variable.

XX male syndrome

XX male syndrome occurs when there has been a recombination in the formation of the male gametes, causing the SRY-portion of the Y-chromosome to move to the X-chromosome. When such an X-chromosome contributes to the child, the development will lead to a male, because of the SRY gene.

Genetic genealogy

In human genetic genealogy (the application of genetics to traditional genealogy) use of the information contained in the Y-chromosome is of particular interest since, unlike other genes, the Y-chromosome is passed exclusively from father to son. Mitochondrial DNA, maternally inherited, is used in an analogous way to trace the maternal line.

Non-mammal Y-chromosome

Many groups of organisms in addition to mammals have Y-chromosomes, but these Y-chromosomes do not share common ancestry with mammalian Y-chromosomes. Such groups include Drosophila, some other insects, some fish, some reptiles, and some plants. In Drosophila melanogaster, the Y-chromosome does not trigger male development. Instead, sex is determined by the number of X-chromosomes. The *D. melanogaster* Y-chromosome does contain genes necessary for male fertility. So XXY *D. melanogaster*

are female, and *D. melanogaster* with a single X (X0), are male but sterile. There are some species of Drosophila in which X0 males are both viable and fertile.

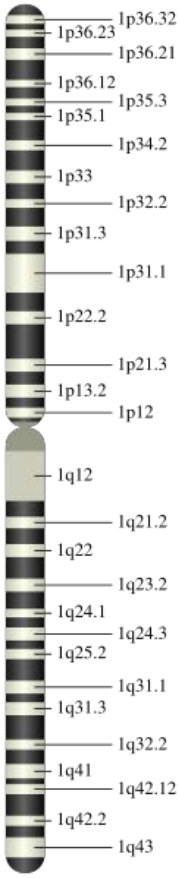
ZW-chromosomes

Other organisms have mirror image sex chromosomes: the female is "XY" and the male is "XX", but by convention biologists call a "female Y" a W chromosome and the other a Z chromosome. For example, female birds, snakes, and butterflies have ZW sex chromosomes, and males have ZZ sex chromosomes.

Chapter 10

Chromosome 1 (Human) and Chromosome 2 (Human)

Chromosome 1 (Human)



Map of Chromosome 1

Chromosome 1 is the designation for the largest human chromosome. Humans have two copies of chromosome 1, as they do with all of the autosomes, which are the non-sex chromosomes. Chromosome 1 spans about 247 million nucleotide base pairs, which are the basic units of information for DNA. It represents about 8% of the total DNA in human cells.

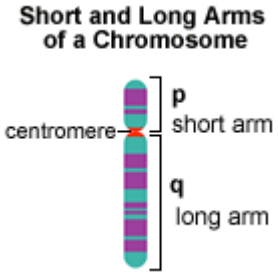
Identifying genes on each chromosome is an active area of genetic research. Chromosome 1 is currently believed to have 4,220 genes, exceeding previous predictions based on its size. It was the last completed chromosome, sequenced two decades after the beginning of the Human Genome Project.

The number of single nucleotide polymorphisms (SNPs) is about 740,000.

Genes

The following are some of the genes located on chromosome 1:

p-arm



Short and long arms

- ACADM: acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain
- COL11A1: collagen, type XI, alpha 1
- CPT2: carnitine palmitoyltransferase II
- DBT: dihydrolipoamide branched chain transacylase E2
- DIRAS3: DIRAS family, GTP-binding RAS-like 3
- ESPN: espin (autosomal recessive deafness 36)
- GALE: UDP-galactose-4-epimerase
- GJB3: gap junction protein, beta 3, 31kDa (connexin 31)
- HMGCL: 3-hydroxymethyl-3-methylglutaryl-Coenzyme A lyase (hydroxymethylglutaricaciduria)
- KCNQ4: potassium voltage-gated channel, KQT-like subfamily, member 4
- KIF1B: kinesin family member 1B
- MFN2: mitofusin 2
- MTHFR: 5,10-methylenetetrahydrofolate reductase (NADPH)
- MUTYH: mutY homolog (E. coli)
- NGF: Nerve Growth Factor
- PARK7: Parkinson disease (autosomal recessive, early onset) 7

