Genetics

Genetics

HORWITZ

Contents

1. Genetics | Fundamentals of Medical Science & Research

Contents

recessive or imprinted disorder

SLO 4 Explain how new mutation accounts for seemingly sporadic occurrences of genetic disorders

SLO 5 Explain how germline mosaicism causes recurrent autosomal dominantly transmitted disease from unaffected parents

SLO 6 Describe how lethal autosomal dominant mutations still give rise to disease

SLO 7 Describe genetic anticipation and explain how it alters patterns of inheritance

Complex genetics

SLO 1 Define what is meant by "complex genetics"

SLO 2 Describe normal genetic variation in the human population

SLO 3 Explain how natural selection influences disease gene frequencies

SLO 4 Explain how meiosis increases genetic diversity

SLO 5 Explain the concept of linkage disequilibrium and haplotypes

SLO 6 Describe the origins of the haplotype block structure of the genome

SLO 7 Explain how "genome-wide association analysis" (GWAS) is used to identify common genetic variants in the population that contribute to risk for common disease and how these variants differ from those responsible for Mendelian disorders

Cancer genetics

SLO 1 Describe how a proto-oncogene transforms into an oncogene

SLO 2 Explain the two-hit model of tumor suppressor gene inactivation

SLO 3 Explain how inherited DNA repair deficiency leads to cancer

Chromosomal abnormalities

SLO 1 Describe the cytogenetic organization of the human genome

SLO 2 Compare and contrast the clinical laboratory tests (karyotype, FISH, and microarrays) used for detecting chromosomal abnormalities

SLO 3 Describe human chromosomal abnormalities and how they are inherited

SLO 4 Distinguish constitutional from acquired chromosomal

abnormalities

SLO 5 List clinical scenarios for which a chromosomal abnormality should be considered

Frontiers of genetic diagnosis and therapy

SLO 1 Compare and contrast indications and limitations of different genetic tests

SLO 2 Compare and contrast sources of DNA used for genetic testing

SLO 3 Be prepared to explain possible outcomes of genetic test results to patients and their families

SLO 4 Identify forms of inherited and acquired disorders potentially amenable to gene therapy and genome editing technologies

2. Genetics Introduction

Session Learning Objectives and a quick synopsis (see main text for explanations):
The goal of this chapter is to provide a brief overview of genetic concepts.

SLO 1 Explain the different ways genetics contributes to disease risk

Distinguish single-gene Mendelian inheritance (autosomal dominant, autosomal recessive, or sex-linked recessive) from complex polygenic inheritance (which contributes to most common diseases). Differentiate germline versus somatic mutation.

SLO 2 Record and interpret a human pedigree using standard symbols

Learn to draw and read the genetic shorthand for a family tree with squares (males) and circles (females), filled (affected) and open, and so forth, to trace the inheritance of genotypes and phenotypes.

SLO 3 Define the relationship between a gene and its phenotype

Define gene, locus, genotype, phenotype, allele, and distinguish homozygosity from heterozygosity.

SLO 4 Apply probabilities to independent vs. dependent genetic events

For simple genetic outcomes, learn when to multiply (independent events) and when to add (mutually exclusive events) probabilities.

Main text

The human genome is comprised of DNA, organized into genes, and carried on 46 chromosomes. Molecular biology encompasses how the DNA is transcribed into RNA and how RNA is translated into proteins. Human genetics is the study of how our genome and its variants give rise to our human form and its variation and how

the genomes and variations of one generation are recombined and passed on to the next generation—inheritance.

SLO 1 Explain the different ways genetics contributes to disease risk

Our genes and chromosomes are constantly subject to change. Genetic variations that are due to changes in the "germline" genome will be passed from one generation to the next. They are called hereditary variations. They contrast with genetic variations acquired post-zygotically in any of the vast number of cells not contributing to gametes. Such mutations are called somatic mutations. They cannot be passed between generations. Most diseases are not primarily hereditary. Nevertheless, genetics is hugely important in human disease, and for almost any disorder, inherited risk factors, acting in isolation or in concert with the environment, can be identified.

Normal genetic variation versus disease. Most genetic variation is responsible for differences in traits distinguishing one person from the next but can also influence vulnerability to disease. Moreover, genetic factors likely have played different biological roles in different places and times throughout history as our environment has changed. Genetic variation that at one time may have conferred reproductive advantages may be harmful for modern life.

Single-gene Mendelian disorders. "Single gene" or "Mendelian" disorders are ones in which mutations in just one gene can be passed down with a predictable pattern of disease from one generation to the next. Such disorders often have dramatic clinical consequences. About 2-5% of the population has a Mendelian disorder. Many, such as cystic fibrosis, can be apparent at birth whereas others, such as Huntington Disease, are more likely to appear with aging. We will be distinguishing between single-gene Mendelian disorders and complex disorders, arising from common genetic variation of smaller effect size distributed throughout the genome and working in concert with the environment.

To understand inheritance, we need to remember that each gene lies on a specific chromosome, either on a sex chromosome or on

one of the 22 pairs of autosomes (non-sex chromosomes). Normally, there are two similar copies (alleles) of each autosomal gene. For an autosomal dominant disorder, only one of the two copies need sustain a pathogenic mutation in order to cause disease, whereas, for an autosomal recessive disorder, mutations in both copies are required to cause disease. During the formation of a gamete, one of the two alleles is randomly included in each egg or sperm. Since the distribution is random, we have to think about probability theory to calculate the risk of transmitting a disorder. The risk of transmitting an autosomal dominant disorder from affected parent to child is 1/2 and is independent of the sex of the parent. The risk of transmitting an autosomal recessive disorder when each parent is a carrier is 1/4. Sex-linked recessive disorders arise from mutations on the X chromosome. Females have two X chromosomes and are seldomly affected. However, for female carriers, the risk of conceiving an affected male (XY chromosome composition) or a carrier female (XX chromosome composition) is 1/2. (This is just a brief introduction. We will be talking about each of these forms of Mendelian inheritance in greater detail later.)

In general, for Mendelian disorders, there is a diversity of mutations within the causative gene across the affected population. The mutation responsible for the disorder in one family may not be the same as one causing disease in a different family.

Exceptions are ancestral "founder" mutations that may be over-represented among certain populations. The autosomal recessive beta-chain hemoglobin mutation that causes sickle cell anemia is one example. In the carrier state, it confers relative resistance to malaria that bestows a selective advantage throughout regions of the world where malaria is prevalent. Another example involves three Ashkenazi Jewish founder mutations in the *BRCA1* and *BRCA2* DNA repair factors, which are frequent in contemporary populations simply because in an early population bottleneck for this community there were only a few founding ancestors. Apparently, by chance they happened to possess these three mutations.

Overall, disease alleles for Mendelian disorders are rare although they may be more prevalent for certain diseases in certain populations. This is in contrast to complex disorders, where the disease-associated alleles are common, albeit of much weaker effect.

By and large for Mendelian disorders, most pathogenic mutations alter the amino acid sequence of the polypeptide encoded by the gene or lead to its loss of production. For complex disorders, mutations often have less dramatic effects on gene function, are better tolerated and therefore prevalent across the population, and consequently exert their effects more subtly, often by influencing gene regulatory regions.

Complex disease. In contrast to Mendelian disorders, most common diseases are the result of additive effects of common genetic variants conferring variable degrees of risk interacting with lifestyle and environment—so called "complex" disorders. ("Multifactorial" and "polygenic" are synonyms for "complex.") Examples include hypertension, diabetes mellitus, atherosclerotic vascular disease, autoimmune disorders, etc. For any disease, there are typically dozens of variant alleles for any of multiple associated genes, some of which increase risk and some of which decrease risk. Each variant may interact with environmental factors, such as diet and tobacco smoking, in different ways. Common diseases often cluster in families because these variants, along with environmental factors and the behaviors associated with them, are shared between relatives, although in less predictable ways than the clear-cut patterns apparent for Mendelian disorders. Some disorders, for example hypercholesterolemia, can result from either rare single-gene disorders inherited from just one parent or, more commonly, due to inheritance as a complex trait involving multiple genetic variants inherited from both parents and influenced non-genetic factors, such as diet.

Chromosomal and other genomic disorders. Chromosomes can undergo physical alterations. These changes can be microscopically visible or molecularly detectable duplications, deletions, insertions,

inversions, translocations of segments between chromosomes, or other sorts of structural rearrangements that can disrupt the protein-coding and/or regulation of multiple genes at once. Copy number variation is a type of structural variation in which a stretch of DNA is duplicated or deleted. Autosomal chromosomal abnormalities and copy number variants can produce dramatic physical phenotypes. Abnormalities of sex chromosomes typically have less severe consequences. About half of all first-trimester spontaneous abortions (miscarriages) contain chromosomal abnormalities.

Somatic genetic variation. Mutations and chromosomal abnormalities can occur in the "germline" and will typically be present in all the cells of the body. Or they can be "acquired" ("somatic") such that they arise post-zygotically and are restricted to a subpopulation of cells. Somatic genetic variation is re

sponsible for cancer but contributes to other disorders, as well as aging. Some inherited cancer predisposition syndromes depend upon an inherited gene for cancer risk, with the actual onset and tissue distribution of the resulting cancer determined by the occurrence of somatic mutations in the same or other genes.

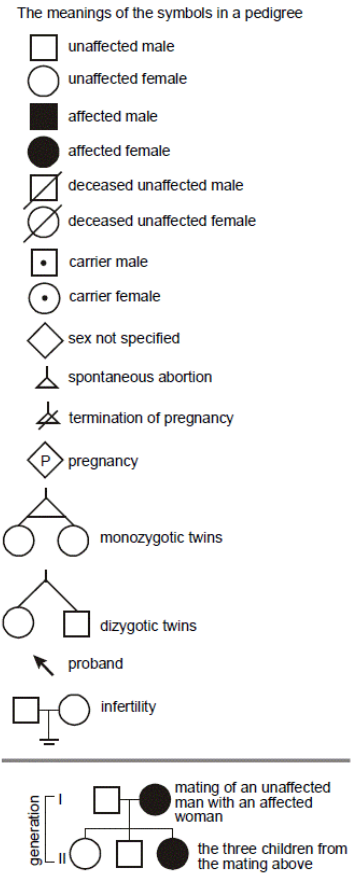
Genetic testing dependson the type of genetic variation. Mendelian disorders exhibit predictable patterns of inheritance and can often be directly tested for by gene sequencing tests. Evaluating the genetic underpinnings of common disease is more challenging and requires genome-wide based approaches. Although not typically used in a clinical setting, such testing has become popular among consumers, where considerable confusion by patients and providers surrounds their use. Chromosomal and copy number variant disorders use a unique collection of technologies for their diagnosis. The identification and the role of acquired mutations in diagnosis and treatment of cancer continues to evolve and is increasingly relied on to personalize molecular therapies. Historical routes to the discovery of the molecular genetic basis of these different sorts of disorders has varied. Genetics is one of the most rapidly changing

fields, contributes to the practice of multiple areas of medicine, and makes increasingly greater contributions to improving and individualizing the care of patients.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=241#h5p-11

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=241#h5p-12

SLO 2 Record and interpret a human pedigree using standard symbols



To simplify recording family history, we make use of symbols. Males are squares. Females are circles. Death is noted with a slash. Horizontal lines connect reproductive partners. Vertical lines indicate parent-child relationships. Diamonds are used when sex is unspecified. Putting a number in the middle of the symbol can denote multiple individuals. The notation for a fetus or pending pregnancy is a diamond with a "P" in the middle. Spontaneous abortion and termination of pregnancy are often shown with a small triangle, without and with a slash, respectively. The "proband" (usually the patient) is the individual who brought the family to attention and is highlighted with an arrow. Dizygotic (fraternal) twins have a forked line connecting them, while monozygotic (identical) twins are joined with an additional horizontal line. The symbols are shaded to indicate a disease. When more than one disease is present, the symbols may be sectored like pieces of pie. We will employ these symbols in this course. For a typical family history, recorded in a medical record, pedigrees of this sort had been considered overkill; however, readily available modules for the electronic health record can automatically

construct a detailed pedigree from a short questionnaire completed by patients.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=241#h5p-13

SLO 3 Define the relationship between a gene and its phenotype

IMPORTANT DEFINITIONS:

Gene – The basic hereditary unit, initially defined by phenotype. By molecular definition, a DNA sequence required for production of a functional product, usually a protein, but rarely an untranslated RNA.

Locus – Literally a "place" on a chromosome or DNA molecule, used fairly interchangeably with "gene" and sometimes used to refer to a collection of closely spaced genes.

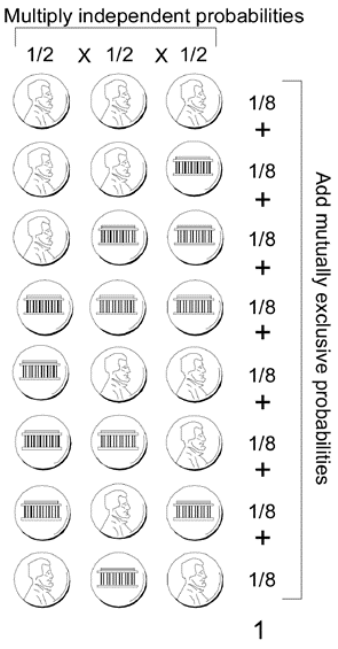
Genotype – An individual's genetic constitution, either collectively at all loci or more typically at a single locus or even at a particular position within a gene.

Phenotype – Observable expression of genotype as a trait or disease.

Allele – One of the alternate versions of a gene present in a population. An over-simplification is that there are just two, "normal" (wild type) and mutant. Of course, for any gene there are typically multiple different wild type alleles, distinguishable by clinically inconsequential "polymorphisms," and many different potential mutant alleles.

Homozygosity vs. **Heterozygosity** – For all autosomal (non sex chromosome) genes, there are two alleles, one inherited from the mother and one inherited from the father. If both versions are the same (both wild type or both mutant), we use the term

"**homozygous**," whereas if each allele is different (one wild type and one mutant) we use the term "**heterozygous**."



Genes were originally defined by Gregor Mendel, a nineteenth century Austrian monk, who formulated the basic rules of inheritance while observing peas, as the theoretical unit responsible for traits ("phenotype"). He correctly deduced that each parent contributes one copy of each of the offspring's two copies of a gene for most traits and that there are different versions of a gene, known as "alleles," within a population. Some of these alleles function "dominantly," in that they are sufficient to cause the phenotype regardless of the allele contributed by the other parent. Other alleles act "recessively" and require that the opposite parent contribute a similarly abnormal allele for the phenotype to be manifest.

SLO 4 Apply probabilities to independent vs. dependent genetic events

Mendelian inheritance involves three simple statistical rules:

- Multiply probabilities for independent events.
- Add probabilities for mutually exclusive events.
- The total probability for all possible outcomes must sum to one.

One example is a coin toss. The probability of heads or tails is one-half and is independent of any prior tosses. "Chance has no

memory." Even if there had been a run of heads, the next toss is no more or less likely to be a tails than if there had been a run of tails—unless something is wrong with the coin! The likelihood of having two consecutive heads is $1/2 \times 1/2 = (1/2)^2 = 1/4$; the probability of having three tails in a row is just $1/2 \times 1/2 \times 1/2 = (1/2)^3 = 1/8$. When you toss a coin, the two possible outcomes are heads or tails; therefore, the sum of the probabilities of each outcome must add up to a total probability of one. So, for a coin tossed three times consecutively, each of the eight possible outcomes occurs with an equal probability of 1/8.

What's the chance that a pregnant XX female will bear an XY male conception? That's an easy one; it's about 1/2. Same question but for an XX female conception? Same answer (1/2). Since these are mutually exclusive probabilities, we know that the probability of conceiving either an XY male or an XX female must be one. Reassuringly, that also happens to be the sum of 1/2 plus 1/2. (For the sake of this exercise to illustrate simple probability concepts we consider only binary, chromosomally defined sex and exclude other chromosomal and non-chromosomal sex variation.)

What's the probability that a pregnant XX female will conceive two consecutive XY males? Since these are independent events, we know that it must be 1/2 times 1/2, which equals 1/4. What about two XX females in a row? Same thing. How about one XY male and one XX female? For two children you can only have either two XY males, two XX females, or one of each, so the probability of an XY male and an XX female must equal the probability of not having two males or two females (since these are mutually exclusive events summing to one), and that is merely $1 - 1/4 - 1/4 = 1/2$. But why is it twice as likely to have an XY male and an XX female as having two XY males or two XX females? That's because there's two different sequences in which to have an XY male and an XX female. The XY male could come first, followed by an XX female, and the chance of that sequence happening is also $1/2 \times 1/2 = 1/4$. Or, the XX female could arrive first, followed by an XY male, and the chance of that sequence happening is the same.

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://uw.pressbooks.pub/fmrgenetics/?p=241#h5p-14

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://uw.pressbooks.pub/fmrgenetics/?p=241#h5p-15

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://uw.pressbooks.pub/fmrgenetics/?p=241#h5p-16

Feedback:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://uw.pressbooks.pub/fmrgenetics/?p=241

3. Single gene inheritance patterns and Prevalence, Heterozygote Frequency, Genetic Risk

Learning Objectives for these two sessions and a quick synopsis:
The overall goals are to distinguish the patterns of different forms of Mendelian inheritance (autosomal dominant, autosomal recessive, and sex-linked recessive inheritance) and identify concepts common to all forms of Mendelian inheritance.

SLO 1 From a family history, identify the likely pattern of inheritance and predict risk for other family members

Based on the sex and generation of the affected individuals in a family, the goal is to predict the most likely form of inheritance, which, in turn, can help narrow the differential diagnosis for clinical signs and symptoms of a particular disorder.

SLO 2 Describe the expected pattern of inheritance within a family for an autosomal dominant disorder

Fully penetrant autosomal dominant disorders are passed from parent to child with a 1/2 probability, and do not depend upon the sex of the parent or child. Consequently, multiple generations within a family are typically affected. There are two copies of each autosomal gene; autosomal dominant inheritance requires that only one parentally contributed allele be mutated to produce a phenotype.

SLO 3 Explain penetrance and situations where incomplete penetrance might be expected

Penetrance refers to whether the inherited allele produces a phenotype or not. Factors resulting in incomplete

penetrance include age, sex, environment, and sometimes other inherited factors.

SLO 4 Explain intuitively, if not quantitatively, how Bayes' theorem allows additional information to be incorporated into risk evaluations

Bayes theorem is a mathematical concept for incorporating educated guesses and anecdotal observations into probability calculations. It is important for all of medicine, not just genetics, and informs our approach for everything ranging from ordering the right test to formulating a differential diagnosis. Although its mathematical formulation is challenging, it can be applied conceptually.

SLO 5 Explain variable expressivity and what accounts for it

Variable expressivity is the concept that different people can have different manifestations of the same genetic disorder. The same factors contributing to incomplete penetrance help explain this phenomenon, but with the additional consideration that when comparing people from different families, there may be different disease alleles that differ between families.

SLO 6 Describe the expected pattern of inheritance within a family for an autosomal recessive disorder

Autosomal recessive inheritance typically requires that each parent be an unaffected carrier for the disease. Consequently, the disease is often restricted to just one generation in a family. When each parent is a carrier, fully penetrant autosomal recessive disorders are passed from parents to a child with a 1/4 probability and do not depend upon the sex of the child. It occurs more commonly in families in which there has been consanguinity (inbreeding) or whose ancestry descends from a small founding population. There are two copies of each autosomal gene; autosomal recessive inheritance requires that both

parentally contributed alleles be mutated to produce a phenotype.

SLO 7 Use the Hardy-Weinberg law to calculate disease incidence based on carrier frequency and vice-versa

The frequency of disease alleles in a population can be used to determine the incidence of disease in a population. This information is important for providing genetic counseling. It also explains why the beneficial effects for unaffected heterozygous carriers of certain recessive disease alleles, who outnumber those who are affected, can lead to their persistence in a population despite their potential for producing disease.

SLO 8 Describe the expected pattern of inheritance within a family for a sex-linked recessive disorder

Genes responsible for sex-linked recessive disorders reside on the X-chromosome, which is involved in biological sex determination. Consequently, penetrance is a function of sex. Males, with only copy of an X-chromosome gene, are typically affected when they inherit a mutant allele from an unaffected carrier mother, who possesses one disease-associated and one normal allele.

Main text

SLO 1 From a family history, identify the likely pattern of inheritance and predict risk for other family members

This learning objective really encompasses all of this chapter. Phenotypes may be inherited through one of several Mendelian patterns. We consider each in turn and introduce related concepts along with each form of inheritance. In general, multigenerational inheritance involving either sex suggests autosomal dominant inheritance. One mutant copy of the gene is sufficient to cause disease with an autosomal dominant disorder. Multiple affected individuals within a generation suggest autosomal recessive inheritance. In contrast to autosomal dominant disorders, one wild type copy of the gene is sufficient to prevent occurrence of an autosomal recessive disorder. Both copies of an autosomal gene

must be mutated to cause disease with this form of inheritance. Suspicion for autosomal recessive inheritance should be heightened when there is inbreeding (consanguinity) or in populations descended from a small number of ancestors. Sex-linked recessive disorders usually affect males and are inherited from carrier mothers.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-17

SLO 2 Describe the expected pattern of inheritance within a family for an autosomal dominant disorder

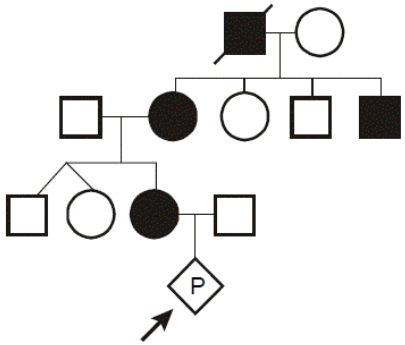
Autosomal dominantinheritance. In autosomal dominant inheritance, the responsible gene is on an "autosome," a chromosome that is not the X or Y sex chromosome. The mutation acts dominantly in that the normal allele is insufficient to compensate for the mutant allele. Heterozygotes with one copy of the disease allele and one normal allele are thus affected. Since one mutant allele has to overcome the effects of a normal allele simultaneously present within the cell, the molecular or cellular effect of the mutation is either "dominant negative," where the mutant gene product adversely affects the wild type gene product, or "toxic gain of function" where the mutant gene product gains a new and damaging property, such as the mutant protein misfolds and "gums up" (to use non-technical jargon) the inner workings of the cell, or occasionally "haploinsufficiency" where half the level of expression from just the normal allele is not sufficient for normal gene activity.

Note that homozygotes for the disease allele are generally rare. For some diseases they are much more severely affected (to the

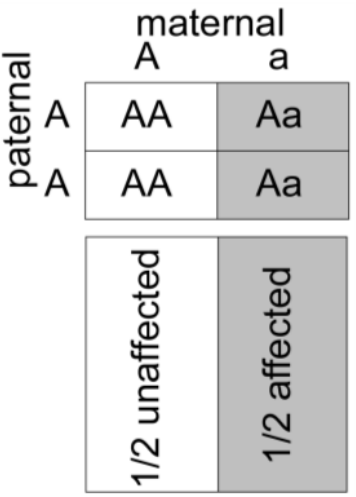
extent that they may even be considered to have a different disorder), while for others there is no difference in phenotype with the heterozygous state. In general, you don't need to worry about homozygotes for the diseased allele in autosomal dominant disorders, because of their rarity. "Codominance" refers to traits where homozygous wild type, heterozygotes, and homozygous mutant genotypes exhibit three distinct phenotypes, with the heterozygotes exhibiting severity somewhere in-between homozygous wild type or mutant. It can also be a feature of traits, such as blood types, that are not typically associated with disease.

Characteristics of autosomal dominant inheritance

- The phenotype appears in every generation, each affected person having an affected parent (except with reduced penetrance, new mutation, germline mosaicism, or anticipation).
- Each child of an affected parent has a 50% risk of inheriting the trait.
- Unaffected family members do not transmit the phenotype to their children (except with reduced penetrance, new mutation, germline mosaicism, or anticipation).
- Males and females are equally likely to transmit the trait, to children of either sex. In particular, there is male-to-male transmission (in contrast to sex-linked recessive inheritance).
- New mutations are relatively common, sometimes accounting for up to half or more of all patients, depending on the fitness



of the trait.



Example: Marfan syndrome. Marfan syndrome is a hereditary disease affecting the skeleton, eyes, and cardiovascular system. Skeletal manifestations are comprised of disproportionately long extremities including long fingers and toes (spider-like "arachnodactyly"), sternal chest deformity ("pectus excavatum" or simply "pectus"), and lateral curvature of the spine ("scoliosis"). Ophthalmologic abnormalities consist of near-sightedness ("myopia") and lens dislocation ("ectopia lentis"). The major cardiovascular abnormality is a risk for aortic aneurysm and dissection. Mutations in the fibrillin (*FBN1*) gene on chromosome 15 are the most common cause of Marfan syndrome. Note that gene names are italicized; human gene names are written in all capital letters, whereas for other organisms, only the first letter is capitalized.

In this pedigree, every affected individual has a 1/2 probability

of transmitting Marfan syndrome to each of their offspring. Thus, the chance that the fetus will inherit Marfan syndrome from the affected mother is 1/2.

One useful tool for illustrating Mendelian inheritance is the "Punnett" square. Here it can be seen that the affected mother's genotype is A/a, where "a" represents the mutant allele for the gene causing Marfan syndrome and "A" represents a "normal" or non-disease-causing allele. We put the maternal genotype horizontally on top of the Punnett square and note that during maternal meiosis there is an equal (and hence 1/2) probability of transmitting either the a or A allele to the oocyte. We put the father's genotype vertically on the side of the Punnett square. His genotype is denoted as A/A, since it is inferred that because he is unaffected he has two normal alleles of the fibrillin gene. Again, during paternal meiosis the chance of segregating either normal allele to the spermatozoa is equal and is 1/2. Since paternal and maternal meiosis are independent events, then we just multiply the individual probabilities to determine the probability that both will happen. From the Punnett square, then, we can see that there are four possible outcomes, each with probability of $1/2 \times 1/2 = 1/4$. The probabilities for all of the possible outcomes sums to one, since we know that one of these must actually happen. Two of the outcomes result in a conception that inherits Marfan syndrome with the genotype A/a, whereas two of the outcomes yield a conception that does not inherit Marfan syndrome (A/A). The probability that the conception will have inherited Marfan syndrome is therefore the number of the squares in the diagram that yield genotype A/a (two) divided by the total number of possible outcomes (four), which is 1/2. The probability that the conception will not have inherited Marfan syndrome is similarly the number of squares in the diagram that produce the genotype A/A (two squares), divided by the total number (four squares) or 1/2. Again, the sum of these mutually exclusive events is one.

