Sarah Stauffer · Aaron Gardner
Wilko Duprez · Dewi Ayu Kencana Ungu
Philip Wismer

Labster Virtual Lab Experiments
Basic Genetics

Recommended
in Germany

Springer Spektrum

Labster Virtual Lab Experiments: Basic Genetics

Sarah Stauffer · Aaron Gardner ·
Wilko Duprez ·
Dewi Ayu Kencana Ungu · Philip Wismer

Labster Virtual Lab Experiments: Basic Genetics

Springer Spektrum
Labster

Sarah Stauffer
Labster Group ApS
København K, Denmark

Aaron Gardner
Labster Group ApS
København K, Denmark

Wilko Duprez
Labster Group ApS
København K, Denmark

Dewi Ayu Kencana Ungu
Labster Group ApS
København K, Denmark

Philip Wismer
Labster Group ApS
København K, Denmark

ISBN 978-3-662-57998-5 ISBN 978-3-662-57999-2 (eBook)
https://doi.org/10.1007/978-3-662-57999-2

Library of Congress Control Number: 2018961190

Springer Spektrum
© Labster ApS under license to Springer Verlag GmbH 2018
This work is subject to copyright. All rights are reserved by © Labster ApS under license to Springer Verlag GmbH, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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 “Basic Genetics” textbook, which is part of the “Labster Virtual Lab Experiments” series.

This book will help you to learn the fundamental concepts of basic genetics while applying your newly acquired knowledge in a virtual lab environment. In each chapter you will be introduced to one of five virtual lab simulations and the true-to-life missions that you will encounter when playing the simulations. Study the theory section presented in each chapter carefully and you will be fully prepared to master the challenging tasks awaiting you in the virtual lab!

Finally, you will find the 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.

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 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. Sarah Stauffer, Dr. Aaron Gardner, Dr. Wilko Duprez, Dewi Ayu Kencana Ungu, Philip Wismer and Silvia Tjong.

About the Content of this Textbook

Genetics is the study of how living organisms pass on traits or characteristics to the next generation, and the mechanisms that control variations in these traits. The study and manipulation of the genetic code is now a cornerstone of life science research and beyond; from identifying single DNA nucleotides, to reading an organism's entire genome to observing the spread of genes through a population.

Mendelian Inheritance

In the first chapter of this book, you will learn how Gregor Mendel kick-started genetic research in the late 19th-century by painstakingly investigating how specific traits were inherited across generations. Mendel's basic postulates are still widely used today, and in the Mendelian Inheritance simulation you will learn to apply two of them to a real-world situation. Can you help a color-blind patient understand if his children will also be affected?

Polymerase Chain Reaction

After learning how to apply Mendel's postulates you will learn about one of the most important techniques used in modern labs: the polymerase chain reaction (PCR). PCR is a widely used molecular biology technique whereby a single copy (or a small number) of a DNA fragment can be massively amplified for later analysis. PCR is widely used across disciplines but in the Polymerase Chain Reaction simulation you'll help investigators identify a murderer based on their DNA fingerprint.

Animal Genetics

Combining Mendelian inheritance with molecular biology techniques such as PCR is an incredibly powerful way of investigating population genetics, and identifying which genes are linked with specific traits. In the Animal Genetics simulation you will investigate a mutation on a cattle farm which leads to double-muscling. Can you identify in which gene the mutation resides and design a rapid test to detect it?

Gene Expression

While mutation is an important driver of genetic variation, there are many other mechanisms employed by our cell, giving rise to the huge variety of tissues contained in our bodies. It is estimated that our DNA contains approximately 20,000 genes. But not all genes are expressed equally; this variation gives rise to the different tissues of our body, determines variation between individuals and can even lead to the development of certain diseases or disorders. In the Gene Expression simulation, you will perform a cutting-edge next-generation sequencing (NGS) experiment to identify a candidate gene which may be linked with obesity.

Gene Regulation
Finally, you will combine all of your genetic knowledge and learn about how variation in gene expression is a tightly regulated process, with numerous factors increasing or decreasing expression in a highly complex fashion. While incredibly complicated, an understanding of this regulation can lead to some amazing discoveries, such as that described by Shinya Yamanaka whose study of stem cell biology won him the Nobel Prize for physiology in 2012. In the Gene Regulation simulation, you will replicate Yamanaka's groundbreaking work, generating induced pluripotent stem (iPS) cells, then attempt to differentiate these cells for transplantation.

At the end of your time in the virtual lab simulations, you should have a solid understanding of genetics; from its first discovery through to the cutting-edge science being performed today, and the techniques you will learn are fundamental to life science research and beyond.

How to Access the Virtual Simulations?

You can access the five 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 five 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.