- PINK1: PTEN induced putative kinase 1
- PLOD1: procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase 1
- TSHB: thyroid stimulating hormone, beta
- UROD: uroporphyrinogen decarboxylase (the gene for porphyria cutanea tarda)

q-arm

- ASPM: a brain size determinant
- F5: coagulation factor V (proaccelerin, labile factor)
- FMO3: flavin containing monooxygenase 3
- GBA: glucosidase, beta; acid (includes glucosylceramidase) (gene for Gaucher disease)
- GLC1A: gene for glaucoma
- HFE2: hemochromatosis type 2 (juvenile)
- HPC1: gene for prostate cancer
- IRF6: gene for connective tissue formation
- LMNA: lamin A/C
- MPZ: myelin protein zero (Charcot-Marie-Tooth neuropathy 1B)
- MTR: 5-methyltetrahydrofolate-homocysteine methyltransferase
- PPOX: protoporphyrinogen oxidase
- PSEN2: presenilin 2 (Alzheimer disease 4)
- SDHB: succinate dehydrogenase complex subunit B
- TNNT2: cardiac troponin T2
- USH2A: Usher syndrome 2A (autosomal recessive, mild)

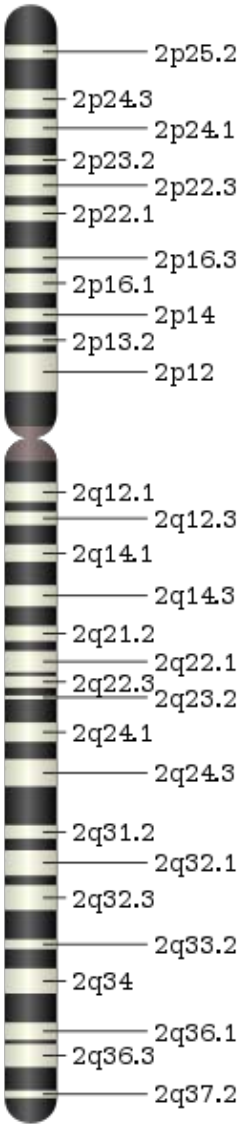
Diseases & disorders

There are 890 known diseases related to this chromosome. Some of these diseases are deafness, Alzheimer disease, glaucoma and breast cancer. Rearrangements and mutations of chromosome 1 are prevalent in cancer and many other diseases. Patterns of sequence variation reveal signals of recent selection in specific genes that may contribute to human fitness, and also in regions where no function is evident. The following diseases are some of those related to genes on chromosome 1 (which contains the most known genetic diseases of any human chromosome):

- 1q21.1 deletion syndrome
- 1q21.1 duplication syndrome
- Alzheimer disease
- Alzheimer disease, type 4
- Breast cancer
- Brooke Greenberg Disease (Syndrome X)
- Carnitine palmitoyltransferase II deficiency
- Charcot-Marie-Tooth disease
- Charcot-Marie-Tooth disease, type 1
- Charcot-Marie-Tooth disease, type 2
- collagenopathy, types II and XI

- congenital hypothyroidism
- Deafness, autosomal recessive deafness 36
- Ehlers-Danlos syndrome
- Ehlers-Danlos syndrome, kyphoscoliosis type
- Factor V Leiden thrombophilia
- Familial adenomatous polyposis
- galactosemia
- Gaucher disease
- Gaucher disease type 1
- Gaucher disease type 2
- Gaucher disease type 3
- Gaucher-like disease
- Gelatinous drop-like corneal dystrophy
- Glaucoma
- Hemochromatosis
- Hemochromatosis, type 2
- Hepatoerythropoietic porphyria
- Homocystinuria
- Hutchinson Gilford Progeria Syndrome
- 3-hydroxy-3-methylglutaryl-CoA lyase deficiency
- Hypertrophic cardiomyopathy, autosomal dominant mutations of TNNT2; hypertrophy usually mild; restrictive phenotype may be present; may carry high risk of sudden cardiac death
- maple syrup urine disease
- medium-chain acyl-coenzyme A dehydrogenase deficiency
- Microcephaly
- Muckle-Wells Syndrome
- Nonsyndromic deafness
- Nonsyndromic deafness, autosomal dominant
- Nonsyndromic deafness, autosomal recessive
- Oligodendroglioma
- Parkinson disease
- Pheochromocytoma
- porphyria
- porphyria cutanea tarda
- popliteal pterygium syndrome
- prostate cancer
- Stickler syndrome
- Stickler syndrome, COL11A1
- trimethylaminuria
- Usher syndrome
- Usher syndrome type II
- Van der Woude syndrome
- Variegate porphyria