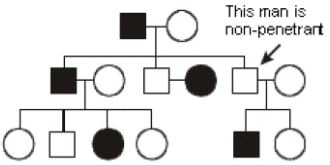
An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-18

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-19

SLO 3 Explain penetrance and situations where incomplete penetrance might be expected

Penetrance vs. Expressivity. The phenotype may not appear in all individuals or, if it is present, may not be the same in different individuals. The former concept refers to "penetrance," while the latter relates to "expressivity."

Penetrance – Whether or not a mutation in a gene has any phenotypic expression at all. In contrast to expressivity, severity is not taken into account with the concept of penetrance, which is kind of an "all or none" description.

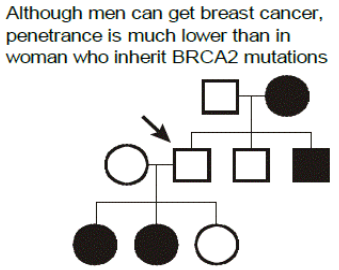


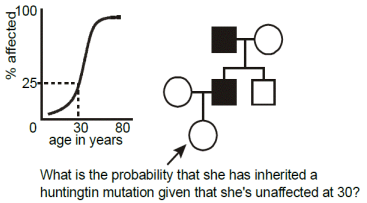
Example of incomplete penetrance:Factor V Leiden deficiency. Factor V is a component of the blood-clotting cascade. The Leiden mutation (named after the city in the Netherlands where it was discovered) is an amino acid missense substitution that impairs the function of blood clotting factor V, leading to resistance to activated protein C. Heterozygotes for the factor V Leiden allele develop a hypercoagulable state that

leads to a risk for developing venous blood clots. As many as 40% of individuals of European-American ancestry presenting with deep venous thrombosis (DVT) may be factor V Leiden heterozygotes. Because the factor V Leiden allele confers hypercoagulability in the heterozygous state, it is inherited in an autosomal dominant fashion. The clinical phenotype, DVT, will manifest over the lifetime of only about 10% of heterozygotes, however. We call this phenomenon incomplete penetrance, in that not everyone who inherits the mutation will have clinical manifestations of the disease.

Can you think of reasons for incomplete penetrance? Other thrombosis risk factors include sedentary activity (such as airliner travel or a postoperative patient confined to bed), pregnancy, the presence of a central venous catheter commonly used for chemotherapy or long-term antibiotic administration, oral contraceptive use, cancer, and other genetic factors contributing to hypercoagulability. Thus, it is a combination of other genes and environmental factors differing between individuals that accounts for incomplete penetrance.

***Example of sex-dependent penetrance*: BRCA2-associated hereditary breast and ovarian cancer.** BRCA2 is one gene responsible for hereditary breast and ovarian cancer. Males who belong to a BRCA2 breast cancer family can also develop breast cancer, albeit infrequently compared to females. With BRCA2 mutations, the penetrance for breast cancer is greatly reduced in males, an example of sex-dependent penetrance.





Example of age-dependent penetrance: Huntington disease. Huntington disease is an autosomal dominant neurodegenerative disease characterized by a dance-like movement disorder known as "chorea." Anatomically, there is focal degeneration of the caudate nucleus in the brain. Mutations in the gene, *HTT*, are responsible for the disease. The mutations are always expansions of polyglutamine-encoding CAG trinucleotide repeat tracts (more on this later). The disease demonstrates age-dependent penetrance, in that, unlike some illnesses where the phenotype is present at birth, clinical manifestations, and hence diagnosis, rarely occur before adulthood. In this pedigree, we wish to calculate the probability that a woman who is presently 30 years-old inherited Huntington disease from her affected father, knowing that at this age she is asymptomatic. From published studies, we have available data correlating the age of onset of symptoms and signs of Huntington disease in individuals heterozygous for mutations of a certain-sized polyglutamine repeat. How do we go about applying this data toward the risk calculation posed in this example pedigree? To do so we need to discuss another aspect of probability, known as Bayes' theorem.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-20

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-21

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-22

SLO 4 Explain intuitively, if not quantitatively, how Bayes' theorem allows additional information to be incorporated into risk evaluations

Bayes' Theorem. Bayes' theorem, named for an eighteenth-century mathematician and cleric, offers a formal means to change probability estimates to take into account new information. The process involves changing a "prior probability" based on new data ("conditional probability"). You then calculate a "posterior probability," a revised estimate.

There are two frequent applications of Bayes' theorem in clinical genetics: calculating the probability that someone inherited an autosomal dominant disease demonstrating age-dependent penetrance when they are at a given age and remain unaffected (our present example) and calculating the probability that someone is a carrier of a sex-linked or autosomal recessive disease after they have already had some number of unaffected children. You'll encounter Bayes' theorem in multiple other contexts throughout medical training. Among other applications, Bayes' theorem is particularly useful for explaining why screening tests are best used

in a population where there is an elevated risk for disease prevalence. Even for the most sensitive and specific screening tests, unless the pre-test probability is elevated, the frequency of false positives will greatly exceed the frequency of true positives. For this course, you will not need to use Bayes' theorem to calculate exact probabilities; however, we hope that you will gain an intuitive appreciation for its application and understand how it can be used to adjust probabilities in one direction or another.

Returning to the case, it seems reasonable that the longer one who is at risk of inheriting Huntington disease lives without developing symptoms, then the less likely it is that they inherited it. For a 30-year-old person, in the above example, the risk is reduced from 1/2 to 3/7.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-23

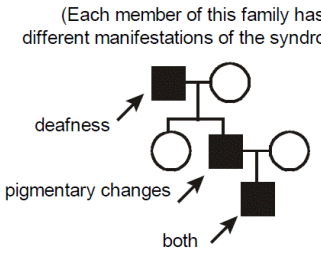
An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-24

SLO 5 Explain variable expressivity and what accounts for it

Expressivity – refers to the degree of expression of the phenotype. Unlike penetrance, expressivity takes into consideration varying breadth and/or severity of the clinical features of a disease. The term is most often used in the context of "variable expressivity."

Variable expressivity refers to mutations in the same gene having different clinical consequences, depending upon the person, and can be with reference to severity of one particular manifestation of disease or to the variability of the range of involvement of different tissues and organ systems.

Pleiotropy – Variable expressivity is not to be confused with pleiotropy. Pleiotropy refers to a mutation in a gene that results in multiple phenotypic consequences in diverse tissues.



Example ofa pleiotropic disorder exhibitingvariable expressivity: Waardenburg Syndrome. Each member of this family has different manifestations of Waardenburg syndrome. One has deafness, another has skin and hair pigmentary changes, the third has both deafness and pigmentary changes. Waardenburg syndrome therefore exhibits pleiotropy. Waardenburg syndrome is an autosomal dominant disorder caused by mutations in the gene encoding the PAX3 transcription factor. There is a cellular defect in the migration of neural crest cells during embryogenesis with a resultant phenotype affecting pigmentation and nervous system development. Note that the melanosome pigment cells are derived from the neural crest. Given that neural crest migration contributes to the formation of several organs, it's easy to understand how mutation of PAX3 has pleiotropic effects. It's similarly easy to envision one source for variable expressivity of Waardenburg syndrome between different affected individuals. Simply, not everyone will have exactly the same mutation. Individuals from different families will have different mutant alleles ("allelic heterogeneity"). In the particular case of Waardenburg syndrome and many other inherited diseases, however, even individuals within the same family, who would of course be expected to have the exact

same mutant allele, demonstrate variable expressivity, with a differing intensity and spectrum of disease.

Can you think of some explanations? The answers are basically the same as those accounting for incomplete penetrance (but with an additional wrinkle): somewhat weaker effects of so-called "modifying" genes (often called the "genetic background" of an individual) and differing environmental exposures between the individuals. For example, deafness could be influenced by occupational exposure to noise. The additional wrinkle is that there is probably also just an underlying randomness to development, in that neural crest cells migrate somewhat randomly during embryogenesis, so even monozygotic (identical) twins exhibit variable expressivity for Waardenburg syndrome.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-25

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-26

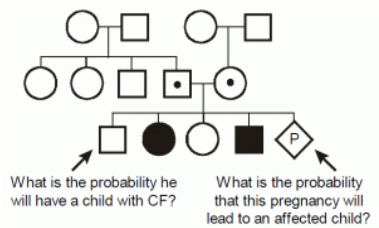
SLO 6 Describe the expected pattern of inheritance within a family for an autosomal recessive disorder

Autosomal recessive inheritance. Recessive inheritance implies that both alleles of an autosomal gene must be defective. Autosomal recessive inheritance is often the consequence of a loss-of-function

mutation, molecularly resulting from inactivation of the gene. Genes can be inactivated through a variety of different types of mutations. Affected individuals are homozygous for the disease allele and are typically the children of parents who are both unaffected heterozygous "carriers" of the disease allele.

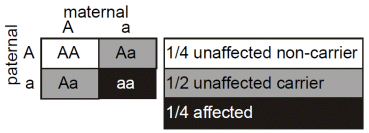
Characteristics of autosomal recessive inheritance

- If it appears in more than one family member, typically it is seen only within one sibship (children of the same parents), not in other generations.
- The recurrence risk for each sib of the proband is 25%.
- More common with consanguinity, especially for rare diseases.
- Usually, males and females are equally likely to be affected.
- New mutation is almost never a consideration.



Example: cystic fibrosis. Cystic fibrosis (CF) is among the most common autosomal recessive diseases in the European-American and certain other populations. It results from mutations in *CFTR*, a gene encoding a transmembrane chloride ion channel. The defect in the chloride channel leads to viscous mucous production which, in turn, leads to pleiotropic pathology in primarily three organ systems. Most serious are the pulmonary complications. The bronchioles become progressively dilated, inelastic, and mucous impacted ("bronchiectasis"), and the lungs become recurrently then chronically infected with *Pseudomonas*, *Burkholderia*, and other bacterial species. Pancreatic exocrine insufficiency is also frequently seen in CF patients, resulting in gastrointestinal malabsorption. Azoospermia with male infertility is apparent, although, except for the encumbrances of a chronic disease, women with CF can be fertile. There is good "genotype-phenotype" correlation in CF, in that the particular combination of mutant

alleles tends to determine both the severity and spectrum of the disease. 70% of the mutant alleles in the European-American population are the "ΔF508" mutation corresponding to a three nucleotide, in-frame deletion that ablates chloride ion channel activity. The mutation deletes a phenylalanine residue (abbreviated as "F" in the single letter code) at amino acid position 508 of the CFTR protein required for trafficking of the channel to the cell surface. The classic ΔF508 homozygote has severe lung disease as well as pancreatic exocrine deficiency.



In this pedigree, we wish to know what is the probability that the fetus in the pending pregnancy inherited CF, given that this couple has already had two children affected with CF. Since CF is an autosomal recessive disease, we can infer that both parents are asymptomatic heterozygote carriers of CF mutations. They thus have one normal allele for *CFTR* and one disease allele. We optionally put a dot in the middle of their pedigree symbols to denote the fact that they are inferred to be obligate heterozygotes. We can set up a Punnett square. In this case we place the carrier mother horizontally on the top and denote her genotype as A/a, where A corresponds to the normal ("wild type") allele and *a* represents the mutant allele. We place the father vertically on the left with the same A/a genotype. Since there is a 1/2 probability of segregating either the *a* or A allele into the sperm or egg during meiosis in the dad or mom, respectively, then we can see that there are four possible outcomes each with equal probability of 1/4. Since there are two different ways to produce a carrier state, the probability that the fetus will be a carrier is 1/4 + 1/4 = 1/2. The probability that the fetus will have genotype a/a and inherit CF is 1/4, and the probability that the fetus will be homozygous for two normal alleles (A/A) is also 1/4. The sum of these mutually exclusive and complete outcomes is one.

What is the probability that the unaffected brother will be a carrier? You might be tempted to say 1/2, but that is incorrect. We

know that he does not have CF (the black smaller square), so we can eliminate this from the four possible outcomes and cross this off of our Punnett square. This leaves only three possible outcomes in the Punnett square, and we must reset these outcomes so that they sum to one. Since we know that they all originally had equal probabilities, each of the three remaining small squares in our Punnett square should now have a probability of 1/3. Thus, because we know him to be unaffected, the probability that he is a carrier is 2/3. (This is almost certain to be a quiz or board question.)

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-27

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-28

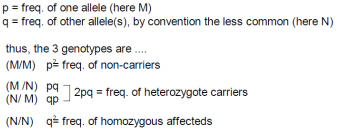
An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-29

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-30

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-31

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-32

SLO 7 Use the Hardy-Weinberg law to calculate disease incidence based on carrier frequency and vice-versa



Hardy-Weinberg law. The Hardy-Weinberg law focuses on populations rather than individuals. It is a formalization of the concept that the frequency of alleles in a large population will be constant from one generation to the next, under assumptions that mating is random (the genotype doesn't influence mate selection) and that the genotype has no selective effect on success at producing offspring. The last three equations below express this law. Again, the sum of the three possible genotypes will equal 1.0.

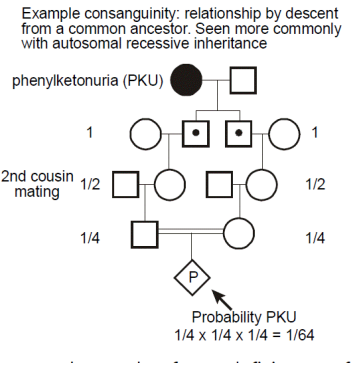
A practical implication of the Hardy-Weinberg law is that we can calculate the probability that individuals are carriers of a recessive gene mutation from the prevalence of the disease in the population. Importantly, if these predictions do not match observed population frequencies then we can infer that the mutation does alter genetic fitness (i.e., likelihood of reproduction).

Let's refer back to the example CF pedigree and ask what is the probability that the brother will have a child with CF. We previously determined that he has a 2/3 chance of being a carrier. We next need to calculate the probability that his mate will be a carrier of CF. We can do this from only knowing the prevalence of the disease in the population, assuming that the mutations do not alter genetic fitness. (Prevalence is the total number of individuals in a population with the condition divided by the total population.) Let's say that his mate is of broadly European-American ancestry. The prevalence of CF in the European-American population at birth is about 1/2,000. Since everyone with CF must have the homozygous mutant genotype (N/N), we know that $q^2 = 1/2{,}000$. Thus, q, the frequency of the mutant alleles in the population, equals the square root of 1/2,000, which is about 0.022. We can solve for p explicitly since we know that the sum $p + q = 1$, but we don't really need to do this for these kinds of calculations.

It turns out that since q is usually so small, meaning p is usually so close to one, we can just approximate $p \approx 1$ when using the Hardy-Weinberg equation.

The frequency of heterozygote carriers in the population is thus just 2pq, which is approximately equal to 2q, which is 0.044. So about 4.4% of the European-American population are CF carriers, corresponding to about 1/23 European-American people. (Note that even for a somewhat rare autosomal recessive disease like CF,

heterozygote carriers are fairly common, because 2pq >> q2.) Thus, to calculate the probability that the brother in the above pedigree will have a child with CF, it is his chance of being a carrier (2/3) times the chance that, if he were a carrier, that he would transmit the mutant allele during the meiosis producing his sperm (1/2) times the chance that his mate, randomly selected from the European-American population, would also be a carrier (1/23) times the chance that—should she be a carrier—that she would transmit the mutant allele during the meiosis producing her egg (1/2). And the answer is 2/3 × 1/2 × 1/23 × 1/2 = 0.008 or 0.8%. CF generally adheres to Hardy-Weinberg equilibrium. Negative effects on genetic fitness are generally restricted to the small q2 population of affected homozygotes with the mutation, and, if anything, the much larger 2pq population of unaffected heterozygote carriers may actually have a reproductive advantage, for various postulated reasons.



Example: hemochromatosis. It should be emphasized that incomplete penetrance and variable expressivity, as well as pleiotropy, are all seen with autosomal recessive disease (or sex-linked recessive inheritance, which we will discuss shortly), just as is the case with autosomal dominant inheritance. A good example of an autosomal recessive disease that demonstrates both incomplete penetrance and variable expressivity (in addition to pleiotropy) is hemochromatosis. Hemochromatosis, or "iron overload syndrome" is among the most common recessive diseases in individuals of European-American descent. It results from mutations in a gene, *HFE*, involved in iron transport via binding to the transferrin receptor. A single allele (C282Y), resulting in a cysteine-to-tyrosine missense amino acid substitution at the 282nd position of the

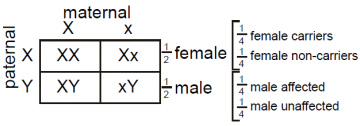
polypeptide, accounts for about 88% of all disease-associated alleles in this population. (A milder allele, H63D, accounts for most of the rest of the mutations in this population.) The disease clinically manifests with cirrhosis and consequent risk for hepatocellular carcinoma, characteristic arthritis, bronzing of the skin, cardiomyopathy, diabetes mellitus, and, in males, testicular atrophy. Elevation of the serum iron carrier protein, ferritin, serves as a laboratory marker of disease. Hemochromatosis is simply treated, if diagnosed early enough, with phlebotomy (blood draw), since red blood cells represent the major form of storage of iron in the body. It used to be definitively diagnosed through liver biopsy and corroborating laboratory studies, including increased levels of serum ferritin and transferrin, but DNA diagnostic studies are now the gold standard.

Not everyone who is homozygous for the most common mutant allele will develop symptoms of hemochromatosis. Thus, this disease demonstrates incomplete penetrance. The reasons for this are that not everyone is exposed to the same environmental and genetic factors. The environmental factors would include dietary iron intake. Other genetic factors regulating iron metabolism modify hemochromatosis risk. Furthermore, this disease demonstrates age-dependent penetrance, merely because it takes some time for the toxic accumulation of iron in organs. The disease also demonstrates sex-dependent penetrance. Since females are more likely to menstruate and have an average lower red blood cell mass, their total body iron stores tend to be less than those of non-menstruating males, and they are less likely to develop organ toxicity. Furthermore, the disease demonstrates variable expressivity, in that not everyone who has hemochromatosis will exhibit similar clinical severity or patterns of organ system involvement. For example, obese individuals and those with other hereditary risk for diabetes mellitus might be more likely to develop this complication of the disease. Individuals who consume ethanol may be more likely to sustain liver damage. Finally, individuals who are “compound heterozygotes” for H63D/C282Y alleles (as opposed

to C282Y homozygotes), typically have milder disease, while H63D homozygotes are often unaffected.

Consanguinity. "Consanguinity" refers to relationship by descent from a common ancestor (also known as "inbreeding"). Consanguinity is a concern in autosomal recessive disease, because if this is a rare disease (due to an infrequent allele), then the disease will occur more commonly in individuals whose parents are related. Consanguinity is noted on a pedigree by two horizontal lines between the male and female partner.

Always consider the possibility of consanguinity in the approach to care of a patient with autosomal recessive diseases, especially when the disease is rare.



Example: phenylketonuria. In this example, we can see how there is an increased probability of a recessive disease, here phenylketonuria (PKU), occurring in a pregnancy resulting from a consanguineous mating. PKU most commonly results from deficiency of phenylalanine hydroxylase, which is involved in the metabolism of the amino acid phenylalanine. This leads to the accumulation of neurotoxic levels of phenylalanine. Clinical manifestations include intellectual disability, seizure disorder, a "mousy" odor, and hypopigmentation of skin and hair. Children who inherit the metabolic defect, but who avoid dietary exposure to high concentrations of phenylalanine, however, are essentially normal. PKU screening at birth has therefore become routine. It is important to ask about the possibility of consanguinity when taking a family history, especially when considering the possibility of an autosomal recessive disease. For the case of PKU, its occurrence is relatively uniform across the globe; however, an increased frequency has been seen in individuals from the Romani population. The diagnosis of a rare disease in an ethnically isolated or otherwise consanguineous pedigree should make one think about the possibility of autosomal recessive inheritance.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-33

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-34

SLO 8 Describe the expected pattern of inheritance within a family for a sex-linked recessive disorder

Sex-linked recessive. Sex-linked recessive inheritance occurs when the mutated gene resides on the X chromosome and acts recessively. This is also known as "X-linked recessive inheritance." Since females have two copies of the X chromosome, in order to be affected, both copies of the responsible gene must correspond to a disease-causing allele. This is much less likely than in the situation for males, in which there is only a single X chromosome. Males are therefore considered "hemizygous" for the X chromosome. Males inherit their single X chromosome from their mother, so sex-linked recessive disease follows a maternally inherited distribution pattern in the family.

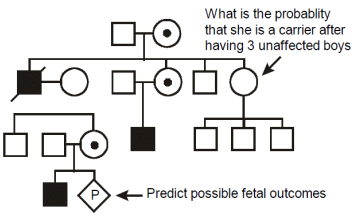
Characteristics of sex-linked recessive inheritance

- Males are more commonly affected than females.
- The gene responsible is transmitted from an affected male through his female offspring, who are seldomly affected. Each female offspring is an obligatory heterozygous carrier. Each of

the carrier female offspring's male children has a 50% chance of inheriting it.
- No male-to-male transmission occurs.
- The affected males in a pedigree are usually related through females.
- Heterozygous female carriers are usually unaffected, but some may express the condition with variable severity (called skewed "Lyonization" or "skewed X-chromosome inactivation").

Example: hemophilia A. Hemophilia A is among the most common of sex-linked recessive diseases. It results from a deficiency of factor VIII, a component of the blood clotting cascade. Most of the bleeding is into joint spaces or into the gut. Males who have a factor VIII mutation are affected because they are hemizygous for the gene, having only one copy of the X chromosome where the gene resides. Females who have a factor VIII mutation are usually unaffected. Although they may have reduced levels of factor VIII activity (resulting from the phenomenon of X chromosome inactivation, described below), this is usually sufficient to prevent development of clinical symptoms.

Here is the pedigree of an affected family. What is the probability that the fetus in the pending pregnancy will have inherited hemophilia A? We can assume that the pregnant mother is a carrier. She has already had one affected male child, and there is an extensive history of hemophilia A in males in the family. (As you'll see later, if there were only this one affected male child, we could not exclude the possibility that he is the result of a new mutation.) Since the mother is a carrier, then we can deduce that she has one normal factor VIII allele and one mutant factor VIII allele. The probability that she will transmit the mutant allele during oogenesis is therefore 1/2. The probability that the fetus will inherit the mutant



factor VIII allele is therefore also 1/2. In this case, the sex of the fetus is important. If the fetus inherits an X chromosome from the father and is female, then she will probably not be affected with hemophilia A (although there are some exceptions that we will come to). So, we need only be concerned with the possibility that the fetus inherits a Y chromosome from the father and is a male—an event with probability of 1/2. The product of these two independent events, the 1/2 probability that the mother will contribute her mutant factor VIII allele to the egg and the 1/2 probability that the father will contribute a Y and make the conceptus a male, is 1/4, and this is the probability that the fetus will be a male inheriting hemophilia A. Again, one can draw out a Punnett square, but we should start to be able to see these things by now without the aid of such didactic tools.

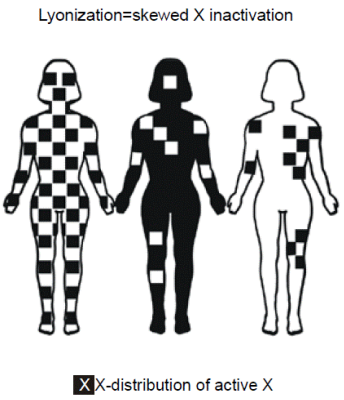
Now let's ask a more difficult question. What is the probability that the woman indicated by an arrow in the pedigree is actually a carrier of hemophilia A after giving birth to three unaffected children? At birth, she was at 1/2 risk of having inherited the mutant allele from her carrier mother. However, as she keeps giving birth to unaffected males, we might begin to doubt that she is a carrier. The more unaffected males she produces, then the less likely it is that she actually inherited carrier status for hemophilia A. It is almost as if she is revealing to us her hand of cards by letting us know what remains in the deck, card by card. To precisely calculate her risk requires use of Bayes' theorem. We won't do so here because you won't need to perform these calculations for the course. However, in case you are interested, the risk that she is a carrier after having given birth to three unaffected males is reduced to 1/9.

To try to get an intuitive grasp for this question, think of the extreme case. What if she were initially at 1/2 risk to be a carrier, but she had 25 unaffected male children? Of course, at this point we would begin to seriously doubt that she inherited the mutant factor VIII allele.

X *chromosome inactivation.* In XX female individuals, one of the

two copies of the X chromosome is largely inactivated, in a process known as "Lyonization."

Lyonization – a term used for the random inactivation of one of the X chromosome in each cell of a female is named after Mary Lyon, who discovered X inactivation. The word is most often used when there is a "skewed" or "unfavorable" pattern of X chromosome inactivation, such that the female is at least partly affected for a sex-linked recessive disorder.

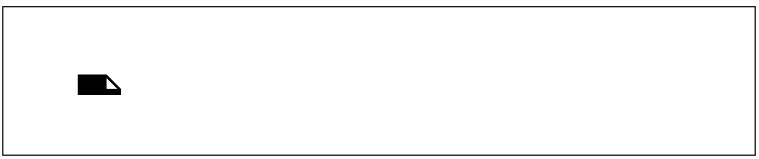


Females possess two copies of X chromosome genes since they have two X chromosomes. In contrast, males have only one copy of genes residing on the X chromosome since they only have one X chromosome. The Y chromosome, for the most part, does not contain any of the same genes. The evolutionary rationale for inactivating one of the X chromosomes in females is presumably as a mechanism of "dosage compensation" to ensure that both sexes have the same number of functioning alleles for genes on the X chromosome. The particular X chromosome that is transcriptionally silenced in any given cell well be condensed, making it visible even during "interphase" of the mitotic cell cycle, when chromosomes are ordinarily elongated and not microscopically visible. The name for the condensed microscopic appearance of the X chromosome is the "Barr body." Females ordinarily have one Barr body per cell, whereas males have zero. Since X chromosome inactivation in humans is randomly distributed in early embryonic development throughout tissues, skewing of inactivation might result in a segmental pattern of distribution of expression of the mutant gene and resulting clinical phenotype as shown in the middle and right figures. Some women could have some locally uneven patterns of X inactivation, merely

due to random statistical variation. There are a few genes in females, however, that do escape X inactivation, even though the rest of the chromosome is shutdown.