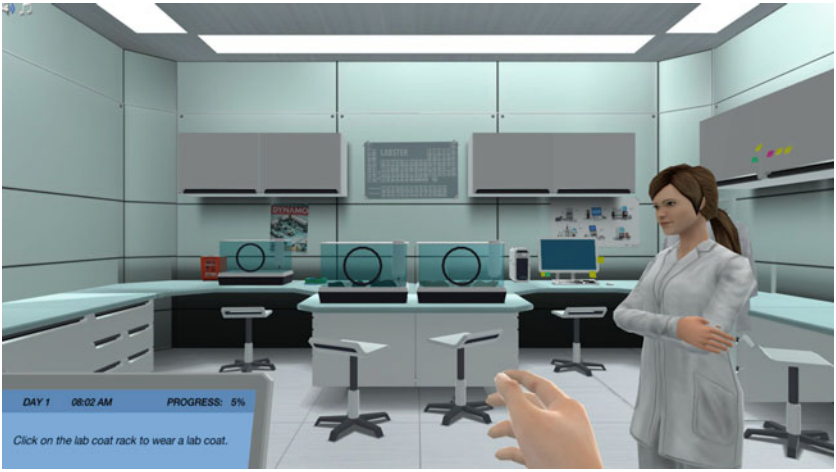
Please be aware that the six month period starts once you sign in for the first time.

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

Contents

1

Mendelian Inheritance



DAY 1 08:02 AM PROGRESS: 5%

Click on the lab coat rack to wear a lab coat.

S. Stauffer et al., *Labster Virtual Lab Experiments: Basic Genetics*,
https://doi.org/10.1007/978-3-662-57999-2_1

1.1 Mendelian Inheritance Simulation

In the Mendelian Inheritance simulation, you will learn the basic principles of Mendelian inheritance. You will see how two of Mendel's postulates can be applied to determine how characteristics are passed from one generation to the next. You will also learn what inheritance looks like at the molecular level. Will you be able to find out why some people inherit color blindness but others don't?

Cross Purebred Mice and Observe Their Phenotypes

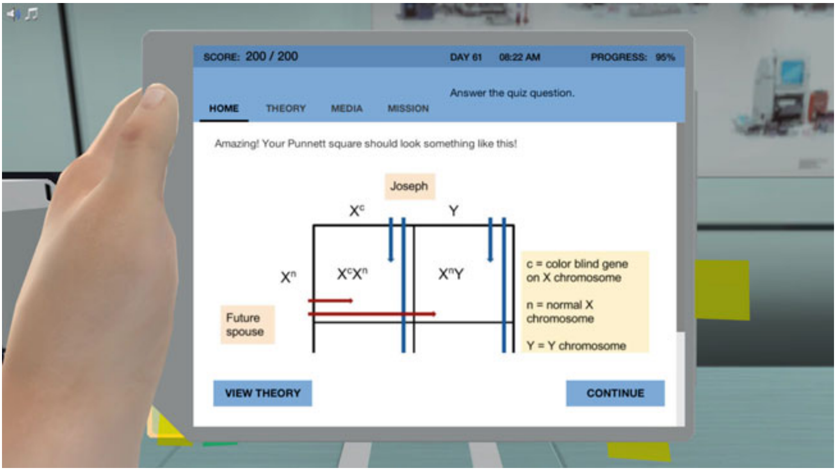
Before investigating the genetics of color blindness in more detail, you will learn step by step how to approach the question. To understand the basic laws of inheritance, you will perform mouse experiments (Fig. 1.1) and observe how specific mouse genes can influence the fur color of the animal. Will you be able to figure out which fur color is dominant, and how this genetic model could apply to color blindness?

Predict the Genotypes in a Patient's Family

With a correct understanding of genetics, we don't always need to perform experiments in the lab. You will discover how Punnett squares and pedigree trees can be used to predict the genetic makeup of a family (Fig. 1.2). Importantly, what is true for the fur color in mice, can also be applied to human color blindness!

Pick up two mice of the same color on Workbench 2 and put them into the Main Cage on Workbench 1.

Fig. 1.1 Breeding experiment setup in the Mendelian Inheritance simulation



Answer the quiz question.

