Aaron Gardner · Sarah Stauffer
Lindsay Petley-Ragan · Philip Wismer
Dewi Ayu Kencana Ungu

Labster Virtual Lab Experiments

Genetics of Human Diseases

Recommended in Germany

Springer Spektrum

Labster Virtual Lab Experiments: Genetics of Human Diseases

Aaron Gardner · Sarah Stauffer · Lindsay Petley-Ragan · Philip Wismer · Dewi Ayu Kencana Ungu

Labster Virtual Lab Experiments: Genetics of Human Diseases

Springer Spektrum

Labster

Aaron Gardner
Labster Group ApS
København K, Denmark

Philip Wismer
Labster Group ApS
København K, Denmark

Sarah Stauffer
Labster Group ApS
København K, Denmark

Dewi Ayu Kencana Ungu
Labster Group ApS
København K, Denmark

Lindsay Petley-Ragan
Labster Group ApS
København K, Denmark

ISBN 978-3-662-58743-0 ISBN 978-3-662-58744-7 (eBook)
https://doi.org/10.1007/978-3-662-58744-7

Springer Spektrum

Illustrations by: Silvia Tjong
Editor-in-Chief: Stephanie Preuß
Cover figure: © Labster ApS

This Springer Spektrum imprint is published by the registered company Springer-Verlag GmbH, DE part of Springer Nature.
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface

Welcome to the "Genetics of Human Diseases" textbook, which is part of the "Labster Virtual Lab Experiments" series.

This book will help you to learn the key concepts underlying genetic disease in humans while applying your newly acquired knowledge in a virtual lab environment. In each chapter you will be introduced to one of four virtual lab simulations and the true-to-life missions that you will encounter when playing the simulations. Study the theory section presented in each of the chapters closely and you will be fully prepared to master the challenging tasks in the virtual lab!

Finally, you will find learning objectives and techniques covered by the virtual lab simulation at the end of each chapter to easily align its content with your exam preparation.

Other topics in the "Labster Virtual Lab Experiments" book series include "Basic Genetics", "Basic Biology" and "Basic Biochemistry" (see section "Further Volumes in the Labster Virtual Lab Experiments Series" at the end of the book).

About Labster

Labster is a company dedicated to developing virtual lab simulations that are designed to stimulate students' natural curiosity and highlight the connection between science and the real world. These simulations have been shown to improve the achievement of learning outcomes among students, by making the learning experience more immersive and engaging. The content of this book was created by the Labster team members Dr. Aaron Gardner, Dr. Sarah Stauffer, Dr. Lindsay Petley-Ragan, Philip Wismer, Dewi Ayu Kencana Ungu, and Silvia Tjong.

About this Textbook

Our genome is an amazing thing. Consistent enough between us that we're all human, but containing so much variation that we're all individual. Some of the variation in our genetic code has no (known) effect, some describes the many ways we can differ for example the color of our hair or eyes, and yet more can predispose us to certain diseases.

In this volume we'll focus on the latter aspect and cover three distinct genetic diseases or disorders, describing their effect and the vital techniques used to detect and diagnose them. But we'll end on a more positive note, describing one of the ways scientists hope to use the power of genetics to fix underlying health issues. Throughout you'll learn about cutting-edge techniques that scientists and healthcare practitioners use on a daily basis, and leave with a new understanding of the Genetics of Human Diseases.

Monogenic Disorders
In the genetic age monogenic disorders could be thought of as the simplest type of genetic disorder. Arising from mutations in a single gene, it is often easy to detect and trace these disorders through a population, and predict the risk of occurrence in future generations. However, while relatively easy to detect they are often far from easy to treat, as you'll see when we investigate cystic fibrosis in the Monogenic Disorders simulation. Therefore, a good understanding of the molecular techniques used to diagnose cystic fibrosis, and how to calculate risk is vital in limiting the impact of the disease on families.

Cytogenetics
Molecular genetics isn't the only tool that scientists and healthcare practitioners have. Analyzing the number and structure of our chromosomes can be hugely informative, and is routinely used to detect certain abnormalities as part of prenatal testing. In the Cytogenetics simulation you'll use a genome-wide high-resolution technique (array comparative genomic hybridization) to check for chromosomal copy number changes. Will you be able to interpret the results and inform a mother-to-be about the health of her fetus?

Medical Genetics
We covered a "simple" monogenic disorder in the first chapter, but can we use these techniques in less clear-cut cases? In the Medical Genetics simulation you'll investigate the case of a potential familial cancer. By using the techniques you've

developed in the previous two chapters you will identify the gene of interest, predict its inheritance pattern and ultimately help a family understand their risk of disease.

Viral Gene Therapy
Finally, we will flip things around and show how scientists are attempting to fix disease-causing regions of our DNA. Viruses are normally something we want to avoid, but by manipulating their genetic code we can make them work for, rather than against us. In the Viral Gene Therapy simulation you'll work on a promising technique to treat a major cause of heart disease by inserting a functional human gene, into living heart tissue. What effect will your experiment have and will you be able to find a cure for this genetic condition?

How to Access the Virtual Lab Simulations?

You can access the four virtual lab simulations included in this book at www.labster.com/springer.

If you have purchased a printed copy of this textbook, you will find a voucher code **on the last page**, which gives you free access to the four simulations for the duration of one semester (six months).

If you are using the e-book version, you can sign up and buy access to the simulations through the same link.

Please be aware that the six month period starts once you sign in for the first time.

If you have any questions about the use of the voucher, you can contact us at customerservice@springer.com.

Contents

Monogenic Disorders

1

A. Gardner et al., *Labster Virtual Lab Experiments: Genetics of Human Diseases*,
https://doi.org/10.1007/978-3-662-58744-7_1

1.1 Monogenic Disorders Simulation

Join a team of doctors and learn about the basics of monogenic disorders and how to perform a genetic risk assessment based on family history and genetic tests. As the name suggests monogenic disorders arise due to mutations or changes in a single gene. Cystic fibrosis (CF) is the most common life-limiting genetic disorder in those of North European descent, estimated to affect 70,000 people worldwide. CF is a monogenic disorder caused by one of over 2000 disease-causing mutations in the *cystic fibrosis transmembrane conductance regulator* (*CFTR*) gene. Importantly, as a recessive disorder, both copies of the gene must carry mutations for the disease to manifest.

Learn About Cystic Fibrosis and how It Is Detected Genetically
Predominantly present within the cell membrane of epithelial cells, the CFTR protein is a chloride channel that regulates epithelial surface fluid secretion in the respiratory and gastrointestinal systems as well as numerous other epithelial tissues. In CF mutations in the *CFTR* gene lead to a reduction, or complete loss, of CFTR function. These mutations can be detected by several techniques, including polymerase chain reaction (PCR) to check for changes in CFTR protein length (Fig. 1.1).

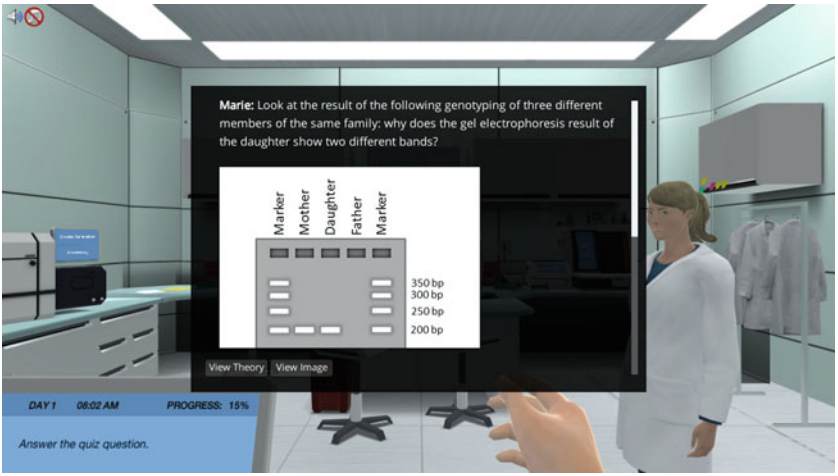


Fig. 1.1 A lab assistant explains how to use PCR and gel electrophoresis to detect mutations in *CFTR* in the Monogenic Disorders simulation

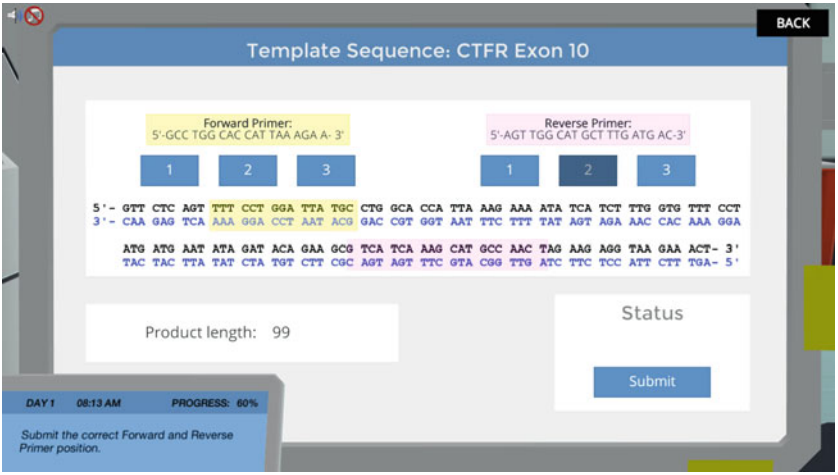


Fig. 1.2 Learn how to design forward and reverse primers to study the *CFTR* gene in the Monogenic Disorders simulation

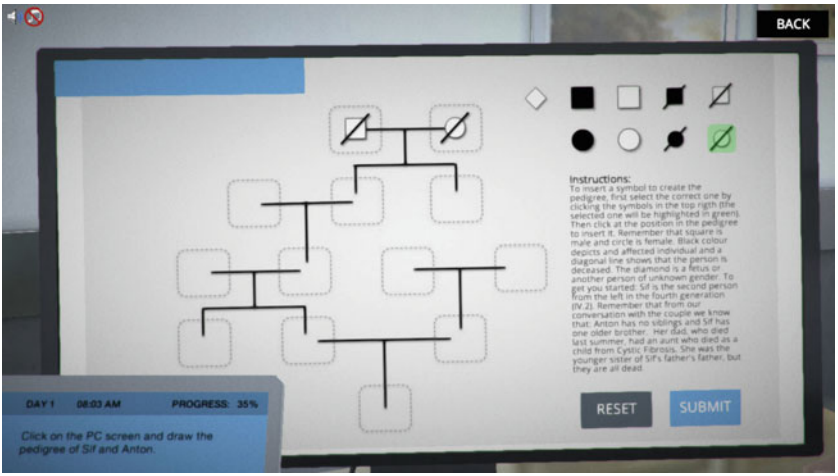


Fig. 1.3 Construct a pedigree tree and calculate the risk that the unborn infant will inherit cystic fibrosis from its parents in the Monogenic Disorders simulation

Design Custom Primers to Investigate *CFTR*
Although CF is a monogenic disorder, there are actually many mutations in the *CFTR* gene that can give rise to the disease, often influencing disease severity. By using PCR and gel-electrophoresis with specially designed primers it is often possible to identify which mutation is present in the *CFTR* gene (Fig. 1.2).

Generate a Pedigree and Predict the Risk of Inheritance
As CF is a genetic disorder it is possible to track the mutation through generations by generating a pedigree, based on genetic analysis of related individuals and the family's medical history (Fig. 1.3). With sufficient knowledge you can even predict the potential risk of future generations inheriting the disease, or carrying a single copy of the gene!

Will you be able to pull together all your knowledge of CF and various key molecular techniques to inform a young couple about the risk of their future children inheriting the disease?

1.2 Monogenic Disorders Theory Content

In this lab, you will learn about the monogenic disorder cystic fibrosis (CF) and how it is inherited from one generation to the next. To understand this you will cover the basics of genetics, key facets of CF and important molecular techniques used to study the disease. The theory content below covers everything you'll need to be successful in the Monogenic Disorders simulation.

Monogenic Disorders

Monogenic disorders are classified as genetic disorders, and their distinctive characteristic is that only one gene is involved in the development of the disease. Examples of monogenic disorders include CF or hemophilia.

Cystic Fibrosis

CF is a type of monogenic disorder caused by mutations in the *cystic fibrosis transmembrane conductance regulator* (*CFTR*) gene, an ion transporter found mainly in epithelial cells. CF is caused by mutations in the *CFTR* gene which lead to a reduction, or complete loss, of CFTR protein function. Due to impaired ion-transport,

the surface liquid which lines many epithelial tissues becomes more viscous, often with an associated build-up of mucus.

Mucus secretion is a normal part of airway physiology, as it helps protect the cells of the airway from mechanical damage and infection. In health this mucus is swept out of the airways by the action of the cilia which line the airway. However, in CF an excess of this mucus can build up as it is not cleared correctly due to a failure of cilial action caused by changes in the viscosity of the surface liquid.

While a monogenic disorder, a major feature of the disease is its heterogeneity, meaning that *CFTR* genes from many different patients with CF contain different types of mutations, with over 2000 currently recorded. However, the most common mutation in the *CFTR* gene (affecting ~60–70% of the population) is a deletion of three nucleotides, resulting in a loss of the amino acid phenylalanine (F) at position 508 in the protein. It is therefore commonly called *F508del* (also ΔF508).

Historically, the major cause of mortality associated with CF was the patient's inability to absorb nutrients from food, due to a lack of digestive enzymes in the intestine. This is caused by the build-up of mucus, which prevents the enzymes from passing into the intestine from the pancreas. As such individuals would usually die early in childhood. With a better understanding of the disease and nutritional supplementation, digestive issues associated with the disease are now typically well managed in people with CF.

Therefore, lung disease associated with CF is now the major issue facing individuals today. The airways become clogged with a thick, viscous mucus limiting the individual's ability to breathe. This mucus also represents an attractive breeding ground for bacteria leading to chronic bacterial infections and associated inflammation. This in turn leads to remodeling of the airways resulting in the loss of lung function. This loss and the ongoing bacterial infections are often the major cause of mortality in people with CF these days.

In the Monogenic Diseases simulation we focus on CF, therefore all of the content below is discussed in relation to CF. For a more general description of genetics and associated simulations please see our other publication "Labster Virtual Lab Experiments: Basic Genetics" (Stauffer S et al. [2018] Labster Virtual Lab Experiments: Basic Genetics, 1st edn. Springer, Heidelberg, ISBN 978-3-662-57999-2).

Gene

A gene is a section of DNA and is the basic unit of heredity in living organisms. In CF we are interested in the *CFTR* gene.

Genes encode information to construct functional proteins, i.e. the *CFTR* gene encodes for the CFTR protein. This code is "read" through the process of transcription, which creates a complementary mRNA strand of the gene. The mRNA strand is then "translated" in the cell's ribosomes where this code is used to build a chain of amino acids, ultimately constructing a protein.

Humans are thought to have between 19,000–20,000 genes in total, and together all of these genes are referred to as the genome.

Locus

We use the term locus to refer to the location or position of a gene (or significant sequence) on a chromosome. *CFTR* is located on the long arm of chromosome seven, specifically at position q31.2.

Allele

An allele is one of several alternative forms of the same gene, where the DNA sequence differs slightly. Sometimes, different alleles can result in different observable phenotypic traits, such as changes in pigmentation. However, most genetic variation results in little or no observable variation. Rarely, changes in the DNA sequence will result in alleles which impact on an individual's health, for example CF is caused by changes in the DNA sequence of the *CFTR* gene leading to a reduction or loss of CFTR protein function.

Most multicellular organisms are diploid. This means that they possess two alleles for each gene, one inherited from each parent. If the two alleles are the same, the organism is said to be homozygous with respect to that gene. If they are different we call the organism heterozygous with respect to that gene (see zygosity).

Phenotype

Phenotype refers to an organism's set of observable traits. The phenotype results from the expression of an organism's genotype, inherited epigenetic factors, and environmental conditions.

Traits can be observable (e.g. eye color) or physiological (e.g. lactose intolerance), they can also be more severe and lead to diseases such as CF.

Genotype

The genotype consists of an organism's DNA segments or genes, which make up its phenotype. In the case of CF, the *CFTR* alleles an individual carries will determine whether they develop the disease or not.

Mutation

From a genetic point of view the term mutation can refer to any permanent change in a DNA sequence regardless of its impact on the genome or the individual. However, in clinical terms a disease-causing mutation refers to a permanent change in a DNA sequence which leads to the development of a disease or disorder. Mutations can vary from a single base changes, or point mutation, up to chromosomal mutations which include the duplication or deletion of entire chromosomes, or chromosome regions, which are covered in Chap. 2 of this book.

Central Dogma of Molecular Biology

During gene expression DNA is transcribed into mRNA which is then translated into an amino acid chain which goes on to form a mature protein. During translation the nucleotides which make up the mRNA strand are 'read' in batches of three, known as codons, which encode for specific amino acids. Some specific codons are termed stop codons, rather than codon for an amino acid these stop codons instead terminate translation.

Point Mutations

Point mutations describe a genetic mutation where a single nucleotide base is changed, inserted or deleted in a DNA sequence. This change in DNA leads to a change in the mRNA which can in turn lead to changes in the amino acid chain produced. There are three common outcomes to point mutations; frameshifts, synonymous mutations and non-synonymous mutations:

- **Frameshift** – Insertions and deletions lead to frameshift mutations, where the normal reading frame for the mRNA is shifted, the affected codon and all sub-

sequent codons will therefore be 'read' differently. For example the sequence:

'AGU CAG AGA'

encodes for the amino acids serine-glutamine-arginine. An insertion of an '*A*' nucleotide between the marked 'G' and 'U' nucleotides of the first codon would therefore result in the following sequence:

'AG*A* UCA GAG A..'

which encodes for arginine-serine-glutamic acid. Frameshift mutations commonly result in significantly altered proteins often with dramatically different lengths due to the change in reading-frame altering stop codon location.