At a molecular level, the inactive X chromosome expresses a unique gene, XIST, encoding a non-translated RNA. The XIST RNA physically accumulates along the inactive X chromosome causing inactivation of most of the genes on this chromosome and leading to its condensation into a Barr body. Somewhat paradoxically, the XIST gene is not expressed from the active X chromosome (whereas the other genes on this copy of the X chromosome are expressed). On the active X chromosome, the promoter of the XIST gene is heavily methylated at CpG sequences, thereby silencing its expression.

Consider factor VIII in hemophilia A. Most of the factor VIII production occurs in the liver where it is secreted into the peripheral blood stream. Extreme skewing in X inactivation is uncommon for hemophilia A. This is because it does not really matter whether there is a locally uneven pattern of X inactivation in the liver cells. On balance, close to half of the hepatocytes will be producing factor VIII, and it will all get averaged out upon secretion into the blood. A completely different situation pertains to a protein that is not secreted, but rather whose effect is local to the cell that produces it. A good example is **Duchenne muscular dystrophy**, a sex-linked autosomal recessive severe form of muscular dystrophy primarily affecting males but for which females are sometimes symptomatic. In the case of Duchenne muscular dystrophy, the defect is in a protein, dystrophin, which connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane. If there were a skewed pattern of X inactivation in a particular muscle, then we might expect that muscle to be weak and there to be a focal, segmental pattern of weakness.



An interactive H5P element has been excluded from this version of the text. You can view it online here: https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-35

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://uw.pressbooks.pub/fmrgenetics/?p=254#h5p-36

Feedback:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://uw.pressbooks.pub/fmrgenetics/?p=254

4. Other types of inheritance

Session Learning Objectives and a quick synopsis:
While most single gene disorders obey straightforward rules of inheritance, complicated loopholes in Mendel's laws modify how some clinical traits are inherited: Genes encoded in the mitochondria are transmitted maternally. Uniparental gene expression for certain genes is due to imprinting, in which only one parental allele is expressed, or uniparental disomy, in which both alleles are derived from the same parent. New mutations give rise to the sudden appearance of a trait in a family and also account for germline and somatic mosaicism. The phenomenon of anticipation leads to worsening clinical severity or earlier age of onset for some disorders from one generation to the next due to mutations caused by unstable repetitive DNA sequences or shortening of the ends of chromosomes (telomeres).

SLO 1 Describe features of mitochondrial inheritance

Mitochondrial DNA is inherited from maternal mitochondria (with rare exception), meaning that inheritance of disorders caused by mutations in mitochondrial genes is from mother to child.

SLO 2 Distinguish maternal from paternal imprinting

A handful of autosomal genes are imprinted. Maternal imprinting means that the allele of a particular gene inherited from the mother is transcriptionally silent, and the paternally inherited allele is active. If a mutation in a maternally imprinted gene causes a dominant disorder, then that disorder cannot be inherited from the mother because the maternal allele is not expressed, anyway. A disease-causing mutation in such a gene can only be inherited from the father (who may or may not be affected, depending upon which of his parents he inherited the mutation from). Paternal imprinting is the opposite; the paternally inherited

allele is silenced, and the maternally inherited allele is active. If a mutation in a paternally imprinted gene causes a dominant disorder, then that disorder cannot be inherited from the father because the paternal allele is not expressed, anyway. A disease-causing mutation in such a gene can only be inherited from the mother (who may or may not be affected, depending upon which of her parents she inherited the mutation from). An unaffected father who inherited a maternally imprinted disease-causing gene mutation from his mother has a 1/2 probability of transmitting the disorder to a child, regardless of the sex of the child. An unaffected mother who inherited a paternally imprinted disease-causing gene mutation from her father has a 1/2 probability of transmitting the disorder to a child, regardless of the sex of the child.

SLO 3 Explain how uniparental disomy can give rise to an autosomal recessive or imprinted disorder

Uniparental disomy (also known as isodisomy) occurs when both alleles for a gene are inherited from the same parent. In other words, that particular allele is duplicated, and the allele from the opposite parent is absent. While several mechanisms can cause this to occur, typically it results from a chromosome nondisjunction event early during embryogenesis. In that case, both chromosomes are inherited from the same parent.

If an allele responsible for an autosomal recessive disorder resides on a chromosome duplicated due to isodisomy, then it will cause that disorder because the affected individual will be homozygous for the mutation.

If a gene affected by uniparental disomy is one that is ordinarily maternally imprinted and it is the maternal allele that is duplicated, then there will be no expression of that gene (because the paternal allele from which it is normally expressed is lost), resulting in disease. If a gene affected by uniparental disomy is one that is ordinarily paternally

imprinted and it is the paternal allele that is duplicated, then there will be no expression of that gene (because the maternal allele from which it is normally expressed is lost, resulting in disease).

SLO 4 Explain how new mutation accounts for seemingly sporadic occurrences of genetic disorders

Mutations must begin somewhere. When they occur for the first time in a gamete, then the conceptus is the first in the family to inherit that disease. The likelihood of this happening is inversely proportional to the severity of the disease. A disease that is lethal in childhood can only occur sporadically because those who have it do not live to reproductive age.

SLO 5 Explain how germline mosaicism causes recurrent autosomal dominantly transmitted disease from unaffected parents

Another time when new mutations occur is in one of the cells of a developing embryo. Depending upon which type of tissue the affected cell gives rise to, the patient may or may not experience much in the way of disease, or disease manifestations may be restricted to a certain part of the body. However, if the mutant clone gives rise to testes or ovaries, then it has the potential to be inherited through the resulting sperm or egg.

SLO 6 Describe how lethal autosomal dominant mutations still give rise to disease

Some mutations are so deleterious that if all the cells of the body contained the mutation then the embryo could not survive. Nevertheless, a mutation can occur in a particular cell during embryogenesis, leading to a subset of tissue containing the cell with the mutation, thereby restricting disease manifestations to certain parts of the body.

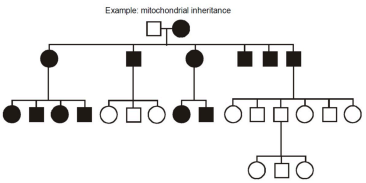
SLO 7 Describe genetic anticipation and explain how it alters patterns of inheritance

Genetic anticipation refers to worsening severity or earlier

age of onset for a disease from one generation to the next. It is caused by mutations involving some genes containing an unstable repetitive element that intergenerationally increases in length, often in a parent of origin specific manner. Anticipation can also be caused by progressive shortening of telomeres, which are repetitive sequences at the ends of chromosomes, when there are mutations in genes encoding telomerase, which is responsible for maintaining telomere length, particularly in the germ cell compartment.

Main text

SLO 1 Describe features of mitochondrial inheritance



Maternally inherited transmission occurs when the defective gene is encoded in the mitochondrial genome. The mitochondria are organelles in the cytoplasm that perform aerobic energy metabolism. Because the cytoplasmic volume of an oocyte and their number of mitochondria is so vastly larger than for spermatozoa, mitochondria are inherited nearly exclusively from the oocyte during the formation of the zygote making them nearly exclusively maternal in origin. Mitochondria contain their own circular DNA genome plus RNA polymerases and protein translation apparatus including ribosomes and tRNA. That genome has only sixteen thousand base pairs and uses a genetic code that is slightly from the nuclear genome. Most children of an affected female are usually affected, to varying degrees.

Disorders caused by mitochondrial gene mutations typically involve nerves and muscle. Examples include **myoclonus epilepsy with ragged red fibers (MERRF)**; **mitochondrial myopathy, encephalopathy, lactic-acidosis, and stroke-like episodes (MELAS),** and **Kearns-Sayre syndrome**. Often they arise from point mutations, meaning that they result from substitution, deletion, or insertion of a single DNA base pair. These diseases tend to have

some degree of clinical overlap, and the first two are pretty much as described by their lengthy names. In particular, ragged red skeletal muscle fibers, evident with a special stain, is a diagnostic feature of mitochondrial disorders.

Heteroplasmy vs. Homoplasmy – In mitochondrial inheritance, the disease is conventionally thought to be inherited exclusively from the mother, and usually all of an affected mother's children also inherit the disorder. However, sometimes children can escape inheriting a mitochondrial disorder or can be affected to variable degrees, depending upon how the mutation is distributed within the heterogeneous population of hundreds of mitochondria inherited from the mother. Additionally, a non-penetrant mother may sometimes transmit disease to her children. Therefore, the proportion of mutant mitochondrial molecules contributes to the penetrance and expressivity of some mitochondrial disorders. This concept is referred to as "heteroplasmy." The contrasting term, "homoplasmy," refers to a cell containing a uniform population of mitochondrial DNA, either entirely wild type or entirely mutant; correspondingly, the child inherits a uniform population of mitochondrial genomes from the mother.

Not all the proteins responsible for mitochondrial function are encoded by the mitochondrial genome; some are products of ordinary, nuclear genes, and this has an interesting implication. A rare but intriguing class of mitochondrial disorders are the result of mutations in the mitochondrial genome, but the inheritance is "Mendelian." In these unusual disorders, defects in nuclear genes encoding enzymes involved in mitochondrial DNA replication lead to frequent acquired mutations in the mitochondrial genome in many cells. These disorders typically present with neuromuscular features, similar to other mitochondrial disorders.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-37

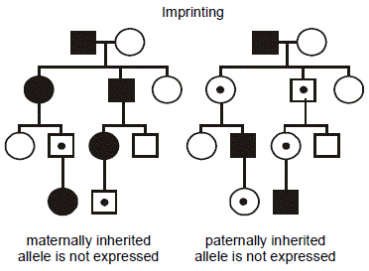
An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-38

SLO 2 Distinguish maternal from paternal imprinting

Normally, genes are equally well expressed from both parental alleles. But for a small number of genes, about 75 or so in humans, expression occurs only from either the maternal or paternal allele. They are said to be imprinted. For a subset of imprinted genes, only the maternal allele is expressed. For a different subset of imprinted genes only the paternal allele is expressed. For disease-causing mutations in imprinted genes, inheritance occurs only from one parent, because the gene is expressed only from either the maternally or paternally inherited allele. The most extreme examples of imprinting are the tumors, hydatidiform mole and ovarian teratoma. Hydatidiform mole results from fertilization of a polar body (an anucleate biproduct of female meiosis, more on this later) by an X-chromosome-containing sperm; the result is all placenta without an embryo. In contrast, ovarian teratoma results from an egg spontaneously becoming diploid and acting as if it has been fertilized; it gives rise to embryonic differentiation, but no placenta. Both are 46,XX and therefore chromosomally identical. The only difference between these two (and a normal female conceptus) involves the differential expression of just a handful of imprinted genes required for normal development.

With imprinting, gene expression occurs from only one parent's allele (the maternal or the paternal, depending upon the gene). For all known imprinted genes, the particular gene and the particular parental allele that is expressed or silenced is constant from one person to the next. Disorders associated with imprinted genes can only be inherited from either the mother or the father, depending upon the gene causing the disorder.

In the figure, consider two hypothetical, identically structured pedigrees. (We put a dot in the middle of the symbols for individuals who possess the mutant allele yet are unaffected.) The family on the left transmits a disease caused by a maternally imprinted gene where only the paternally inherited allele is expressed (i.e., transcribed). Consequently, if the mutation is inherited from the father, then only the mutant gene is expressed; since the normal, maternal allele is not expressed, then the phenotype appears. The reverse situation is considered in the pedigree on the right: here the disease results from mutation of a paternally imprinted gene where only the maternally inherited allele is expressed. Thus, when the mutation is inherited from the mother, there is no compensatory expression of the normal, paternal allele, and the phenotype emerges. Note that there is still a 1/2 probability that the mutant, vs. normal, allele will be transmitted at each meiosis, and so imprinting is superimposed on an autosomal dominant pattern of inheritance. Further note that it is immaterial whether the parent is affected or not—only the sex of the transmitting parent determines whether inheritance of the imprinted gene will cause the disease.



To reiterate, "paternal imprinting" specifically means that that the paternal allele for that gene is silenced and that only the maternal allele is expressed, and "maternal imprinting" means that the

maternal allele for that gene is silenced and only the paternal allele is expressed.

Prader-Willi and Angelman syndrome are the best-known examples of diseases resulting from imprinting. And, they add another wrinkle to the complexity of imprinting; these two diseases result from oppositely imprinted genes in the same region of chromosome 15. Prader-Willi is inherited through mutations in the paternal homolog of this region of chromosome 15; Angelman syndrome results from inheritance of identical or different mutations in the maternal homolog of this chromosomal region.

In Prader-Willi and Angelman syndromes, an identical "chromosomal microdeletion" (also known as "copy number variant") of several genes in a critical region of chromosome 15 can have a different disease outcome depending on which parent it is inherited from. Prader-Willi syndrome is clinically characterized in the neonatal period by failure to thrive, hypotonia, and mild to moderate intellectual disability. Later in childhood, these individuals have difficulty with satiety and have a notorious appetite. (One source for advice about this illness advises: "Refrigerators, cupboards, and garbage cans need to be locked.") Consequently, individuals with Prader-Willi are quite obese. The oppositely imprinted counterpart of Prader-Willi syndrome is Angelman syndrome and consists of a somewhat small body habitus with severe intellectual disability and a marionette-like scissoring gait.

The molecular genetics of Prader-Willi and Angelman syndrome are complicated. Suffice it to say that there is an "imprinting center," a small region within the imprinted domain on chromosome 15 that exerts a more regional effect on gene expression from this portion of the chromosome. The imprinting mechanism relies on alternatively spliced short transcripts and site-specific DNA methylation in this region. There are at least two nearby and oppositely imprinted genes, *SNRPN*, encoding a protein that functions in pre-mRNA processing and may contribute to tissue-specific alternative splicing, and *UBE3A*, an E3 ubiquitin ligase

regulating proteasome-mediated proteolytic degradation of other proteins. Mutations in SNRPN appear sufficient to cause Prader-Willi syndrome, when inherited from the father. Mutations in UBE3A cause Angelman syndrome, when inherited from the mother. For some patients, other mutations in this region may contribute. A deletion large enough to remove both SNRPN and UBE3A can cause both Prader-Willi and Angelman syndrome to appear in the same family, depending upon the sex of the transmitting parent.

The biological role of imprinting remains unclear. Most of the handful of genes that are known to be imprinted are generally responsible for growth effects in fetal life. Some have argued that this represents a battle of the sexes at the molecular and cellular level, in which the father imprints these genes in an "on" position to insure that they are expressed in the conception that he has fathered so that he will have big, presumably healthier, kids, whereas the mother imprints these genes in the "off" position, so that her fetus will not grow too large and endanger her own well-being.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-39

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-40

SLO 3 Explain how uniparental disomy can give rise to an autosomal recessive or imprinted disorder

Uniparental disomy occurs when both copies of a particular chromosome (or even only part of a chromosome) are inherited from just one parent, with no contribution from the other parent.

Given that the chromosome 15 region responsible for Prader-Willi and Angelman syndromes is imprinted, it is not surprising that uniparental disomy can also be a mechanism for these two disorders. In particular, maternal uniparental disomy for chromosome 15, in which there is no contribution of essential genes only expressed from paternal alleles, is one cause of Prader-Willi syndrome.

Uniparental disomy does not require imprinting in order for it to be a genetic cause of disease. In fact, it was first discovered in a situation where an autosomal recessive disease was found in a child where only one of her parents was a carrier for the mutant allele. In that case, a child had cystic fibrosis. Only her mother was a heterozygous carrier of a mutant allele for the *CFTR* gene on chromosome 7, and her father was homozygous wild type. Molecular investigations revealed that both of the child's copies of chromosome 7 were of maternal origin, thereby conferring upon the child two mutant alleles for *CFTR*. This child also had excessively short stature. Presumably, uniparental disomy also had an effect on one or more other genes on chromosome 7 capable of governing skeletal growth, either through duplicating a variant allele that behaved recessively or through an imprinting effect.

The frequency with which uniparental disomy occurs is difficult to estimate, since if disease-causing alleles of the disomic genes are not involved, or the disomic region is not imprinted, then its occurrence will likely be clinically inconsequential and go undetected. Probably the most common mechanism for generating uniparental disomy is the "rescue" of a trisomic conception. We will revisit this topic when we discuss chromosomal disorders.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-41

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-42

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-43

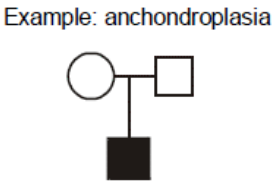
An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-44

An interactive H5P element has been excluded from this

version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-45

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-46

SLO 4 Explain how new mutation accounts for seemingly sporadic occurrences of genetic disorders



Example: achondroplasia. What are the explanations for this pedigree in which two unaffected parents give birth to a child with achondroplasia? Achondroplasia is an autosomal dominant disorder of short-limbed short stature resulting from a single base pair mutation leading to amino acid substitution in the fibroblast growth factor receptor 3 (*FGFR3*) gene. In this family, there are two normally statured and unaffected parents who have a child with achondroplasia. We can exclude incomplete penetrance in the parents because achondroplasia is known to be completely penetrant. We should consider three possible explanations. First, non-paternity (or perhaps less commonly, non-maternity) is always a possibility. A second possible explanation that we will discuss shortly is germline mosaicism. Germline mosaicism has been documented to occur for achondroplasia, but it is a rare phenomenon. At this point we have no reason to invoke it as a consideration for this pregnancy, but if this couple should have two

children with achondroplasia then this would be the most probable explanation. The third possibility, however, given the current pedigree with just one affected child, is the most probable, and that is the child represents a new mutation (also known as "de novo" mutation). In other words, the mutation occurred for the first time in this family's history, in the affected child.

Whenever a child with a highly penetrant autosomal dominant disease is born to two unaffected parents, then the possibility of new mutation should be considered. When a second affected child is born to the same couple, then germline mosaicism becomes a stronger possibility.

Typically, the mutation would occur in a gamete (egg or sperm) that gave rise to that individual. The mutation could also occur in the zygote or very early post-zygotically such that it was present in the majority of the cells of the embryo. If it arose a few cell divisions later in the embryo, then it would be distributed in a mosaic fashion in the body. New mutations represent about a third of all cases of achondroplasia, and, therefore, a family like this is not uncommon. It turns out that for achondroplasia, the overwhelming majority of the new mutations occur on the paternal chromosome and appear to be associated with advanced paternal age.

It is estimated that every child has about 100 new genomic DNA sequence changes, not found in either parent. The majority occur on paternally inherited chromosomes. Most are inconsequential because functional gene sequences constitute only a small portion of the genome and, for those few new mutations that do reside within a gene, some are not pathogenic or behave recessively, in which case a heterozygote would be unaffected.

In general, advanced paternal age is associated with an increased risk for new mutations in a single gene, Mendelian disorder, while, as we shall discuss later, advanced maternal age is associated with a risk for chromosomal disorders resulting from meiotic non-disjunction. The reason for this is because mutations are often linked to DNA replication, occurring during cell division. The formation of sperm is ongoing throughout life. However, the

production of oocytes requires no further mitotic cell division beyond early fetal life.

For serious autosomal dominant illnesses, new mutations account for 100% of cases. Why should it be so high? The answer is because the genetic "fitness" (successful reproduction) is reduced. Individuals affected with serious disease either die before reproductive age or are otherwise too sick to reproduce. To account for a reasonable assumption of a constant disease incidence in the population over time, the alleles lost to lethal events must be replaced by new mutations.

For example, heterozygous mutations in the gene, *ELANE*, encoding the serine protease neutrophil elastase are the most common cause of **severe congenital neutropenia**. Neutropenic individuals have a low number of circulating neutrophils (phagocytic white blood cells) and succumb to opportunistic infections or are, at the least, otherwise too sick to become parents. Consequently, nearly all mutations in *ELANE* responsible for severe congenital neutropenia arise newly, and there is seldom a family history of neutropenia.

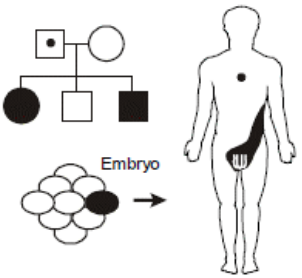
It is reasonable to suppose that the incidence of a genetic disease should remain constant in a population over time. Since lethal alleles will be lost in the population on account of poor genetic fitness, it stands to reason then that the rate at which new mutation generates new disease alleles will equal the rate at which existing alleles are lost from the population. Based on this premise, one-third of all those affected with a severe sex-linked recessive disease will represent new mutations, without a prior family history of that disorder. Let's look at the situation in sex-linked recessive diseases that are lethal, where fitness is zero. All the alleles in an affected male are lost from the population in the next generation and must have been replaced with new mutations if the incidence of the disease remains constant over time. Since one-third of all X chromosomes in the population collectively reside in males, then one-third of all lethal alleles should be new mutations. In fact, this

is observed to be true for Duchenne muscular dystrophy and other severe sex-linked recessive disorders.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-47

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-48

SLO 5 Explain how germline mosaicism causes recurrent autosomal dominantly transmitted disease from unaffected parents



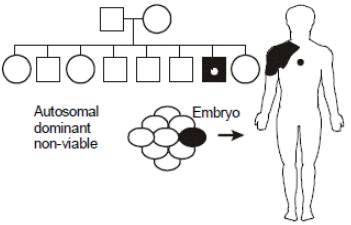
Germline mosaicism occurs when the original mutation arose post-zygotically in the affected parent such that only a fraction of the parent's cells contains the mutation. In particular, the germ cells giving rise to the gametes do contain the mutation, but the mutation is in such minority in the rest of the tissues of the body so as to not produce a recognizable phenotype. It explains the recurrence of a highly penetrant autosomal dominant disease in a family with two unaffected parents. An example occurs with **osteogenesis imperfecta**, an autosomal dominant disorder of collagen producing brittle bones prone to recurrent breakage. A

particularly striking case of germline mosaicism was found when a male without osteogenesis imperfecta fathered two children with the disease by two different females. The collagen mutation could be molecularly identified and was identical in both children but appeared absent from DNA extracted from peripheral white blood cells in the father. However, individual hairs were plucked from the father's head and then subjected to mutational analysis. As it turned out, some of the hair shafts had the mutation and others didn't—demonstrating mosaicism down to the level of the hair follicle.

Germline mosaicism should be considered as a possible explanation for a couple in which neither parent is affected yet has more than one child with a highly penetrant autosomal dominant illness.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-49

SLO 6 Describe how lethal autosomal dominant mutations still give rise to disease



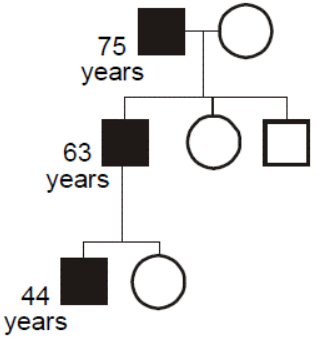
There are mutations in some genes that would result in a non-viable pregnancy. Consequently, they are only seen in a mosaic state in affected individuals, in which the mutation is confined to a small patch of tissue. An example is **McCune-Albright syndrome**. McCune-Albright syndrome is virtually never inherited, as would be expected with

such a disease, and results from activating mutations in the gene GNAS, encoding the alpha subunit of the stimulatory G protein G_s that regulates intracellular cAMP signaling. Clinically, McCune-Albright is characterized by multiple endocrine tumors comprising somatic foci of constitutively activated cells.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-50

SLO 7 Describe genetic anticipation and explain how it alters patterns of inheritance

Anticipation – the clinical observation that an inherited disease displays increasing severity and/or an earlier age of onset with each subsequent generation.



Some heritable disorders demonstrate genetic anticipation, a declining age of onset or worsening severity with each passing generation. Anticipation is a feature of disorders where the gene mutation consists of an unstable short tandem repeat sequence, nearly all of which are developmental or degenerative neurological diseases. Anticipation can also occur with disorders related to preservation of telomere length.

Fragile X syndrome. Fragile X syndrome is among the most common causes of intellectual disability for both males and females. Affected individuals may also have characteristic dysmorphic

features, including long facies (fancy medical term for face) with prominence of the ears, forehead, and jaw. Most postpubertal males with fragile X have enlarged testes ("macro-orchidism"). The disease takes its name from the fact that most affected individuals have a "fragile site" demonstrable at the terminus of the long arm of the X chromosome.

Chromosome fragile sites are gaps or constrictions that are vulnerable to breakage in otherwise condensed chromatin on metaphase chromosomes. They become apparent only during in vitro cell culture under certain conditions. The genomic DNA sequence at fragile sites tends to consist of long runs of di or triplet nucleotide repeats. Fragile sites are considered part of normal chromosome structure, and numerous fragile sites are normally present at defined positions throughout the genome in most people. A few fragile sites, however, are rare and are associated with disease.

The fragile X site, for example, looks like a small bit of the end of the chromosome has broken off and is hanging on thread-like to the rest of the chromosome. While the fragile X site may be vulnerable to breakage, breakage of the chromosome is not what causes fragile X syndrome. Instead, it is caused by aberrant expression of a particular gene due to changes in the length of a triplet repeat sequence that is contained within that gene; the longer repeat also leads to appearance of the fragile site in laboratory-cultured cells.

Like most hereditary illnesses resulting from mutation of a gene on the X chromosome, fragile X is inherited as a sex-linked recessive disorder. Males inherit the disease from a carrier mother. The penetrance for females is rather high, however, and results from skewing of X chromosome inactivation in affected females.