Chromosome 2 (Human)

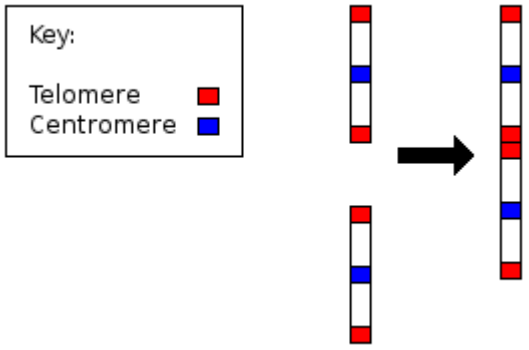


Chromosome 2 is one of the 23 pairs of chromosomes in humans. People normally have two copies of this chromosome. Chromosome 2 is the second largest human chromosome, spanning more than 237 million base pairs (the building material of DNA) and representing almost 8% of the total DNA in cells.

Identifying genes on each chromosome is an active area of genetic research. Because researchers use different approaches to predict the number of genes on each chromosome, the estimated number of genes varies. Chromosome 2 likely contains 1,491 genes, including those of the HOXD homeobox gene cluster.

Evolution

All members of Hominidae except humans have 24 pairs of chromosomes. Humans have only 23 pairs of chromosomes. Human chromosome 2 is widely accepted to be a result of an end-to-end fusion of two ancestral chromosomes.



Fusion of ancestral chromosomes left distinctive remnants of telomeres, and a vestigial centromere

The evidence for this includes:

- The correspondence of chromosome 2 to two ape chromosomes. The closest human relative, the chimpanzee, has near-identical DNA sequences to human chromosome 2, but they are found in two separate chromosomes. The same is true of the more distant gorilla and orangutan.
- The presence of a vestigial centromere. Normally a chromosome has just one centromere, but in chromosome 2 there are remnants of a second centromere.
- The presence of vestigial telomeres. These are normally found only at the ends of a chromosome, but in chromosome 2 there are additional telomere sequences in the middle.

Some argue that chromosome 2 presents very strong evidence in favour of the common descent of humans and other apes. According to researcher J. W. IJdo, "We conclude that the locus cloned in cosmids c8.1 and c29B is the relic of an ancient telomere-telomere fusion and marks the point at which two ancestral ape chromosomes fused to give rise to human chromosome 2."

Genes

The following genes are located on chromosome 2:

- ABCA12: ATP-binding cassette, sub-family A (ABC1), member 12
- ABCG5 and ABCG8: ATP-binding cassette, subfamily A, members 5 and 8
- AGXT: alanine-glyoxylate aminotransferase (oxalosis I; hyperoxaluria I; glycolicaciduria; serine-pyruvate aminotransferase)

- ALMS1: Alstrom syndrome 1
- ALS2: amyotrophic lateral sclerosis 2 (juvenile)
- BMPR2: bone morphogenetic protein receptor, type II (serine/threonine kinase)
- COL3A1: collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)
- COL4A3: collagen, type IV, alpha 3 (Goodpasture antigen)
- COL4A4: collagen, type IV, alpha 4
- COL5A2: collagen, type V, alpha 2
- HADHA: hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit
- HADHB: hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit
- MSH2: mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)
- MSH6: mutS homolog 6 (E. coli)
- NR4A2: nuclear receptor subfamily 4, group A, member 2
- OTOF: otoferlin
- PAX3: paired box gene 3 (Waardenburg syndrome 1)
- PAX8: paired box gene 8
- PELI1: Ubiquitin ligase
- SLC40A1: solute carrier family 40 (iron-regulated transporter), member 1
- TPO: thyroid peroxidase
- TBR1: T-box, brain, 1