- **Synonymous** – The point mutation does not change the amino acid sequence of the protein encoded for by the gene, which typically does not affect the length of the proteins. For example, the codons 'UGU' and 'UGC' both encode for the amino acid cysteine, making the U > C point mutation synonymous. Importantly, diseases can still arise from these mutations as changes in the DNA sequence can have other impacts outside of direct protein coding.
- **Non-synonymous** – Conversely, in this instance the point mutation does result in a change to the amino acid sequence of the protein. Non-synonymous mutations can be further divided into missense and nonsense mutations.
  - **Missense** – Missense mutations arise when the point mutation results in a different amino acid being present, **but not** a stop codon. This amino acid change can dramatically alter the function of the protein but does not usually alter its length. For example the codon 'UGU' encodes for a cysteine amino acid, whereas the codon 'UGG' encodes for tryptophan. In this case the U > G point mutation produces a non-synonymous missense mutation.
  - **Nonsense** – Nonsense mutations arise when the point mutation results in a stop codon being present, which results in a truncated and usually non-functional protein. For example the a 'UGU' to 'UGA' point mutation results in the cysteine amino acid changing into a stop codon. In this case the U > A point mutation produces a non-synonymous nonsense mutation.

Zygosity

So-called diploid organisms such as mammals carry two alleles – one gene variation inherited from each parent. Zygosity indicates the similarity of the two alleles

for one particular trait, i.e. whether the DNA sequence at a particular locus are the same or different.

Three different genotypes at a single locus are found in diploid organisms:

- **Heterozygous** – refers to the state of having two different alleles for a particular trait.
- **Homozygous** – refers to the state of having two identical alleles for a particular trait.
- **Hemizygous** – refers to the state of having only one allele for a particular trait. Males are said to be hemizygous, in that they only have one allele for any X-linked characteristic (see X-linked mode of inheritance).

Dominant and Recessive Alleles

When an allele is **dominant**, one copy of the gene is enough to produce the respective phenotype. In order to exhibit a **recessive** phenotype however, an individual must have two copies of the gene.

Dominant Trait

A dominant trait is defined as a trait which confers the same phenotype regardless of whether a diploid individual has one or two copies of the respective allele.

Recessive Trait

A recessive trait is a trait that is only expressed if the diploid organism carries two identical alleles. A recessive trait is "latent" or non-expressed when the individual is heterozygous, that is, if it carries one recessive allele and one dominant allele. CF is a recessive disease – both of an individual's *CFTR* alleles must carry a disease-causing mutation for the disease to manifest. Those who carry a single copy are not affected, but are often referred to as carriers.

Pedigree Analysis

A pedigree tree is a diagram depicting members of a family, their inter-relations, and their disease/phenotypic status. It provides an overview of the inheritance pattern and frequency for a specific trait or disease.

Each family member is represented by a symbol: circles for females and squares for males. Filled symbols represent affected individuals, and empty ones represent

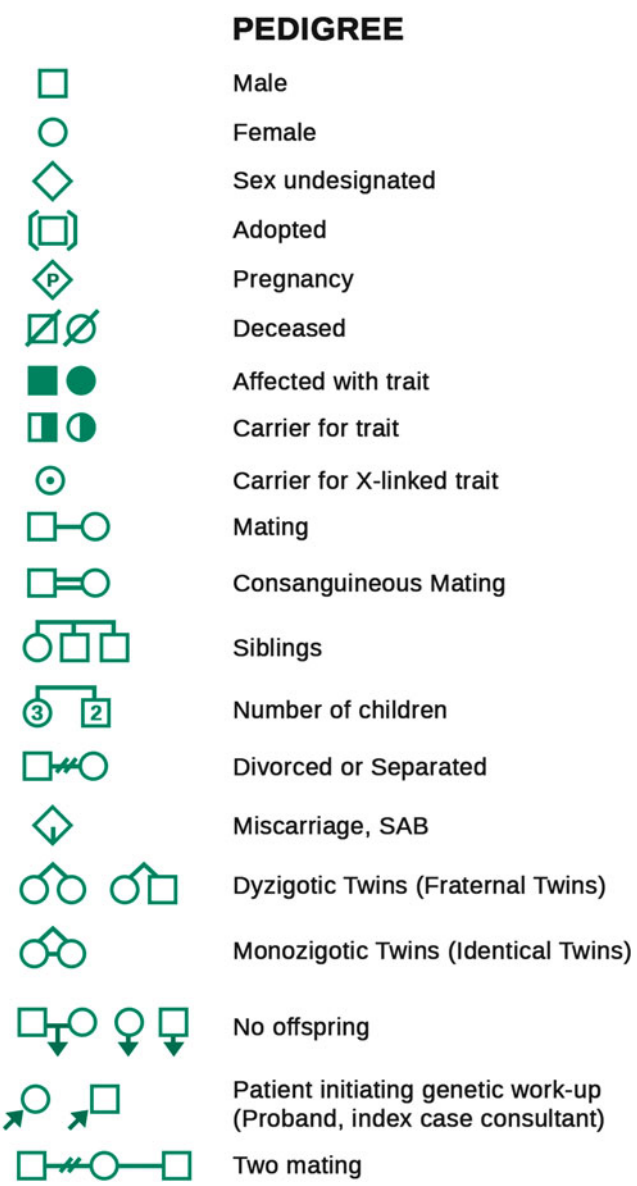


Fig. 1.4 **Explanation of pedigree symbols.** Numerous symbols are used when constructing pedigrees. For diseases such as CF, carrier or affected status can be used to predict risk in future generations

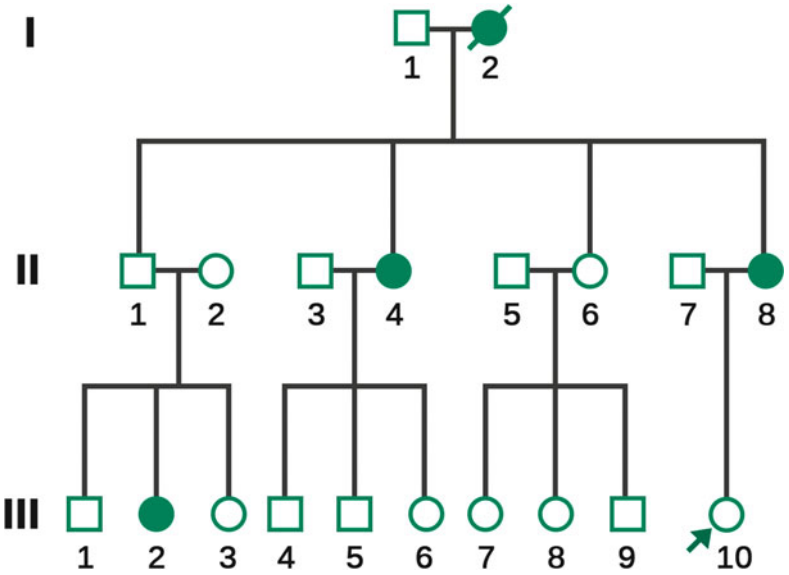


Fig. 1.5 A pedigree showing a phenotype that is passed on over three generations. In this example image we can see that individuals 2 of generation I, 4 and 8 of generation II and 2 of generation III are all affected by the phenotype being investigated. We can also see that individual 2 of generation I is now deceased. The arrow pointing at individual 10 of generation III identifies that they are the proband, or the individual for whom the genetic workup was initiated

un-affected individuals, although there are many other types of symbols used as well (Fig. 1.4).

When constructing a pedigree, reproduction is illustrated by connecting horizontal lines between individuals, with the resulting offspring depicted below. Each generation is assigned a Roman numeral (e.g. I, II, III) while individuals of a generation are sometimes assigned Arabic numerals (e.g. 1, 2, 3) as seen in Fig. 1.5.

Autosomal Mode of Inheritance

The autosomal mode of inheritance refers to the inheritance of genes that are present on autosomes. An autosome is any chromosome not considered as a sex chromosome, or not involved in sex determination.

For instance, a human somatic cell will normally contain 23 pairs of chromosomes (total = 46 chromosomes). 22 of these pairs will be autosomes, and only one of them will be a pair of allosomes (the X and Y chromosomes), also known as the sex chromosomes.

Autosomal Recessive

An autosomal recessive disease, such as CF, is a recessive disease linked to an autosome. The main characteristics observed for these kinds of diseases are:

- The disease may jump generations
- Affected individuals may be inbred
- Males and females are equally affected

Autosomal Dominant

An autosomal dominant disease is a dominant disease linked to an autosome. The main characteristics observed for these kinds of diseases are:

- All generations have affected individuals
- An affected individual has at least one affected parent
- Males and females are equally affected

X-linked Mode of Inheritance

The X-linked mode of inheritance refers to the inheritance of genes that are present on X-, but not on Y-chromosomes. Some examples of X-linked traits/conditions include color blindness, hemophilia and muscular dystrophy.

X-linked Recessive

An X-linked recessive disease is a recessive disease linked to the X chromosome. The main characteristics observed for these kinds of diseases are:

- The disease may jump generations
- Males are more frequently affected

X-linked Dominant

An X-linked dominant disease is a dominant disease linked to the X chromosome. The main characteristics observed for these kinds of diseases are:

- All generations have affected individuals
- Affected males pass the disease on to all their daughters
- Males are always affected

Punnett Square

A Punnett square is a visual representation of a cross. The genotypes of the parents are denoted along the top and the side of the grid. The possible genotypes of the offspring are obtained by combining the different alleles in the grid.

The Punnett square in Fig. 1.6 shows an example of a cross between a heterozygous father and a homozygous dominant mother.

Calculating the Risk of Inheriting a Genetic Disease

In order to assess the risk of a newborn suffering a certain genetic disease, first we need to know the mode of inheritance of the disease. Once we know that, we can assess the carrier risk of the parents. To do that, we will start with the affected individual(s) in the family, and we will calculate the risk of the intermediate family

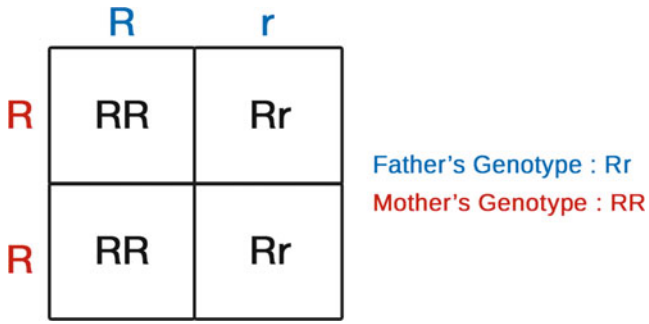


Fig. 1.6 **An example Punnett square.** The Punnett square shows an example of a cross between a heterozygous father (R and r) and a homozygous dominant mother (R and R). Typically an uppercase letter is used to denote a dominant allele, and a lowercase letter a recessive allele. As offspring will receive one allele from the mother and one from the father (shown around the edge of the square) it is possible to construct a diagram showing all possible variants (shown in the center of the square). In this case, there will be a 1:1 ratio of homozygous and heterozygous offspring

members until we reach the parents. We can use a Punnett square to calculate the probabilities of carrying the different alleles, using the mode of inheritance to rule out the options that are not compatible with the disease under study. To calculate the risk, you need to know a bit of probability!

Risk Calculation of Autosomal Dominant Inheritance

In diseases following autosomal dominant inheritance, the disease is manifesting in individuals carrying only one allele with the disease-causing mutation. These patients are heterozygous for the mutation. When one of the parents is affected by the disease each of their offspring have a 50% risk of inheriting the mutation and the disease (Fig. 1.7). An exception to this rule is in the cases where the disease displays reduced penetrance: this means that individuals carrying the disease-causing mutation do not necessarily develop the disease symptoms, however, they may still pass on the mutation to their offspring, and they may become affected.

Risk Calculation of Autosomal Recessive Inheritance

In recessive disorders the symptoms are only seen in a person if both alleles of the gene carries mutations; or, in other words, if the person carries no normal, functional allele of the gene. This means that heterozygous individuals are healthy, but if both parents are heterozygous carriers of the disease there is a risk that their offspring receive mutated alleles from both and thus become homozygous and affected. When assessing the risk that a child or fetus will be affected by an

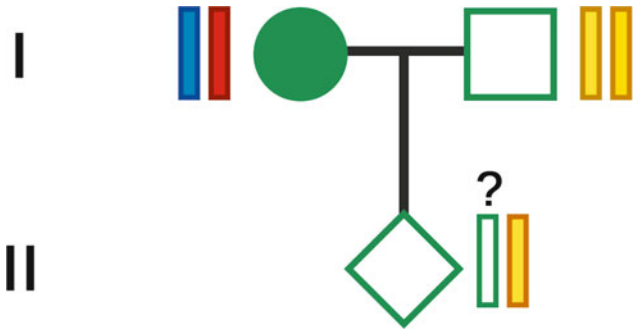


Fig. 1.7 A pedigree showing autosomal dominant inheritance. In this example the mother carries both a dominant and recessive allele, and is therefore affected. Whereas the father carries two recessive alleles and is unaffected. The child will receive one recessive allele from the father, but could receive either a recessive or dominant allele from the mother. The child therefore has a 50% risk of being affected

autosomal recessive genetic disease we therefore have to start by calculating the risk that each of the parents are healthy carriers. In order to do so we look at the pedigree (Fig. 1.8): do we know of any ancestors who were affected with the disease? If not, we need to know the general carrier frequency in the population, but if any ancestors were affected the risk calculation will start with them. If the parent of a person is affected the carrier risk of all offspring is 100%, as the homozygous parent in all cases will pass on a mutated allele of the gene. This of course provided that the other parent carries two normal alleles.

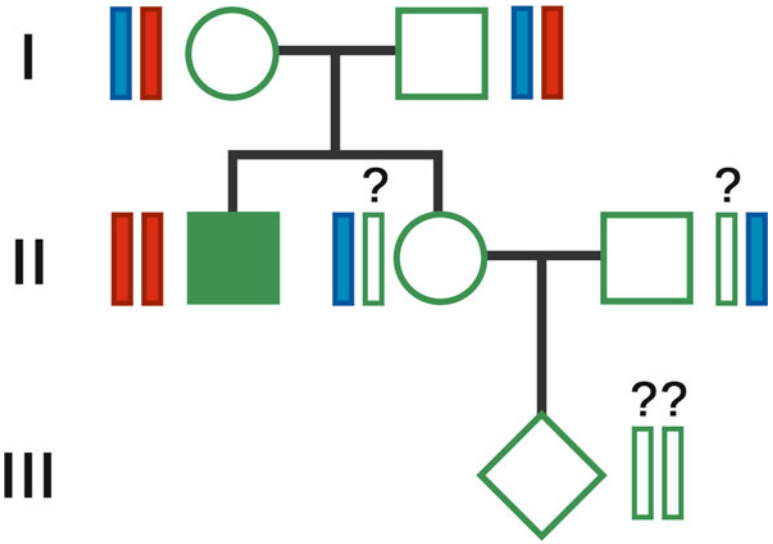


Fig. 1.8 A pedigree showing autosomal recessive inheritance. Working out carrier status for recessive disorders can be difficult. In this example only the son in generation II is affected. We therefore know that both his unaffected parents in generation I must both be carriers in order for the son to receive two recessive disease alleles. As the daughter and her partner are unaffected we know that each must therefore carry a non-disease allele, but the status of the other allele remains unknown. In the case of the daughter we can calculate her exact risk of carrying a disease allele. As she could inherit either type of allele from either parent there are four different outcomes; affected, carrier from mother, carrier from father, not a carrier. We can discount the first option, and from the remaining three outcomes two involve her carrying a risk allele, she therefore has a 2/3 chance of being a carrier. For her partner as we don't know his pedigree we would assign what is known as the population risk. Therefore if the daughter and her partner wish to have a child its risk of being affected could be calculated as $(1/2 \cdot 2/3) \cdot (1/2 \cdot \text{population risk})$

If a sibling is affected, for example the son in generation II of Fig. 1.8 above, the calculation of the carrier risk in all unaffected siblings goes as follows: if one sibling is affected we know that both parents most likely are heterozygous carriers (if they are healthy), meaning that they both may pass on the mutation. The combinations of alleles that can be found in their offspring are heterozygous with a disease allele from the mother, heterozygous with a disease allele from the father, homozygous with two normal alleles or homozygous with two mutated alleles. However, if the sibling in question is healthy we can rule out the fourth combination, homozygous with two mutated alleles, as this would cause the individual to be affected (which he/she is not). This leaves three possible combinations of alleles in question for this person, and in two of these the individual carries one mutated allele, i.e. is a heterozygous carrier. In other words, if a sibling is affected the carrier risk of the healthy siblings is 2/3. Once you have assessed the carrier risk of a person you calculate the carrier risk of that person's offspring by dividing the risk by two, reflecting the fact that each time a potential heterozygous person passes on one of that person's alleles to a child there is a 50% chance that it is the normal allele that is passed on. As a rule of thumb, when we know the disease status we assess the risk if healthy individuals are heterozygous carriers; when we don't know whether the child or fetus is affected, we calculate the risk that this individual is homozygous and affected. Once you have assessed the carrier risk of the parents of the child or fetus, you calculate the disease risk by multiplying the risk that the mother passed on a disease-causing allele with the risk that the father did. If both parents are heterozygous this risk is $1/2 \cdot 1/2 = 1/4$.

Risk Calculation for X-linked Dominant Inheritance

In this rare mode of inheritance the disease gene is located on the X chromosome, and both men and women carrying the disease-causing mutation will develop the disease. The risk of the offspring will depend on the sex of the affected parent as well as the sex of the child. An affected woman will have a 50% risk of passing on the mutation and the disease to both daughters and sons (Fig. 1.9).

However, an affected man will pass on the X chromosome with the mutation to all of his daughters but none of his sons, meaning that disease risk in daughters is 100% and in sons 0% (Fig. 1.10).