The inheritance of fragile X is not, however, quite as straightforward as it is for other sex-linked recessive diseases. Consistent with anticipation, the likelihood of being affected with fragile X is dependent on the position of the individual in the pedigree. Individuals appearing in a generation subsequent to one

in which somebody is already known to have the disease are at a higher risk for having affected children.

Fragile X results from mutations in a gene, *FMR1*, at the location of the fragile site on the X chromosome. The protein product of the gene binds RNA and appears to function in the nuclear export of RNA. The mutation involves "expansion" of a tandemly repeated CGG DNA sequence. In normal individuals there are about 8 to 50 copies of a CGG trinucleotide repeat in the 5' untranslated region (5'-UTR) containing the promoter regulating *FMR1* expression. Individuals with fragile X syndrome have a mutation comprised of about 200 to 1,000 copies of this repetitive trinucleotide sequence. There is a third category of individual known as "premutation carriers," who are unaffected with fragile X, but who are at risk for having children or later descendants with fragile X, who have a repeat length that's somewhere in-between. Affected individuals with the "full mutation" have greatly reduced transcriptional expression of the fragile X gene from the chromosome containing the expanded repeat. The multiple CpG sequences resulting from repeat expansion provide more targets for DNA methylation. As is usual with cytosine methylation in the promoter region of a gene, transcription is switched off, thereby resulting in loss of expression of that allele. There are extremely rare individuals who have fragile X syndrome resulting, not from expansion of CGG repeats, but rather as a result of an inactivating point mutation in the coding sequence of the gene, confirming that fragile X syndrome is caused by loss of expression of *FMR1*.

Genetic testing, typically using the polymerase chain reaction (PCR) or other methods to measure the length of the CGG repeat tract, is available for the fragile X syndrome and has replaced the much less sensitive and older approach based on cell culture to demonstrate extrusion of the fragile site on the X chromosome.

Although complicated, the inheritance of fragile X syndrome is predictable.

The important concept to appreciate is that once CGG repeat sequences (or other repeat sequences associated with other

diseases) have expanded to a certain length, they become unstable during meiosis and are at risk of expanding to an even greater length in each subsequent generation. There is some threshold number of repeats that, once crossed, turns a premutation into a full mutation capable of causing the disease. For most disorders caused by expansion of short repeat sequences, meiotic expansion of the repeat tends to occur only when inherited from the mother (as is the case for fragile X) or from the father, depending upon the disorder.

Expansion of the fragile X CGG triplet repeat from a normal length to the premutation length almost never happens. The premutation appears to have been inherited from just a few common ancestral founders in the general population.

On the other hand, expansion of a premutation-length repeat to a full mutation is rather common but can only occur through a female meiosis. For unexplained reasons, for each particular disease caused by a repeat expansion, the expansion is characteristically more likely to occur when inherited from a parent of one sex or the other.

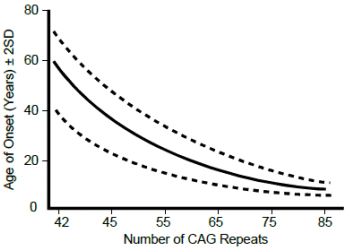
In general, for fragile X syndrome and other disorders caused by unstable short repeats, the risk that a premutation will expand to a full mutation is dependent upon the length of the premutation repeat in the parent from whom it is inherited. The greater the length of the premutation, then the more likely it is to expand during meiosis. In general, however, for fragile X syndrome, we can assume there to be about an 80% probability that a premutation repeat will expand to a full mutation length during female meiosis. However, when the premutation repeat length reaches 90 or more triplets, then expansion to a full mutation happens with near certainty during female meiosis.

The penetrance of fragile X in males with the full mutation is 100%. In females with the full mutation, the penetrance is about 50% (as a result of skewed Lyonization).

Premutation carriers exhibit different symptoms than those with a full mutation. Female fragile X premutation carriers tend to have

premature ovarian failure (i.e., early onset of menopause), and male fragile X premutation carriers tend to develop tremors and ataxia (poorly coordinated movement) with advancing age, a disorder known as **fragile X–associated tremor/ataxia syndrome (FXTAS)**. Paradoxically, whereas the full CGG repeat mutation tends to extinguish expression of the *FMR1* gene, the premutation tends to elevate its expression.

Huntington disease. Huntington disease is an autosomal dominant, adult-onset, severe neurodegenerative disorder. It is characterized in the initial stages by mild cognitive symptoms with progressive choreiform (involuntary dance-like) movement disorder. At the neuroanatomic level, there is degeneration of the caudate and putamen in the basal ganglia of the brain. The responsible gene, *HTT*, is expressed ubiquitously throughout the brain. Molecular pathogenesis of the disorder remains a subject of investigation, but the protein encoded by *HTT* plays a role in vesicle trafficking and endocytosis.

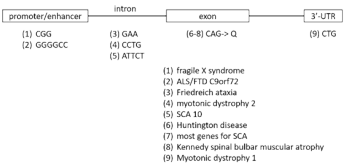


The mutation in every individual with Huntington disease is a variably sized tandem expansion of the DNA triplet repeat sequence CAG, contained in the protein-coding sequence of the gene, and encoding the amino acid glutamine. Individuals without Huntington disease normally have fewer than 35 CAG repeats, although the repeat length is polymorphic among the normal population. Individuals with 40 or more CAG repeats usually develop Huntington disease. The protein encoded by *HTT* from individuals with a CAG repeat tract expansion will then have an elongated polyglutamine tract within the polypeptide.

The age of onset of the disease can be largely explained by the number of repeats an individual possesses—the greater the number of repeats, then the earlier the age of onset. Just as with fragile X

syndrome, once the number of repeats passes a critical threshold, then the repeat tract becomes unstable and is at risk for further increases in length with every subsequent meiosis. Unlike the case for fragile X syndrome, meiotic repeat expansion, especially large expansions, occurs preferentially during *paternal* meiosis.

For most diseases where the mutation is caused by an unstable short DNA repeat there is a strong inverse correlation between the length of the repeat and the age of onset or severity of the disease.



Presymptomatic diagnostic testing can determine if one has inherited a disease-causing mutation before symptoms are apparent. The test for Huntington disease is simple and consists of a PCR assay to determine the length of the CAG repeat tract. Ethical issues relating to the use of the test are not so simple, however. It is recommended that individuals at risk for inheriting Huntington disease who undergo molecular genetic testing for presymptomatic diagnosis receive appropriate pre-test genetic counseling and post-test psychological support. Similar recommendations apply to presymptomatic diagnostic testing for any disorder with potentially profound impacts on health and longevity.

Other diseases caused by unstable short tandem repeats. There are several human genetic disorders caused by short tandem DNA repeat expansion. While they all play by a unique set of complicated rules, their inheritance patterns are for the most part predictable, there is a set of common clinical features, and genetic testing can precisely define risk.

Characteristics of diseases caused by short tandem repeat expansion

- The short tandem DNA repeats, depending upon the disease gene in which they reside, can be located in upstream regulatory regions, within an intron, within untranslated

regions of mRNA or in the coding sequence where they are translated (in particular for CAG repeats, translated into runs of the amino acid glutamine).

- Disease-associated alleles can exist in "pre-mutation" and "full-mutation" states, with the two states being differentiated by the length of the expanded tract of repeats.
- The repeat length at most disease loci is variable in the normal population. Expansion from a normal length repeat to a premutation repeat happens only infrequently, but once a premutation has occurred, there is a high likelihood of further meiotic expansion to a full mutation state.
- In general, expansion from premutation to full mutation shows a parent-specific effect (either maternal or paternal) for each particular disorder. The expansion occurs during meiosis.
- The length of the repeat tract inversely correlates with age of onset of the disease and accounts for anticipation, as revealed by an earlier age of onset or worsening severity with consecutive generations.
- At the molecular level, the repeats can be a site of epigenetic modification, such as promoter hypermethylation, and thereby lead to reduced or increased production of mRNA transcript or, if contained in a protein-coding sequence, altered proteins.
- Depending upon the particular gene and its associated disorder, pathogenesis may be caused by too much or too little of the normal gene product, a toxic mRNA transcript that can aggregate or sequester RNA-binding proteins, or a toxic protein capable of aggregating.

Anticipation with disorders of telomeres. Finally, another completely different mechanism for genetic anticipation occurs with disorders involving replication of telomeres, the structures at the ends of chromosomes. Due to the fact that DNA synthesis precedes only in the 5' to 3' direction, the ends of chromosomes represent a particular problem for DNA replication, and a special complex of proteins along with an RNA-priming template comprising

telomerase and associated factors is required for maintenance of telomere length in stem cells, including germ cells. Mutations in any number of genes, including *TERT*, which encodes the protein component of telomerase, and *TERC*, which produces a non-protein-coding RNA, can cause the disorder **dyskeratosis congenita**. For most somatic cells, including somatic stem cells, there is progressive shortening of telomeres with each cell division. When telomeres become too short, cells can no longer divide. Telomere length therefore dictates the replicative lifespan of cells. In dyskeratosis congenita, telomere length in both somatic cells, as well as germ cells giving rise to sperm or eggs, is not maintained. Consequently, telomeres shorten with each passing generation in these disorders, resulting in short telomeres in all cells of a person inheriting the disorder. Cells with reduced telomere length cannot undergo as many cell divisions. The most rapidly dividing populations of cells are the most severely affected, including hair, skin, nails, bone marrow, and epithelial cells lining the lungs, leading to premature aging, aplastic anemia and bone marrow failure, and fibrotic lung disease, respectively. Moreover, the telomere length continues to shorten between generations, thereby leading to genetic anticipation with intergenerational worsening of symptoms occurring at progressively earlier ages of onset.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-51

An interactive H5P element has been excluded from this

version of the text. You can view it online here: https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-52

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://uw.pressbooks.pub/fmrgenetics/?p=263#h5p-53

Feedback:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://uw.pressbooks.pub/fmrgenetics/?p=263

5. Complex Traits, Meiosis, Genome Wide Association Studies (GWAS)

Session Learning Objectives and a quick synopsis:
Nearly all diseases have a genetic component, though most are not highly penetrant, single gene, Mendelian disorders. Instead, common diseases are the result of many common genetic variants with small effect size that are spread throughout the genome. Some of these variants increase risk while others lower risk. The sum of these inherited factors, in combination with lifestyle/environment (which can also be shared among families and populations) influence susceptibility to most diseases.

SLO 1 Define what is meant by "complex genetics"

Most common diseases are inherited through "complex genetics," and are not caused by just one highly penetrant gene mutation. Instead, they result from a combination of common population variants in multiple genes, acting in concert with environmental factors.

SLO 2 Describe normal genetic variation in the human population

When comparing the genomes of any two people, most of the differences are due to single nucleotide polymorphisms (SNPs). On average, there is a SNP differentiating any two unrelated people about every thousand base pairs or so. Any two unrelated people differ by about 5-10 million SNPs. Most of the SNPs and other types of sequence variants in the genome that differentiate people are of ancient human origin and occur commonly within and across populations.

SLO 3 Explain how natural selection influences disease gene frequencies

Human populations are subject to evolutionary pressures. Deleterious mutations tend to not be maintained due to their harmful effects, whereas beneficial mutations are selected for. However, the same mutation can be beneficial or harmful, depending upon gene dosage (heterozygous vs. homozygous) and environmental context. A classic example is sickle cell disease, an autosomal recessive disease that nevertheless confers malarial resistance to asymptomatic carriers and is therefore maintained in the population.

SLO 4 Explain how meiosis increases genetic diversity

Recombination and independent assortment during meiosis produces new gene combinations in offspring, thereby adding to population diversity with each new generation.

SLO 5 Explain the concept of linkage disequilibrium and haplotypes

Genetic variants close by one another on the same chromosome tend to be inherited together, because the probability of a recombination event occurring between them during meiosis is low—a phenomenon known as "linkage disequilibrium." Resulting genetic variants commonly inherited across a chromosomal region shared between people within a family or at the population level define a "haplotype."

SLO 6 Describe the origins of the haplotype block structure of the genome

Over time, meiotic recombination whittles down the length of each haplotype and reduces ancestral contributions to small chromosomal "blocks" of a few thousand contiguous base pairs. Because these chromosomal regions cannot further recombine, they reach a lower size limit in which all the genetic variants contained within the region are inherited together. Our genomes therefore represent a mosaic derived from ancient ancestors, in which the contributions from any ancestor

from more than just a few generations ago is indistinguishable.

SLO 7 Explain how "genome-wide association analysis" (GWAS) is used to identify common genetic variants in the population that contribute to risk for common disease and how these variants differ from those responsible for Mendelian disorders

GWAS is an experimental tool used to identify common variants with small effect sizes, collectively contributing to disease risk. It involves a simple association test to see if particular SNPs representative of certain ancestral haplotype blocks are found more commonly in people with a certain trait compared to controls.

Main Text:

SLO 1 Define what is meant by "complex genetics"

In contrast to Mendelian disorders such as sickle cell anemia, Marfan syndrome, or cystic fibrosis, in most cases common diseases such as hypertension, atherosclerotic vascular disease, diabetes mellitus, and mental illness cannot be explained by mutations in a single gene. Instead, they are usually associated with the effects of multiple genes interacting in combination with lifestyle and environmental factors. (However, for almost any common disease, such as hypercholesterolemia or cancer, there are the somewhat rare families who are an exception and inherit the disorder due to a causative mutation in a single gene.) This is what is meant by "complex genetics" or "multifactorial" or "polygenic" disorders.

Heritability. For a complex trait, its "heritability," which is the fraction of phenotype variability attributable to underlying genetic variation, can be measured through several different epidemiological approaches. One such approach is to study twins.

Twin studies. About 1/100 births are twins. Dizygotic (DZ), or fraternal twins, result from the fertilization of two different eggs, whereas monozygotic (MZ), or identical twins, are the product of a division of the developing embryo. Dizygotic twins share 50% of their DNA (just as do ordinary siblings) whereas monozygotic twins share 100%. Geneticists use the concept of concordance in

twin studies, where if both twins share the same trait or both are unaffected they are said to be concordant. Measuring the difference in concordance frequency for a trait between monozygotic and dizygotic twins can separate genetic from environmental factors (under the assumption that the environments are the same for both monozygotic and dizygotic twins).

Heritability of some common traits. For example, the concordance frequency for bipolar affective disorder for monozygotic twins is 79%. Hence, in a pair of monozygotic twins, if one twin is affected then 79% of the time the other twin will be, too. In contrast, the concordance frequency for bipolar affective disorder is only 24% for dizygotic twins. For a highly penetrant autosomal dominant disorder unaffected by environment, concordance for monozygotic twins would be 100%, and concordance for dizygotic twins would be 50%. Thus, it can be concluded that, at the population level, there is a strong genetic contribution to the risk for bipolar affective disorder, but it stops short of what would be expected for a single-gene disorder, and there is also evidence for contributions from environmental factors. In contrast, the concordance frequency for measles, an infectious disease caused by a single-stranded RNA virus, for which there could nevertheless be some degree of genetic risk for vulnerability, is 95% in monozygotic twins and 87% in dizygotic twin pairs. This supports the fact that it is largely the shared environment between twins, rather than the extra 50% identity in genetics, that is mainly responsible for measles infections.

Threshold concept for complex inheritance. The threshold model for complex genetics posits that there is a certain cumulative "red line" of environmental and genetic risk factors that must be crossed to develop a particular disease.

For a complex disease, a threshold level of environmental and genetic risk factors must be crossed to develop the disease. Not all genes contribute equally to risk in all people for a complex disease.

Sex-dependent thresholds. The threshold for developing the disease may differ between males and females. For example,

coronary artery disease is a complex disease resulting in rare circumstances from the influence of a single gene (such as mutations in the LDL receptor in familial hypercholesterolemia), but more often from a polygenic contribution (i.e., multiple genes influencing cholesterol, triglycerides, homocysteine, etc.) as well as environmental factors (such as diet, tobacco smoking, and exercise). For any given level of genetic and environmental risk, however, males have more coronary artery disease, at least in part because of the protective effects of estrogen in females.

There is an interesting consequence of the fact that the risk threshold differs between the sexes: the recurrence risk in children will be greater for the children of the less susceptible sex.

For example, if one's mother has coronary artery disease, that person is at greater risk for developing coronary artery disease (regardless of their sex), than if their father had coronary artery disease. Since females have a higher threshold for developing coronary artery disease, they must therefore have more risk factors (genetic and environmental), overall, and are consequently likely to specifically have more genetic risk factors that one is capable of inheriting. A person is therefore more likely to inherit more genetic risk factors for coronary artery disease from one's mother than they would if it were their father who were affected with coronary artery disease.

What is needed to understand the basis of complex genetic disorders. Before we can understand the molecular genetic basis of complex traits, we need to dive a little more deeply into some fundamental aspects of human genetics, as it relates to the sources of human genetic variation and how genetic variation arising in any one individual may be represented across a population in subsequent generations over time.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-54

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-55

SLO 2 Describe normal genetic variation in the human population

The sequence of our genome contributes to what makes us unique—and what we have in common with other human beings.

The human "diploid" genome is about 6.6 Gb (billion base pairs, "gigabases"). Recall that there are 23 pairs of human chromosomes, one member of each pair being inherited from each parent. Consequently, the "haploid" genome, defined as the genome in a gamete (egg or sperm), is 3.3 GB. For most genes there is one maternal and one paternal copy except for the small circular mitochondrial genome (16,569 base pairs) and a few genes unique to the X chromosome and far fewer genes unique to the Y chromosome.

The most common type of genetic difference between any two people involves single base-pair differences, otherwise known as a "single nucleotide polymorphism" (or, more commonly, "SNP").

Mutation – A change in the DNA sequence usually conferring a deleterious effect (at least in a medical context). It may be present in the germline causing an inherited disorder or it may be acquired post-zygotically in somatic tissues, for example, in cancer.

Polymorphism – A DNA sequence difference usually of no pathological consequence. The difference between a mutation vs.

polymorphism gets into rhetorically muddy waters when it comes to sequence differences contributing to traits, such as blue eyes vs. brown, with little meaningful medical significance. In these circumstances, the use of the term "variant" may be preferred.

Variant of undetermined (or unknown) significance (VUS) – This is a term nearly always used in the context of genetic testing. Genetic testing often turns up a VUS. Such a DNA sequence difference may be unique to the tested individual, or at least not commonly found in other people, and its properties may not be clearly pathogenic and/or not previously shown to segregate with disease. Finding a VUS on genetic testing is another way of saying, "we found something but don't have sufficient information to determine its clinical implications, if any."

Any two unrelated people differ by about 5-10 million SNPs. On average, there is a SNP differentiating any two unrelated people about every thousand base pairs or so.

The vast majority of SNPs and other types of sequence variants in the genome that differentiate people are of ancient human origin and occur commonly within and across populations. Nevertheless, every individual still possesses about 25,000-50,000 rare single nucleotide variants, in addition to about ten times as many small insertion/deletion variants. These "private" variants may be unique to any particular person and are shared only with their family members. When comparing people with ancestral origins from different parts of the world, there is actually greater genetic diversity within a local population than when comparing across geographically distant populations. Because most human history transpired before the migration out of Africa, most, but not all, genetic variants arose before that time and are distributed across people of all ancestries.

Genome – The complete sequence of DNA present in a cell or organism. The haploid human genome is ~3.3 GB (billion base pairs).

Exome – The protein-coding portion of the genome, which constitutes ~1% of the human genome or ~30 Mb (million base pairs, "megabases"), spread out across ~180,000 exons from ~20,000 total

genes. Since most Mendelian disorders (but not complex diseases) are the result of mutations within the protein-coding portion of genes, clinical genetic testing for challenging diagnostic questions often relies on exome sequencing, as opposed to genome sequencing. Exome sequencing has several advantages: It is less costly. It returns less inconsequential data, including polymorphisms, requiring analysis. Based on the "next-generation" massively parallel technologies in use today, any given nucleotide is read multiple times, assuring greater accuracy than can be achieved with whole genome sequencing, where breadth of coverage occurs at the expense of depth of coverage. The high depth of coverage can be exploited to help identify potential large-scale deletions or insertions ("copy number changes").

Sources of genetic variation. Although exposure to environmental mutagens, such as ionizing radiation, can alter DNA sequences, most changes occur spontaneously. Since most cells are "somatic" and do not contribute to gamete formation, these DNA sequence changes are not passed down from one generation to the next—although they may contribute to cancer. DNA changes occurring in germ cell precursors giving rise to eggs and sperm do, however, result in heritable changes to the "germline."

The most common type of SNP in both somatic cells and in the germline is a change of a C:G base pair to a T:A base pair and reflects the fact that methylcytidine, an epigenetic mark occurring in CG sequences (usually referred to historically as "CpG" with the "p" representing the phosphodiester bond), spontaneously hydrolytically deaminates to thymine. Another common type of spontaneous mutation reflects oxidative damage to DNA. DNA damage occurs tens-to-hundreds of thousands of times per cell per day. Most damage is repaired by specific DNA damage response pathways, but the process is not completely perfect, resulting in a mutation. Many different mutational signatures can be linked to particular environmental carcinogens (e.g., tobacco smoke) or cell endogenous processes, such as noted for oxidative damage, making

it possible in some cases to link particular environmental agents with the types of cancer they contribute to.

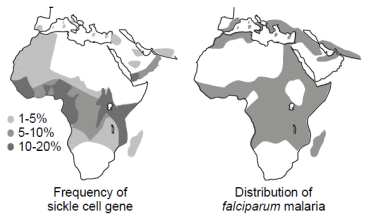
Once a germline DNA variant arises in any given individual, the extent to which it is propagated further among that person's descendants depends upon both random events as well as the population's size and natural selection if the variant has functional consequences. The random sampling of gametes during sexual reproduction leads to "genetic drift," which is a stochastically introduced fluctuation in the population frequency of a genetic variant. When a population size is reduced (goes through a "bottleneck"), particular genetic variants may be subsequently over-represented through "founder effects," merely because a modern population can be traced back to a small number of founding individuals. Finally, some DNA sequence changes alter gene activity and produce new phenotypes that may be advantageous or deleterious and are therefore subject to natural selection.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-56

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-57

SLO 3 Explain how natural selection influences disease gene frequencies

Genetic "fitness" refers to the relative reproductive success imparted by a particular genetic variant. For example, a mutation producing a severe phenotype, such as an autosomal dominant Mendelian disease resulting in fatality during childhood, has zero reproductive fitness. Fitness also reflects external factors, such as the environment.



Heterozygote advantage. Selective pressures operate differently for autosomal recessive diseases than they do for dominant diseases. A low fitness for a recessive disease has much less effect on removing a disease allele from the population than it would for a dominant disease. This is because most of the alleles will exist within a reservoir of unaffected heterozygotes. On the other hand, selective pressures can influence the heterozygote population, which is large in comparison to the homozygous affected population. If the heterozygotes have a selective advantage compared to those who are homozygous for normal alleles, then the frequency of the disease allele will increase. Since heterozygotes are far more common than affected homozygotes, for any given disease allele, the benefit to the relatively large heterozygote population may be more than enough to make up for any deleterious effect upon fitness in the comparatively smaller homozygous population. This is the reason why sickle cell anemia and other hemoglobinopathies as well as thalassemia are so common in regions of the world where malaria is endemic. **Sickle cell anemia** is an autosomal recessive disease caused by an amino acid missense substitution in the beta chain of hemoglobin (designated Hb S). Heterozygous carriers of the sickle cell trait are relatively resistant to *Plasmodium falciparum* malarial infection and its complications. A similar heterozygote advantage is thought to explain the

maintenance of recessive alleles for cystic fibrosis in the population. There is evidence that individuals heterozygous for cystic fibrosis mutations are relatively resistant to *cholera* and *Salmonella* enteritis.

From the Hardy-Weinberg law, the heterozygote advantage results because $2pq >> q^2$, and explains the high carrier frequency for certain autosomal recessive disease alleles that confer disease resistance in a heterozygote state.

Carrier selection also occurs for sex-linked recessive disease where unaffected carriers are similarly more abundant than affected individuals, as in the case of **glucose-6-phosphate dehydrogenase (G6PD) deficiency**, which follows a geographic distribution similar to that of sickle cell disease and appears to confer similar resistance to malarial infection and its complications. G6PD participates in the hexose monophosphate shunt of glycolysis. In fact, since both Lyonization (X inactivation) and carrier selection apply in the case of G6PD deficiency, one can make a striking observation in microscopic examination of the peripheral blood smear of heterozygous female carriers infected with malaria. If the smear is stained to histochemically detect G6PD, then Lyonization will be discernable at the level of single cells—some erythrocytes will express G6PD and others won't. Interestingly, the malarial parasites will be seen preferentially in the red blood cells that express G6PD and be relatively scarce in red blood cells that lack G6PD activity. A deficiency of G6PD activity results in hemolytic anemia upon exposure to oxidants, including sulfa antibiotics, quinine-based antimalarials, and fava beans. Ironically, problems with G6PD first became widely recognized when GI's of African ancestry serving in Korea in the 1950s suffered disproportionately compared to individuals of other ancestries from side-effects of therapeutic treatment with prophylactic antimalarial drugs.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-58

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-59

SLO 4 Explain how meiosis increases genetic diversity

In order to understand complex genetics, one must also understand the concept of genetic linkage. And to understand linkage, one must know about meiosis. Of course, meiosis is central to other concepts in genetics, such as the origins of constitutional chromosomal imbalances, which we will cover later.

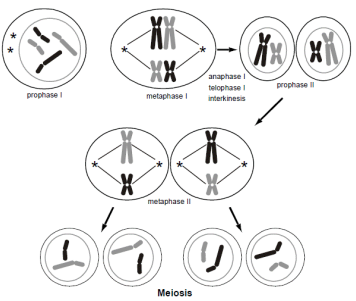
Meiosis – is the cell division process in which "haploid" gametes are formed from "diploid" germ cells. Meiosis addresses the problem of preventing genome size from doubling at each generation in a sexual organism. Meiosis increases genetic diversity by randomly re-sorting chromosomes from the parents, the male making sperm and the female making eggs (i.e., the grandparents of the future conception). It is also the time when "recombination" between each parental chromosome homolog takes place, further increasing genetic diversity, by recombining the chromosomes themselves such that they become composites of each set of grandparents.