Related diseases & disorders

The following diseases are related to genes located on chromosome 2:

- Autism
- Alport syndrome
- Alström syndrome
- Amyotrophic lateral sclerosis
- Amyotrophic lateral sclerosis, type 2
- Congenital hypothyroidism
- Ehlers-Danlos syndrome
- Ehlers-Danlos syndrome, classical type
- Ehlers-Danlos syndrome, vascular type
- Fibrodysplasia ossificans progressiva
- Harlequin type ichthyosis
- Hemochromatosis
- Hemochromatosis, type 4
- Hereditary nonpolyposis colorectal cancer
- Infantile-onset ascending hereditary spastic paralysis
- Juvenile primary lateral sclerosis
- Long-chain 3-hydroxyacyl-coenzyme A dehydrogenase deficiency
- Maturity onset diabetes of the young type 6
- Mitochondrial trifunctional protein deficiency

- Nonsyndromic deafness
- Nonsyndromic deafness, autosomal recessive
- Primary hyperoxaluria
- Primary pulmonary hypertension
- Sitosterolemia (knockout of either ABCG5 or ABCG8)
- Synesthesia
- Waardenburg syndrome

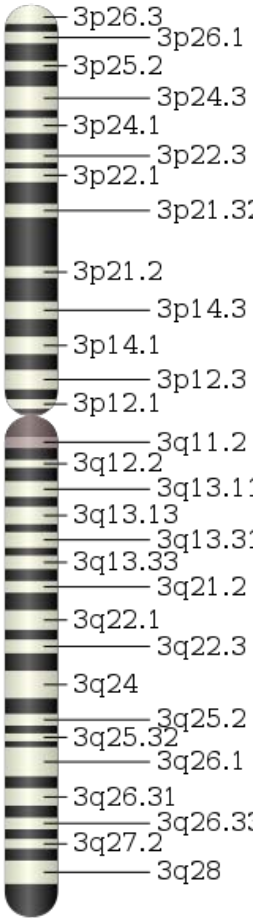
Intelligence

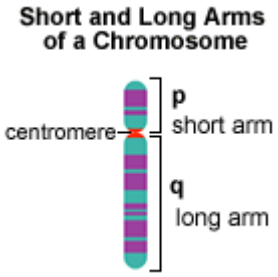
Recent studies suggest that genes on chromosome 2 may play an important role in human intelligence.

Chapter 11

Chromosome 3 (Human) and Chromosome 4 (Human)

Chromosome 3 (Human)





Short and long arms

Chromosome 3 is one of the 23 pairs of chromosomes in humans. People normally have two copies of this chromosome. Chromosome 3 spans almost 200 million base pairs (the building material of DNA) and represents about 6.5 percent of the total DNA in cells.

Identifying genes on each chromosome is an active area of genetic research. Because researchers use different approaches to predict the number of genes on each chromosome, the estimated number of genes varies. Chromosome 3 likely contains between 1,100 and 1,500 genes.

Genes

The following are some of the genes located on chromosome 3:

p-arm

- ALAS1: aminolevulinate, delta-, synthase 1
- BTD: biotinidase
- CCR5: chemokine (C-C motif) receptor 5
- CNTN4: Contactin 4
- COL7A1: Collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)
- MITF: microphthalmia-associated transcription factor
- MLH1: mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)
- OXTR: oxytocin receptor
- PTHR1: parathyroid hormone receptor 1
- SCN5A: sodium channel, voltage-gated, type V, alpha (long QT syndrome 3)
- SLC25A20: solute carrier family 25 (carnitine/acylcarnitine translocase), member 20
- TMIE: transmembrane inner ear
- VHL: von Hippel-Lindau tumor suppressor

q-arm

- CPOX: coproporphyrinogen oxidase (coproporphyria, harderoporphyria)
- HGD: homogentisate 1,2-dioxygenase (homogentisate oxidase)