Risk Calculation of X-linked Recessive Inheritance

In diseases caused by X-linked recessive inheritance we often see affected sons of unaffected mothers, due to the fact that a woman carrying a disease-causing mutation in the heterozygous state is healthy (since she has one normal allele). Whereas her son, if inheriting the mutation from her, will be hemizygous for the

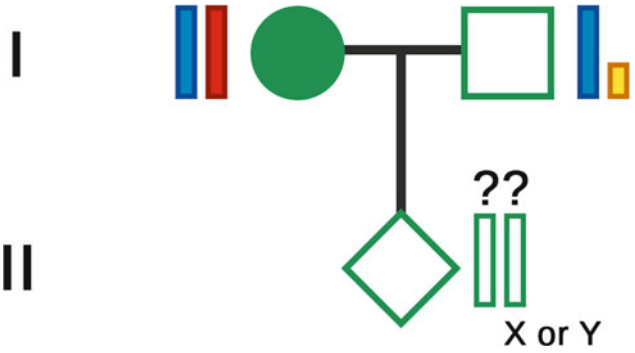


Fig. 1.9 A pedigree showing X-linked dominant inheritance with the mother carrying the disease-causing mutation. In this example the mother carries an X-linked dominant disease allele, and a non-disease allele, and is therefore affected. The father carries a single non-disease allele, we can therefore discount him from any risk analysis relating to the child. Therefore any children, regardless of sex, will be at a 50% risk of inheriting the X-linked disease allele and being affected

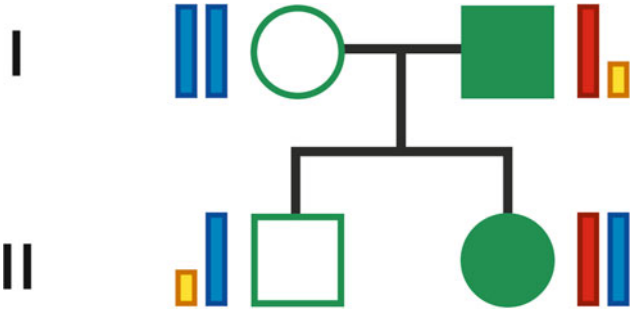


Fig. 1.10 A pedigree showing X-linked dominant inheritance with the father carrying the disease-causing mutation. In this example the father carries an X-linked dominant disease allele and is therefore affected. The mother carries two non-disease alleles. As the father will always pass on his Y-chromosome to any sons, and they will receive their X-chromosome from their mother, male children will have no risk of being affected. However, any daughters will receive the father's X-linked dominant disease allele, along with a non-disease allele from their mother, female children will therefore all be affected

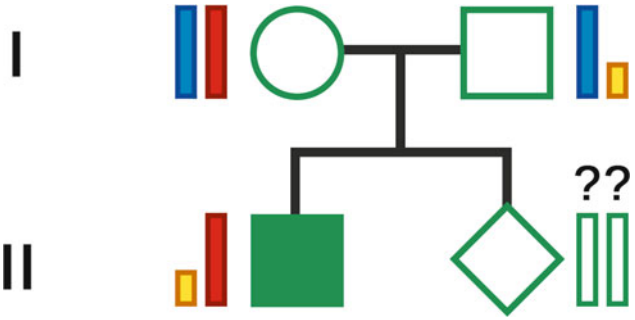


Fig. 1.11 A pedigree showing X-linked recessive inheritance with the sibling carrying the disease-causing mutation. If one sibling is affected, the unaffected mother must be a carrier, thus 50% of sisters will be carriers too and 50% of brothers affected

mutation and affected (as he only has one X chromosome and this carries the mutation). The risk that the son of a heterozygous woman becomes affected is thus 50% (Fig. 1.11).

The daughter of a woman heterozygous for an X-linked recessive disorder will not develop the disorder (unless we have one of those rare cases where her father has the same disorder), but her carrier risk can be calculated by dividing her

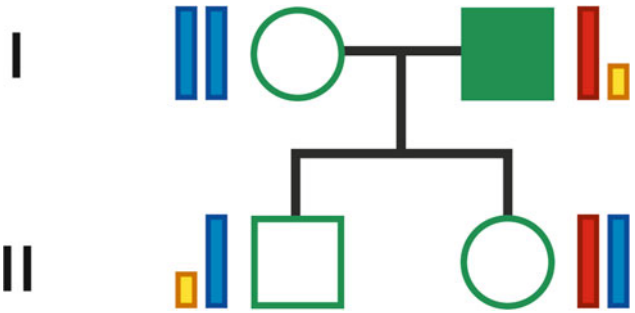


Fig. 1.12 A pedigree showing X-linked recessive inheritance with the father carrying the disease-causing mutation. In this example the father carries an X-linked recessive disease allele, and is therefore affected. The mother carries two non-disease alleles and is so unaffected. As the daughter receives an X chromosome from each parent she will become a carrier, whereas the son will receive a Y-chromosome from his father and a non-disease X-chromosome from his mother

mother's carrier risk by 2. If a man is affected with an X-linked recessive disorder he will pass on the mutation to all of his daughters, but not to any of his sons, as they receive a Y chromosome from him (Fig. 1.12).

Diagnosis of Monogenic Disorders

To diagnose a monogenic disorder, the first step is to draw the family pedigree. This will indicate the most likely mode of inheritance and the risk of being affected. Then, the most feasible (easiest, cheapest, most reliable) way to investigate if a patient carries a certain mutation is to analyze the sequence of the gene involved in the suspected disease (genotyping).

Genotyping

Genotyping is the process of identifying differences in the genotype between one organism and another, or compared to a reference gene sequence. It provides information about the alleles the parental generation has passed on to its offspring.

Typical methods of genotyping include polymerase chain reaction (PCR) followed by gel electrophoresis, or DNA sequencing. Genotyping is particularly important when studying genes and gene variants that are associated with a disease.

DNA Sequencing

Sequencing is a technique used for "reading" the precise order of nucleotides in a DNA fragment. Small DNA fragments, whole genes, or even genomes can be sequenced.

The first DNA sequencing technologies were based on the chain-termination technique developed by Fred Sanger and colleagues in 1977. This method is very similar to PCR, but it involves only one primer, which anneals close to the region of interest at the 3′ end of the DNA template (Fig. 1.13). During the sequencing reaction, a mixture of normal nucleotides (deoxynucleotide [commonly nucleotides] e.g. dATP, dTTP, dCTP, and, dGTP) and "stop"-nucleotides (dideoxynucleotide e.g. ddATP, ddTTP, ddCTP, and ddGTP) are added to the DNA template. The dideoxynucleotides lack the 3′ hydroxyl group found on normal deoxynucleotides,

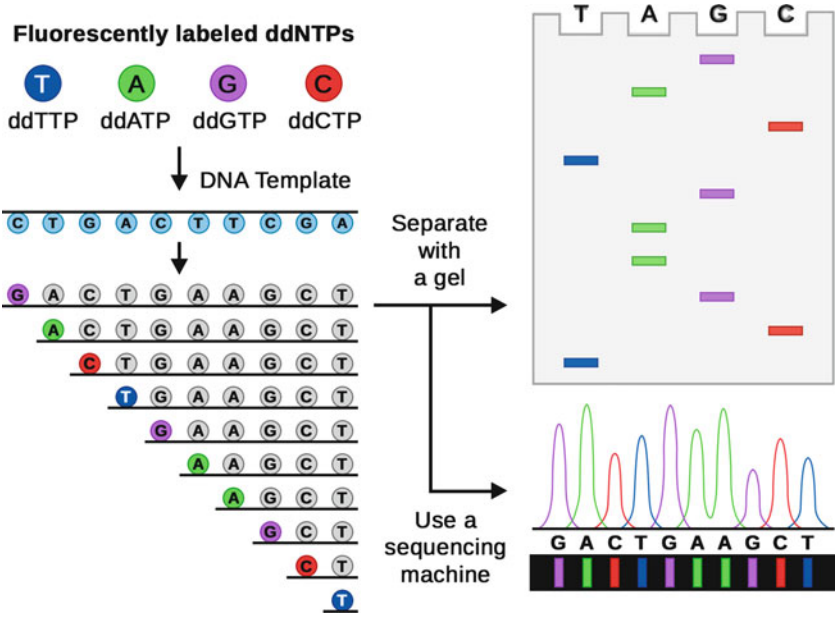


Fig. 1.13 Schematic overview of Sanger (chain-termination) method for DNA sequencing. DNA fragments of all possible lengths are produced during the elongation cycles, each terminated with a colored stop nucleotide. When the fragments are separated according to their size on a gel, the order of colors detected will indicate the precise DNA sequence

this therefore means no further nucleotides can be added to the chain. Each type of stop-nucleotide is labeled with a specific color, allowing them to be identified.

In every cycle the target DNA is replicated by a DNA polymerase until a stop-nucleotide is added, which stops further elongation (chain-termination). After 35 cycles, a large number of fragments of all possible lengths are produced. These fragments are run in a specialized acrylamide gel where their length and the color of the "end-bases" are detected. Because the fragments are separated based on their size in the gel, the labeled nucleotides are detected one by one and thus the precise DNA sequence in the fragment can be reconstructed.

The purpose of sequencing is, for example, to predict the protein sequence of a gene, to compare species on a sequence level (genes or genome), or to search for a mutation.

Importantly, while DNA sequencing of this type is still routinely performed it has in many cases been supplanted by newer technologies which are collectively known as Next Generation Sequencing. While differing in exact methodology these techniques broadly rely on sequencing many shorter DNA fragments simultaneously (sometimes in the range of multiple billions of fragments), and then using complex algorithms to computationally stitch all the fragments back together. These techniques have revolutionized life sciences by allowing multiple entire genomes to be sequenced in days, rather than months or even years.

Polymerase Chain Reaction

PCR is a method used to prepare billions of copies of specific DNA sequences, or in other words, to amplify a DNA sample. It is often necessary to have a larger number of copies of a specific DNA sequence found in a typical sample for further DNA analysis (for example, for DNA fingerprinting or genotyping).

The PCR is highly specific, meaning that it will only produce copies of the desired sequence from the template (sample) DNA. This specificity is ensured by

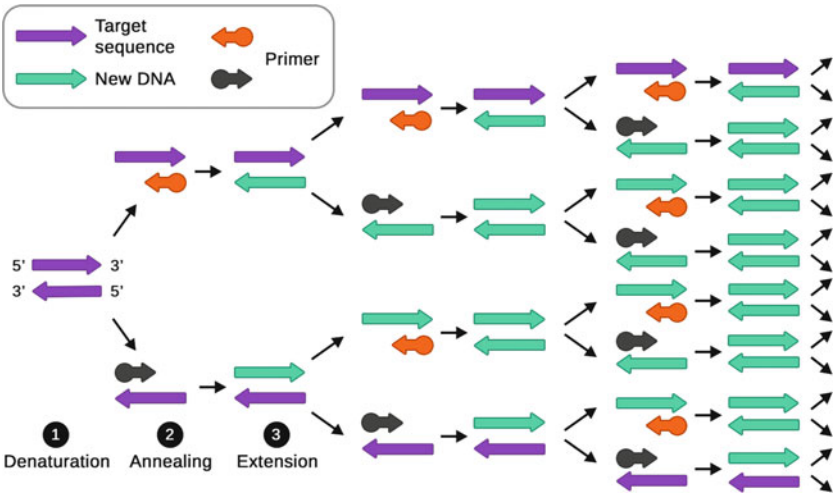


Fig. 1.14 **PCR experiment:** PCR consists of three steps: 1. Denaturation, 2. Annealing, and 3. Extension. The steps are repeated many times (often up to 30 cycles), producing billions of DNA copies of specific regions

the primers, which are designed to be complementary to the template strand and anneal to specific regions on each side of the DNA region of interest (target region).

All cycles are performed without intervention in a PCR machine, also called thermocycler, which can be programmed to change the temperature automatically after each step. By the end of one cycle, parts of the initial DNA strands will have doubled in number (Fig. 1.14). Therefore, after 30 cycles, at least one billion copies of the target sequence will be present in the tube.

For a more thorough review of PCR please see publication "Labster Virtual Lab Experiments: Basic Genetics" (Stauffer S et al. (2018) Labster Virtual Lab Experiments: Basic Genetics, 1st edn. Springer, Heidelberg, ISBN 978-3-662-57999-2).

PCR Experiment

To prepare billions of DNA copies, many repeated cycles of DNA synthesis are performed in one PCR tube. Each cycle includes three distinct steps defined by the temperature (Fig. 1.14):

1. **Denaturation step (95 ºC)**: At this high temperature, the hydrogen bonds holding together the two DNA strands are broken, and the DNA strands fall apart. The single-stranded DNA template is now available for copying.
2. **Annealing step (5–10 ºC below the primer with the lowest melting temperature [T_m], generally 50–65 ºC)**: At the annealing temperature short DNA pieces called primers bind at complementary sites of the template DNA. The primers define the target sequence, which is the specific region of DNA that will be copied. The annealing temperature is calculated from the primer composition. Typically the optimal annealing temperature for each primer is calculated.
3. **Extension step (72 ºC):** At 72 ºC, an enzyme called DNA polymerase is responsible for copying DNA. It recognizes the 3′ end of a primer bound to a template strand and starts copying the template DNA in the 5′ direction.

Primers

Primers are short fragments of DNA or RNA used to start DNA synthesis by a DNA polymerase. They are typically 18–25 nucleotides in length and will bind (anneal) to a complementary region of single-stranded DNA, called the template strand. They mark the point where DNA synthesis will begin. When a primer is bound,

the polymerase can also bind to the DNA at the 3′ end of the primer and then copies the DNA strand in a 3′ → 5′ direction.

By using specific primers PCR can be used to localize mutations within genes. For example, we can use primers which will amplify regions of the *CFTR* gene in order to detect which *CFTR* mutation is present.

PCR Mutation Analysis

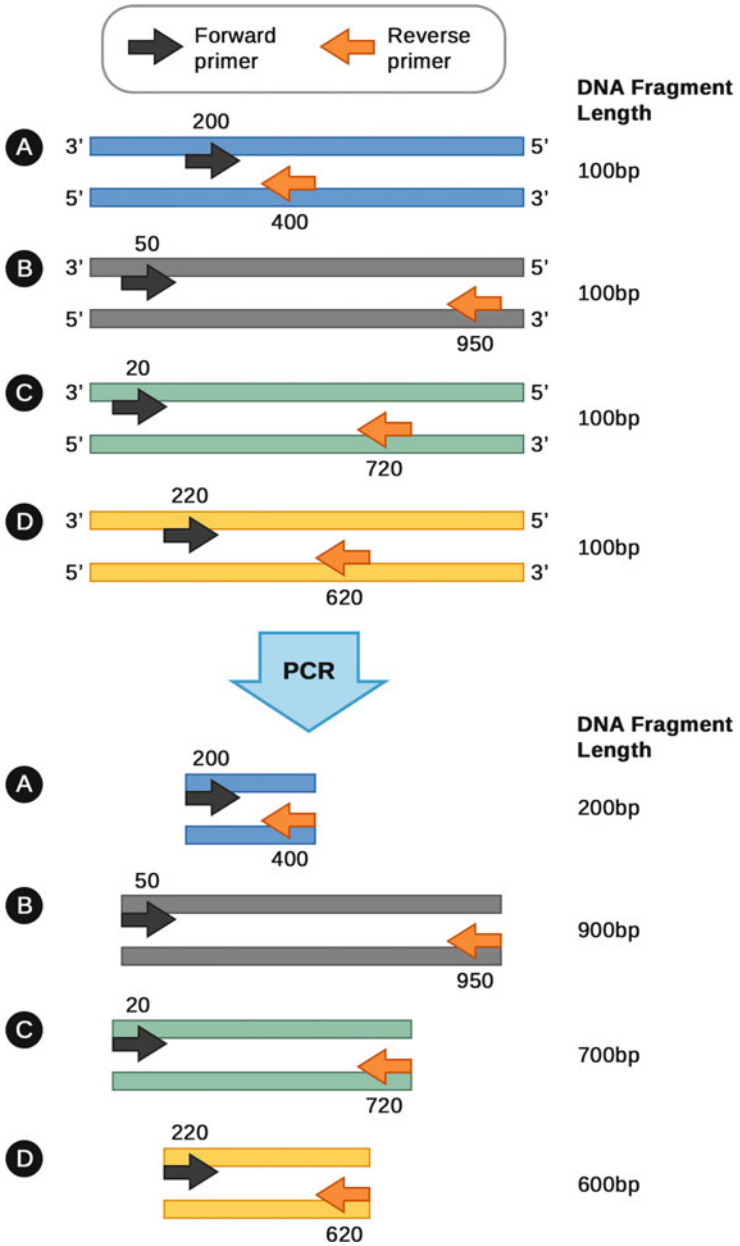
The result of a PCR experiment is billions of copies of the DNA region flanked by the primers. The size of the amplified fragments is determined by the primers (Fig. 1.15). When certain *CFTR* mutations are present the products formed will be of different sizes, allowing researchers to determine which mutations are present. To visualize these fragments other techniques, such as gel electrophoresis, need to be utilized.

Gel Electrophoresis

Gel electrophoresis is a widely used method to separate charged macromolecules (DNA, RNA, or proteins) of different sizes and to estimate their molecular size. It is based on the principle that once an electric field is applied, negatively charged macromolecules are attracted to the positive pole, and separate according to their length in a matrix gel, such as acrylamide. Shorter macromolecules pass more easily through the pores of the matrix and so travel further than longer macromolecules. This technique is often used to separate DNA or RNA molecules, for example, in the case of DNA profiling or to study RNA integrity.

Gel electrophoresis is also commonly used to separate PCR-amplified DNA fragments or to isolate and extract DNA fragments of a specific size.

Fig. 1.15 Examples for different PCR product lengths. Following a PCR experiment, the PCR product A will be 200 bp, PCR product B will be 900 bp long, C will be 700 bp, and D will be 400 bp long. The length of the PCR product is defined by the position at which forward and reverse primer bind to the target DNA sequence


Forward primer
Reverse primer
DNA Fragment Length
A
200
3'
5'
400
5'
3'
100bp
B
50
950
100bp
C
20
720
100bp
D
220
620
100bp
PCR
DNA Fragment Length
200bp
900bp
700bp
600bp

Capillary Electrophoresis

Capillary electrophoresis (CE) is a type of electrophoresis that is performed in submillimeter diameter capillaries and in micro- and nanofluidic channels.

Protein Truncation Test

A protein truncation test is a powerful method to evaluate DNA mutations that result in protein truncation *in vitro*. Such mutations are often the most severe in CF as they result in a complete lack of protein.