Mistakes in meiosis are responsible for major chromosomal abnormalities.

Meiosis is fundamental to sex. Sex increases opportunity for genetic diversity, and therefore makes it more likely that at least some individuals within a population will be able to survive in a changing environment.

Meiosis is divided into two sequential stages, meiosis I and meiosis II. Meiosis I and meiosis II are then each subdivided into several substages that share names with each other and with the substages of somatic cell division ("mitosis").

Meiosis accomplishes two things. The first is to prevent a doubling in the quantity of chromosomes from one generation to the next.



Normally, for every pair of chromosomes in every individual, one chromosome came from the mother's egg and the other from the father's sperm. During the longest part of the cell cycle (interphase), the chromosomes are not microscopically visible in a cell. Only during cell division and gametogenesis do the chromosomes condense enough to become microscopically visible (predominantly during the metaphases). The condensation facilitates their mitotic or meiotic segregation. During the formation of an egg or a sperm, one parental chromosome from each pair is randomly sorted for inclusion in the gamete. The way this is accomplished is that the 23 pairs of chromosomes form a physical pair (called a "synapse") and then line up along the midline of the cell. In this figure, for simplicity, we are only showing two different chromosomes, distinguishable by their size. We are shading each chromosome, black vs. gray, to distinguish maternal from paternal contributions. Each pair of chromosomes lines up in a random orientation (referred to as "independent assortment"), so that in one pair the

maternally inherited chromosome might be on the left and the paternally inherited chromosome on the right, while in another pair the opposite may be true. Then, the cell divides into two, so that only one member of the pair sorts into each of the two daughter cells. Consequently, there are a very large number of ways that the parental chromosome homologs can be distributed into each gamete, even in the absence of chromosomal recombination. This is known as meiosis I, and at this point there are now just 23 chromosomes per cell.

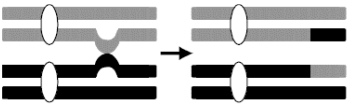
The chromosomes had completed a round of DNA replication, similar to mitosis, before entering into meiosis I. Each chromosome in the synaptic complex therefore has two "sister chromatid" arms that are side-by-side. Since there is a pair of "homologous" chromosomes in each synapse, each with two sister chromatids, there are a total of four chromatids in the synapse. The result is that the two daughter cells produced at the conclusion of meiosis I have just 23 chromosomes (a haploid quantity of chromosomes) but each chromosome is present in a duplicated form. Each daughter cell must then undergo a second event (meiosis II, which is similar to meiosis I except that there are only 23 instead of 46 chromosomes). During meiosis II, the sister chromatids are pulled apart, and each of the 23, now non-duplicated chromosomes with just a single chromatid per each arm, goes into two more daughter cells. The result is that there are now four daughter cells, each with a haploid quantity of non-duplicated chromosomes.

In the formation of a sperm, one diploid cell begins meiosis and four haploid spermatozoa result from it. A significant difference during oogenesis in the female is that meiosis produces only one haploid oocyte. One of the daughter cells resulting from meiosis I is discarded (and consequently does not enter meiosis II), and one of the daughter cells resulting from meiosis II is also discarded. These discarded cells are known as "polar bodies." Rarely, a sperm aberrantly fertilizes a polar body, resulting in a hydatidiform mole which can evolve into choriocarcinoma, the only type of human

cancer where the tumor genome is not that of the individual in whom it arises, but in this case is derived from the partner's sperm.

Another significant way that meiosis differs between the sexes is that for males, meiosis is a continuous activity that begins at puberty and proceeds until death. For females, there are several thousand primitive oocytes arising in the developing ovary of a female embryo and that initiate meiosis I well before birth, but then arrest during embryonic development. Meiosis does not resume until after puberty, and then in only one egg at a time (with the egg that is ovulated during a particular menstrual cycle, at which point meiosis I is completed). Meiosis then again arrests, and meiosis II is not completed until just after fertilization, at which point the second polar body is ejected.

Recombination. The second accomplishment of meiosis is to produce genetically variable gametes through independent assortment (as noted above) and recombination. During genetic recombination, an individual's two parental chromosome homologs actually physically break and recombine to produce a recombinant chromosome, containing combinations of grandparental alleles that were not previously linked in either parent. Recombination occurs in only two of the four chromatids present in the synapse. So, from every meiotic event there are always two potential recombinant chromosomes (that represent reciprocal exchanges) and two non-recombinant chromosomes.



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-60

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-61

SLO 5 Explain the concept of linkage disequilibrium and haplotypes

Genes are not always inherited independently from one another. The circumstance in which genes *are* independently inherited from one another is when they reside on different chromosomes. In this case, they are not physically "linked" to one another, because each chromosome is composed of a different linear molecule of DNA, and each pair of chromosome sorts independently from other pairs of chromosomes during meiosis. However, genes on the same chromosome are physically linked to one another. If there were no recombination, then genes residing on the same chromosome would always co-segregate with each other through meiosis.

Nevertheless, as explained above, recombination does occur between two parental chromosome homologs, for any particular chromosome, during meiosis. The consequence is that alleles for two physically linked genes on the same chromosome inherited from the same parent can be dissociated from one another in meiosis.

Recombination more or less occurs randomly across a chromosome. In general, at least one recombination event (a so-called "crossing over" event) occurs per chromosome arm per meiosis. Typically, there are several crossing overs per chromosome arm, involving combinations of any two of the four sister chromatids participating in the meiotic synapse.

In general, the further apart two genes on the same chromosome

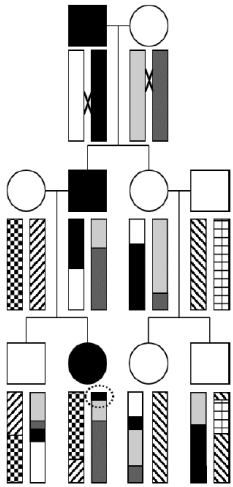
are from one another, then the more likely it is that there will be a recombination event occurring between them; consequently, genes distant from one another along the chromosome arm less often co-segregate with one another. Conversely, if two genes are close together, then it is much less likely that a crossing over event will occur between them. In fact, when two genes are adjacent to one another there is seldom recombination between them, and the two genes consistently co-segregate with one another. Nearby genes that tend to co-segregate with one another are referred to as being in "linkage disequilibrium."

It is not just genes that demonstrate linkage disequilibrium. Any discernible genetic change, including a SNP (the different versions of which are also still referred to as "alleles") can exhibit this property. Closely adjacent clusters of genes or other genomic markers or landmarks, including SNPs, tending to co-segregate with one another on the same parental chromosome, are referred to as a "haplotype."

Linkage disequilibrium – alleles for genes or genomic landmarks, such as a SNP, close to one another (in physical proximity on the same chromosome) tend to co-segregate through families with one another, since it is unlikely that a recombination event will occur between them.

Haplotype – refers to a group of alleles from closely linked loci that are usually inherited together as a unit and that exhibit linkage disequilibrium.

Ultimately, recombination is a biochemical process involving DNA and proteins that can bind to DNA and physically recombine double-stranded molecules. Consequently, over time, recombination will whittle down the length of a haplotype but it most likely reaches a minimal size limit, in length defined by base pairs of DNA, that is both a consequence of the number of generations that have elapsed since the initial haplotype was defined, as well as the physical limits of recombination.



To illustrate the concept of how haplotypes, which are blocks of linkage disequilibrium, are created by recombination during meiosis, let's examine this three-generation pedigree transmitting a highly penetrant autosomal dominant trait, where the affected individuals are filled-in with black.

The rectangles represent a chromosome arm. Note that each "ancestral" haplotype (which in this case corresponds to the entire chromosome) is uniformly shaded or patterned. Each of the four people who "marries into" the family introduces a new, fresh, ancestral haplotype, denoted with a unique pattern.

The "×" symbol indicates a site of recombination (or crossing over) in the first generation. Notice that the black and white ancestral haplotypes physically recombine in the affected male in the first generation and produce a new recombinant haplotype where the ancestral segment containing the causative mutation is shortened and inherited by the affected male child. His female partner also has a recombination event, and the chromosome containing a mixture of ancestral haplotypes is similarly inherited by their affected child. As the second generation choose mates of their own, new ancestral

haplotypes are introduced, here represented by unbroken patterns. Recombination continues, and each ancestral haplotype is further reduced in size with each passing generation.

Can you determine the position on the ancestral chromosome where the responsible gene must reside? Hint: it's circled! As you can see, that is the shortest haplotype inherited by all three affected individuals, and none of the unaffected individuals possess this haplotype at that region of the chromosome.

Linkage disequilibrium can be used to date the emergence of an allele in the population. For example, consider the CCR5-Δ32 allele, which affords resistance to human immunodeficiency virus (HIV) infection. CCR5 is a chemokine receptor protein present on the surface of macrophages. The Δ32 mutation deletes 32 nucleotides and creates a frameshift that prematurely leads to termination of the altered reading frame. Homozygotes for the Δ32 allele of CCR5 are advantageously resistant to infection by HIV. Heterozygotes appear to have intermediate levels of resistance to HIV. Chromosomes that have the Δ32 mutation on them all have the exact same alleles for various markers that flank the CCR5 gene extending on a small stretch on either side. From this haplotype sharing it can be deduced that everyone possessing this allele must be descended from a common ancestor. From theoretical studies it can be inferred that the length of the haplotype held in common between these modern-day individuals must have taken about 30 or so generations to reach this shortened length through progressive meiotic recombination events. It can thus be deduced that this mutation first arose about 700 years ago (assuming something like 23 years/generation) and is hypothesized to have coincided with the Plague in Europe. It would appear that this mutation was a relatively random event existing in one or a few related individuals in the Middle Ages and presumably became more common in the population because it conferred some degree of resistance to the Plague and quite incidentally, also HIV, which probably had not yet infected humans at that time. *Yersinia pestis*, the Plague bacillus, utilizes the CCR5 molecule to infect macrophages, as does HIV,

explaining why mutant CCR5 would confer resistance to both pathogens.

Genetic Linkage Analysis. The above diagram showing co-segregation of a progressively shortening haplotype with a disease phenotype within a family actually depicts the process of "genetic linkage analysis," which was the historical approach used to identify genes responsible for most Mendelian disorders. However, this was during a time before the Human Genome Project determined a reference genome sequence and before the current powerful "next-generation" of massively parallel DNA sequence technologies responsible for economical exome and genome sequencing came into being.

Today there are more powerful and less expensive technologies, namely, ready clinical access to whole exome and whole genome sequencing.

While most single-gene Mendelian disorders have already been identified, there are still a few that continue to be discovered because they are either rare or the spectrum of clinical findings associated with a known syndrome had not been fully appreciated previously. The ability to perform exome or genome sequencing on patients who have defied clinical diagnosis has brought fresh hope to patients and families with heretofore seemingly unexplained diseases.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-62

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-63

SLO 6 Describe the origins of the haplotype block structure of the genome

Let's return to the above figure demonstrating how a haplotype progressively shortens as a result of meiotic recombination with each passing generation. Instead of just observing this phenomenon for three generations, as in the figure, imagine what would happen over a great number of generations, say as many generations as have elapsed from the time when humans first migrated out of Africa and began populating the rest of the world. The haplotype continues to shorten, but it does not become infinitesimally small and instead approaches a minimal limit, probably ultimately restricted in size by the biochemical limits imposed by the recombination process itself.

Therefore, each of our chromosomes is a mosaic of ancient, ancestral haplotypes, averaging from tens of thousands to hundreds of thousands of base pairs in length. These ancient haplotype blocks are distinguishable by the unique variants, mostly SNPs, but also insertions and deletions and other types of polymorphisms, that arose long ago and that define the ancestral haplotype.

And since populations tend to grow exponentially, modern populations are descended from relatively few founders. Therefore, the set of chromosomes amongst all people in contemporary populations achieve their genetic diversity by mixing and matching "blocks" representing a relatively small number of ancient haplotypes. And within a haplotype block, adjacent polymorphisms are all in linkage disequilibrium. That is, they continue to segregate with one another in a pedigree. In many cases, determining the

genotype of just one SNP at a particular location in the human genome is sufficient to uniquely identify the particular ancestral block and infer adjacent sequence along the chromosome for a few hundred thousand base pairs in either direction. Of course, new mutations and polymorphisms have sprung up over time within an ancient haplotype block, but, in general, these are relatively few and far between; when comparing any particular ancient haplotype block shared by any two people possessing that particular block, the similarities in DNA sequence will be far greater than the differences.

The international scientific collaboration that has mapped the haplotype structure of the human genome across populations has produced what has come to be known as the "HapMap." The collection of SNPs that can define a particular haplotype block are sometimes referred to as "tag SNPs." To a very good approximation, determining the genotypes for 300,000 to a million of these tag SNPs (one or a few per each block) can be used to stitch together the sequence of haplotype blocks and infer (or in technical jargon, "impute") the entire genome sequence of any given individual, with the significant exception of polymorphisms and mutations that have arisen in any particular family in more recent generations.

Which particular haplotype blocks persist in the population has to do with chance (genetic drift), population bottlenecks, and any potential beneficial or deleterious influences that genetic variants contained within the block might confer (natural selection).

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-64

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-65

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-66

SLO 7 Explain how "genome-wide association analysis" (GWAS) is used to identify common genetic variants in the population that contribute to risk for common disease and how these variants differ from those responsible for Mendelian disorders

Let's recap what we've learned so far about complex genetics:

- There is a significant underlying genetic component to common disease, distributed at multiple loci across the genome, each with weak effect.
- Most of the genetic variation in the population is ancient and therefore common, in that any particular variant is shared by a large number of people.
- There is a block-like structure to the human genome such that an individual's genome represents a composite of a large but ultimately finite number of blocks measuring from a few tens or hundreds of thousands of nucleotides in length.
- Any given block is distinguished by a collection of SNPs unique

to that block, so that if we determine the genotype of only one or maybe a few of these so-called tag SNPs we can, more or less, be confident of the DNA sequence of the entire haplotype block.

How can we put this all together to define the molecular genetic basis of common diseases? The answer is a "genome-wide association analysis" (GWAS).

Genome-wide association analysis (GWAS) – In a nutshell, a GWAS operates under the hypothesis that common genetic variants contribute to risk for common disease. GWAS consists of the following steps. A large number of cases and controls are assembled, with hundreds of thousands to millions of subjects typically now participating in any single GWAS. For each subject, cases and controls included, genotypes are determined for about 300,000 to a million or so tag SNPs dispersed across the genome. (Genotyping SNPs can be performed rapidly and economically using a variety of high throughput technologies.) Then a simple association analysis, using a chi-square statistical test, is employed to determine which SNPs are found more frequently or less frequently in cases compared to controls. Those SNPs occurring more frequently in cases are deduced to confer risk for disease, whereas those that occur more frequently in controls are considered protective.

Note that because each SNP only tags a haplotype block, genetic variation anywhere within that ancestral haplotype block may be responsible for the observed effect on disease risk. In other words, the tag SNP is chosen only for convenience, and the fact that it is associated with a particular disease does not necessarily imply that particular variant is causative. It could be another SNP or other polymorphism in linkage disequilibrium within that haplotype block or it could be a combination of polymorphisms acting in concert with one another. Similarly, the haplotype block may contain several genes. Once an association with a particular genetic region has been established it may not be possible to definitively identify what gene is contributing to the effect, and it is even more challenging to

identify what particular genetic variant (or combination of variants) within a given gene is responsible.

GWAS has now been successfully used in many large studies to identify common variants (present in about 5-50% of individuals within a population) with small effect sizes (increasing risk by about 20-50%) for virtually every common disease (such as diabetes mellitus, coronary artery disease, chronic kidney disease, and autoimmune disorders like systemic lupus erythematosus), including for those with a quantitative component (for example, hypertension). The results can be biologically interesting. For instance, GWAS of lung cancer and chronic obstructive pulmonary disease (COPD) has—perhaps unsurprisingly—identified genetic variants in nicotinic receptors, suggesting that inborn genetic variation in vulnerability to tobacco addiction contributes to risk for developing smoking-related diseases.

Consumer-level genetic testing. In addition to offering information about genetic ancestry, widely available, low-cost consumer-level genetic testing, such as through "23andMe," reports genotype results for variants originally identified through GWAS along with genotype determination for some common alleles for Mendelian disorders, such as sickle cell disease and common mutations of *CFTR*. They also report findings for other common disorders that blur the boundary between complex disease and single gene Mendelian disorders, such as late-onset **Alzheimer disease** resulting from the *APOE* E4 allele, which behaves as a Mendelian disorder but where the disease-associated allele is very common in the general population.

GWAS results are frequently misinterpreted by patients and physicians alike. In general, Mendelian disorders are uncommon, highly penetrant, and the responsible mutations are restricted to individual families and fairly obviously disrupt the protein-coding sequence of a gene. In contrast, variants contributing to common disorders confer only incremental risk and act additively with one another in concert with lifestyle and environmental exposures and their effect on how a gene functions may be difficult to tease apart.

The first human genome sequence was completed in 2003, required thousands of scientists working in laboratories throughout the world, and cost nearly $3,000,000,000. Direct-to-consumer whole genome sequencing now costs under $300. One can only begin to imagine how this information will revolutionize the practice of medicine and strain the ability of patients and health care providers to make sense of the data.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-67

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-68

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=268#h5p-69

Feedback:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://uw.pressbooks.pub/fmrgenetics/?p=268

6. Cancer genetics

Session Learning Objectives and a quick synopsis:

Most cancer is not inherited. Instead, it usually results from bad luck caused by the accumulation of mutations acquired with aging or environmental exposures such as tobacco. These mutations happen throughout the genome and are generally inconsequential except when they disrupt certain important genes influencing cell growth. There is also a role for complex inheritance of risk factors for most malignancies. Nevertheless, some families do have inherited forms of cancer. The genes responsible for inherited forms of cancer are sometimes also mutated in non-inherited (sporadic) forms of cancer.

SLO 1 Describe how a proto-oncogene transforms into an oncogene

Most proto-oncogenes are ordinarily involved in regulating cell growth. Mutations that unleash a proto-oncogene's cancer-inducing potential increase its activity, such as through its over-expression. Only one allele needs to be mutated. Proto-oncogenes primarily undergo somatic mutations. Only in rare circumstances can they be tolerated in the germline (and consequently all cells of the body) and cause a hereditary cancer predisposition syndrome.

SLO 2 Explain the two-hit model of tumor suppressor gene inactivation

In contrast to proto-oncogenes, tumor suppressor genes tend to exhibit a broader range of activities influencing cell growth. Tumor suppressor genes require loss of their activity in order to promote cancer. Consequently, most mutations inactivate them, and both alleles must be mutated to promote cancer. In inherited cancer predisposition syndromes, one allele is inactivated through a germline mutation, and inactivating mutation of the second allele is

the rate-limiting step that causes cancer (the "two-hit" mechanism). Some of the same genes responsible for inherited cancer predisposition syndromes are also biallelically mutated in sporadic cases of cancer, occurring in individuals without an inherited germline cancer predisposition; however, in this setting both alleles are mutated somatically.

SLO 3 Explain how inherited DNA repair deficiency leads to cancer

Inherited or acquired deficiency of DNA repair factors and other genes required to maintain the integrity of the genome is a cause of cancer. Nevertheless, mutations in DNA repair factors, by themselves, do not promote cell growth disturbances characteristic of cancer. However, a deficiency of DNA repair factors can lead to an accumulation of mutations throughout the genome, including in proto-oncogenes or tumor suppressor genes that do directly regulate cell growth, thereby leading to cancer. Many of these genes also fit into the category of "tumor suppressor" genes, requiring two mutations; when inherited, one allele is mutated in the germline and the other is acquired and restricted to the tumor, whereas, when sporadic, both alleles are mutated somatically.

Main text

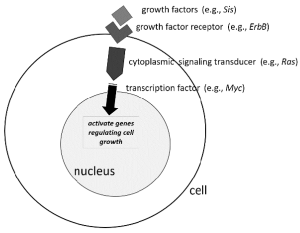
Cancer involves mutation of genes.

There are two generally accepted categories of genes whose mutation gives rise to cancer: oncogenes (known as "proto-oncogenes" before they are mutated) and tumor suppressor genes. Genes involved in DNA repair or that otherwise contribute to the integrity of the genome and gene expression are also frequently mutated in cancer; many of these also behave as tumor suppressor genes, at least with respect to the two-hit mechanism.

The conceptual framework for describing cancer genes and much of the terminology in use today is intricately tied to the experiments through which they were discovered.

SLO 1 Describe how a proto-oncogene transforms into an oncogene

Discoveryof proto-oncogenes. Oncogenes were originally recognized as agents that produced cancer in animals and that could be isolated from viruses with an RNA genome. These viruses—and other "retroviruses" like them—contain an RNA genome and possess the enzyme, "reverse transcriptase," that reverse transcribes their RNA genome into DNA upon infection. Retroviral oncogenes represent a mutant version of a gene contained in the host genome and that the virus has usurped. The non-mutated form of the gene, that is found in the host organism, is known as a "proto-oncogene." Later it was appreciated that somatic mutation of proto-oncogenes can lead to human cancers that are not associated with retroviral infections.

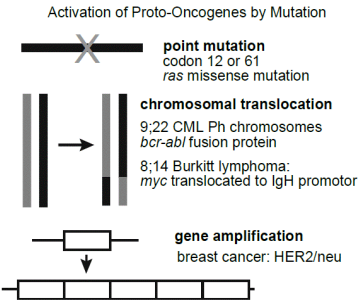


Many mammalian proto-oncogenes are known. In contrast to tumor suppressor genes (as we will soon discuss), the mutations that transform a proto-oncogene into a cancer-causing oncogene act dominantly at the cellular level. That is, one mutant copy (allele) of the oncogene is sufficient to perturb cell growth, and the wild type allele of the proto-oncogene is insufficient to halt the process.

Proto-oncogenes can be tied to cellular pathways regulating growth and differentiation. Mutations that transform a proto-oncogene into an oncogene render the protein it encodes hyperactive. Typically, this is a signaling protein involved in relaying an extracellular signal such as a growth factor to gene regulation in the nucleus.

For example, RAS, frequently mutated in many human tumors, encodes a GTPase that is "switched on" by incoming receptor signals and undergoes GTP/GDP exchange. Activated RAS protein ultimately turns on downstream genes involved in cell growth,

differentiation, and survival. (Humans have three closely related *RAS* genes, *HRAS*, *KRAS*, and *NRAS*, which we will collectively and nonspecifically refer to here as just, "RAS.") Cancer-associated, activating *RAS* mutations, which transform the proto-oncogene into an oncogene, cause the protein to be constitutively switched on all the time, even in the absence of receptor signals. Another frequently mutated human proto-oncogene is *MYC*, encoding a transcription factor that is situated downstream of multiple cell signaling pathways and activates a transcriptional program governing cell growth and division. In humans, oncogenic *MYC* mutations may lead to its overexpression or prevent its normal proteolytic turnover. Mutations in genes regulating any number of steps in cell signaling—levels of a particular growth factor, activation of its receptor, or conveyance of that signal downstream through the cytoplasm and into the nucleus—can lead to unregulated cell growth, a hallmark of cancer.



Several different types of mutations can transform a proto-oncogene into an oncogene. These include point mutation, chromosomal translocations that lead to gene fusion events, chromosomal translocations that lead to over expression by placing the proto-oncogene under the inappropriate control of another gene's promoter or enhancer sequences, and gene amplification in which a gene is tandemly duplicated several times, thereby increasing its expression.

Somatic chromosome 9;22 reciprocal translocation, the so-called "Philadelphia chromosome," is characteristic for chronic myeloid leukemia (CML). It results in a fusion of the protein-coding sequences of the *BCR* and *ABL* proto-oncogenes. It generates a fusion protein with a new tyrosine kinase activity. A highly effective

tyrosine kinase inhibitor imatinib (trade name, Gleevec) was developed to target the BCR-ABL CML-specific tyrosine kinase.

Several different chromosomal translocations juxtapose various proto-oncogenes with the immunoglobulin heavy chain promoter on chromosome 14 and define certain subtypes of non-Hodgkin lymphoma (e.g., t(8;14) in Burkitt lymphoma).

Somatic amplification of the *HER2* gene, encoding the EGF receptor 2 (EGFR2), in breast cancer is a therapeutic target of the monoclonal antibody-based drug, trastuzumab (trade name, Herceptin, among others).

Unlike tumor suppressor genes, mutations in proto-oncogenes are only rarely inherited. One such inherited cancer predisposition syndrome resulting from germline (inherited) transmission of a mutant proto-oncogene is **multiple endocrine neoplasia 2 (MEN2)**, in which inherited mutations of the *RET* proto-oncogene can cause a syndrome with a high frequency of medullary thyroid carcinoma, parathyroid adenoma, and/or pheochromocytoma of the adrenal and related tissue. *RET* also frequently undergoes activating somatic mutation in individuals who sporadically develop medullary thyroid cancer in the absence of an inherited syndrome such as MEN2.

Characteristics of oncogenes

- Identified as transforming genes of animal retroviruses.
- An activated form of a cellular gene (proto-oncogene).
- Act dominantly at the cellular level, which means only one allele need be mutated.
- Mutations are somatic and seldom inherited (one exception is *RET* in MEN2).

An interactive H5P element has been excluded from this

version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=272#h5p-70

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=272#h5p-71

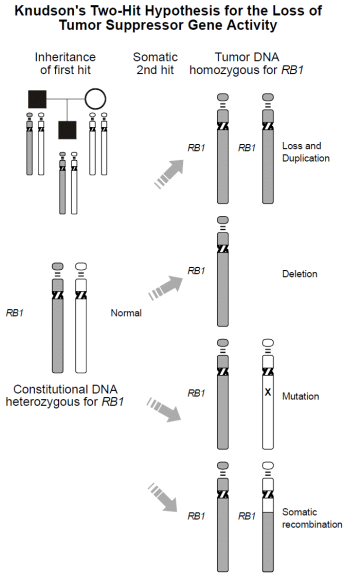
SLO 2 Explain the two-hit model of tumor suppressor gene inactivation

Tumor suppressor genes are genes whose normal function reduces the chances of cancer. Most tumor suppressor genes have been identified from mutations in individuals exhibiting autosomal dominantly inherited predisposition to cancer.