- MCCC1: methylcrotonoyl-Coenzyme A carboxylase 1 (alpha)
- PCCB: propionyl Coenzyme A carboxylase, beta polypeptide
- PDCD10: programmed cell death 10
- PIK3CA: phosphoinositide-3-kinase, catalytic, alpha polypeptide
- RAB7: RAB7, member RAS oncogene family
- RHO: rhodopsin visual pigment
- SOX2: transcription factor
- USH3A: Usher syndrome 3A
- ZNF9: zinc finger protein 9 (a cellular retroviral nucleic acid binding protein)

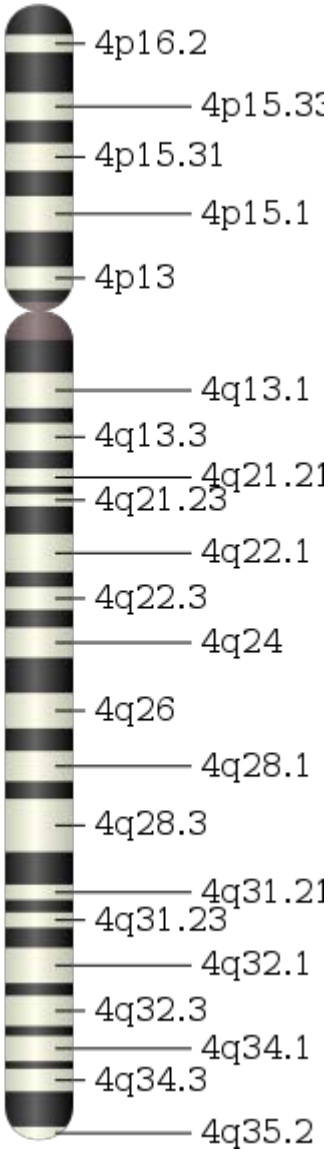
Diseases & disorders

The following diseases are some of those related to genes on chromosome 3:

- 3-methylcrotonyl-CoA carboxylase deficiency
- 3q29 microdeletion syndrome
- Alkaptonuria
- Arrhythmogenic right ventricular dysplasia
- Atransferrinemia
- Autism
- Biotinidase deficiency
- Blepharophimosis, epicanthus inversus and ptosis type 1
- Breast/colon/lung/pancreatic cancer
- Brugada syndrome
- Carnitine-acylcarnitine translocase deficiency
- Cataracts
- Cerebral cavernous malformation
- Charcot-Marie-Tooth disease, type 2
- Charcot-Marie-Tooth disease
- Chromosome 3q duplication syndrome
- Coproporphyria
- Deafness
- Diabetes
- Dopamine receptor
- Dystrophic epidermolysis bullosa
- Endplate acetlycholinesterase deficiency
- Essential tremors
- Glaucoma, primary open angle
- Glycogen storage disease
- Hailey-Hailey disease
- Harderoporphyrinuria
- Heart block, progressive/nonprogressive
- Hereditary coproporphyria
- Hereditary nonpolyposis colorectal cancer
- HIV infection, susceptibility/resistance to
- Hypobetalipoproteinemia, familial

- Leukoencephalopathy with vanishing white matter
- Long QT syndrome
- Lymphomas
- Malignant hyperthermia susceptibility
- Metaphyseal chondrodysplasia, Murk Jansen type
- Moebius syndrome
- Moyamoya disease
- Mucopolysaccharidosis
- Muir-Torre family cancer syndrome
- Myotonic dystrophy, type 2
- Myotonic dystrophy
- Neuropathy, hereditary motor and sensory, Okinawa type
- Night blindness
- Nonsyndromic deafness, autosomal recessive
- Nonsyndromic deafness
- Ovarian cancer
- Porphyria
- Propionic acidemia
- Protein S deficiency
- Pseudo-Zellweger syndrome
- Retinitis pigmentosa
- Romano-Ward syndrome
- Septo-optic dysplasia
- Short stature
- Spinocerebellar ataxia
- Sucrose intolerance
- T-cell leukemia translocation altered gene
- Usher syndrome type III
- Usher syndrome (Finland)
- Usher syndrome
- von Hippel-Lindau syndrome
- Waardenburg syndrome
- Xeroderma pigmentosum, complementation group c

Chromosome 4 (Human)



Chromosome 4 is one of the 23 pairs of chromosomes in humans. People normally have two copies of this chromosome. Chromosome 4 spans more than 186 million base pairs (the building material of DNA) and represents between 6 and 6.5 percent of the total DNA in cells.