So-called nonsense mutations result in a premature stop codon. This means that the resulting protein will be terminated earlier than it is supposed to and can be detected using a protein truncation test. With a protein truncation test, we do not need to use an animal as a model system to synthesize the protein; instead, we have an *in vitro* system where the resulting protein can be synthesized without the need for living cells.

Protein Truncation Experiment

The protein truncation test is composed of four major steps (Fig. 1.16):

- Nucleic acid isolation; either genomic DNA, total RNA, or poly-A RNA.
- Amplification of a specific region of the gene of interest using PCR. In this step, a start codon (ATG) is added to the 5′ end of the amplified DNA, to allow for transcription.
- *In vitro* transcription and translation of the product. If a nonsense mutation is present the protein produced will be truncated compared to the protein produced from the "normal" gene.
- Proteins are detected using gel electrophoresis. A truncated protein will be smaller and will therefore run further into the gel.

Protein Truncation Analysis

Protein truncation tests are often used to analyze disease-related mutations, for example, in CF where certain mutations result in a nonfunctional protein. From protein truncation tests, we can conclude whether there is a mutation in a particular part of the gene that results in protein truncation; however, this does not provide

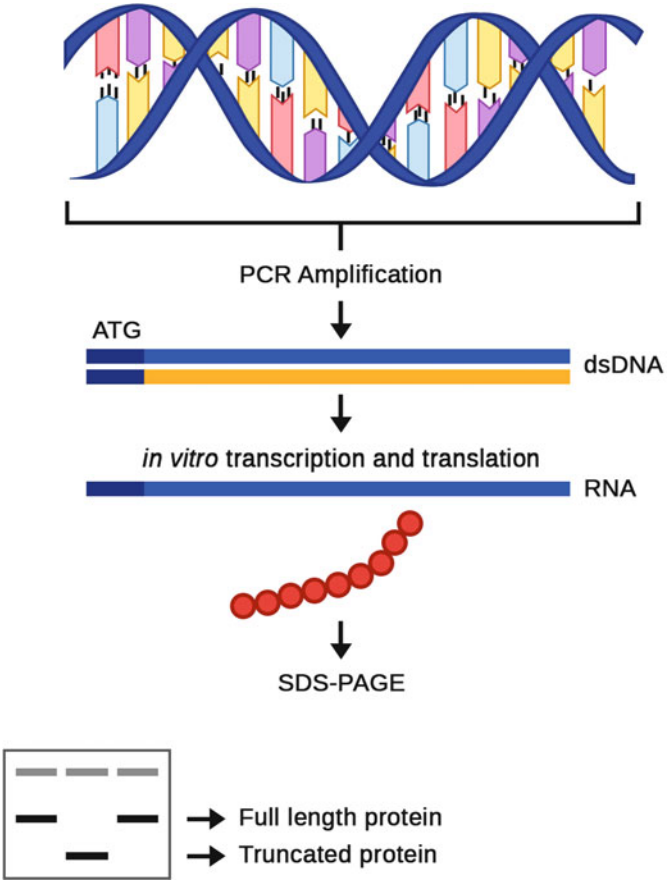


Fig. 1.16 **A schematic diagram showing the different steps of the protein truncation test.** Isolated nucleic acids are tagged with an ATG start codon and then transcribed and translated in vitro. The resulting protein products are then separated by a form of gel electrophoresis known as SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). If a nonsense mutation is present the translation will terminate early, resulting in a truncated protein. This will be detected on the gel as the truncated protein will run further

any information regarding where exactly the mutation is located and what the associated DNA sequence is (for example, it could be a substitution creating a stop codon or a deletion resulting in a frameshift and then a stop codon or others).

1.3 Let's Get Started

Phew, that was a lot of content. But now you should know all about CF and the wide variety of tools that scientists use to detect the disease and its associated mutations, and then predict the risk for future generations. Will you be able to use all this knowledge to inform a young couple about the risk of their unborn child having CF?

Techniques Used in the Lab

- PCR
- Gel electrophoresis
- Capillary electrophoresis

Learning Objectives

At the end of this simulation, you will be able to . . .

- Describe the basic concepts of inheritance
- Build and interpret a pedigree based on family data
- Understand genetic risk assessment and counseling
- Gain insight into the vital work a genetics laboratory performs

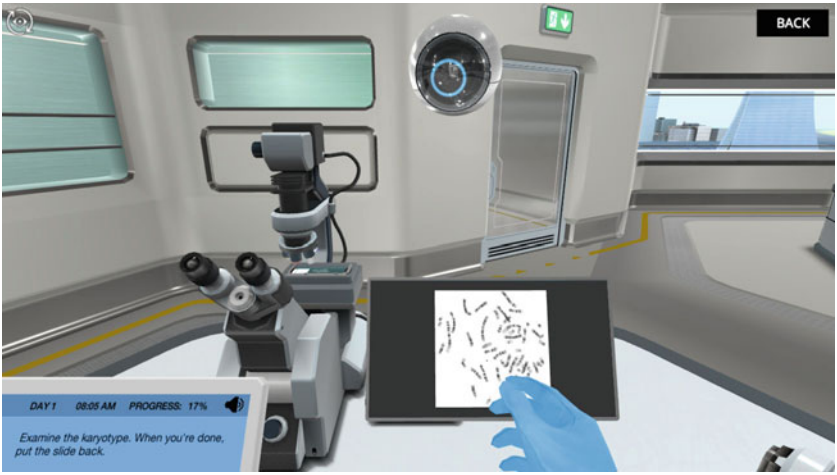
ACCESS THE VIRTUAL LAB SIMULATION HERE www.labster.com/springer **BY USING THE UNIQUE CODE AT THE END OF THE PRINTED BOOK. IF YOU USE THE E-BOOK YOU CAN PURCHASE ACCESS TO THE SIMULATIONS THROUGH THE SAME LINK.**

Further reading

Alberts B et al (2015) The molecular biology of the cell, 6th edn. Garland Science, Abingdon
Hartwell L et al (2015) Genetics: from genes to genomes, 5th edn. McGraw-Hill, Boston
OpenStax (2018) Biology. OpenStax CNX. http://cnx.org/contents/185cbf87. Accessed 1 June 2018
Urry LA et al (2014) Campbell biology, 10th edn. Pearson, Boston

Cytogenetics

2



A. Gardner et al., *Labster Virtual Lab Experiments: Genetics of Human Diseases*,
https://doi.org/10.1007/978-3-662-58744-7_2

2.1 Cytogenetics Simulation

In the Cytogenetics simulation, you begin in an ultrasound examination room where a young mother-to-be receives an abnormal ultrasound result. Your task is to use the power of cytogenetics to identify the underlying cause of the result and to inform her about the possible outcomes for her unborn child.

Assist with a Prenatal Screen and Perform an Amniocentesis
Prenatal ultrasounds are routinely performed to check on fetal development and to identify potential malformations (Fig. 2.1). If any are detected then more in-depth tests, investigating the genetics of the fetus are often performed. These tests require fetal genetic material which can be isolated using techniques such as an amniocentesis, which you will help perform.

Perform Array Comparative Genomic Hybridization
In order to find the underlying cause of the potential abnormality, you will perform array-based comparative genome hybridization (array CGH) using an amniocentesis sample taken from the mother-to-be. You will learn the basic principles of array CGH and how to analyze the results (Fig. 2.2 and 2.3).

Fig. 2.1 Assist the midwife as she performs a prenatal ultrasound in the Cytogenetics simulation

Fig. 2.2 Analyze an amniocentesis sample using array CGH in the Cytogenetics simulation

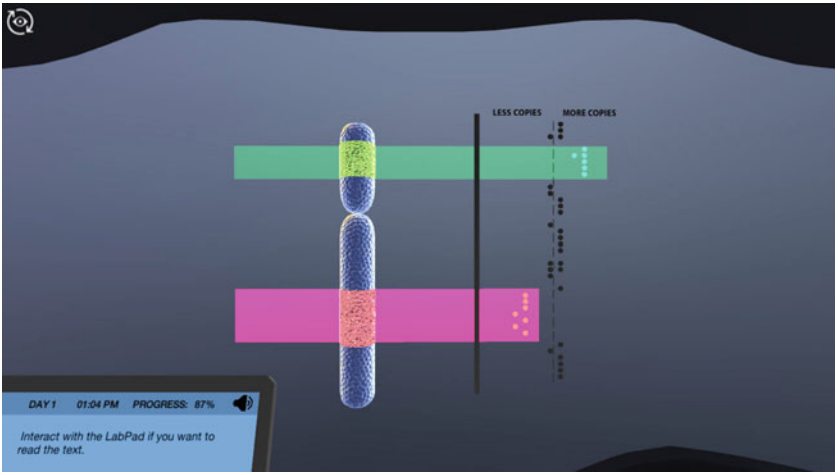


Fig. 2.3 Learn how array CGH works at the molecular level in the Cytogenetics simulation

Identify If a Chromosome Translocation Has Occurred
You will then learn about all the different translocations that can occur and learn about example conditions associated with each. Then, based on the results from the array CGH you will be able to determine if a chromosome translocation has occurred, and determine its exact location and type.

Will you be able to collate all your data and counsel the mother-to-be on your findings, and what they mean for her unborn child?

2.2 Cytogenetics Theory Content

Cytogenetics is the study of inheritance specifically in relation to the structure and function of chromosomes. This key technique is widely used in prenatal screening as a way of identifying any potential fetal abnormalities arising from alterations in chromosome structure. The theory content below will provide you with all the information to successfully complete the Cytogenetics simulation.

Prenatal Screening

A prenatal screen is often offered to mother's as a way of detecting potential fetal abnormalities. Screening can either be routine and offered to everyone, or specifically targeted for example when there is a history of a particular disease in a family.

Prenatal screening tests can include an ultrasound scan and other non-invasive screening such as, cell-free DNA testing.

Pregnant women are typically offered two tests:

1. A screening based on a blood sample at week 8–13 of gestation measuring PAPP- A (Pregnancy-associated protein-A) and free hCG (human Chorionic Gonadotropin), and an ultrasound scan at week 11–14 of gestation where the nuchal (a collection of fluid under the skin at the back of the neck) thickness is measured. A combined risk for Down syndrome (Trisomy 21) is estimated. If a high combined risk is found, an invasive prenatal test (often a chorion villus sample) is offered.
2. A second-trimester ultrasound scan at week 18–20 of gestation. The fetus will be examined for structural abnormalities and developmental defects. If for example a structural malformation is found, an invasive prenatal test (often an amniocentesis) is offered. At the second-trimester ultrasound scan the localization of the placenta is also described.

The material isolated from the invasive tests is then used in prenatal diagnostic tests. These can include a chromosome analysis, testing for changes in chromosome structure or number, or may be DNA based where a particular gene is investigated using a technique such as DNA sequencing.

Cytogenetics

Where ultrasound and other non-invasive tests have detected a potential abnormality, a chorionic villus sample or amniocentesis is often taken for analysis via cytogenetics. Scientists look for alterations in chromosome structure or number which may be linked to the potential abnormality. The best known cytogenetic test is used in the diagnosis of Down syndrome where all or part of a third copy of chromosome 21 is present, which is detected by using karyotyping. However, it can also be used for many other diagnoses including those for a cleft lip or palate.

Recent advances have made it possible to perform some such tests on maternal blood samples, greatly reducing the risk to the fetus. However, some tests still require the more invasive sampling technique.

Amniotic Fluid Sample

In order to perform a chromosome analysis, mitotic cells are required. For fetal analysis these can be obtained by culturing the cells from an amniotic fluid sample. Amniotic fluids are obtained by amniocentesis. Factors that can affect the success of cell culturing are:

- **Sterility:** Amniotic fluid should be kept sterile. All laboratory work involving amniotic fluid must be performed inside a biosafety cabinet. The addition of antibiotics such as penicillin and streptomycin to the amniotic fluid may slow cell growth.
- **Presence of blood:** The presence of blood in the amniotic fluid is undesirable because it can be a sign of maternal cell contamination and it can increase the acidity of the culture and inhibit cell growth.

The amniotic fluid consists of three major cell types, including:

- **Amniotic fluid cells:** This cell type forms the majority of the cells in a new culture. It has a pleomorphic appearance and has no maternal cell equivalent.

- **Epithelial cells:** Epithelial cells typically do not grow well in amniotic cell cultures, where they do grow they have a cuboidal shape.
- **Fibroblast cells:** These cells have a thin and elongated shape, this type of cell predominates in older cultures.

Chromosomes

The word chromosome is derived from the Greek words "chromos", meaning color, and "soma", meaning body. In meiosis and mitosis, our genetic material is condensed into chromosomes that are visible in the microscope after fixation and staining.

Each chromosome has a unique banding pattern that is used for its identification and for the detection of potential structural abnormalities. The chromosome banding pattern is comprised of alternating light and dark bands, that appear along its length after being digested with an enzyme (trypsin) and stained with a dye (Giemsa or Leishman).

Chromosome Structure

The most visible form of chromosomes is the metaphase chromosome. Chromosomes at this stage are used to study chromosome morphology and staining characteristics.

Each metaphase chromosome consists of two sister chromatids, which are attached to the centromere; with the centromere dividing a chromosome into a short arm (p) and a long arm (q). Telomere refers to the specialized DNA repeat sequences located at the end of chromosomes.

Each chromosome can be classified according to the position of the centromere (Fig. 2.4) as:

- **Metacentric**: The centromere is located centrally, so the lengths of the two arms are almost equal.
- **Submetacentric**: The short arm is somewhat shorter than the long arm.
- **Acrocentric**: The short arm is extremely short. The human genome has six acrocentric chromosomes which include: 13, 14, 15, 21, 22 and Y.
- **Telocentric:** The centromere is located at the end of the chromosome. Humans do not normally have this type of chromosome.

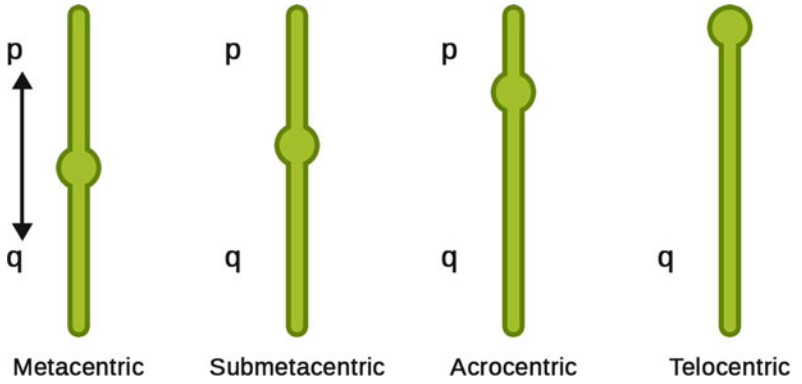


Fig. 2.4 **Centromere locations.** Chromosomes can be classified depending on the position of the centromere. In metacentric chromosomes the centromere is located centrally; this is shifted slightly in submetacentric chromosomes with a slightly shorter p-arm. When the p-arm is very short the chromosome is referred to as acrocentric, and if the p-arm is absent telocentric

Chromosome Abnormalities

Chromosomal abnormalities can be divided into: numerical and structural. They can be inherited from a parent (such as a translocation) or be "de novo" (new to the individual).

Chromosomal abnormalities can be associated with a syndrome. A syndrome involves more organs or features – a complex of symptoms. A syndrome can be sporadic, often with no known cause, or it can be caused by a gene mutation or a chromosomal abnormality. The most common chromosomal syndromes are Down syndrome (trisomy 21), Turner syndrome (45,X), and Klinefelter syndrome (47,XXY). Two common types of abnormalities include numerical and structural.

Numerical Chromosome Abnormalities

Individuals with the correct number of chromosomes for their species are known as **euploid**. In humans, euploidy corresponds to 22 pairs of autosomes and one pair of sex chromosomes, giving a total of 23 chromosome pairs. An individual with an error in chromosome number is described as **aneuploid**. This can include monosomy, where a chromosome is lost, or trisomy where a chromosome is gained.

Monosomic human zygotes missing any one copy of an autosome invariably fail to develop to birth because they have only one copy of essential genes. Most

autosomal trisomies also fail to develop to birth; however, duplications of some of the smaller chromosomes (13, 15, 18, 21, or 22) can result in offspring that survive for several weeks to many years.

Structural Chromosomal Abnormalities

Structural abnormalities occur when regions of a chromosome are significantly altered and can include:

- **Deletion:** A part of the chromosome is missing or deleted.
- **Insertion:** A part of one chromosome has been deleted from its normal location, and inserted into another chromosome.
- **Inversion:** A part of the chromosome has broken off, turned upside down and is reattached in the same place.
- **Duplication:** A part of the chromosome is duplicated, resulting in extra genetic material, for example, the duplication of 9p which is associated with a cleft lip or palate and other issues.
- **Reciprocal translocation:** A part of one chromosome is exchanged with a part of another chromosome.
- **Robertsonian translocation:** An entire chromosome has attached to another at the centromere – in humans these only occur with chromosomes 13, 14, 15, 21, and 22.

Cleft Lip and Palate

A cleft lip and palate is a malformation, which occurs when the lip or roof of the mouth does not completely fuse during the first trimester of fetal development (Fig. 2.5).

Cleft lips and palates are named depending on their severity (Fig. 2.6).

A cleft lip or palate can be diagnosed by a prenatal ultrasound. The upper lip and palate is formed separately at different gestational age ranges. The upper lip develops from fusing tissue of the nose and the center of the face. The formation of the upper lip takes place between the 7th and 9th week of gestation. The palate develops from fusing tissue of the upper lip, gum and inner mouth. The formation of the palate takes place between 7th and 14th week of gestation.

A cleft lip or palate usually detected during the ultrasound that carried out between 18 and 21 weeks of gestation.