Knudson "two-hit" model of tumorigenesis. The prototype tumor suppressor gene is the *RB* gene, first isolated as the cause of familial **retinoblastoma**. In 1971, Alfred Knudson proposed the elegant "two hit" model of tumorigenesis to explain the epidemiology of retinoblastoma. Retinoblastoma is a childhood tumor of the retina. Knudson observed that about 40% of patients with retinoblastoma had an onset in infancy or early childhood. Those cases also tended to be bilateral, more frequently had a parent who also had retinoblastoma, and were often associated with additional malignancy later in life. In contrast, the remaining 60% of cases had a later age of onset, tended to be unilateral, almost never had a family history of the tumor, and were not associated with risk

for additional types of tumors. Knudson correctly hypothesized that the 40% of cases with early onset, bilaterality, and risk for other tumors represented a germline mutation in a particular gene, whereas the 60% of cases with later onset, unilaterality, and no risk of secondary tumors resulted from somatic mutations of the same gene. He coined the term "tumor suppressor gene" for this hypothesized gene.

Knudson's reasoning went something like this: Let's define a tumor suppressor gene as one that, must be *in*activated in order to cause cancer—unlike an oncogene (which had actually not yet been discovered) which must be activated by mutation to cause cancer. Since there are two copies of all autosomal genes, then we should require that both copies of a tumor suppressor gene must be inactivated in order to promote tumor development. Let's say the chance of an inactivating mutation in any particular gene (here the tumor suppressor gene) in any particular cell is one in a million (10^{-6}). What is the probability of mutating both alleles of the same gene in any particular cell? It's just 10^{-6} 10^{-6}, which equals 10^{-12}. That's a very small number, and it's not likely that this will happen. In fact, even if there were a million cells, say in the retina, the probability that one of these cells sustains two mutations in the alleles of same gene—2 hits—is only one in a million. But, what happens if someone inherits a mutation that inactivates one copy of that gene, such that all of the cells of the body, including in the retina and other tissues, have, from the moment of conception onward, already sustained the first hit? Then the probability of the second hit is still one in a million per cell. But, since there are about a million cells in the retina, the chance that the second hit will occur is pretty large—so large, in fact, that retinoblastomas often form during embryonic development and are present at birth. That is what happens with autosomal dominant inheritance of an



inactivating mutation in a single copy of the *RB* gene, whereas somatic mutation of both alleles of *RB* is a rarer occurrence. We predicted the chances of the latter happening to be about one in a million, and that is about how often retinoblastoma actually is found in the general population (i.e., among people who do not possess a germline mutation in one copy of *RB*).

The rest of the differences between sporadic and inherited forms of RB is well explained: The occurrence of multiple tumors in individuals inheriting constitutional mutations in *RB* is because a second hit occurs with high probability. In fact, the second hit is a random event. Some unfortunate patients will therefore develop two or three or more retinoblastomas, while a few lucky individuals will be non-penetrant and never develop a tumor. The elevated risk for other types of tumors in individuals inheriting constitutional mutations in *RB* results from the fact that all of the cells in the body harbor the first hit mutation. Since loss of RB function can lead to the development of tumors in other tissues, individuals with inherited mutations in *RB* are also at high risk for other types of cancer. Finally, since tumor development requires two somatic mutation events in individuals not inheriting a mutant *RB* gene, it makes sense that the longer one lives, the greater the probability of exposure to mutation; this explains the later age of onset of tumor development in non-inherited forms of retinoblastoma.

Since the time of Knudson's work, the *RB* gene has been identified. RB regulates transit through the cell cycle, by complexing with

transcription factors. Somatic mutations of *RB* occur in many different types of tumors, and it turns out to be one of the most important genes in regulating orderly growth and differentiation of tissues.

Some individuals without a family history of retinoblastoma can still have a constitutional mutation in one allele of *RB*. These individuals represent new mutations—not surprisingly, given the poor genetic fitness of a childhood cancer predisposition syndrome. Such a person is subject to all the risks that befall an individual with a hereditary form of retinoblastoma, including the probability of 1/2 of transmitting it to each of their children.

Note that, in contrast to oncogenes, *RB* and other tumor suppressor genes act recessively at the cellular level. Only when function is completely lost, through mutations that inactivate both alleles, is tumor growth promoted.

Remember that proto-oncogenes must gain an activity to become a full-fledged oncogene. Tumor suppressor genes, in contrast, must lose their activity to cause cancer.

Most of the mutations of tumor suppressor genes (either inherited or somatic) are deletions or point mutations that cripple their function. Deletions of tumor suppressor genes are probably most common. Consequently, there is a unique experimental signature of a mutagenic event in the genome that takes out a tumor suppressor gene. The name for this is "**loss of heterozygosity (LOH)**," based on the fact that inactivating mutations of tumor suppressor genes frequently involve deletions or disomy for the mutant allele, which reveal themselves by LOH for polymorphisms that can be used to experimentally distinguish the two alleles.

Following are but a few of several examples of clinically important cancer predisposition syndromes resulting from autosomal dominant inheritance of heterozygous germline inactivating mutations in a particular tumor suppressor gene.

TP53 and Li-Fraumeni syndrome. TP53 is a DNA binding protein whose expression is induced by DNA damage. It was formerly, and often still, referred to as "p53" based on an apparent mass of 53

kilodaltons. Among other things, it regulates decisions regarding programmed cell death (apoptosis). One molecular mechanism to reduce the probability of developing a tumor is to kill off cells that have sustained extensive mutation. TP53 integrates numerous signals monitoring the health of the cell. When faced with overwhelming DNA damage, TP53 will make the decision to commit programmed cell death, in order to avoid malignant transformation of the cell. But, what happens if TP53 isn't working? One can imagine that cells destined to form tumors will then not be eliminated.

Heterozygous germline mutations of *TP53* are the cause of Li-Fraumeni syndrome, named for the two physicians who first characterized the disorder. Li-Fraumeni syndrome is an autosomal dominant disorder involving the inheritance of multiple malignancies, especially breast cancer, sarcomas, brain tumors, leukemia, and adrenocortical carcinoma.

Like *RB*, *TP53* is somatically mutated in many malignancies. In fact, it is thought to be the single most commonly mutated gene in all of cancer.

Some families that would appear to have Li-Fraumeni syndrome but lack germline mutations in *TP53* have been found to harbor heritable mutations in other genes similarly regulating the cell cycle and DNA repair, including *CHEK2*, encoding a protein kinase forming a DNA damage and replication "checkpoint."

Neurofibromatosis 1. Neurofibromatosis 1 (NF1, also known as Von Recklinghausen disease) is a common autosomal dominant inherited syndrome of benign tumors known as neurofibromas occurring in conjunction with other benign tumors of the peripheral and central nervous system. The tumors can occasionally undergo malignant transformation, and the spectrum of malignancy includes neurofibrosarcoma, schwannoma, glioma, pheochromocytoma, and leukemia. Most of the time, though, the tumors remain benign and cause problems related to CNS "mass effects" and painful, disfiguring, or functional impingement of peripheral nerves. There are obvious cutaneous stigmata of the disease: the characteristic

fleshy neurofibromas, hyperpigmented "cafe-au-lait" macules on the skin, raised hamartomatous lesions of the iris (Lisch nodules), and axillary and inguinal freckling (places where even freckled people don't usually have very many). NF1 is caused by inactivating heterozygous germline mutations in the synonymous *NF1* gene, encoding neurofibromin, a GTPase-activating protein (GAP) that regulates RAS activity. Loss of NF1 activity leads to constitutive activation of the RAS signaling pathway. Mutations in a number of other genes encoding components of the RAS activation pathway can produce inherited disorders with clinical overlap to NF1 and are collectively referred to as "**Rasopathies**."

von Hippel-Lindau Syndrome. von Hippel-Lindau syndrome (VHL) is an autosomal dominant disorder of benign vascular tumors of the cerebellum, elsewhere in the brain and retina, pheochromocytoma, and renal cysts transforming to renal cell cancer. Even though they are benign, these vascular tumors cause problems when they enlarge and/or bleed in the brain and retina. As is the rule with most inherited tumor suppressor gene syndromes, both alleles of the gene (also called *VHL*) responsible for familial VHL are nearly invariably mutated in sporadic, non-familial forms of renal cell carcinoma.

Interestingly, mutation is not the only molecular mechanism for inactivating a tumor suppressor gene. It turns out that in some individuals with renal cell carcinoma, a CpG-rich region in the *VHL* gene promoter undergoes cytosine hypermethylation, switching off transcription of *VHL* and thereby leading to loss of expression, even without a frank mutation occurring in the DNA sequence. Somatic promoter hypermethylation (an epigenetic phenomenon, distinct from mutation) is a relatively common mechanism for inactivating a variety of tumor suppressor genes in sporadically occurring tumors.

The methylation status of a gene is fairly stable through cell division cycles. That is, once a gene is hypermethylated and switched off (whether it should occur aberrantly during tumorigenesis or in a developmentally appropriate manner during tissue differentiation, imprinting, or X chromosome inactivation),

the methylation state is generally preserved in the two daughter cells resulting from mitotic cell division. A somatic change in the methylation status of the cell, however, is not preserved during meiosis; thus, the name "epigenetic," as these states are, with possible rare exception, not heritable from one generation to the next.

The VHL protein is involved in signaling the tissue response to hypoxia.

Familial adenomatous polyposis. Familial adenomatous polyposis (FAP) accounts for about 1% of colon cancer. Probably all colon cancers begin as a benign polyp, which subsequently acquires additional mutations activating proto-oncogenes and inactivating tumor suppressor genes as it progressively transforms into a full-blown malignancy. Individuals with FAP usually have their colons carpeted with polyps by no later than their teenage or early adult years. Adenocarcinoma of the colon is therefore inevitable, and prophylactic colectomy in the early adult years is the only treatment option, as there are simply too many polyps to surveil and resect, even with frequent colonoscopy. The gene causing FAP is known as APC (for *adenomatous polyposis coli*), a protein with a complex role in cell cycle progression and extracellular communication and matrix attachment. Loss of APC activity results in chromosomal instability (CIN), leading to accumulation of multiple gross cytogenetic abnormalities of chromosome structure in tumor cells, which in turns leads to widespread activation of proto-oncogenes and inactivation of tumor suppressor genes.

As is the established paradigm with tumor suppressor genes, APC is commonly mutated in sporadic non-inherited forms of colon cancer.

Hereditary breast and ovarian cancer syndrome. Breast cancer is a common disease, and about one in nine of all females in North America will develop it. Breast cancer is inherited with complex genetics. As with other malignancies, most breast cancers appear sporadically, due to common genetic variants with weak effects, environmental exposures, and accumulation of somatic mutations

as we age. A minority of cases, however, is familial and results from highly penetrant single-gene mutations.

Hereditary breast and ovarian cancer syndrome accounts for about 5% of all breast cancer. Familial breast cancer is genetically heterogeneous. *BRCA1* and *BRCA2* were the first two responsible genes to be identified.

In general, familial forms of cancer exhibit several hallmarks, which help differentiate it from coincidental family clustering of sporadic cases, including earlier age of onset and multiple primary tumors.

Thus, for hereditary breast and ovarian cancer syndrome, there tends to be a younger age of onset of breast cancer (often premenopausal, whereas most sporadic breast cancer occurs post-menopause). Bilateral occurrence of a tumor or multifocal occurrence of more than one primary tumor on a single side (just as with inherited forms of retinoblastoma) or multiple tumors arising over time ("metachronous") are all more common in hereditary breast and ovarian cancer syndrome, as well as in other familial cancer predisposition syndromes. The presence of ovarian cancer in a family, which is much rarer than breast cancer, is often suggestive of families inheriting *BRCA1* and, to lesser extents, *BRCA2* mutations. The occurrence of male breast cancer, a rare event, is associated with *BRCA2* and, to a lesser extent, *BRCA1* mutation. Even individuals from breast cancer families who turn out not to have inherited the predisposing gene can still get breast cancer as a sporadic disease, just because breast cancer is so common in the general population.

Females inheriting a mutation in *BRCA1* or *BRCA2* face a lifetime risk of breast cancer of about 50-80%. Males inheriting *BRCA2* mutations are at elevated risk for developing breast cancer (approximately 6% lifetime risk). Females inheriting a *BRCA1* mutation have a lifetime risk of ovarian cancer of approximately 25-50%, whereas the risk is somewhat lower with *BRCA2* mutations. Germline mutations in both *BRCA1* and *BRCA2* confer smaller

elevations in risk for other types of tumors, including prostate cancer in males, melanoma, and pancreatic cancer.

Some populations show founder effects for BRCA mutations. About 2% of people of Ashkenazi Jewish European ancestry are heterozygous for one of three different mutant alleles of either *BRCA1* (two different alleles) or *BRCA2* (one allele). So large is this effect that most familial breast and ovarian cancer in this population can be attributed to one of these three mutations. Other ethnically isolated populations show a similar preponderance of founder mutations. For example, hereditary breast cancer frequently results from the same *BRCA2* mutation in Iceland, where meticulous genealogical records trace it back to a Viking.

Genetic testing for familial *BRCA1* and *BRCA2* mutations is revolutionizing the approach to this disease. Females found to inherit a germline mutation in either gene can be screened more conscientiously and consider risk reducing therapies, including prophylactic mastectomy and/or oophorectomy.

BRCA1 and *BRCA2* encode components of a large protein complex that detects and repairs DNA damage. The complex is referred to as the "Fanconi complex" because many of its components were first discovered as a cause of the genetically heterogeneous autosomal recessive disorder of DNA repair deficiency, known as "Fanconi anemia".

Heterozygous germline loss of function mutation of *BRCA1* and *BRCA2* cause autosomal dominantly inherited adult onset of cancer (breast, ovarian, and, to lesser extents, other types of cancer), whereas homozygosity for the very same loss of function mutations in *BRCA1*, *BRCA2*, or other genes encoding components of the Fanconi DNA repair complex cause autosomal recessively inherited childhood onset of cancer (Fanconi anemia). Can you think of a reason why there is a difference in the ages of onset depending upon whether the mutations in these genes are heterozygous or homozygous in the germline?

Here's why: Heterozygous mutations of *BRCA1* or *BRCA2* lead to half normal levels (haploinsufficiency) of BRCA1 or BRCA2 proteins.

For most proteins, half-normal levels are good enough. BRCA1 and BRCA2 don't directly regulate cell growth, division, or other cancer cell properties. And half-normal levels of BRCA1 or BRCA2 do not markedly impair DNA repair activities of the Fanconi complex. However, once a second hit (i.e., acquired somatic mutation) occurs in the remaining wild type allele of *BRCA1* or *BRCA2*, then DNA repair is severely disrupted because there is no normal BRCA1 or BRCA2 protein left, at all. Consequently, the cell that now has absence of BRCA1 or BRCA2 loses DNA repair proficiency and accumulates somatic mutations at a much higher frequency throughout the genome. Mutations activating proto-oncogenes or inactivating other tumor suppressor genes arise and therefore ultimately produce a tumor. It should be easy to imagine that if someone is conceived with complete deficiency of BRCA1 or BRCA2 (as a result of autosomal recessive inheritance of homozygous mutations), that total loss of this DNA repair pathway occurs immediately and persists throughout life. Moreover, unlike the case for a second hit, which occurs randomly and only affects an occasional cell, every cell in the body will have lost this type of DNA repair capability. Consequently, mutations begin to accumulate in all cells throughout the genome, including in proto-oncogenes and tumor-suppressor genes directly responsible for disrupting normal growth characteristics in a cancer cell. It should be no surprise, then, that cancer can begin in childhood.

Heterozygous germline *BRCA1* and *BRCA2* mutations are responsible for most families with hereditary breast and ovarian cancer syndrome. Nevertheless, many inherited cancer gene syndromes are now known, and it is challenging to clinically distinguish among them. As a result, clinical testing for hereditary breast and ovarian cancer and other types of inherited cancer syndromes has moved away from testing single genes. Instead, clinical genetic testing is moving to large gene panels that apply next-generation DNA sequencing technologies to simultaneously evaluate for germline mutations in dozens of genes associated with

hereditary cancer predisposition syndromes, including breast cancer.

Characteristics of tumor suppressor genes

- Identified as genes responsible for autosomal dominant human tumor syndromes.
- Recessive at cellular level, means both alleles must be inactivated.
- Often, the same tumor suppressor gene inherited in familial forms of the tumor is somatically mutated in non-hereditary, sporadic forms of the same tumor.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=272#h5p-72

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=272#h5p-73

An interactive H5P element has been excluded from this

version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=272#h5p-74

SLO 3 Explain how inherited DNA repair deficiency leads to cancer

Mutations of proto-oncogenes and tumor suppressor genes lead to cancer, but what would happen if, rather than starting with a mutation in one of those genes, you began with a mutation in a gene responsible for maintaining the integrity of the genome? As we have discussed, we might expect that mutation of DNA repair and cell cycle control genes could lead to a cascade of mutations, eventually activating proto-oncogenes and inactivating tumor suppressor genes.

There are basically two types of genes in this category. One set of genes is directly involved in sensing and repairing DNA damage. The DNA molecule frequently undergoes physical damage: Bases are lost from the phosphodiester backbone, bases become chemically modified by exposure to carcinogens and reactive oxygen species, a DNA strand may break, and DNA polymerase can mistakenly insert the wrong base, leading to a "mismatch." Much of this damage can be directly repaired, and that job falls to this first category of genes. The second major category of gene is involved in recognizing that DNA damage has occurred in the cell, temporarily halting the cell division cycle until the damage can be repaired, and then making an assessment as to whether the cell was fixed properly and should either re-enter the cell cycle or commit to programmed cell death. 2001 University of Washington Nobel Laureate Leland Hartwell in his seminal studies of yeast cell division mutants that failed to recognize when DNA damage had occurred introduced this latter "checkpoint" concept. A checkpoint is a temporary arrest in the cell

cycle at which a particular list of safeguards is monitored before going any further.

(Recently, "checkpoint inhibitor" drugs targeting immune checkpoints have come into use for pharmacologic immunotherapy of cancer. These are completely different from cell cycle checkpoints, and the similar terminology should not be confused.)

Several genes involved in DNA repair and regulation of the mitotic cell cycle have been implicated as having a role in initiating sporadic and hereditary cancer.

Lynch syndrome. Lynch syndrome involves susceptibility to numerous types of malignancy, most commonly of the colon and elsewhere in the gastrointestinal tract, the endometrium of the uterus, and the ovary. It is a genetically heterogeneous disease inherited in an autosomal dominant fashion as a result of heterozygous inactivating mutations in one of several genes. The two most common genes responsible for Lynch syndrome are *MSH2* and *MLH1*, with each being mutated in about a third of families with Lynch syndrome. Less often, Lynch syndrome is caused by mutations in *MSH6* and *PMS2*. Proteins encoded by *MLH1*, *MSH2*, *MSH6*, and *PMS2* are all part of a multi-protein complex involved in DNA double-strand mismatch detection and repair.

DNA mismatches represent incorrectly synthesized base pairs that result from misinsertion of the wrong base by DNA polymerase. Genomic instability of repeated sequences, known as "microsatellite instability," is a hallmark of tumors from patients in Lynch syndrome families. Lynch syndrome accounts for about 5-10% of all cases of colon cancer overall.

Lynch syndrome, as well as hereditary breast and ovarian cancer syndrome, really blur the boundaries between tumor suppressor gene disorders and DNA repair deficiency disorders. Both syndromes are inherited in an autosomal dominant disorder. Both are due to heterozygous germline loss of function mutation in a tumor suppressor gene. Both require complete loss of activity through a somatic second hit through a two-hit hypothesis (with an inherited germline mutation and acquired, somatic mutation

restricted to the tumor). Loss of DNA repair activity then leads to accumulation of secondary mutations activating proto-oncogenes and biallelically inactivating tumor suppressor genes.

Autosomal recessive syndromes of deficiency of DNA repair. The first group of DNA repair genes found to be involved in hereditary cancer predisposition syndromes is composed of rare autosomal recessive syndromes of malignancy. They are each characterized by a cellular deficiency in various aspects of DNA repair. The following lists some representative examples of diseases from this category.

Fanconi anemia. As discussed above, Fanconi anemia is an extremely genetically heterogeneous disorder due to autosomal recessive inheritance of homozygous mutations in genes, including *BRCA1* and *BRCA2* (for which autosomal dominant inheritance of heterozygous mutations are a cause of hereditary breast and ovarian cancer syndrome). Individuals with Fanconi anemia typically have a variety of birth defects, such as renal abnormalities and radial bone deformities, and an extremely elevated risk for developing acute myelogenous leukemia and squamous cell carcinoma of the head and neck. Cells from individuals with Fanconi anemia are sensitive to DNA alkylating agents (such as diepoxybutane and mitomycin C) and demonstrate chromosome breakage with unusual chromosomal tetrad formation during mitosis. Treatment of cultured patient cells with either of these agents forms the basis of a once commonly used (but now increasingly obsolete) clinical test for confirming the diagnosis of Fanconi anemia; however, other tests are available, including multigene panels employing next-generation DNA sequencing.

Ataxia telangiectasia. The defective gene in this autosomal recessive disorder, *ATM*, ordinarily recognizes broken chromosomes and participates in a cell cycle checkpoint to allow sufficient time for DNA repair to transpire. Affected individuals have a greatly elevated risk for hematopoietic and other malignancy. They also have a neurodegenerative syndrome characterized by an ataxic movement disorder. Telangiectasias are a cutaneous manifestation of this disease that are distributed on the sclerae and

in a malar pattern across the face. Cells from individuals with ataxia telangiectasia may demonstrate increased chromosome breakage upon exposure to the DNA cross-linking agent bleomycin, which was used for diagnostic testing before the gene was identified and could be genetically tested for.

Xeroderma pigmentosum. This is a genetically heterogeneous autosomal recessive disorder, resulting from mutations in various transcription factors and nucleases. Individuals with xeroderma pigmentosum are at extreme risk for developing skin cancer upon exposure to sunlight due to a cellular defect in "excision repair" of ultraviolet light-induced thymine dimers.

Summary of cancer genetics

- Rare causes of cancer have enlightened the understanding of common tumors:
- Studies of RNA transforming viruses led to the discovery of proto-oncogenes that are somatically mutated (and dominantly activated) in many forms of cancer.
- Studies of cancer families have led to the identification of tumor suppressor genes, showing frequent inactivation (of both alleles) in sporadic cases of similar type.
- Cancer is, in general, a multistep pathway requiring the accumulation of mutations in several proto-oncogenes or tumor suppressor genes. This process can be accelerated if a DNA repair gene is inactivated.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=272#h5p-75

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://uw.pressbooks.pub/fmrgenetics/?p=272#h5p-76

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://uw.pressbooks.pub/fmrgenetics/?p=272#h5p-77

Feedback:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://uw.pressbooks.pub/fmrgenetics/?p=272

7. Chromosomal abnormalities, Aneuploidies, Chromosomal Rearrangements

Session Learning Objectives and a quick synopsis:
Genes reside on chromosomes. Chromosomes can break, lose or gain material, rearrange, and become lost or duplicated. Chromosomal abnormalities disrupt many genes at once. Some chromosomal abnormalities are restricted to the person in which they are found. Some chromosomal abnormalities can be inherited. Some chromosomal abnormalities can be inherited from a "balanced carrier," in which the parent has structurally abnormal chromosomes that have neither lost or gained genetic material but can then rearrange further during meiosis (formation of a sperm or egg) in a way that causes loss and/or gain of genetic material to become "unbalanced".

SLO 1 Describe the cytogenetic organization of the human genome

The diploid human genome consists of about six billion base pairs of DNA organized into 23 pairs of chromosomes, comprised of 22 pairs of autosomes plus an XX or XY pair of sex chromosomes.

SLO 2 Compare and contrast the clinical laboratory tests (karyotype, FISH, and microarrays) used for detecting chromosomal abnormalities

A karyotype consists of a microscopic image of condensed metaphase chromosomes. Only large-scale copy number changes (interstitial duplications and deletions) and structural rearrangements can be visualized. It requires a

source of living cells. Fluorescence in situ hybridization (FISH) is a different microscopic technique that has much higher resolution, can be performed on metaphase or interphase cells, and even fixed (preserved) cells under certain circumstances, but is directed at probing for the integrity of certain chromosomal regions based on an underlying clinical hypothesis. Chromosome microarrays offer a high-resolution global view of the genome and only require a source of DNA rather than intact cells but are incapable of detecting balanced rearrangements. Karyotype and FISH are time and labor intensive, whereas microarrays can be performed in high throughput manner.

SLO 3 Describe human chromosomal abnormalities and how they are inherited

Some are inherited, some occur de novo in the germline, and others are restricted to somatic tissues, particularly cancers. Chromosome number imbalances are often the result of meiotic nondisjunction occurring as a function of maternal age, whereas structural rearrangements can occur via this or other mechanisms. The range of abnormalities includes loss or duplication of chromosomes; translocations; and interstitial insertions, deletions, and inversions. Chromosomal abnormalities disturb many genes, resulting in a severe clinical phenotype. An exception is sex chromosome abnormalities, which are clinically less pronounced. There are virtually unlimited numbers of possible ways that chromosomes can rearrange, meaning that many patients have unique and previously unseen clinical phenotypes. Nevertheless, there are recurrent rearrangements in which patients share common clinical features. Recurrent rearrangements seen in some cancers have diagnostic, prognostic, and treatment significance. Some chromosomal abnormalities result in no clinical abnormalities in parents but can promote deleterious rearrangements occurring

during meiosis. Different diagnostic technologies are required for different types of changes, as noted in SLO2.

SLO 4 Distinguish constitutional from acquired chromosomal abnormalities

Constitutional cytogenetic abnormality: chromosome make-up at birth that may be heritable. Acquired cytogenetic abnormality: typically not present at birth, limited to certain tissues, and not heritable.