Identifying genes on each chromosome is an active area of genetic research. Because researchers use different approaches to predict the number of genes on each chromosome, the estimated number of genes varies. Chromosome 4 likely contains between 700 and 1,100 genes.

Genes

The following are some of the genes located on chromosome 4:

- ANK2: ankyrin 2, neuronal
- CRMP1: Collapsin response mediator protein 1, a member of CRMP family
- CXCL1: chemokine (C-X-C motif) ligand 1, *scyb1*
- CXCL2: chemokine (C-X-C motif) ligand 2, *scyb2*
- CXCL3: chemokine (C-X-C motif) ligand 3, *scyb3*
- CXCL4: chemokine (C-X-C motif) ligand 4, Platelet factor-4, PF-4, *scyb4*
- CXCL5: chemokine (C-X-C motif) ligand 5, *scyb5*
- CXCL6: chemokine (C-X-C motif) ligand 6, *scyb6*
- CXCL7: chemokine (C-X-C motif) ligand 7, PPBP, *scyb7*
- CXCL8: chemokine (C-X-C motif) ligand 8, interleukin 8 (IL-8), *scyb8*
- CXCL9: chemokine (C-X-C motif) ligand 9, *scyb9*
- CXCL10: chemokine (C-X-C motif) ligand 10, *scyb10*
- CXCL11: chemokine (C-X-C motif) ligand 11, *scyb11*
- CXCL13: chemokine (C-X-C motif) ligand 13, *scyb13*
- DUX4: Thought to be inactive but 2010 research shows a key role in FSHD
- EVC: Ellis van Creveld syndrome
- EVC2: Ellis van Creveld syndrome 2 (limbin)
- FGFR3: fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism, bladder cancer)
- FGFRL1: fibroblast growth factor receptor-like 1
- Complement Factor I: Complement Factor I
- HTT (Huntingtin): huntingtin protein (Huntington's disease)
- MMAA: methylmalonic aciduria (cobalamin deficiency) cblA type
- PHOX2B: codes for a homeodomain transcription factor
- PKD2: polycystic kidney disease 2 (autosomal dominant)
- PLK4
- QDPR: quinoid dihydropteridine reductase
- SNCA: synuclein, alpha (non A4 component of amyloid precursor)
- UCHL1: ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)
- WFS1: Wolfram syndrome 1 (wolframin)
- FGF2: Fibroblast growth factor 2 (basic fibroblast growth factor)
- KDR: Kinase insert domain receptor (Vascular endothelial growth factor receptor 2)
- IGJ: linker protein for immunoglobulin alpha and mu polypeptides
- HCL2 (also called RHA or RHC): related to red hair

Diseases & disorders

The following are some of the diseases related to genes located on chromosome 4:

- achondroplasia
- bladder cancer

- Crouzonodermoskeletal syndrome
- Chronic Lymphocytic Leukemia
- Ellis-van Creveld syndrome
- Facioscapulohumeral muscular dystrophy
- Fibrodysplasia ossificans progessiva FOP
- Hemophilia C
- Huntington's disease
- Hemolytic Uremic Syndrome
- Hirschprung's disease
- hypochondroplasia
- methylmalonic acidemia
- Muenke syndrome
- nonsyndromic deafness
- nonsyndromic deafness, autosomal dominant
- Ondine's Curse
- Parkinsons disease
- polycystic kidney disease
- Romano-Ward syndrome
- SADDAN
- tetrahydrobiopterin deficiency
- thanatophoric dysplasia
- thanatophoric dysplasia, type 1
- thanatophoric dysplasia, type 2
- Wolfram syndrome