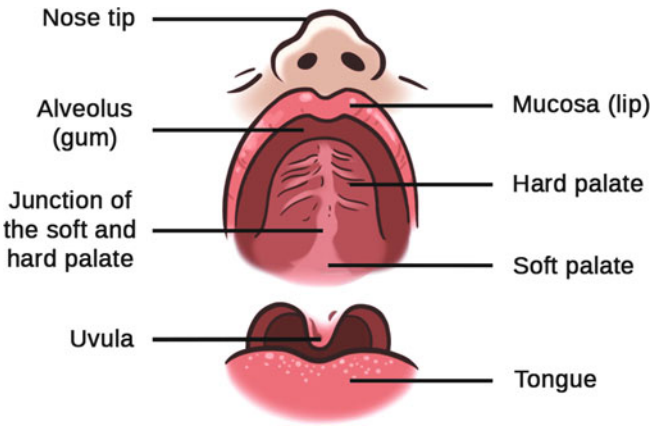


Fig. 2.5 **Anatomy of the lip and palate.** The palate refers to the top of the mouth and is comprised of a hard and soft region

Generally, a cleft malformation is divided into three categories:

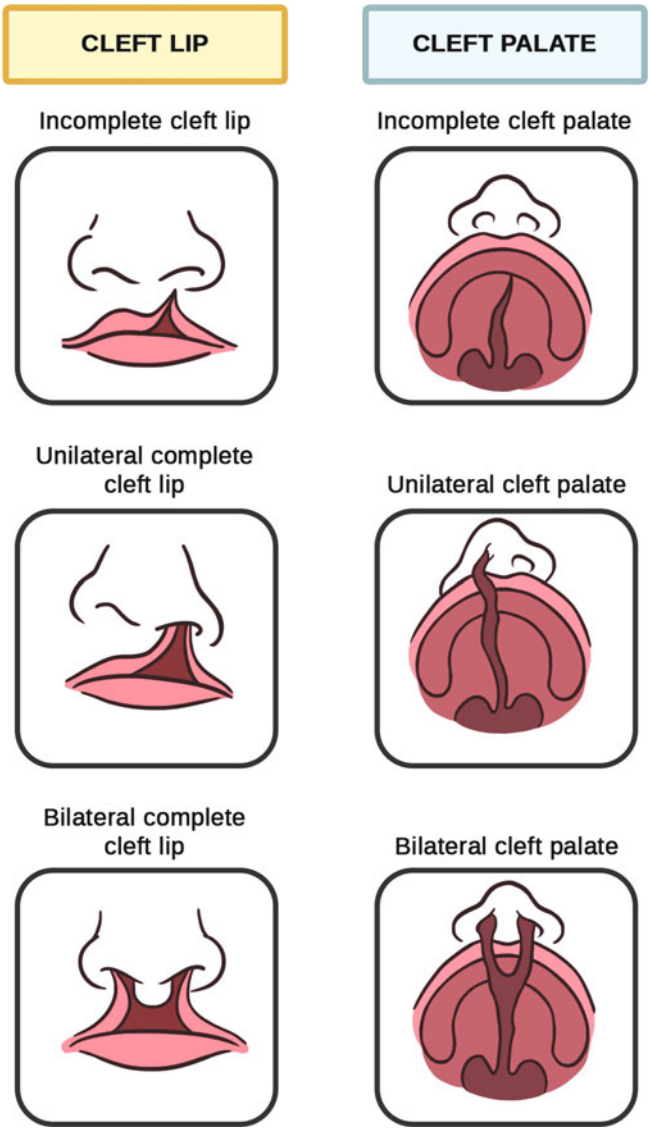
- Isolated cleft lip
- Isolated cleft palate
- Cleft lip and palate

Cleft Lip and Palate Etiology

In most cases the cleft lip and palate is an isolated finding where the child has no other manifestations. Many occur sporadically, meaning that there are no other family members with the phenotype. It is believed that the etiology behind most of these sporadic isolated cases is a combination of environmental factors and

Fig. 2.6 **Anatomy of a cleft lip and palate.** The anatomy of a cleft lip or palate can vary significantly. There are different names given to the cleft lip according to its location and how much of the lip is involved. A cleft on one side of the lip that does not extend into the nose is termed incomplete. A cleft on one side of the lip that extends into the nose is called unilateral complete. A cleft that involves both sides of the lip and extends into and involves the nose is called bilateral complete

►


CLEFT LIP
CLEFT PALATE
Incomplete cleft lip
Incomplete cleft palate
Unilateral complete cleft lip
Unilateral cleft palate
Bilateral complete cleft lip
Bilateral cleft palate

a number of genes with a small contributing effect. We call this multifactorial inheritance. A cleft lip and palate malformation can be treated surgically in early childhood. The recurrence risk (the risk of cleft lip and palate in a sibling) is increased (2–4%) because siblings share some environmental and some genetic factors.

In some families more than one child may have a cleft lip and palate and in some cases the cleft lip and palate is part of a syndrome, meaning that the child has other manifestations. In these cases, one should consider a monogenic or chromosomal etiology. The monogenic diseases which present with a cleft lip and palate include autosomal dominant diseases (e.g. Van der Woude syndrome), autosomal recessive diseases (e.g. Smith–Lemli–Opitz syndrome) and X-linked inherited diseases (e.g. Orofaciodigital syndrome I).

A cleft lip and palate can also be one of several manifestations in numerical chromosome disorders (e.g. Trisomy 13 and Trisomy 18), in microdeletion syndromes (e.g. 22q11 deletion syndrome) and in unique unbalanced chromosome disorders. If a genetic etiology is found, the parents should be appropriately counseled about any risks. If the mutation is identified, the parents can be offered prenatal genetic diagnostic testing for any future pregnancies.

Cleft Lip and Palate Diagnosis

If the prenatal genetic diagnostic tests are normal and no other malformations have been seen at the ultrasound scan, the cleft lip palate is most likely an isolated finding. The couple will be offered consultation with a team of experts to discuss and determine the best treatment plan. The team of experts includes a plastic surgeon, speech therapist, dentist, orthodontist, otolaryngologist, audiologist, geneticist, pediatrician and feeding specialist.

However, in approximately 10–15% of individuals with cleft lip and/or palate, it is a part of a syndrome. If a genetic etiology is established by prenatal genetic diagnostic test (e.g. a chromosome abnormality by array-based comparative genome hybridization (array CGH) and/or karyotyping) the couple can be informed about this specific genetic disease including its natural history, prognosis, potential treatment, recurrence risk, etc.

Array CGH compares the patient's genome against a reference genome and identifies the differences between the two genomes. These differences are recognized as genomic imbalance (for example a deletion/loss or a duplication/gain) in the patient.

In the array CGH it is expected to detect some small DNA segments that are present in more or less than two copies. Some of these are Copy Number Variations (CNV). A CNV is a DNA segment (minimum 1000 base pairs) that is present in varying number of copies in different individuals. They are in most cases harmless but can in rare cases be associated with a disease.

Array CGH Preparation

Two types of DNA need to be prepared for array CGH. Those are DNA samples containing the DNA to be tested, and a DNA reference as a comparison (Fig. 2.7). A prenatal array CGH can be carried out using DNA extracted from cells obtained from amniotic fluid or a chorionic villi samples. A typical DNA extraction is performed using a column with silica-gel membrane where DNA will bind specifically. All contaminants will pass through the silica-gel membrane and be washed away.

However, it is important to note that prenatal samples are at risk of being contaminated by maternal cells. The presence of maternal cells can mask a potential fetal chromosomal abnormality when analyzing the array CGH results. Maternal cell contamination (MCC) poses a risk of both false positive or false negative prenatal misdiagnosis.

Array CGH Analysis

A microarray scanner is used to detect the fluorescence signal. Three fluorescent colors can be observed; yellow indicates a normal or no change in the sequence, green indicates a duplication while red indicates deletion. The signals are converted into a graph illustrating the ratio of sample DNA compared to reference DNA.

Karyotyping

Karyotyping is a process of pairing and ordering all the chromosomes of an organism, thus providing a genome-wide snapshot of an individual's chromosomes. Although not performed in the simulation it is often used in real labs to validate the findings from assays such as array CGH.

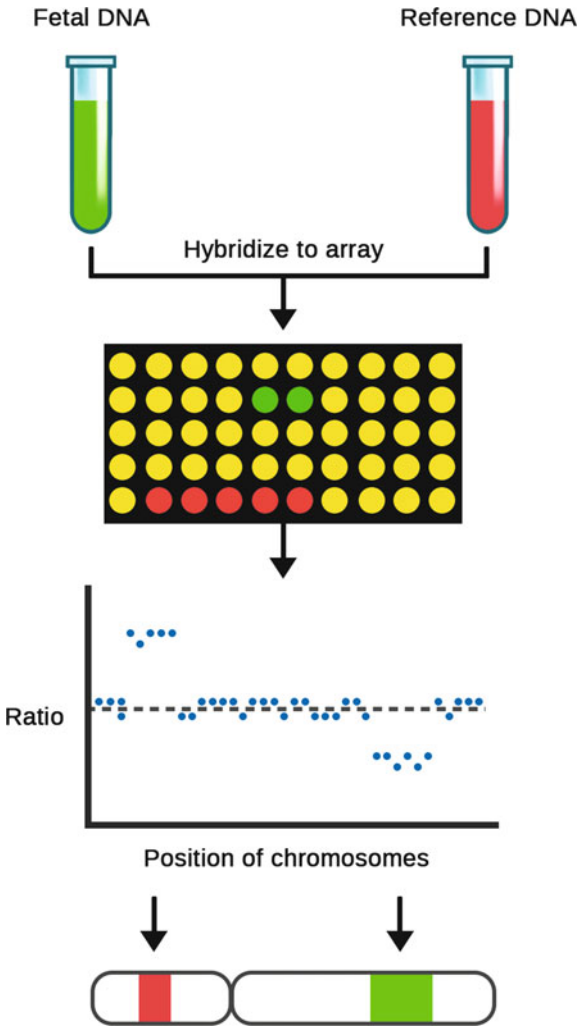


Fig. 2.7 **The array CGH process.** Firstly the DNA is treated with enzymes to cut it into smaller pieces. Then, sample and reference DNA are labeled with different colored fluorescent probes. The reference DNA is labeled with Cy5, a red fluorescent dye, while the sample DNA is labeled with Cy3, a green fluorescent dye. Both samples are mixed and co-hybridized to the array. The array contains several hundred thousand probes (e.g. oligonucleotides) representing sequences from all chromosomes

A karyotype is an appearance of the chromosomes from a cell. A karyogram is a photograph of the chromosome from a cell, arranged in its 22 homologous pairs plus the two sex chromosomes.

The karyotype is prepared from mitotic cells that have been arrested in the metaphase when the condensed chromosomes are visible. Before performing karyotyping, one has to collect the cell samples. The major steps involved are as follows:

1. Cell arrest and harvest
2. Hypotonic treatment
3. Fixation
4. Staining

Karyotype Nomenclature

A karyotype is described by an abbreviation containing the total number of chromosomes, a comma, and the sex chromosomes of an individual. If the individual carries a chromosomal abnormality, this can be described in the karyotype abbreviation. As a general rule, only abnormalities are indicated in a karyotype (apart from the total number of chromosomes and the sex chromosomes).

The nomenclature of band assignment and chromosome aberration is according to the International System for Human Cytogenetic Nomenclature (ISCN) shown in Table 2.1.

Trisomy 9p Syndrome

An extra copy of the short arm of chromosome 9 is called a duplication of 9p. Patients with duplication of 9p have Trisomy 9p syndrome.

More than 150 patients with trisomy 9p have been identified since 1970. Major clinical features include small stature, mental retardation, microcephaly/brachycephaly and dysmorphic features with deep-set eyes a downturned mouth and low-set ears among others. A cleft lip or palate is also often associated with the duplication.

Table 2.1 Symbols used in describing karyotypes. The symbols detailed above are commonly used when describing karyotypes. For example, a normal male would be referred to as 46,XY. Whereas, 47,XX,+21 refers to a female with trisomy 21, i.e. Down syndrome and as a more complex example 46,XX,dup(5)(p14p15.3) would refer to a female with a duplication of the short arm of chromosome 5 from bands p14 to p15.3

Symbol	Description
1–22	Autosome
X,Y	Sex chromosome
(+) or (−)	When placed before an autosomal number, this indicates that the chromosome is an extra or is missing
cen	Centromere
inv	Inversion
dic	Dicentric
p	Short arm of the chromosome
q	Long arm of the chromosome
t	Translocation
del	Deletion
ins	Insertion
dup	Duplication
ter	Terminal
der	Derivative
:	Break
::	Break and join
→	From . . . to

2.3 Let's Get Started

Cytogenetic analysis may not feel as cutting-edge as some of the molecular analyses performed, but it is a vital diagnostic test, especially when used in prenatal testing. Using your new knowledge will you be able to inform the mother-to-be about the issues her child faces?

Techniques Used in the Lab

- Amniocentesis
- Array comparative genomic hybridization

Learning Objectives

At the end of this simulation, you will be able to . . .

- Define medical genetic counseling, especially in the case of prenatal diagnostics
- Describe and perform an array-based comparative genome hybridization (array CGH)
- Identify the various types of chromosome translocation and describe associated conditions

ACCESS THE VIRTUAL LAB SIMULATION HERE www.labster.com/springer **BY USING THE UNIQUE CODE AT THE END OF THE PRINTED BOOK. IF YOU USE THE E-BOOK YOU CAN PURCHASE ACCESS TO THE SIMULATIONS THROUGH THE SAME LINK.**

Further reading

Alberts B et al (2015) The molecular biology of the cell, 6th edn. Garland Science, Abingdon

Hartwell L et al (2015) Genetics: from genes to genomes, 5th edn. McGraw-Hill, Boston

Miron PM (2012) Preparation, culture, and analysis of amniotic fluid samples. Curr Protoc Hum Genet 98(1):8–4

O'Connor C (2008) Karyotyping for chromosomal abnormalities. Nat Educ 1(1):27

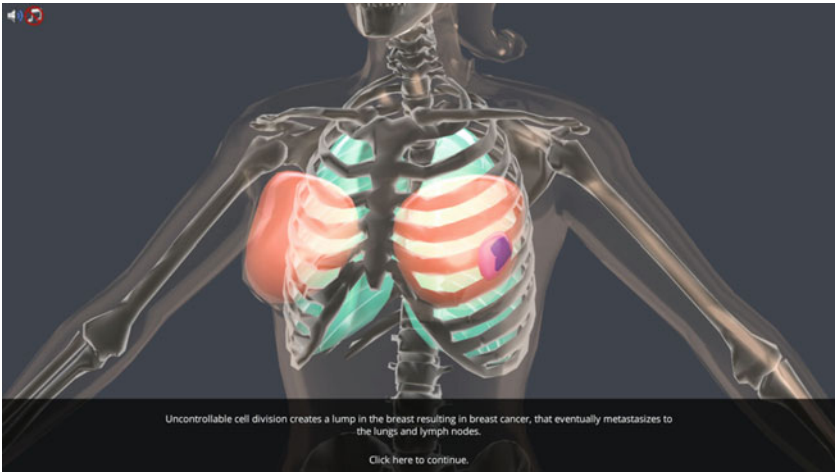
OpenStax (2018) Biology. OpenStax CNX. http://cnx.org/contents/185cbf87. Accessed 1 June 2018

Redon R, Fitzgerald T, Carter NP (2009) Comparative genomic hybridization: DNA labeling, hybridization and detection. Methods Mol Biol 529:267–278

Urry LA et al (2014) Campbell biology, 10th edn. Pearson, Boston

Medical Genetics

3



A. Gardner et al., *Labster Virtual Lab Experiments: Genetics of Human Diseases*,
https://doi.org/10.1007/978-3-662-58744-7_3

3.1 Medical Genetics Simulation

In the Medical Genetics simulation we approach a more complex genetic disease. Monogenic disorders such as CF, and the chromosome abnormalities discussed in the Cytogenetics chapter often have a high penetrance, meaning that a high proportion (or all) of those with the genetic abnormality will develop the disease. As our understanding of genetics and the human body increases, we are discovering that many complex diseases include a genetic component as a predisposition. Here you'll learn about a partially heritable form of breast cancer and the advances in modern medicine that allow us to detect and treat such cancers.

Construct a Pedigree and Perform a Linkage Analysis
Your first task will be to construct a pedigree based on the information you gather about the family. This will allow you to determine whether or not traits are hereditary, and importantly in which manner they are inherited. You will then visit the hospital and talk with a doctor to learn about hereditary cancer, the Knudson two-hit hypothesis and attempt to identify the specific gene causing hereditary breast cancer in a family.

Next, you will move to the laboratory to begin experimenting (Fig. 3.1). You will perform a linkage analysis using four genetic markers that are located close

Fig. 3.1 Use PCR in the Medical Genetics simulation to determine the patient's genotype

to the *BReast CAncer 1* (*BRCA1*) and *BReast CAncer 2* (*BRCA2*) genes. You will then perform PCR to amplify the microsatellite markers and analyze their genotype using gel electrophoresis. Analyzing the genotypes from family members, you will be able to determine which gene is associated with hereditary breast cancer in this family.

Assess the Impact of *BRCA* Mutations

Many breast cancers are associated with mutations in the *BRCA1* or *BRCA2* genes that result in protein truncation and uncontrollable cell division (Fig. 3.2). In order to check if the family's hereditary mutation results in a protein truncation, you will perform a protein truncation test (PTT) which compares the size of the protein synthesized from the patient's DNA versus a healthy control. By analyzing the results using polyacrylamide gel electrophoresis, you will be able to conclude whether the patient has a truncated protein.

Sequence the Patients DNA

After receiving the PTT results, you will need to perform a validation experiment to find out the exact mutation causing the truncated protein. You will sequence the DNA of the gene of interest and analyze the results (Fig. 3.3). Will you be able to identify the exact cause of the patient's cancer and determine if their family is at wider risk?