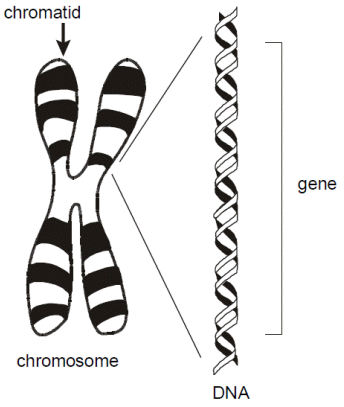
SLO 5 List clinical scenarios for which a chromosomal abnormality should be considered

Indications for suspecting a chromosome abnormality include multiple congenital malformations, intellectual disability, and growth failure, an overall clinical picture that resembles a specific known syndrome involving a chromosomal abnormality, recurrent pregnancy loss, infertility, short stature or primary amenorrhea, ambiguous genitalia, and hematologic malignancies and certain solid tumors.

Main text

1/150 newborns have a chromosome abnormality. At least 50% of first trimester spontaneous abortions ("miscarriages") have a chromosome abnormality. Approximately 2% of all pregnancies in women over 35 years-old have a chromosomal abnormality. Acquired chromosomal abnormalities are a common feature of malignancies.

SLO 1 Describe the cytogenetic organization of the human genome



To review, the diploid human genome consists of about six billion base pairs of DNA organized into 23 pairs of chromosomes. Each chromosome consists of a single linear continuous strand of DNA complexed with histone and other proteins. This combination of DNA and protein is arranged in structures of progressively greater degrees of organization. The double helix is wound into a nucleosome fiber, that in turn is coiled into a solenoid, then assembled into chromatin, and finally the chromatid.

Chromosome – A single DNA molecule condensed on a protein scaffold in the nucleus of a cell. Each chromosome contains hundreds or thousands of genes. Humans have 22 pairs of "autosomes" and one pair of sex chromosomes (X and Y), for a total of 46.

Chromatid – One of two identical parallel DNA strands in a mitotic cell following DNA replication.

Why do we need chromosomes? Although only one twenty-millionth of a meter-wide, the extended length of the human genome would reach several meters unless there were a means for compactly folding and organizing the DNA to fit within the confines of a cell.

Cytogenetics. Cytogenetics refers to the study of chromosomes, their structure, and their inheritance.

An interactive H5P element has been excluded from this

version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-78

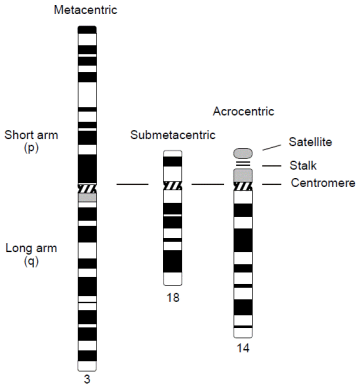
An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-80

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-79

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-81

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-82

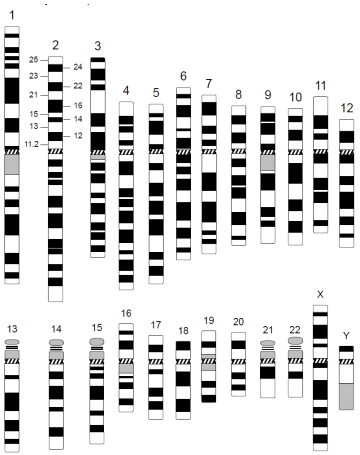
SLO 2 Compare and contrast the clinical laboratory tests (karyotype, FISH, and microarrays) used for detecting chromosomal abnormalities



Karyotype. Microscopic staining and visualization of chromosomes (a "karyotype") is usually performed on dividing cells. Peripheral blood leukocytes are most often used for routine chromosome analysis because of the ease of obtaining the specimen. Other tissues used for chromosome analysis include bone marrow samples (used mainly for the diagnosis of hematologic malignancies), solid tumors, skin (used frequently to detect mosaicism), and amniotic fluid cells and chorionic villi (for prenatal diagnosis of chromosomal disorders). The cells are grown in culture for a short time, then treated with the microtubule polymerization inhibitor colchicine to arrest the cells during the metaphase of mitosis, when chromosomes are maximally condensed. The cells are lysed, the chromosomes are fixed to a glass slide, then stained. Photomicrographs are obtained and images of individual chromosomes are cut out and pasted together for comparison.

Chromosomes are classified into different groups on the basis of centromere position and chromosome size. The centromere is

a standard cytological landmark that divides the chromosome into two arms—p (for petite), or short arm, and q (because it's the next letter of the alphabet) for the long arm. Human chromosomes are classified by centromere position into three types—metacentric (arms of approximately equal length), submetacentric (arms of unequal length), and acrocentric (centromere near one end of the chromosome). The p arm of the acrocentric chromosome is composed of a satellite and stalk region that contains copies of ribosomal RNA genes. Because there are a half-dozen acrocentric chromosomes, ribosomal RNA genes exist in vast redundancy, so deletion of a p arm from an acrocentric chromosome (mainly through chromosome translocation) is usually not clinically meaningful.



In addition, the chromosomes are numbered and arranged according to size, from largest to smallest, and position of their centromere, except for the sex chromosomes, which are always placed in the lower right-hand corner. This ordered configuration of chromosomes is called a karyotype. Note that chromosome numbers were assigned in the era before DNA sequencing became available. While they are supposed to be numbered from longest to shortest, we now know that chromosome 21 is actually shorter than chromosome 22.

Karyotype – Microphotographic array of stained chromosomes ordered according to their length and the relative position of their centromeres. Cells to be karyotyped must be obtained when they are still living and capable of dividing. Karyotypes are not capable

of resolving chromosomal abnormalities smaller than a few million base pairs in length.

Each chromosome has a characteristic banding pattern when stained. Cytogenetic analysis in the laboratory can produce various levels of resolution based upon the quality and origin of the chromosome preparation. Typically, the greater the number of bands that can be distinguished, the better the resolution.

An idealized drawing of the human karyotype is referred to as an "ideogram." The ideogram shown here represents a resolution of 400 bands. Notice that the bands have been numbered with a standardized system, which allows one to define structural and numerical abnormalities. Each arm is first divided into regions (outer numbers) and then each region into specific bands (inner numbers). Numbering is from the centromere outward on each arm, the region number, and finally the band number. For example, band 2p21 (pronounced "two-pee-two-one") indicates chromosome 2, short arm, region 2, band 1.

A standardized system of karyotype nomenclature has been defined. The system of notation is as follows: first indicate the total number of chromosomes, followed by a comma and then by the chromosome constitution. For example, 46,XX for female and 46,XY for male. If an abnormality is present, it is described using specific notations. For example, 47,XY,+21 (male with trisomy 21 Down syndrome).

Routine and high-resolution chromosome analysis permit the detection of numerical and structural chromosome abnormalities, but only to the extent to which we can actually see these changes under the microscope. Small chromosomal deletions, translocations and "marker" (fragmentary) chromosomes are difficult and sometimes impossible to identify with routine analysis. Even under the best conditions, a karyotype cannot resolve chromosomal changes smaller than 5-15 million base pairs.

Fluorescence in situ hybridization (FISH). Fluorescence in situ hybridization (FISH) provides the ability to simultaneously assess molecular and cytologic information. FISH analysis also allows for

the identification of small deletions, translocations, and other chromosomal rearrangements too small to be detectable by ordinary karyotype. Various FISH protocols can distinguish chromosomal abnormalities as small as a few thousand base pairs. Another distinct advantage of FISH is that it can be performed on interphase cells, when the chromosomes are maximally unwound and extended. Unlike with a karyotype where cells are arrested when chromosomes are condensed during mitosis, FISH can be performed without a round of cell division immediately upon obtaining a sample of blood or other tissue, and results can consequently be returned to patients more quickly.

FISH is a technique in which a labeled chromosome-specific DNA segment (probe) is hybridized with chromosomes and then visualized under a fluorescence microscope. Several types of FISH probes are available depending upon the application in question. Probes to sequences repeated in the centromere and that are unique to each chromosome produce intense, tight signals that are easy to count in interphase nuclei and can therefore be used to quickly differentiate certain trisomies, such as trisomy 21 Down syndrome.

Microarrays. An approach for evaluating for small chromosomal abnormalities involves the use of microarrays. Two main platforms in use today are array comparative genomic hybridization (aCGH) and SNP arrays. The two approaches are similar and employ a "chip" onto which thousands of oligonucleotides are individually immobilized onto spots in an array. Genomic DNA is then hybridized to the array. If there is a deletion, then there will be an absence of binding to the spotted oligonucleotides corresponding to that position in the genome. If there is a genomic duplication, then there will be excess binding to the corresponding positions. Microarrays have several advantages compared to karyotype analysis and even FISH. Unlike karyotype and FISH, a microscope is not incorporated into the analysis. They therefore offer greater resolution. Purified DNA can also be analyzed whereas karyotypes and FISH require intact chromosomes typically obtained from growing cells. There

are advantages in cost and throughput, as well. One weakness with microarrays is that they cannot detect "balanced" rearrangements, such as translocations or inversions, where the linear arrangement and relative chromosomal placement of genes are disturbed but the overall content of the genome otherwise remains intact.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-83

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-84

SLO 3 Describe human chromosomal abnormalities and how they are inherited

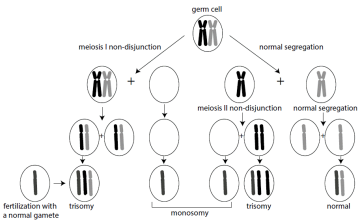
Aneuploidy – refers to an extra or missing chromosome. It is a common type of numerical chromosomal abnormality and can involve either autosomes or sex chromosomes. Trisomy 21 Down syndrome is an example of an autosomal aneuploidy. 47,XXY Klinefelter syndrome and 45,X Turner syndrome are examples of sex chromosome aneuploidies.

Chromosomal abnormalities may be divided into two broad categories—numerical and structural abnormalities. Numerical abnormalities are simply abnormalities of chromosome number,

typically involving just one chromosome (i.e., "aneuploidy," such as trisomy 21 Down syndrome). Structural abnormalities include chromosomal rearrangements, deletions, and duplications.

Numerical abnormalities. Numerical abnormalities consist of those due to extra or missing chromosomes (aneuploidy) as well as multiples of the haploid chromosome number (polyploidy). Aneuploidy may involve either autosomes or sex chromosomes. Most common are monosomies (45 chromosomes) and trisomies (47 chromosomes).

Most aneuploid conditions are the result of nondisjunction, the failure of chromosomes engaged in a synaptic complex to separate during meiosis. At the top of the figure, each parental homolog of the same chromosome is depicted in a synaptic complex. Note that nondisjunction can occur at either meiosis I or meiosis II. When it occurs at meiosis I (left side of the figure), both parental homologs are included in the first division products, but they then segregate normally, except for the fact that there are two copies of one chromosome. Conversely (right side of figure), meiosis I can proceed normally, followed by nondisjunction in meiosis II. When that happens, two copies of the same parental homolog segregate into the gamete. In either case, however, following fertilization with a gamete containing a normal chromosome complement, the resulting zygote will either be trisomic or monosomic for the chromosome that did not disjoin. For simplicity, we are ignoring the impact of recombination in the nondisjunction process. In reality, the effect of recombination can be used to determine whether the nondisjunction event occurred in meiosis I or meiosis II. From such studies, it has been found that the overwhelming majority of cases of nondisjunction occur during meiosis I of maternal gametogenesis.



Most autosomal aneuploidy is the result of maternal meiosis I non-disjunction, which is associated with advanced maternal age.

Autosomal aneuploidy. There are three survivable trisomy syndromes that may result in a live birth, trisomy 21 (Down syndrome), trisomy 18 (Edwards syndrome), and trisomy 13 (Patau syndrome).

Down syndrome (Trisomy 21). The most common autosomal trisomy is Down syndrome, with an incidence of 1/800 live births. It is the most common cause of moderate intellectual disability. Individuals with Down syndrome have characteristic findings including hypotonia, unique features involving the face and head (flat nasal bridge, upslanting palpebral fissures, epicanthal folds, lightly colored "Brushfield" spots of the iris, flat midface, small dysplastic ears, protruding tongue, brachycephaly—"short" head, excess nuchal skin), hand (short and broad with fifth finger mid-phalanx hypoplasia and a single transverse palmar—"simian"—crease), congenital heart disease (most commonly endocardial cushion defects involving persistence of the atrioventricular canal), duodenal atresia, as well as an increased risk of developing leukemia, hypothyroidism, and early onset Alzheimer disease. Most individuals with Down syndrome have trisomy 21 due to nondisjunction. Maternal nondisjunction is typical (usually meiosis I error), and Down syndrome is associated with advanced maternal age (AMA).

It is probably no coincidence that trisomy 21 involves the shortest chromosome (therefore having among the fewest genes) and is the least severe of chromosomal trisomies; presumably, this is because fewer genes are triplicated compared to trisomies involving larger chromosomes.

In a minority of cases (approximately 2-4%), Down syndrome is the result of an unbalanced "Robertsonian" translocation leading to trisomy of the long arm of chromosome 21. A Robertsonian translocation is a special type of chromosome involving fusion of the long arms of two acrocentric chromosomes. In such cases, one parent is a Robertsonian translocation carrier, which predisposes

to offspring with an unbalanced chromosome complement. The chromosomes may segregate abnormally in the offspring of a Robertsonian translocation carrier.

Even though the diagnosis of Down syndrome can be made fairly confidently based on the history and physical examination alone, it is important to always obtain a karyotype on a child with a clinical diagnosis of Down syndrome, both for confirmation and to determine if the child is the result of a Robertsonian translocation—in which case either parent could be a carrier and future pregnancies would be at risk for recurrence.

Characteristics of an autosomal chromosomal imbalance are the triad of developmental delay/intellectual disability, growth retardation, and multiple congenital abnormalities that often results in early fetal or neonatal demise. In contrast, sex chromosome aneuploidy tends toward comparatively mild phenotypes.

Uniparental disomy. As previously discussed in the session on other forms of inheritance, uniparental disomy (sometimes referred to as isodisomy) is the inheritance of both copies of a chromosome (or part of a chromosome) from just one parent, rather than the usual situation of inheriting one copy from each parent. Recall that uniparental disomy causes problems in two situations. First, if one parent is a heterozygous carrier for a recessive disease, then duplication of that chromosome will cause the conception to be homozygous even when the other parent is not a carrier. Second, if the individual has uniparental disomy for a chromosome or portion of a chromosome that includes an imprinted gene (such that it is not expressed from the parental sex corresponding to the duplicated (disomic) chromosome), then there will be loss of expression of that gene (because a chromosome from the other parent who would ordinarily express the imprinted gene is not present). Probably the most common mechanism for uniparental disomy is "rescue" of a conception that was initially trisomic. In this situation, the early embryo, probably consisting of no more than a few cells, is not-viable. However, if there is a *mitotic* non-disjunction event that kicks out one of the extra chromosomes, then that particular cell

of the embryo may divide to clonally expand and give rise to a developing embryo. Note that in such a trisomic cell, either one of the duplicated chromosomes or the parental chromosome present in just a single copy can be lost. If it is the latter, then the result will be uniparental disomy.

Sex chromosome aneuploidy. Monosomy and trisomy of the sex chromosomes also occurs not uncommonly. The following are examples of sex chromosome abnormalities.

Turner syndrome (45,X). The incidence of Turner syndrome is approximately 1 in 5,000 live female births. Individuals with Turner syndrome, caused by near or complete absence of an X chromosome, have many features in common including short stature, lymphedema (resulting in a "puffy" appearance) of hands and feet at birth, webbed neck, broad chest with widely-spaced nipples, cubitus valgus (widened carrying angle of the arms), congenital heart disease (coarctation of aorta being most common), kidney abnormalities and gonadal dysgenesis resulting in a lack of secondary sex characteristics, and amenorrhea and infertility in the vast majority of females. Turner syndrome females have normal intelligence but may have difficulties with spatial perception.

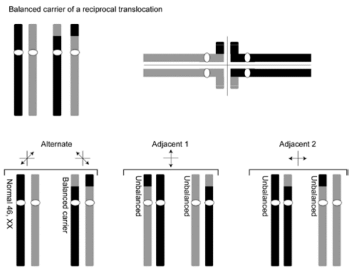
Klinefelter syndrome (47,XXY). Males with Klinefelter syndrome have an additional X chromosome. The incidence of this sex chromosome abnormality is 1 in 1,000 live male births. These males tend to have tall stature, "eunuchoid" habitus, gynecomastia, and testicular atrophy. They typically do not have intellectual disability but may have some learning disorders. The extra X chromosome is maternal in origin in most cases, and there is an increased incidence of Klinefelter syndrome with advanced maternal age. Many of the clinical problems can be ameliorated with testosterone therapy. The extra X chromosome is typically inactivated and forms a Barr body.

Structural chromosomal abnormalities. Structural chromosome abnormalities may be balanced or unbalanced. By "balanced," we mean that the rearrangements do not produce a net loss or gain of genomic material and usually cause no phenotypic effect. Rarely, the "breakpoints" for a structural chromosomal abnormality may

disrupt a gene, but that is the exception, given that genes are relatively sparsely distributed across the entire genome. However, individuals with a balanced rearrangement have an increased risk for producing unbalanced gametes, which, in turn, if fertilized, produce conceptions with a complement of chromosomes containing partial deletions and/or duplications. Examples of balanced rearrangements include reciprocal and Robertsonian translocations (as noted, a specific type of translocation involving acrocentric chromosomes) as well as chromosomal inversions.

Translocations. Translocations involve the exchange of genetic material between nonhomologous chromosomal regions (or different regions, such as the long and short arms of the same chromosome).

Reciprocal chromosomal translocations. Reciprocal translocations involve the breakage of nonhomologous chromosomes with reciprocal exchange of chromosome material.



The carrier of a reciprocal translocation is usually phenotypically normal. However, carriers have an increased risk for producing unbalanced gametes and abnormal phenotype in their offspring.

When chromosomes undergo translocation they have to form complicated structures to achieve paired synapses during meiosis. An example is shown of the meiotic segregation patterns that are possible in the carrier of a balanced, reciprocal translocation involving two chromosomes, colored black and gray and showing how they pair during meiosis. In this example, the parent carries a reciprocal translocation. Notice that the two translocation chromosomes form an unusual "quadriradial" synapse, which may be resolved in one of three ways. Alternate segregation yields one gamete with a normal complement of chromosomes and another gamete containing the pair of reciprocally balanced

translocation chromosomes. Each of these gametes should produce an individual with a normal phenotype following fertilization by a normal gamete of the opposite sex. Either of the adjacent segregation patterns, however, yields unbalanced gametes. Each gamete contains a partial trisomy and a partial monosomy and is therefore expected upon fertilization to produce an abnormal phenotype.

It would appear that the risk of producing an abnormal offspring would be quite high (2/3 or 67%); however, due to prenatal loss of chromosomally abnormal offspring, the empiric risk of an unbalanced offspring from a parent carrying a balanced reciprocal translocation is much lower (approximately 5-15%).

Chromosomalinversions. Inversions within a chromosome result from two breaks on a chromosome with inversion of the segment and reinsertion at its original site. Approximately 1 in 1,000 individuals carries an inversion. Inversions are balanced structural rearrangements and therefore inversion carriers are usually phenotypically normal (no loss or gain of genetic material). However, during meiosis a chromosome with an inversion cannot normally pair with a structurally intact chromosome, and carriers are at risk to produce offspring with deletions or duplications and resultant abnormal phenotype. Inversions are classified as to whether they do or do not span the centromere because the two types behave differently during meiosis, with attendant implications for reproductive outcomes.

Chromosomaldeletions. Deletions within a chromosome are caused by a break in the chromosome with resultant loss of chromosome material. Deletions result in an abnormal phenotype due to the loss of one or more genes.

A well-documented deletion syndrome is **Cri-du-Chat syndrome (5p-)**, involving loss of the terminus of the short arm of chromosome 5. Infants with Cri-du-Chat syndrome have a distinctive "cat-like" cry as well as microcephaly, profound intellectual disability, growth retardation, and a unique facial appearance.

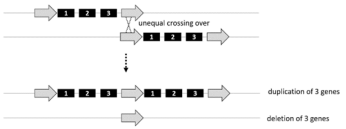
Some deletions, as well as duplications, are too small to see under

the light microscope and hence go undiagnosed by karyotype. They are only detectable by molecular cytogenetic techniques such as FISH or microarrays. These conditions are referred to as "microdeletion" (or "microduplication," "contiguous gene," or "copy number variant (CNV)" disorders).

22q11.2 deletion syndrome (previously going by several different names including CATCH22 syndrome, DiGeorge syndrome, and velocardiofacial syndrome) is probably the most common microdeletion syndrome, with an estimated incidence of about 1/3,000 live births. It results from a deletion of the long arm of one chromosome 22 at band q11.2, spanning several megabases of DNA in which several dozen genes are located. Affected individuals may have cardiac abnormality, abnormal facies, thymic aplasia, cleft palate, hypocalcemia/hypoparathyroidism, and elevated risk for schizophrenia, in addition to other problems. Many of these problems can be tied together developmentally due to abnormal anatomic development of the fourth pharyngeal arch in the embryo. There is variability of the phenotype. 22q11.2 deletion often arises de novo in an affected individual whose parents are unaffected and who lack the causative chromosomal abnormality, but it can also be inherited in an autosomal dominant fashion, in which case affected individuals have a 50% chance of having a child with this disorder with each pregnancy.

One reason for variability in microdeletion syndromes is that different individuals will each have a deletion of varying extent sometimes involving different flanking genes. Nevertheless, a common collection of overlapping phenotypes, resulting from overlapping regions of deletion, can be identified for the patients, as a group.

Chromosomal duplications. Duplications within a chromosome represent tandemly duplicated regions of genetic material resulting in a partial trisomy and also lead to clinical phenotypes.



Generation of copy number variants by unequal meiotic crossing-over. Deletions and duplications of chromosome material may result from unequal crossing-over during meiosis or may occur in the offspring of reciprocal translocation carriers.

Unequal crossing-over during meiosis is a major mechanism for the de novo generation of deletions and duplications. They often occur at repetitive sequences, which are found throughout the genome and can serve as recombination hotspots during meiosis. Ordinarily, crossing-over during meiotic recombination takes place within a pair of homologous chromosomes that line up exactly, base pair-by-base pair. As the figure illustrates, however, when there is a repetitive sequence (open arrow), the two chromosomes within the pair may align out of registration. If that happens, the resulting recombination event can create a deletion or duplication. The same regions that are deleted can also be duplicated. For example, there is a 22q11.2 duplication disorder.

Copy number variant disorders therefore very often arise de novo through this mechanism. Once they are generated, then they are transmitted with autosomal dominant genetics.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-85

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-86

SLO 4 Distinguish constitutional from acquired chromosomal abnormalities

In any discussion of cytogenetics, it is important to first distinguish between "constitutional" and "acquired" chromosomal abnormalities. Constitutional refers to the genetic make-up at birth. These abnormalities are usually present in every cell of the body, are often inherited, and are at risk of being transmitted to children. Acquired (somatic) chromosomal abnormalities usually occur after birth (or at least post-zygotically, after conception), are in a limited distribution of the cells of the body (often times just in a tumor, for example), are not inherited, and are not at risk for being transmitted to future generations.

Acquired chromosome abnormalities may occur in various types of hematologic malignancies and solid tumors.

An example of this type of phenomenon is the reciprocal **t(9;22)** translocation between chromosomes 9 and 22 seen in patients with **chronic myelogenous leukemia (CML)**. This translocation alters the position of two proto-oncogenes (*BCR* and *ABL*) resulting in formation of a fusion gene product with novel catalytic activity. The derivative chromosome 22 is referred to as the "**Philadelphia chromosome**," and is a marker for CML. The *BCR-ABL* fusion gene product creates a novel kinase that is uniquely targetable by an effective drug therapy, the first-in-class for which is imatinib (tradename, Gleevec), which has considerably improved prognosis for CML patients.

Chromothripsis. This is a recently discovered bizarre and spectacular somatic cell phenomenon, predominantly described in

cancer cells, in which multiple chromosomes shatter into thousands of smaller pieces and are imperfectly reassembled, leading to an abnormal number of chromosomes and numerous segmental chromosomal translocations, deletions, and insertions throughout the genome.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-87

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-88

SLO 5 List clinical scenarios for which a chromosomal abnormality should be considered

So, when is it appropriate to order a chromosome analysis? The following are examples of individuals who may have a chromosome abnormality or carry a rearrangement that may predispose to offspring with a chromosome abnormality:

- Individuals with multiple congenital malformations, intellectual disability, and growth failure (typical characteristics of an autosomal chromosomal imbalance).
- Confirmational testing of suspected chromosome abnormalities based upon their clinical phenotype, for example

to confirm a clinical suspicion of 22q11.2 deletion syndrome or Down syndrome.
- Couples with recurrent pregnancy loss (suspecting that one parent may carry a balanced rearrangement).
- Infertility without evidence of a pregnancy loss
- Females with short stature or primary amenorrhea (possible X chromosome abnormality, i.e., Turner syndrome).
- Males with infertility (possibly Klinefelter syndrome or Y chromosome abnormality).
- Patients with ambiguous genitalia, where genetic definition of sex may be helpful in conversations related to gender choice.
- Patients with hematologic malignancies (to aid in diagnosis, prognosis, and treatment of a specific malignancy based upon the particular acquired chromosome abnormality) and certain solid tumors.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-89

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=279#h5p-90

Feedback:

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://uw.pressbooks.pub/fmrgenetics/?p=279

8. Genetic Therapy Frontiers

Session Learning Objectives and a quick synopsis:

Advances in genetics are rapidly and profoundly transforming medical diagnosis, treatment, and ethics. This session offers a brief overview of emerging genetic technologies, many of which have been recently introduced into practice.

SLO1 Compare and contrast indications and limitations of different genetic tests

Cytogenetic tests are useful for evaluating for structural or copy number changes affecting a large number of genes when there are clinical indications that a patient may have a chromosomal abnormality. Evaluation of specific genes or use of large gene panels is appropriate when a differential diagnosis can be formulated. Previously undiagnosable, one-of-a-kind patients often benefit from whole exome or whole genome sequencing.