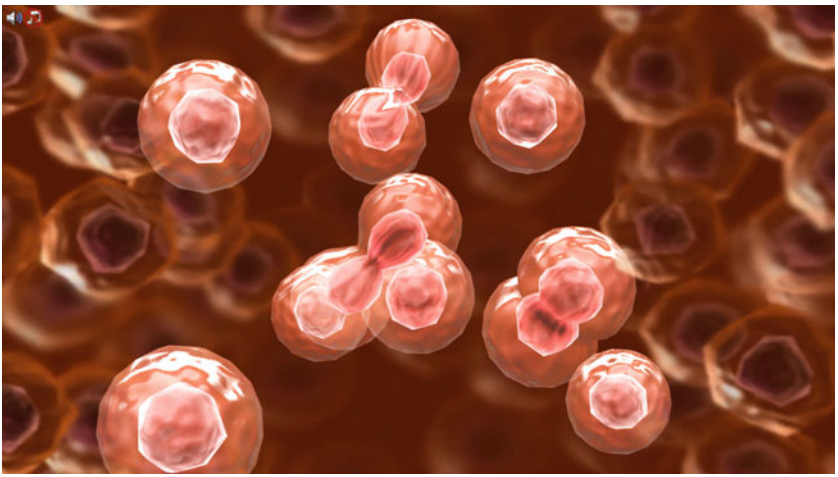


Fig. 3.2 Observe how cancer de-regulates cell division in the Medical Genetics simulation

Fig. 3.3 Sequence the gene of interest to validate your findings in the Medical Genetics simulation

3.2 Medical Genetics Theory Content

Medical genetics is a rapidly evolving field, based on our evolving understanding of the impact of genetics on our health. By identifying genes or mutations related to disease, we can begin to identify the risk of the same disease occurring in close relations, potentially allowing doctors to preemptively treat patients. The theory content below discusses all topics covered in the Medical Genetics simulation.

Cancer

Cancer is a term used to describe a group of diseases involving unregulated cell growth. A hallmark of cancer is cells that multiply uncontrollably, thereby creating malignant tumors that spread to parts of the body. Please note that cancer and the word tumor are not equivalent. A tumor can be either benign or malignant. Benign tumors are tumors that do not invade other parts of the body; they are localized in a specific organ and usually have a much better prognosis. On the other hand, malignant tumors are dangerous because they have the capability to spread to and invade other organs. They can enter the bloodstream or the lymphatic system and

spread distally to other parts of the body; this process is called metastasis. For example, it is common for breast cancer to metastasize into other parts of the body including axillary lymph nodes, the lungs, and the brain.

Tumor Suppressors and Oncogenes

Think of tumor suppressors as the genetic brakes that prevent cancer development in normal cells and oncogenes as genetic gas pedals that accelerate cancer development. There are several functions of tumor suppressor genes, such as ensuring cell cycle controls and checkpoints, repairing DNA damage, promoting apoptosis, and preventing epithelial to mesenchymal transition (EMT). When tumor suppressor genes are inactivated, these protective functions are impaired which can result in cells containing damaged DNA to proliferate uncontrollably, thereby forming tumors and eventually metastasizing to other parts of the body.

There are several ways in which tumor suppressor genes can be deactivated, including through epigenetic regulation (promoter hypermethylation), gene deletion, or mutations resulting in protein truncation. Most tumor suppressor genes are recessive, meaning both alleles need to be inactivated for the phenotype to appear. With only one of the alleles still active, the tumor suppressor gene can still perform its functions. The genes *BReast CAncer 1* (*BRCA1*) and *BReast CAncer 2* (*BRCA2*) are tumor suppressors, along with many other examples including *P53*, *PTEN*, *RB*, and *APC*.

Oncogenes, on the other hand, are dominant; a mutation in only one of the alleles is sufficient for cells to evade apoptosis and continue proliferating. In cancer, a proto-oncogene is a normal gene that, when mutated, results in a gain of function and can then be referred to as an oncogene. Oncogenes can be activated in several ways, including mutation, epigenetic regulation, chromosomal translocation, or gene duplication. The result is a hyperactive or highly expressed oncogene that overcomes the functions of tumor suppressor genes and promotes cancer development. Examples of oncogenes include *MYC*, *RAS*, *TGFB*, *HER2*, and *ERK* among others.

Sporadic Vs. Heritable Cancer

Cancer is a genetic disease that is caused by aberrations in an individual's DNA. This also means that the aberrations can be passed on from the parents to their offspring, resulting in a heritable cancer. These individuals are born with one de-

fective allele that is inherited from one of their parents. Due to their recessive nature, defective tumor suppressor genes are more commonly inherited than oncogenes since both alleles need to be nonfunctional for cancer to develop.

This was first explained by Alfred G. Knudson in 1971 in Knudson's two-hit hypothesis (Fig. 3.4). Knudson studied retinoblastoma (a cancer that develops in the retina) that occurs in very young children. He discovered that these children inherited a defective tumor suppressor gene called *RB1* from one of their parents.

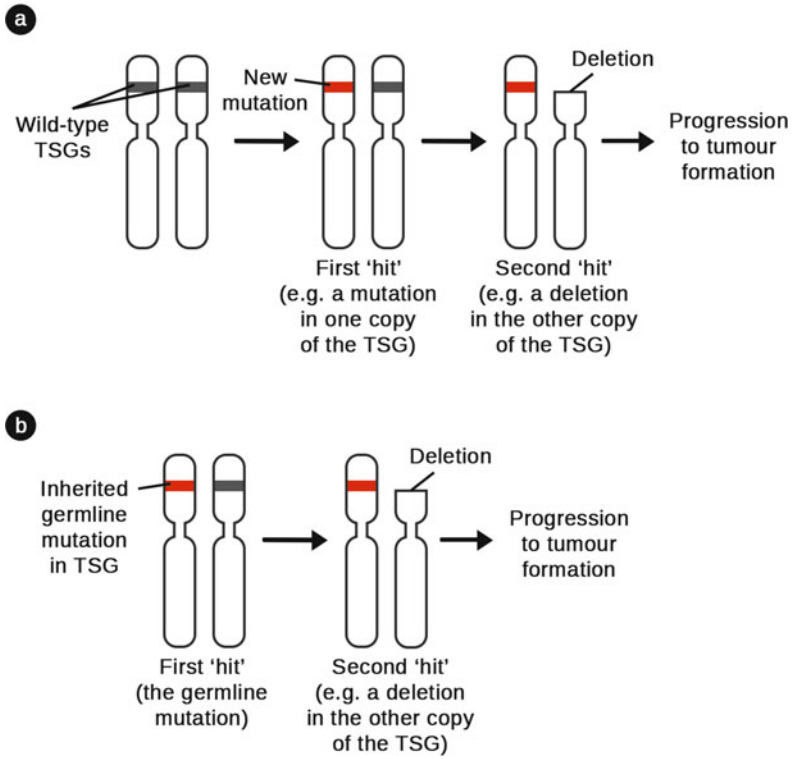


Fig. 3.4 Illustration of Knudson's two-hit hypothesis of tumor suppressor inactivation in familial cancer. The two hit-hypothesis requires two mutations to arise. In **a** both of these hits occur during the individual's life, for example, a mutation (first hit) in one copy of the tumor suppressor gene (TSG) followed by a deletion later on (second hit). **b** In individuals where the first 'hit' has already occurred due to a germline mutation they only require a single additional hit to go on to develop cancer. Such individuals therefore typically develop cancer at an early age

The inherited defective gene is referred to as the first hit, while the second hit is an event that occurs during their life. This could be due to a random mutation occurring during cell division or due to an external factor, such as solar radiation that results in the mutation of the second allele. Because these individuals have already inherited the first hit, they develop cancer at a much earlier age than other individuals who are born with both healthy alleles and therefore need to acquire two hits during their life.

Breast Cancer

Breast cancer is classified as a carcinoma, a type of cancer that is derived from epithelial cells. The National Cancer Institute at the NIH ranks breast cancer as the second most common cancer after prostate cancer, with an estimated 230,000 new cases diagnosed in 2013 and almost 40,000 deaths. In the US, approximately one in eight women will develop breast cancer at some point in her life, the rate being just under 12%. Importantly, breast cancer does not only occur in women; men are also susceptible to breast cancer, although the rate of occurrence is much lower.

Approximately 5%–10% of breast cancers can be linked to gene mutations that are inherited from one of the parents, with mutations in the *BRCA1* and *BRCA2* genes being the most common. Women with *BRCA1* or *BRCA2* mutations have up to an 80% risk of developing breast cancer during their lifetime, and they are more likely to be diagnosed at a younger age (before menopause). This was explained previously in the sporadic vs. heritable cancer section. An increased ovarian cancer risk is also associated with these genetic mutations. Mutations in *BRCA1* or *BRCA2* can be passed on from either the mother or the father.

BRCA1 is located on chromosome 17 at position 17q21, whereas *BRCA2* is located on chromosome 13 at position 13q12.3. These genes play a significant role in breast cancer as mutations in these two genes. Hundreds of cancer-associated mutations in *BRCA1/2* have been documented, and they are not only specific to breast and ovarian cancer.

BRCA and DNA Repair

Tumor suppressor genes can be further classified into gatekeepers and caretakers. Gatekeeper genes control inhibition of cell growth making them the guardians of the cell cycle. Caretakers, as the name suggests, are genes that take care of genome stability. Mutations in DNA and chromosomal rearrangements result in genome

instability. The caretaker genes promote genome stability by maintaining and repairing DNA damage. They are involved in nucleotide excision repair, base excision repair, non-homologous end joining (NHEJ) repair pathways, mismatch repair pathways, and telomere metabolism. Both *BRCA1* and *BRCA2* are tumor suppressor genes that are subclassified as caretakers. *BRCA1* is involved in both NHEJ repair and homologous recombination, while *BRCA2* is only involved in homologous recombination. *BRCA1* and *BRCA2* help perform DNA repair by working with other important proteins such as KU proteins that are involved in NEHJ and Rad51 proteins that are involved in homologous recombination (Fig. 3.5).

Genetic Testing

We can perform genetic testing to determine if a certain trait is inherited in a family. In 1990, Dr. Mary-Claire King performed linkage analysis at UC Berkeley in families that had a history of breast cancer and narrowed down the location of a significant gene to chromosome 17q21 1. This gene was later named as *BRCA1*.

Linked genes are located in close proximity with each other and tend to be inherited together. During meiosis, the two sister chromatids are separated and crossing over can occur, resulting in homologous recombination, where material from one sister chromatid is transferred to the other chromatid (Fig. 3.6). This event gives rise to genetic variations that differentiate one individual from the other offspring. When two genes are located close to each other, they have a high chance of being inherited together because they have a high chance of being exchanged together during homologous recombination.

On the other hand, if two genes are located at opposite ends of the chromosome, they will not be inherited together more than the expected 50% of the time (due to random sorting). In this case, one gene will be exchanged during homologous recombination and transferred to the other sister chromatid while the other gene will stay in the original chromatid. Because only one chromosome (pair of chromatids) is passed to the offspring and the two genes are now located in different chromosomes, the offspring will only inherit one of the two genes.

Linkage analysis is very useful to determine the approximate location of a gene of interest, for example, one that is suspected to cause a disease or medical condition. To identify the location of this gene, we can use markers or specific DNA sequences with known exact locations. We often use microsatellites or SNPs (single nucleotide polymorphisms) as markers. By observing which combination of specific markers and alleles is consistently present in affected individuals, we can then narrow down the approximate location of the gene.

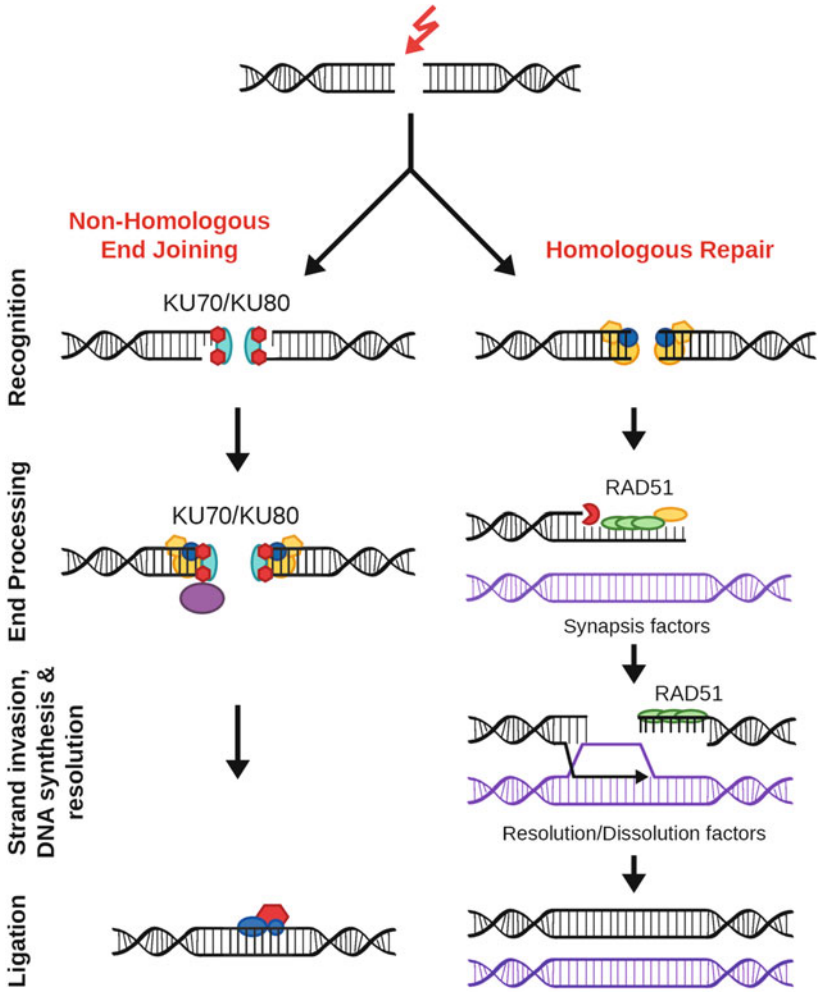


Fig. 3.5 **DNA repair pathways.** Both non-homologous end joining (NHEJ) and homologous repair (HR) are responsible for maintaining genomic stability by repairing broken DNA strands. HR is assumed to be error free because it uses a template. Whereas when a double-strand DNA lesion is repaired by NHEJ there is no validating DNA template present so it may result in a novel DNA strand formation with loss of information, or the introduction of a mutation which may lead to the development of cancer

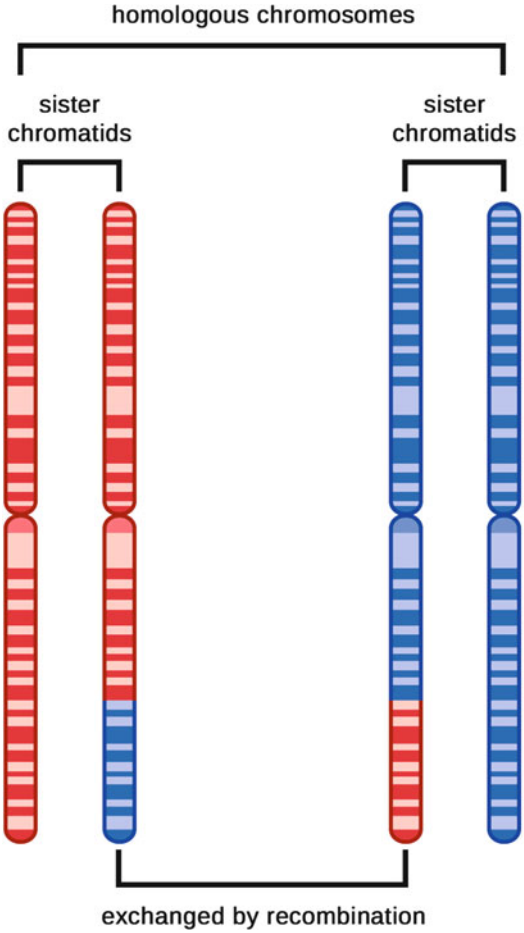


Fig. 3.6 Homologous recombination during meiosis giving rise to new genetic variations. Homologous recombination is a type of genetic recombination in which nucleotide sequences are exchanged, during meiosis, between two similar or identical molecules of DNA giving rise to new genetic variations

We need to analyze many individuals in the family over several generations to identify a gene's location because we need to find a specific allele that is always present in affected individuals and never present in healthy individuals. Since families are genetically related, there is less genetic variation within a family making it best to compare individuals within a close family rather than comparing alleles to unrelated healthy individuals.

Microsatellites

Microsatellites are short repeating DNA sequences of two to six base pairs. The number of repeats can vary among individuals and are used to distinguish between different alleles. For example, if allele 1 has four repeats, allele 2 has five repeats, allele 3 has six repeats then researchers would easily be able to identify each specific allele. SNPs can also be used as markers in linkage analysis; however, there are only two alleles for SNPs. For example, C/T, which means that allele 1 has C, while allele 2 has T. Due to the higher number of variability in microsatellites than in SNPs, it is easier to perform linkage analysis using microsatellites.

One of the most common examples of microsatellites is $(CA)_n$, where n represents the number of CA repeats. The number of repeats can vary between 3 to 100 times, thus creating a high number of alleles that are present in the population. Remember that this high number of alleles for one gene is present in the population and not in the individual! Individuals always have only two copies of each chromosome in their cells meaning that there are always two alleles present. When an individual has a different allele on each chromosome, for example $(CA)_{10}$ and $(CA)_{15}$ he/she is said to be heterozygous for the microsatellite. On the other hand, when an individual has the same allele on both chromosomes, he/she is said to be homozygous for the microsatellite. We can analyze the number of repeats of the microsatellite by first, amplifying it using PCR and then, analyzing the length using gel electrophoresis.

3.3 Let's Get Started

In the previous two simulations you've worked on disorders with a relatively high penetrance. Here you'll have to combine your previous knowledge and work to discover exactly which gene is linked with a familial form of breast cancer. Will you be able to identify the specific mutation and inform the patient's wider family about their breast cancer risk?

Techniques Used in the Lab

- PCR
- Gel electrophoresis
- Protein Truncation Test
- DNA sequencing

Learning Objectives

At the end of this simulation, you will be able to . . .

- Perform linkage analysis based on your understanding of Mendelian genetics
- Analyze a PCR experiment
- Understand the basics of breast cancer, tumor suppressor, oncogenes and *BRCA1/2*

ACCESS THE VIRTUAL LAB SIMULATION HERE www.labster.com/springer **BY USING THE UNIQUE CODE AT THE END OF THE PRINTED BOOK. IF YOU USE THE E-BOOK YOU CAN PURCHASE ACCESS TO THE SIMULATIONS THROUGH THE SAME LINK.**

Further reading

Alberts B et al (2015) The molecular biology of the cell, 6th edn. Garland Science, Abingdon

Hartwell L et al (2015) Genetics: from genes to genomes, 5th edn. McGraw-Hill, Boston

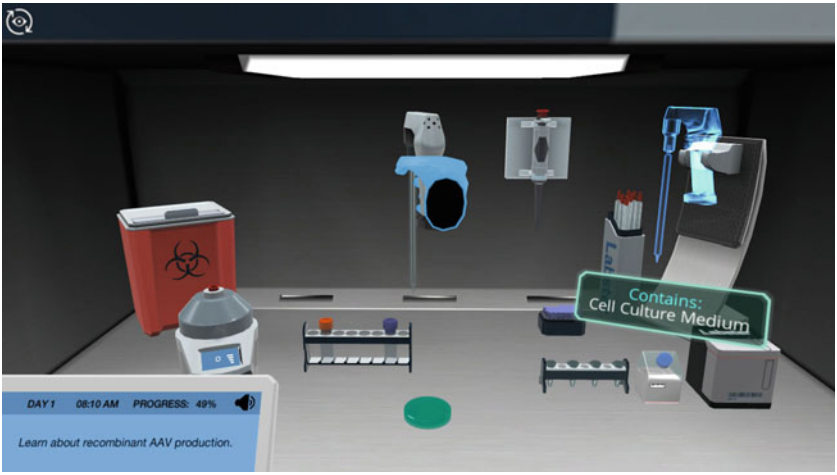
OpenStax (2018) Biology. OpenStax CNX. http://cnx.org/contents/185cbf87. Accessed 1 June 2018

Shah R, Rosso K, Nathanson SD (2014) Pathogenesis, prevention, diagnosis and treatment of breast cancer. World J Clin Oncol 5(3):283–298

Urry LA et al (2014) Campbell biology, 10th edn. Pearson, Boston

Viral Gene Therapy

4



A. Gardner et al., *Labster Virtual Lab Experiments: Genetics of Human Diseases*,
https://doi.org/10.1007/978-3-662-58744-7_4

4.1 Viral Gene Therapy Simulation

In the previous three chapters we've covered the impact our genetics can have on our health. In this final chapter we reverse this and look at the exciting field of gene therapy. In the Viral Gene Therapy simulation, you will learn about the use of modified viruses and how we can manipulate their genetic code to fix mutations inside our own bodies. Based on recent exciting research findings you'll investigate the role of genetics in heart failure. Will you be able to design a virus that can help to improve the symptoms of heart failure in patients?

Using Viruses to Treat Heart Failure
Heart failure is one of the major health issues facing Western populations. Dietary and lifestyle causes are well understood; however, a significant number of individuals are at risk based purely on their genetics. If we know which gene and mutation is linked with a disease, then we should be able to "fix" this via gene therapy, using the infectious power of viruses to improve our health (Fig. 4.1).

Design, Produce and Test Your Virus in the Lab
While gene therapy sounds simple on paper, there are lots of obstacles standing in the way of a viable treatment. The first is in identifying a suitable virus and

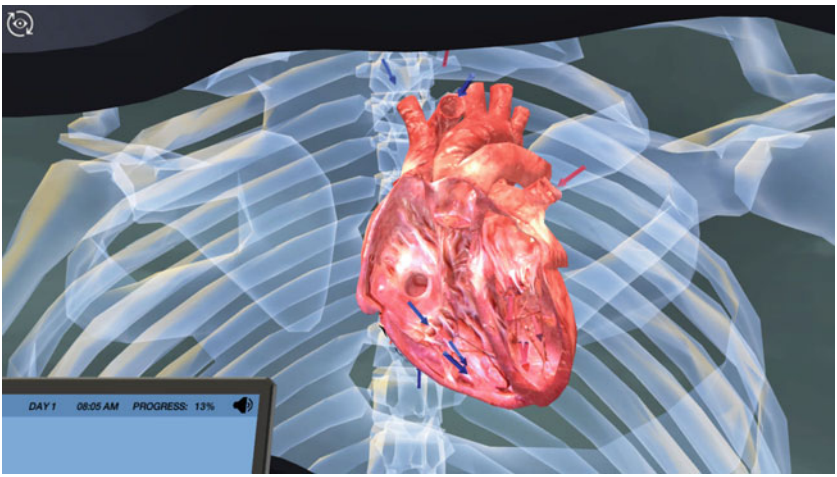


Fig. 4.1 Learn about the heart and how heart failure is a growing risk in the Viral Gene Therapy simulation

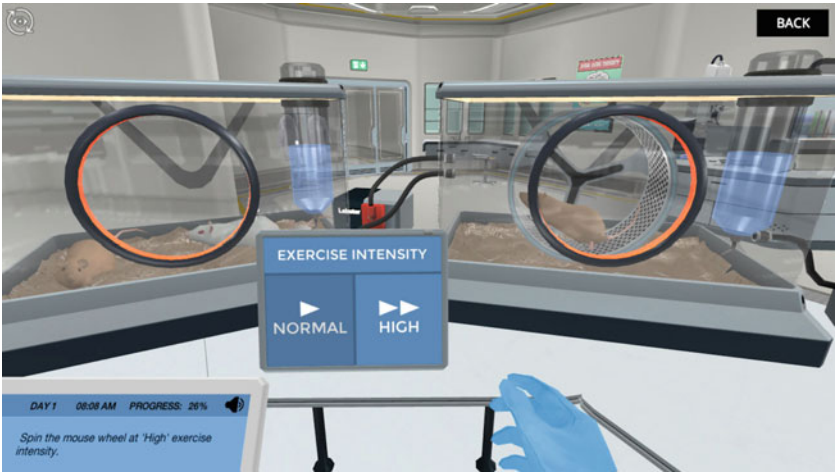


Fig. 4.2 Design and test your virus to treat heart failure in a suitable model in the Viral Gene Therapy simulation

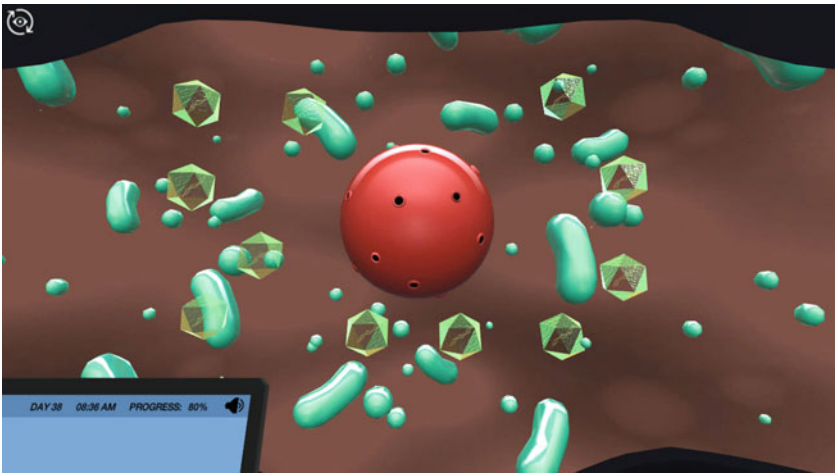


Fig. 4.3 Observe your virus in action in the Viral Gene Therapy simulation

deciding on the best way to insert the genetic material within. Then, the efficiency of the produced viruses needs to be tested in a suitable model system, in this case, a mouse model (Fig. 4.2). Since this is a virtual lab, no real animals will be harmed in this experiment, and results are available immediately instead of the usual five weeks!

Test the Efficacy of Your Gene Therapy Treatment
Once you have designed your gene therapy treatment and produced your viruses, the next step is to see if it has a beneficial effect (Fig. 4.3). Will your virus cure a mouse model of heart failure and do you think it would make a suitable therapeutic treatment for humans?

4.2 Viral Gene Therapy Theory Content

Does using viruses as a form of therapy sound like science fiction to you? You'd be surprised, but it's an increasingly popular technique used to deliver functional genes into patients. In this lab, you will learn about the use of modified viruses and how to equip them with therapeutic genes. The content below covers all the related theory needed to complete the Viral Gene Therapy simulation successfully.

Gene Therapy

Gene therapy is a technique for treating disease by altering the patient's genetic material. Most often, gene therapy works by introducing a healthy copy of a defective gene (also called therapeutic gene) into the patient's cells. Different vectors exist that carry and deliver the therapeutic gene into the patient's cells. Viral vectors are most commonly used, but also non-viral methods exist, such as injection of naked DNA or DNA complexes.

Therapeutic Gene

Therapeutic genes are used in gene therapy as functional copies of disease-causing genes. They are delivered as nucleic acid polymers (RNA or DNA) that replace the mutated or lost gene in the patient treated.

Table 4.1 Baltimore classification. Viruses can be classified into one of seven groups based on the structure of their genetic material and their method of replication. Ds = double-strand, ss = single-strand, (+) = positive-sense, (−) = negative-sense

Group	Name	Description	Example
I	dsDNA	Double-stranded DNA	Herpesviruses
II	ssDNA	Single-stranded DNA	Parvoviruses
III	dsRNA	Double-stranded RNA	Reoviruses
IV	(+)ssRNA	Single-stranded RNA, positive sense strand	Flaviviruses
V	(−)ssRNA	Single-stranded RNA, negative sense strand	Orthomyxoviruses
VI	ssRNA-RT	Single-stranded RNA, positive sense strand, DNA as replication intermediate	Retroviruses
VII	dsDNA-RT	Double-stranded DNA, RNA as replication intermediate	Hepadnaviruses

Viruses

Viruses are small infectious agents that can only replicate inside the living cells of other host organisms. When not replicating in a host cell viruses exist as stable particles also known as virions. The size of most viruses vary from 20 to 300 nm and hence can only be seen using an electron microscope. Viruses are built from a protein coat known as the capsid which protects the genetic material carried within. Some viruses, such as retroviruses, carry a lipid envelope around the capsid. The type of genetic material found within viruses can vary widely and is often used to classify them into one of seven groups as for example in the Baltimore classification, named after David Baltimore (Table 4.1).

Viral Vectors in Gene Therapy

In gene therapy, viral and non-viral vectors are used to deliver a therapeutic gene to a patient's cells.

Viruses are the perfect tool for nucleic acid delivery as they naturally evolved to insert their genetic material into cells to replicate. Some viruses, such as retroviruses, even insert their genes into the host cell's genome. In order to ensure the safety of the viral vectors, either the whole viral genome is replaced by the therapeutic gene or the parts that are disease-causing are deleted from the viral vector (Fig. 4.4).

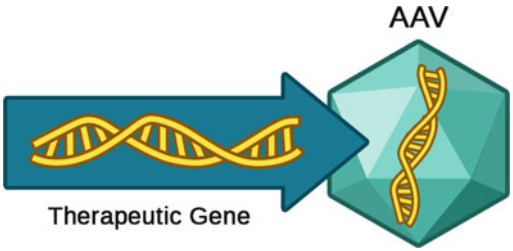


Fig. 4.4 Incorporation of a therapeutic gene into the viral vector adeno-associated virus (AAV). By incorporating a therapeutic gene into a vector, in this case AAV, it is then possible to deliver this gene to the body, or specific organ or tissue of an affected individual

Most commonly used viral vectors are:

- Adenoviruses
- Adeno-associated virus (AAV)
- Retrovirus
- Herpes simplex virus
- Vaccinia virus

Adeno-associated Virus (AAV)

The adeno-associated viruses (AAV) are small, single-stranded DNA viruses that belong to the family of *Parvoviridae*.

AAV infect humans and other primates; however, they are not known to cause any disease. The lack of pathogenicity makes AAV attractive as viral vectors for gene therapy.

AAV are characterized by an icosahedral capsid and a genome size of 4.7 kb. The genome contains two open reading frames, encoding for *rep* and *cap*. *Rep* is composed of four genes encoding for viral proteins required for replication. *Cap* contains VP1, VP2, and VP3, which together build the virus capsid.

AAV in Gene Therapy

The desired therapeutic gene cDNA is inserted between two inverted terminal repeats (ITRs) that aid the packaging of the viral genome into the virus capsid.

Once infected with a recombinant AAV, the introduced therapeutic gene will not integrate into the host genome but remain as episomal concatemers in the cell nucleus. Consequently, AAV DNA is lost upon cell division as the episomal DNA is not replicated. Since AAV doesn't integrate into the human genome, it does not present the risk of random insertion and mutagenesis, which makes AAV a more predictable viral vector for gene therapy compared to retroviral vectors.

In addition, AAV shows a great relative selectivity for heart muscle cells and is therefore the preferred gene delivery system for targeting the heart, as for example in the case of treating heart failure patients (more details in the following sections of this chapter).

Compared to other viral vectors, AAV has a rather small cloning capacity and requires the complete replacement of the 4.7 kb genome. This allows for the incorporation of only small therapeutic genes.

The application of AAV as a viral vector for gene therapy is limited by the fact that some patients were unnoticedly infected with AAV and therefore produced neutralizing antibodies against the virus, which impairs the efficiency of the treatment.

Retroviral Vectors

Retroviruses are lipid-enveloped viruses comprised of linear single-stranded RNA genomes of 7 to 11 kb. The main features of retroviral vectors are the reverse transcription of the viral RNA genome into DNA, and stable integration into the host DNA. Lentivirus is a genus of retroviruses. Example of lentiviruses is human immunodeficiency virus (HIV).

At the end of all retroviral genomes there are two long terminal repeat (LTR) sequences. The LTR sequences become the border of the *gag*, *pol* and *env* genes. Lentiviruses also encode for *tat* and *rev*. The function of each gene is as follows:

- ***gag:*** structural protein
- ***pol:*** nucleic acid polymerase/integrases
- ***env:*** surface glycoprotein
- ***tat:*** regulatory protein for gene expression
- ***rev:*** regulatory protein accessory genes

Retroviral Vectors in Gene Therapy

Among virus types that are generally used in clinical gene therapy, the retrovirus is the leading delivery system due to its efficiency.

In order to increase vector biosafety, the retroviral vector gene is modified by removing up to six genes essential for HIV replication and pathogenesis. The overall principle of lentiviral vector gene transfer is described as follows:

- The genes that encode the structural proteins (*gag*, *pol*, and *env*) are deleted from the virus vector (plasmid).
- The viral vector carries the gene of interest and psi packaging. The virus vector is co-transfected with packaging plasmids that carry the viral structural genes to the packaging cell line. The recombinant lentivirus is then produced in the cell line before being extracted.
- The recombinant lentivirus particle is transduced into target cells, where reverse transcription, foreign gene integration, and foreign gene ectopic expression take place.

The advantages of retroviral vectors include:

- Stable integration into the host genome
- High-level expression
- Capability to infect a broad variety of target cell types
- Ability to carry foreign genes up to 8 kb

However, one disadvantage of using retroviruses as vectors in gene therapy is the potential for insertional mutagenesis by random viral integration into the host DNA. When this occurs in tumor suppressor genes it can lead to the development of cancer.

Targeting in Gene Therapy

One challenge in the development of the gene therapy approach using viral vectors is the specific targeting of the vector to a tissue inside the body, since an untargeted vector can potentially lead to serious side-effects.

One possibility to prevent those side-effects is the engineering of the viral surface proteins so that they specifically bind to cell receptors that are only expressed in the tissue to be targeted. This approach is also called pseudo-typing.

In addition, some viral vectors are known to naturally infect specific tissues, such as adeno-associated virus (AAV) 2 that presents natural tropism towards skeletal muscles, neurons, vascular smooth muscles, and liver cells.

Finally, the gene therapy drug can be injected or given intravenously to ensure that a specific tissue in the body is targeted.

Another layer of specificity can further be achieved by expressing the therapeutic gene from a tissue-specific promoter.

The Heart

The heart is a fist-sized organ which sits in the middle of the chest, and slightly to the left. Forming part of the circulatory system it functions to pump blood around the body by "beating", or more accurately by undergoing a controlled series of contractions known as the cardiac cycle. Comprised of four chambers, the upper left and right atria and the lower left and right ventricles, through which blood flows in a single direction.

The Cardiac Cycle

The cardiac cycle is the period of time that begins with contraction of the atria and ends with ventricular relaxation. The cardiac cycle includes all events associated with the blood flow through the heart during one complete heartbeat.

Deoxygenated blood flows into the right atrium and is pumped into the right ventricle, before it is in turn pumped into the pulmonary capillaries of the lung for oxygenation. Oxygenated blood returns from the lungs into the right atrium before passing into the right ventricle and then into the main circulatory system. When a chamber is actively contracting it is said to be in systole, whereas a relaxing chamber is said to be in diastole.

A single cardiac cycle therefore begins with atrial systole and continues to ventricular systole, atrial diastole, and ventricular diastole. Interruption to the cardiac cycle can be fatal and is a major form of heart failure.

Ejection Fraction (EF)

The ejection fraction (EF) describes the amount of blood that is pumped out of the ventricles with each contraction of the heart (Fig. 4.5). The resulting percentage

$$\frac{\text{Amount of blood pumped out of the ventricle}}{\text{Total amount of blood in ventricle}} = \textbf{Ejection fraction (\%)}$$

Fig. 4.5 Calculation of the ejection fraction (EF). The EF is calculated by dividing the amount of blood pumped out of a ventricle, by the total amount of blood in the ventricle. If a large proportion of blood remains in the ventricle, the EF will be low and this indicative of poor heart health

indicates heart performance and helps to diagnose heart failure or other types of heart disease. Normal EF values for a healthy heart are 50 to 75%.

Heart Failure

Heart failure, sometimes also called congestive heart failure, describes a condition in which the heart is not able to pump sufficient blood to cover the body's needs. In particular, the heart wall, or myocardium, is dilated in heart failure patients.

Heart failure is a fatal condition and the most common reason for hospitalization of people above the age of 65. The causes of heart failure include coronary artery disease, high blood pressure, heart attacks, or conditions that overwork the heart.

Typical symptoms of heart failure are shortness of breath, fluid retention (e.g. swollen ankles, legs, abdomen), reduced ability to exercise, weakness and fatigue, rapid or irregular heartbeats. Heart failure can be triggered for example during physical exercise when the heart is under strenuous effort.