SLO2 Compare and contrast sources of DNA used for genetic testing

DNA taken from cells representative of the patient's germline constitution, easily obtained from blood or via skin biopsy, is useful for genetic testing. Tumor tissue can be used to identify cancer-specific mutations. DNA from tumor tissue can sometimes be detected in peripheral blood, urine, or feces. Amniotic fluid, placental tissue, or fetal DNA circulating in maternal serum can be used for prenatal genetic diagnosis.

SLO3 Be prepared to explain possible outcomes of genetic test results to patients and their families

Genetic tests can return with a positive result, a negative result, a variant of undetermined/unknown significance, or a worrisome incidental finding unrelated to the problem that motivated the patient to seek clinical attention. Patients

need to be informed of the potential outcomes, and physicians need to understand how to react to such findings.

SLO4 Identify forms of inherited and acquired disorders potentially amenable to gene therapy and genome editing technologies

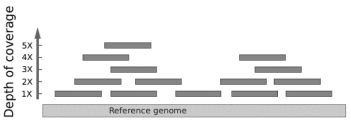
Autosomal and sex-linked recessive disorders are amenable to "gene addition" forms of gene therapy. Autosomal dominant disorders require that the mutant allele be inactivated. In vivo and ex vivo therapeutic technologies involving viral delivery of genes, genome editing, and oligonucleotide modification of gene expression and mRNA splicing continue to evolve.

Main text

SLO1 Compare and contrast indications and limitations of different genetic tests

Several technologies are in current use for performing genetic tests.

Sanger DNA sequencing. Sanger DNA sequencing (also known as dideoxy DNA sequencing, due to its use of chain-terminating dideoxynucleotide substrates) is sort of an analog technology in which many copies of identical molecular fragments of DNA are analyzed at the same time. There is a single readout of results, in the form of an electropherogram tracing. Sanger sequencing works best when there is a limited genetic differential diagnosis for the disorder being considered, such that only a few genes or a known mutation are being evaluated. For example, Sanger DNA sequencing can be used to confirm diagnosis of a suspected hemoglobinopathy, which is often first evaluated by hemoglobin electrophoresis.



Next-generation DNA sequencing. Next-generation DNA sequencing is sort of a digital technology where many different DNA fragments of about a thousand base pairs or so can be more economically sequenced at once. The entire population of sequenced DNA

molecules are evaluated individually, yet as an overlapping composite in order to interpret the sequence at any given base pair. Next-generation DNA sequencing can be used to evaluate multiple genes at once, including even the entire exome (i.e., the ~1% of the genome contained in exons) or whole genome. Because each nucleotide position is evaluated by multiple short DNA sequence reads to a high "depth of coverage" (or "read depth"), there is a potential to identify non-germline somatic mutations occurring in cancer or in mosaic disorders. For example, a single base position might be read more than 100 times. If the mutation is a heterozygous single base substitution, then, on average, about 50% of the reads should reveal the variant. If a variant base is found to occur at a significantly lower frequency, then that might be taken as evidence of mosaicism or clonal heterogeneity if studying a tumor. The latest advancement in DNA sequencing is "long read" technology, which determines the sequence of individual DNA molecules spanning tens to hundreds of thousands of nucleotides. Compared to conventional "short read" next-generation sequencing, long read sequencing offers several advantages: it enhances the detection of mutations within repetitive regions of the genome that are often misinterpreted by short read methods, and it can distinguish whether multiple variants are located on opposite parental chromosomes (in *trans*) or the same chromosome (in *cis*), which is particularly valuable when evaluating potential autosomal recessive conditions.

Gene panels. Next-generation DNA sequencing is now invariably used for large gene panels. For example, **retinitis pigmentosa**, an inherited form of progressive blindness, is extremely genetically heterogeneous. There are more than 100 genes known to cause the disorder, so a gene panel test would be appropriate in this situation.

Exome and whole genome sequencing. Next-generation DNA sequencing is the only technology available for analysis of the exome, which constitutes ~1% of the human genome or ~30 Mb (million base pairs), spread out across ~180,000 exons from ~20,000 total genes. Exome (or whole genome) sequencing is appropriate

when the disorder is so unique that there is literally no good guess as to what the patient may have.

Limitations of DNA sequencing. DNA sequencing has several limitations. Regardless of Sanger sequencing vs. next-generation DNA sequencing, if two different mutations are found in a particular gene known to be associated with an autosomal recessive disorder, it cannot be readily determined whether they are occurring in *trans* (one mutation on one parental chromosome and the other mutation on the other parental chromosomal homolog as would be required for autosomal recessive inheritance with each parent being a carrier) or in *cis* (both on the same parental chromosome with the other parental chromosome being normal). The only way to "phase" the distribution of variants discovered through DNA sequencing is to sequence the parents and determine how they segregate within the family or else resort to research studies involving the physical isolation of a single DNA molecule, which is generally outside the scope of readily available clinical laboratory testing. Long read sequencing, which is increasingly being adopted clinically, can also determine the phasing of variants. For this reason, when performing exome sequencing on an individual where the mode of inheritance is unknown (and that is frequently the case since exome sequencing is usually reserved for patients where the diagnosis is uncertain), then it is best to sequence both parents (known as "trio" sequencing), if available, at the same time as the patient.

Neither Sanger sequencing nor next-generation sequencing is especially good at detecting insertion or deletion mutations, and next-generation sequencing is particularly poor at resolving short repetitive sequences such as those encountered with neurodegenerative disorders like the CAG repeat responsible for Huntington disease. Instead, for the latter, PCR with electrophoretic separation of the products is usually performed to simply evaluate the length of the repetitive tract. Increasingly, long read DNA sequencing is also being utilized.

Cytogenetic testing. Although next-generation DNA sequencing is widely available, cytogenetic tests such as karyotyping, FISH, and

microarray remain essential because they assess chromosomal structure rather than DNA sequence. Karyotyping provides a genome-wide overview of large-scale chromosomal abnormalities, but its resolution is limited to changes larger than a few million base pairs. FISH (fluorescence in situ hybridization) is not suited for genome-wide screening and typically requires a specific clinical hypothesis, but it is valuable for confirming known microdeletion or microduplication syndromes and for detecting tumor-specific rearrangements of diagnostic or prognostic significance, such as the t(9;22) Philadelphia chromosome in chronic myeloid leukemia. Microarrays have largely replaced FISH for genome-wide assessment of copy number variants, such as deletions and duplications seen in disorders like 22q11.2 deletion syndrome. However, microarrays cannot detect "copy-neutral" structural changes—such as balanced translocations or inversions—because, unlike karyotyping or FISH, they do not visualize chromosome architecture.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=283#h5p-4

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=283#h5p-3

SLO2 Compare and contrast sources of DNA used for genetic testing

Genetic testing requires DNA, and it can be obtained from a variety of sources.

***Constitutional* DNA.** DNA representative of all the cells of the body and hence corresponding to the "germline" is conveniently obtained from peripheral blood ("peripheral" to distinguish it from blood cells obtained from the bone marrow), via venous phlebotomy, or from saliva. Sometimes dermal fibroblasts isolated by a skin biopsy are used in place of blood; one situation would be to screen for a germline mutation that might cause leukemia, where acquired mutations might also be present in white blood cells.

***Fetal* DNA.** Fetal DNA can be obtained by amniocentesis (where fetal cells are directly sloughed off into the amniotic fluid) or by chorionic villus sampling (CVS), in which a placental biopsy is performed. More recently, screening tests employing next-generation DNA sequencing to analyze cell-free fetal DNA circulating in maternal serum (noninvasive prenatal testing test (NIPT)) have become routine for prenatal screening for fetal aneuploidy, such as Down syndrome, increasingly replacing an older "triple screen" test that analyzed maternal serum levels of alpha fetal protein, human chorionic gonadotrophin, and unconjugated estriol.

***Cell-free* DNA.** As is the case during pregnancy where fragments of fetal DNA circulate in maternal serum, degraded DNA fragments are exuded by cells, including cancer cells, undergoing apoptosis or necrosis. There is active research into using next-generation DNA sequencing technology to sensitively detect cancer-associated mutations in cell-free DNA, obtained from serum, urine, or stool. Dubbed a "liquid biopsy," the concept is that activating proto-oncogene or inactivating tumor suppressor gene mutations may be detectable in cell-free DNA at the earliest stages of cancer. An emerging challenge is to identify the tissue of origin of cell-free tumor DNA, based on tissue-specific epigenetic marks that may influence how the circulating fragments are partially digested.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=283#h5p-5

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=283#h5p-6

SLO3 Be prepared to explain possible outcomes of genetic test results to patients and their families

Variants of uncertain significance.

How does a physician determine whether a genetic variant is significant or disease-causing? Variants that clearly disrupt protein-coding—such as nonsense mutations that introduce premature stop codons or mutations that alter exon-intron splicing—are strong candidates for being deleterious. In contrast, missense mutations (which change one amino acid to another) are more difficult to interpret. Their potential impact can be estimated based on whether the affected residue lies within a functionally important domain, is evolutionarily conserved across species, or involves a nonconservative amino acid substitution (e.g., switching from acidic to basic). Numerous computational tools and meta-predictors help evaluate such variants, but their accuracy is limited.

Beyond functional predictions, population frequency data and

prior clinical observations are also critical. A variant that is rare in the general population but recurrent among affected individuals is more likely to be pathogenic. Family segregation analysis can further support causality: if the variant is present in an affected parent, that supports its relevance; if an unaffected parent carries it, the variant is more likely benign. A de novo variant—not found in either parent—is especially compelling, as new protein-altering mutations are relatively rare and often pathogenic.

Variants of uncertain significance (also referred to as variants of unknown significance—either way abbreviated as "VUS") remain a frequent source of uncertainty for clinicians and patients alike. This challenge is even greater in individuals from populations underrepresented in genetic databases, where distinguishing between pathogenic mutations and benign polymorphisms becomes more difficult, especially in individuals from underrepresented ancestries.

Incidental findings. Genetic testing is increasingly moving toward large gene panels and exome sequencing because many disorders can be caused by a large number of genes. It may be more economical to simply perform an exome analysis rather than to devise and continually revise an ever-growing list of genes responsible for a particular clinical phenotype. Yet, by casting a wider net, the number of identified genetic variants grows larger.

Consider a case where exome sequencing is performed on a child with a clinically undiagnosed disorder of intellectual disability and congenital anomalies. Exome testing returns with a positive result and leads to a diagnosis. However, unexpectedly, exome sequencing identifies a pathogenic mutation in a gene predisposing to cancer, such as a heterozygous mutation in TP53, responsible for Li-Fraumeni syndrome. Should these results be reported? What if the parents expressly said at the time testing was performed that they did not desire to learn about anything else lurking in the genetic data? One could conceivably argue that since there are no clear data that early cancer detection in Li-Fraumeni syndrome improves clinical outcome, there is little harm in not revealing this

information. But what if, instead, there was an incidental finding of a heterozygous mutation in a gene, such as *KCNQ1*, encoding a voltage-gated potassium channel that causes **long QT syndrome**, an inherited disorder predisposing to lethal cardiac arrhythmias? Sudden cardiac death can be avoided by implantation of an automatic defibrillator.

Controversy surrounds the issue of return of incidental genetic findings to patients. One criterion for distinguishing whether a particular incidental finding should or should not be reported is whether it's "actionable" (i.e., something could be done about it, whether that be screening for cancer, implanting an automatic defibrillator, or some other intervention depending upon the disorder). Currently, one professional body, the American College of Medical Genetics, identifies 49 genes, *KCNQ1* and *TP53* included, which if incidentally found to contain a mutation on genetic testing (in particular, exome sequencing) performed for other indications should be returned to the patient (or the parents of a child), regardless of stated desires.

It is therefore important that patients be informed of the possibility that genetic testing could lead to identification of variants of undetermined significance or incidental findings that no one may have been expecting.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=283#h5p-7

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=283#h5p-8

SLO4 Identify forms of inherited and acquired disorders potentially amenable to gene therapy and genome editing technologies

Gene therapy. Gene therapy refers to the genetic manipulation of the somatic genome for the treatment of diseases.

It is important to emphasize that targets chosen for gene therapy are somatic cells, not those contributing to gamete formation. In fact, at least for now, a deliberate effort is made to avoid modification to sperm or eggs, in order to prevent permanent changes to the human genome, thereby forever altering the future of our species.

Recently, troubling news came from China, where doctors employed CRISPR-based genome editing (a method described further below) to purposefully introduce the CCR5-Δ32 mutation into human embryos toward the goal of conferring HIV resistance and increasing intelligence (the latter being another reported association for this variant). Children alive today are said to have resulted from these experiments. Those responsible received worldwide opprobrium. Some of myriad concerns include the possibility that CRISPR will introduce mutations in other genes ("off-target" effects) and that there may be untoward consequences of introducing even genetic variation thought to have favorable effects. Moral and ethical implications of manipulating the human germline are profound.

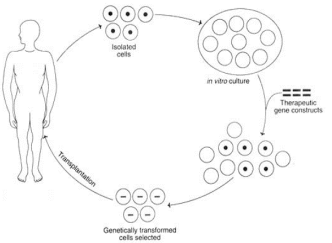
While much of the focus of gene therapy has been directed at the treatment of inherited disorders, ongoing efforts address acquired

disorders, such as cancer or HIV infection. Gene therapy is a field very much in flux, as new gene delivery systems and genome-editing technologies continue to evolve. Recent years have seen success and scores of FDA (United States Food and Drug Administration)-approved therapies, but not to be overlooked is the history of hyped claims, a rush to proceed while sidestepping safety concerns, financial conflicts of interest, and bad behavior inside and outside of the laboratory and clinic. For our purposes, treatment of a topic of such wide scope is necessarily brief. We will review some of the general principles guiding the selection of diseases amenable to gene therapy and methods for gene delivery and genome modification.

Treatment of inherited disorders. A first consideration is what genetic disorders might be appropriate for gene therapy. Autosomal recessive and sex-linked recessive disorders, in general, make for suitable targets for "gene addition" approaches to gene therapy because they usually result from a reduced or absent amount of activity of the gene product. Many metabolic disorders are simply due to an absence of the enzymatic activity required for catalyzing a step in a metabolic pathway. Therefore, supplying a normal copy of the deficient gene could conceivably make up for insufficient activity resulting from reduced amounts of protein encoded by the mutant gene. In contrast, for an autosomal dominant disorder, the mutation is heterozygous, meaning that the cell already possesses a wild type copy of the gene. Consequently, simply introducing a normal copy of the gene, in order to generate greater levels of wild type protein, may not work in the same way that it will for a recessive disorder. Instead, approaches to gene therapy for an autosomal dominant disorder may depend on inactivating expression of the mutant allele.

The next consideration is whether to target cells in vivo, as opposed to isolating cells from a patient, modifying them in vitro, and then returning them to the patient—a so-called "ex vivo" approach to gene therapy. It would be ideal if therapeutic gene delivery could occur entirely in vivo. Unfortunately, to achieve

therapeutic correction in a target tissue, a fairly large number of cells need to be modified. A tractable target for gene therapy is to perform ex vivo modification of hematopoietic stem cells in which a patient's own cells, once therapeutically altered, can be autologously re-engrafted into the patient's bone marrow. Of course, this is primarily of benefit for disorders involving the bone marrow.



The following sections on gene therapy introduce many specific examples of a disease and approaches being developed for treatment. The important take-away is to recognize how understanding of genetics offers new opportunities for designing therapeutic measures. The actual diseases and proposed details for therapy are less important than the concepts and logic that you can anticipate seeing in future treatments.

Gene therapy viral vectors. Most vectors used to introduce a foreign gene, or to modify an endogenous gene, are based on viruses.

Early on, retroviruses were introduced for gene therapy. A disadvantage to retroviral vectors is that there is a risk that the retrovirus will disrupt a gene at the site of insertion into the host genome and thereby lead either to that gene's loss of activity or to inappropriate expression. If it were a tumor suppressor gene or proto-oncogene then one could imagine that this may lead to cancer. In fact, this risk has materialized in several high-profile trials of gene therapy involving treatment of two different sex-linked recessive immunodeficiency disorders, **X-linked severe combined immunodeficiency (X-SCID)**, in which there are few T and NK (natural killer) lymphocytes, and **Wiskott-Aldrich syndrome**, characterized by eczema, thrombocytopenia, immune deficiency, and bloody diarrhea. In those trials, enhancer elements

of the retrovirus activated expression of proto-oncogenes at the site of their integration, curing the immunodeficiency but causing leukemia.

Adeno-associated virus (AAV) is a single-stranded DNA virus that requires adenovirus as a helper virus to productively infect cells. The native AAV has an unusual property in that it tends to integrate into the targeted human cell's genome specifically at a certain locus on chromosome 19. However, AAV modified for use as a vector has lost this site specificity. AAV vectors infect non-dividing cells. At present, AAV has attracted much attention for both ex vivo and particularly in vivo gene therapy because it appears to provide prolonged and moderately high levels of expression of the therapeutic gene. The size limit of the gene that can be packaged is somewhat smaller than that which can be delivered by retroviruses and adenovirus. Its single-stranded genome may also make it valuable in genome editing strategies that make use of targeted homologous recombination. Most current FDA-approved gene therapies utilize AAV vectors.

Leber congenital amaurosis is a genetically heterogeneous congenital form of blindness. One genetic etiology results from homozygous mutations in the gene encoding RPE65, a retinal enzyme contributing to the "photo cycle" by regenerating visual pigment involved in the detection of light by rods and cones. FDA-approved gene therapy (voretigene neparvovec, sold under the brand name Luxturna) consists of subretinal injection of an AAV vector expressing *RPE65* cDNA—a "cDNA" is a compact, intronless version of the gene retrotranscribed from its mRNA. It costs $425,000 per eye. (There is little need for most physicians to remember the complicated brand or generic names for genetic therapies.)

Spinal muscular atrophy is an autosomal recessive degenerative neuromuscular disorder most commonly caused by homozygous mutation of *SMN1*, a gene encoding a transcription factor-associated protein required for motor neuron survival. It can now be treated with intravenous infusion of an AAV vector expressing

SMN1 (onasemnogene abeparvovec-xioi, brand name Zolgensma). Its current cost is $2.1 million. Shortly following its FDA approval in 2019, allegations of data manipulation surfaced, and top drug company executives were fired.

Duchenne muscular dystrophy is a severe X-linked recessive disorder that primarily affects males and involves progressive degeneration of skeletal and often cardiac muscle. The disorder is caused by mutations in the *DMD* gene, encoding dystrophin. Dystrophin is a gigantic gene, spanning over two million base pairs, containing 79 exons, and taking RNA polymerase 16 hours to transcribe. In 2023, the FDA approved an AAV-based gene therapy (delandistrogene moxeparvovec, brand name Elevidys) delivered by intravenous infusion. Because the size of *DMD*, even as an intronless cDNA, vastly exceeds the vector capacity of AAV, gene therapy makes use of an engineered "mini" version of dystrophin, about one-third the size of the normal protein while still retaining function. Approval was granted despite a phase 3 clinical trial that did not demonstrate statistically significant improvement in motor function one year after treatment. The therapy is administered as a one-time dose intended to provide lifelong benefit, with a cost of $3.2 million per dosage. By 2025, nearly 1,000 patients had received the treatment. However, after two patients died in 2025 of acute liver failure—a known risk of AAV-based gene therapy—the manufacturer temporarily suspended distribution of the drug.

***Oligonucleotide-based therapies targeting* RNA.** RNA-targeted therapeutics use oligonucleotides that bind to specific RNA sequences, but their mechanisms of action and effects differ. Small interfering RNAs (siRNAs) and antisense oligonucleotides (ASOs) promote degradation of target mRNA, reducing the expression of harmful gene products such as mutant proteins or viral transcripts. In contrast, steric-blocking oligonucleotides act without degrading RNA; they interfere with cellular machinery or modulate splicing to restore or alter protein production in genetic disorders. Despite their high specificity due to sequence-based targeting, a major challenge for these therapies remains the efficient delivery of

oligonucleotides into cells. This category of drug is distinct from mRNA-based therapies that are intended to result in translation of an exogenously delivered protein, such as the mRNA-based vaccines developed for COVID-19.

Several such drugs have been FDA-approved, including for **transthyretin amyloidosis**, an autosomal dominant disorder that results from heterozygous mutations in the transthyretin retinol transporter that misfold and form amyloid deposits in heart, brain, and other tissue. Even wild type transthyretin can misfold in people who do not possess mutations in the gene encoding it and cause end organ damage as it accumulates with aging. Currently two siRNA-based therapies target hereditary and wild type forms of the disease, patisiran (Onpatto) and vutrisiran (Amvutta), which use lipid nanoparticles or covalent linkage to a carbohydrate moiety, respectively, to facilitate delivery of the oligonucleotide to the liver, where transthyretin is synthesized before its secretion into the bloodstream. Two additional ASO therapies are inotersen (Tegsedi) and eplontersen (Wainua), which are chemically modified in different ways to promote stability and delivery to the liver.

Two other RNA-based drugs are now in use to treat **hypercholesterolemia**. One (mipomersen (Kynamro)) is an ASO that silences expression of apolipoprotein B, a component of low-density lipoprotein (LDL) cholesterol produced in the liver and thereby directly lowers LDL levels by preventing its formation. The other (inclisiran (Leqvio)) is an siRNA that targets the mRNA encoding proprotein convertase subtilisin-kexin type 9 (PCSK9), an enzyme negatively regulating levels of the LDL receptor. The insight for why reducing levels of PCSK9 should be therapeutic came about from genetic studies showing that people with PCSK9 mutations that reduce its activity have lower LDL levels.

Somewhat similar are therapies employing chemically modified DNA-based oligonucleotides that can re-direct the cellular splicing machinery to skip exons containing deleterious mutations. For example, in **Duchenne muscular dystrophy**, some mutations in the gene encoding dystrophin create frameshifts or nonsense codons.

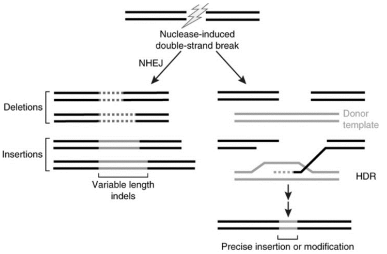
If the exon containing the mutation is deleted from the final transcript, by splicing around it during maturation of the mRNA, even though there may be a large in-frame deletion, the overall protein is still sufficient to function mostly normally, compared to the effect of the mutation which may be to either truncate the protein at the point the mutation occurs or, even worse, cause the entire transcript to decay due to nonsense-mediated decay. There are currently four different FDA-approved oligonucleotide therapies, each targeting mutations in different exons (exons 45, 51, or 53). These therapies will not, however, be effective for patients with mutations elsewhere in the gene.

Oligonucleotide have also been engineered to promote exon inclusion. The gene, *SMN1*, whose mutations are responsible for **spinal muscular atrophy** as noted previously, has a nearly identical and adjacent gene paralog, *SMN2*, resulting from an evolutionary gene duplication event. *SMN2* differs by just a single translationally silent nucleotide substitution that alters its splicing. As a result, *SMN2* is alternately spliced, skipping an internal exon required for full activity, and ordinarily only produces small amounts of intact functional protein, identical to SMN1. The number of *SMN2* copies in the genome varies between zero and eight among different people in the population, due to highly polymorphic copy number variation in this region. It was observed that disease severity for patients with *SMN1* mutations is inversely proportional to the number of copies of *SMN2*, meaning that residual expression of full length SMN2 can compensate for SMN1 mutations. Based on this observation, an FDA-approved therapy (nusinersen (Spinraza)) was designed in which an oligonucleotide targeting an intron blocks the splicing signal that causes the exclusion of the exon in *SMN2*, thereby therapeutically increasing expression of full length *SMN2*, producing a protein identical to SMN1.

Genome editing. The human genome can be site-specifically modified. The general approach is to employ an engineered endonuclease (a DNase that cleaves internal to the linear molecule). The broken DNA molecule can then repair itself, usually by

nonhomologous end-joining (NHEJ). When the broken ends are ligated back to together, a small deletion or insertion of variable length usually is generated. That by itself may be sufficient to, say, inactivate a dominant allele for an inherited disorder.

But, the approach can be taken even further. If another short piece of DNA is provided during the repair process, it can be employed as a substrate for directing the repair of the broken strand, through a process known as homology-directed repair (HDR). By supplying the appropriate template for homology-directed repair, humans can write what they want into the genome.



The breakthrough achieved with CRISPR is that DNA binding specificity is directed by an RNA template that guides the endonuclease to the cut-site where genetic modification is desired. Oligonucleotides can be rapidly, efficiently, and cheaply synthesized. The protein that binds the guide RNA and also cuts the DNA is a modified form of a naturally occurring bacterial protein, Cas9, that normally functions as an RNA-guided endonuclease. It seems that nature had evolved this technology in bacteria first, as a sort of molecular immune system against bacteriophages and DNA-based molecular parasites.

Genome editing technology continues to evolve. In one iteration, "prime editing" allows for the templated changes to be encoded in the guide RNA and requires only a single stranded DNA break, which is more efficient and less prone to introducing additional mutations compared to a double strand break with HDR. Even newer technologies employ modified forms of CRISPR to edit bases directly in DNA or RNA. Targeting the mitochondrial genome has been experimentally demonstrated. Some of the editing machinery can be delivered as mRNA, using the same technologies as involved with COVID-19 vaccination.

The first FDA-approved therapy employing gene editing is

Casgevy (exagamglogene autotemcel), used for the treatment of **sickle cell disease**. The therapy involves harvesting a patient's hematopoietic stem cells, editing them ex vivo to disrupt an erythroid-specific enhancer sequence required for expression of the *BCL11A* gene, and reinfusing the modified cells. *BCL11A* is a transcriptional repressor of fetal hemoglobin (HbF); by silencing its expression in erythroid cells, HbF production is reactivated, which inhibits sickling and reduces disease severity. This one-time, autologous treatment offers a transformative therapeutic approach for a debilitating inherited disorder.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=283#h5p-9

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://uw.pressbooks.pub/fmrgenetics/?p=283#h5p-10

Feedback:

One or more interactive elements has been excluded

from this version of the text. You can view them online here:
https://uw.pressbooks.pub/fmrgenetics/?p=283