Treatment of Heart Failure

Conventional treatments of heart failure include ACE-inhibitors, beta-blocker, diuretic, and Angiotensin II Receptor Blockers (ARB's). However, most of the commonly available drugs have high side effects, and many patients still have a poor life expectancy despite treatment. Consequently, new types of treatments are needed.

Viral gene therapy using AAV as a vector and *SERCA2a* as a therapeutic gene demonstrates a promising new therapeutic approach in the field of heart failure (Fig. 4.6).

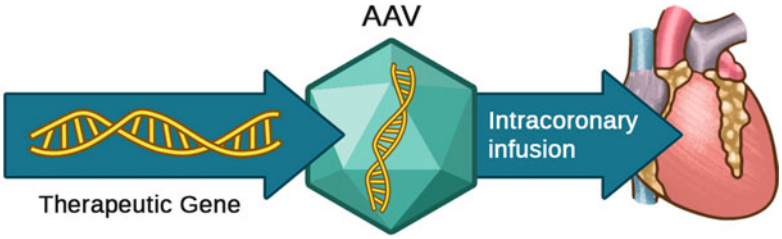


Fig. 4.6 **In the case of gene therapy treatment of heart failure, the therapeutic gene is directly targeted to the heart, for example by intracoronary infusion**

SERCA

Sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA), is a calcium ATPase-type P-ATPase. It resides in the sarcoplasmic reticulum (SR) in myocytes and regulates re-uptake of Ca^{2+} into the SR. It thereby regulates the contractile properties of muscle cells in the heart and thus the cardiac cycle. Three major paralogs exist:

- SERCA1
- SERCA2
- SERCA3

SERCA2a was identified to specifically regulate the Ca^{2+} cycling in cardiomyocytes, particularly Ca^{2+} removal, which triggers myocardial relaxation. SERCA2a is therefore considered a critical factor in the progression of heart failure (Fig. 4.7).

Recombinant AAV Production for Gene Therapy

Several systems exist to produce recombinant AAV. The most common system is based on a co-transfection of HEK293 cells (as described below) as the packaging cell line (Fig. 4.8 and 4.9). Following lipid-based co-transfection, cells are harvested by lysis and the virus is purified for infection of target cells.

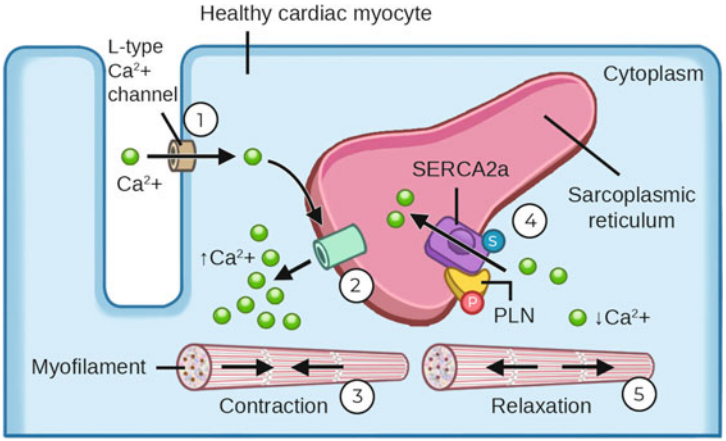


Fig. 4.7 Schematic representation of the function of SERCA2a in healthy cardiac myocytes. During diastole, Ca^{2+} enters the cell (1) and triggers the release of large amounts of Ca^{2+} from the sarcoplasmic reticulum (SR) (2). This causes the myofilaments to contract (3). At the same time, SERCA2a is released from its inhibitor PLN and shuttles Ca^{2+} back into the SR (4), which allows the myofilaments to relax (5)

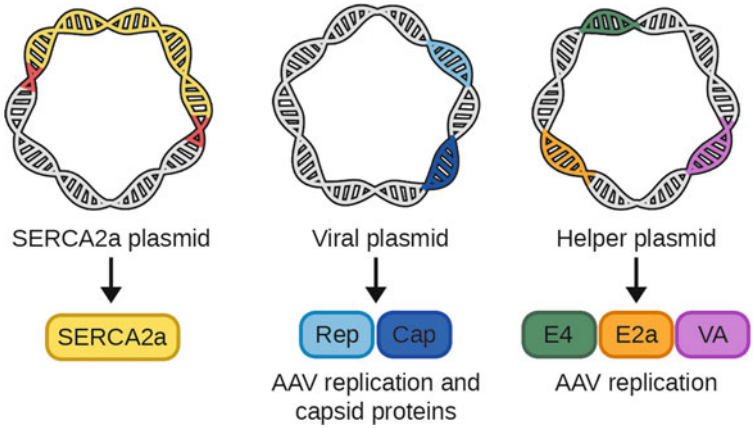


Fig. 4.8 Plasmids required for AAV production. The viral plasmid encodes for the replication (Rep) and Capsid (Cap) proteins. The helper plasmid encodes for the essential factors E4, E2a, and VA. Without those factors, no replication would occur. Finally, a third plasmid is co-transfected the therapeutic gene cDNA (shown here in yellow)

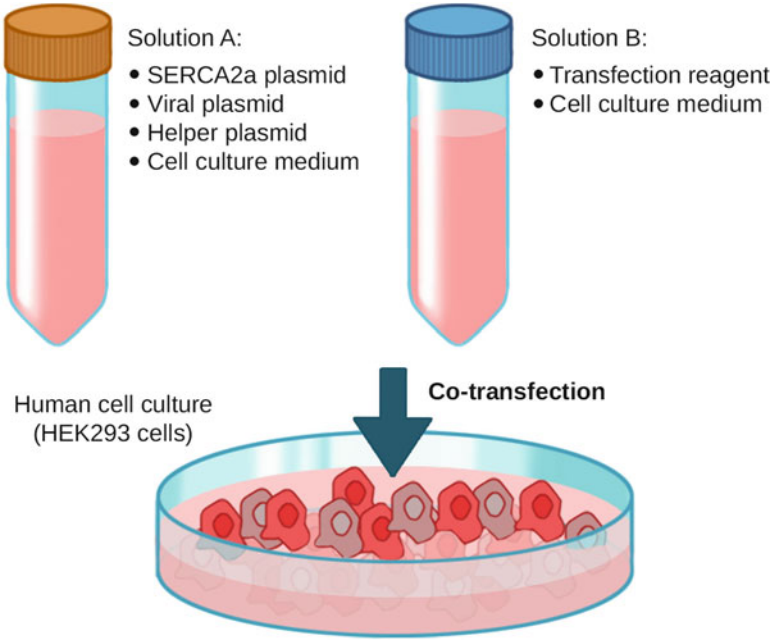


Fig. 4.9 **The co-transfection procedure.** The three plasmids described in Fig. 4.9 are then mixed with a transfection reagent such as a lipid, and then transfected into a suitable cell line, such as HEK293 cells. The cells are lysed and the final virus is purified for future use

Lipid-based Transfection

Transfection is a process that introduces nucleic acids into eukaryotic cells to modify gene expression, for the study of gene function and gene products. This technology was developed to enable nucleic acids such as DNA to enter the cells. Due to their negative net charge, which is the same as that in the cell membrane, nucleic acids cannot enter without external aid. Therefore, researchers need chemical reagents or physical stimuli to allow the nucleic acids to enter the cells. Currently, the most broadly used method is the lipid-based transfection method.

Lipid-based transfection uses cationic lipids which then combine with nucleic acids to form a "transfection complex" which will enter the cell (Fig. 4.10). Although the molecular mechanism is not fully understood yet, it is known that the transfection complex electrostatically interacts with the cell membrane, entering via endocytosis to subsequently free the nucleic acid inside the cell.

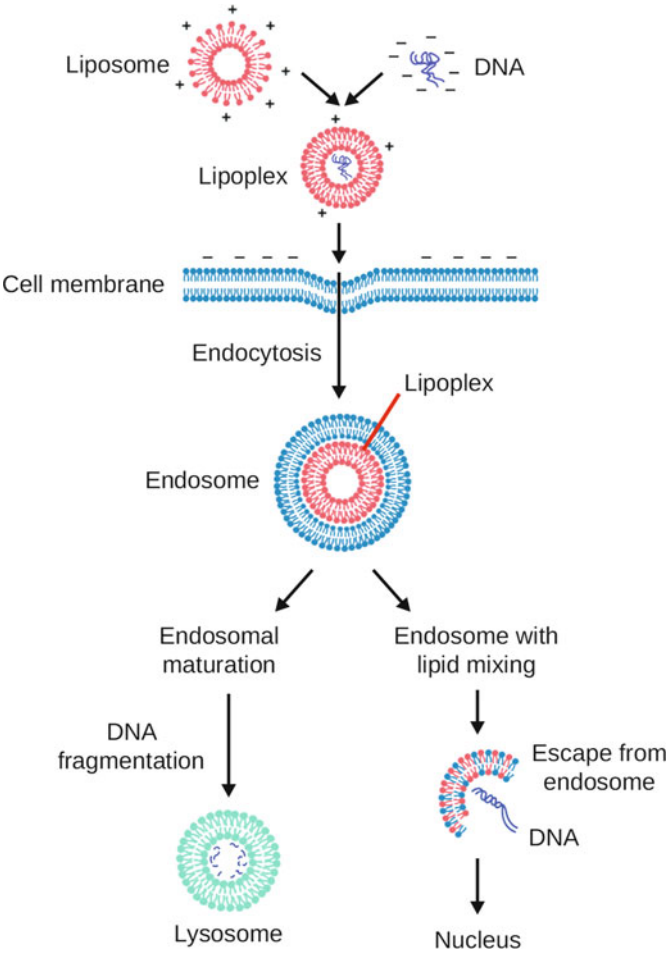


Fig. 4.10 **Complexing of DNA with a transfection reagent.** Lipids can be used to efficiently deliver genetic material such as DNA through the cell membrane. DNA is complexed with a liposome, a spherical structure comprised of a specific lipid. This lipid masks the negative charge of the genetic material thus allowing it to pass through the cell membrane. Once inside the cell the lipoplex can either be degraded by the cells lysosomal machinery, or the DNA can be released inside the cell. The cells usual transcriptional and translational machinery can then produce any proteins encoded for by the DNA molecule

This method is fast and uses easy protocols which do not require media changes. It also achieves high efficiency and expression performance and is applicable to a broad range of cell lines.

Use of Animal Models in Research

Animal models are used in research to study diseases and to test new treatments before given to humans as part of clinical trials.

The criterion used to select the animal model depends on the objective of the study, but one of the most important considerations is the degree of conservation of the studied process in humans. Thus the similarity of the animal model is very important when aiming to apply the conclusions of the experiment to humans.

The most common animal models utilized in research are mice and rats. The genetic and physiological similarities between mice and humans coupled with short generation times and low maintenance costs make it the ideal model organism.

There is a whole catalog of mouse strains that were genetically modified to study certain diseases that would naturally not occur in mice. The data obtained from these studies is an additional advantage of mice over other mammals.

Knock-out Mice

A knock-out mouse is a laboratory mouse in which researchers have inactivated, or "knocked out" an existing gene by replacing it or disrupting it with an artificial piece of DNA. The loss of gene activity often causes changes in a mouse's phenotype, which includes appearance, behavior and other observable physical and biochemical characteristics.

Knocking out the activity of a gene provides valuable clues about what that gene normally does. Observing the characteristics of knock-out mice gives researchers information that can be used to better understand how a similar gene may cause or contribute to disease in humans. Examples of research in which knockout mice have been useful include studying and modeling different kinds of cancer, obesity, heart disease, diabetes, arthritis, substance abuse, anxiety, aging and Parkinson disease. Knock-out mice also offer a biological context in which drugs and other therapies can be developed and tested.

Transmission Electron Microscopy

Electron microscopy (EM) has been proven to be an invaluable tool in the discovery and characterization of viruses because it is the only method to visualize the ultrastructure of viral particles.

In Transmission electron microscopy (TEM), a specific EM technique, the electrons pass through the sample like light through a shadow puppet screen. Dense structures absorb a lot of electrons and create a dark spot on the resulting image, just like the shadow of a puppet blocking the light. A TEM image is always black and white; staining techniques only allow increasing the density of certain structures and thereby making them appear darker. To bundle the electrons, the TEM contains strong magnets that are analogous to lenses in the light microscope. To efficiently illuminate the specimen with an electron beam, the sample slice needs to be very thin, and the body of the TEM has to be evacuated.

4.3 Let's Get Started

Wow! What an adventure, you've learned so much about genetic diseases but we want to end on a positive note. Let's use the infectious power of viruses for good and use them to "fix" disease-causing mutations in our genetic code. The field of gene therapy is incredibly exciting and represents a revolution in healthcare, will you be able to contribute to its development and help design a therapy targeting heart failure?

Techniques Used in the Lab

- Co-transfection of mammalian cells
- Viral vector production
- Electron microscopy

Learning Objectives

At the end of this simulation, you will be able to . . .

- Explain the use of gene therapy for the treatment of heart failure
- Explain the causes of heart failure

- Design a viral-mediated gene therapy approach
- Define "therapeutic gene"
- Describe the anatomy and function of the heart from a healthy person vs. a heart failure patient
- Produce replication defective recombinant adeno-associated virus

ACCESS THE VIRTUAL LAB SIMULATION HERE www.labster.com/springer **BY USING THE UNIQUE CODE AT THE END OF THE PRINTED BOOK. IF YOU USE THE E-BOOK YOU CAN PURCHASE ACCESS TO THE SIMULATIONS THROUGH THE SAME LINK.**

Further reading

Alberts B et al (2015) The molecular biology of the cell, 6th edn. Garland Science, Abingdon
Flint SJ et al (2015) Principles of Virology, 4th edn. Taylor & Francis Inc, Enfield
Hartwell L et al (2015) Genetics: from genes to genomes, 5th edn. McGraw-Hill, Boston
OpenStax (2018) Biology. OpenStax CNX. http://cnx.org/contents/185cbf87. Accessed 1 June 2018
Shareef MA, Anwer LA, Poizat C (2014) Cardiac SERCA2A/B: Therapeutic targets for heart failure. Eur J Pharmacol 724:1–8

Further Volumes in the Labster Virtual Lab Experiments Series

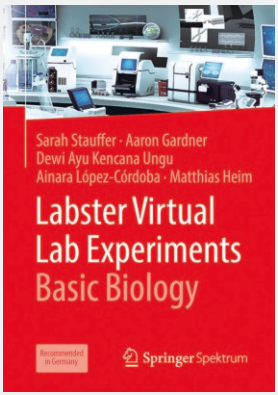
Stauffer S et al. (2018) Labster Virtual Lab Experiments: Basic Genetics, 1st edn. Springer, Heidelberg, ISBN 978-3-662-57999-2

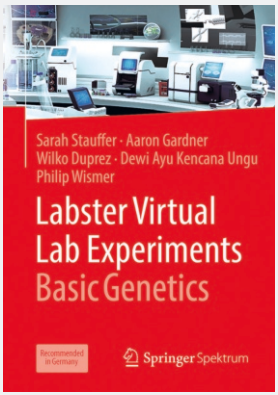
Stauffer S et al. (2018) Labster Virtual Lab Experiments: Basic Biology, 1st edn. Springer, Heidelberg, ISBN 978-3-662-57996-1

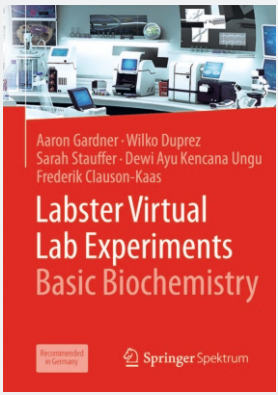
Gardner A et al. (proof stage) Labster Virtual Lab Experiments: Basic Biochemistry, 1st edn. Springer, Heidelberg, ISBN 978-3-662-58499-6

A. Gardner et al., *Labster Virtual Lab Experiments: Genetics of Human Diseases*,
https://doi.org/10.1007/978-3-662-58744-7

Labster Virtual Lab Experiments







- each book includes 4–5 computer simulations with true-to-life challenges
- you can try out cutting edge techniques and equipment in the games
- your perfect way to prepare for exams and lab rotations

S. Stauffer, A. Gardner, D.A.K. Ungu, A. López-Córdoba, M. Heim
Labster Virtual Lab Experiments: Basic Biology
1st ed. 2018, IX, 77 p., 48 illus. in color.
Book + eBook. Textbook for German language market.
eBook
19,99 € (D) | 19,99 € (A) | *CHF 20,00
Hardcover + eBook
29,99 € (D) | 30,71 € (A) | *CHF 32,50
ISBN 978-3-662-57995-4

S. Stauffer, A. Gardner, W. Duprez, D.A.K. Ungu, P. Wismer
Labster Virtual Lab Experiments: Basic Genetics
1st ed. 2018, IX, 66 p., 40 illus. in color.
Book + eBook. Textbook for German language market.
eBook
19,99 € (D) | 19,99 € (A) | *CHF 20,00
Hardcover + eBook
29,99 € (D) | 30,71 € (A) | *CHF 32,50
ISBN 978-3-662-57998-5

A. Gardner, W. Duprez, S. Stauffer, D.A.K. Ungu, F. Clauson-Kaas
Labster Virtual Lab Experiments: Basic Biochemistry
1st ed. 2019, Approx. 70 p., 47 illus., 45 illus. in color.
Book + eBook. Textbook for German language market.
eBook
19,99 € (D) | 19,99 € (A) | *CHF 20,00
Hardcover + eBook
32,81 € (D) | 33,59 € (A) | *CHF 37,99
ISBN 978-3-662-58498-9

* € (D) sind gebundene Ladenpreise in Deutschland und enthalten 7% MwSt; € (A) sind gebundene Ladenpreise in Österreich und enthalten 10% MwSt. CHF sind unverbindliche Preisempfehlungen und enthalten die landesübliche MwSt. Programm- und Preisänderungen (auch bei Irrtümern) vorbehalten. Es gelten unsere Allgemeinen Liefer- und Zahlungsbedingungen.