Amazing! Your Punnett square should look something like this!

c = color blind gene on X chromosome

n = normal X chromosome

Y = Y chromosome

Fig. 1.2 LabPad in the Mendelian Inheritance simulation introducing the concept of Punnett squares

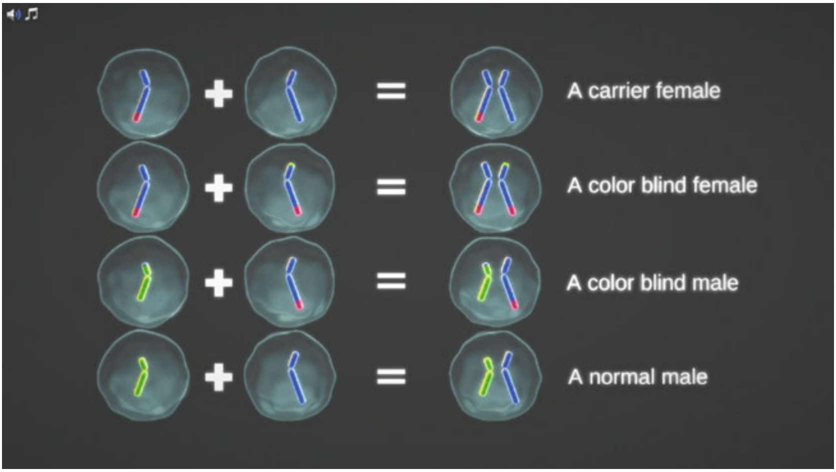


Fig. 1.3 3D animation in the Mendelian Inheritance simulation, visualizing inheritance on the cellular level

Experience Inheritance at the Cellular Level
After you understand the basic principles of inheritance, you will dive deeper into the laws of inheritance as you watch how cells divide to become gametes inside the reproductive system (Fig. 1.3). By unraveling the laws of X-linked inheritance you will discover why color blindness affects more men than women.

1.2 Mendelian Inheritance Theory Content

In the 19th century, Gregor Mendel investigated how specific traits were inherited across generations of pea plants. His pioneering work forms the basis of all modern day genetic research, and the basic model of inheritance now shares his name. The concepts covered in this Mendelian Inheritance theory content will cover all the topics discussed in the simulation, and will give you a firm understanding of the basics of genetics, with later chapters covering specific molecular aspects of genetics.

Genotype

Genotype refers to an organism's DNA, genes and alleles, which determine a particular phenotype, or trait. For example, a mouse may have a genotype which gives rise to a phenotype of black fur rather than brown fur. Or a person may have a genotype which causes color blindness.

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 or environmental conditions, or a combination of the three.

Traits can be observable (for example the color of mouse fur or human eye color) or physiological (for example color blindness or lactose intolerance).

Allele

An allele is one of several alternative forms of the same gene at the same genetic locus. Different alleles can lead to different observable phenotypic traits.

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.

Alleles can also be said to be dominant or recessive to each other. Using mice, for example, fur color may be controlled by a single gene; if the allele for black fur is dominant, a mouse carrying this allele will have black fur regardless of what the other allele codes for. In order to not have black fur, the mouse must have two recessive alleles. We can investigate dominance and inheritance using a Punnett square.

Hybridization

Hybridization refers to the process of mating between two individuals from different breeds which are genetically distinct.

The mating of two purebred parents that express different traits for only one characteristic is called a monohybrid cross. The example in Fig. 1.4 shows a cross between two pea plants with yellow and green seeds, respectively. This technique is often used to breed animals or plants with desirable traits; for example, plants which produce more fruit, or are more tolerant to harsh weather.

Punnett Square

A Punnett square is a visual representation of a genetic cross and is used to make predictions about the outcome of crossing experiments (Fig. 1.5). The maternal and paternal genotypes 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.

Pedigree

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 (Fig. 1.6).

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

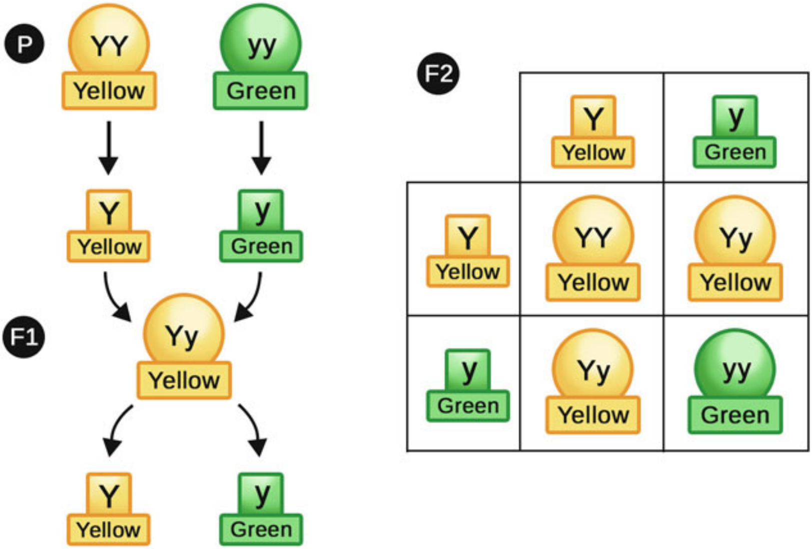


Fig. 1.4 **Monohybrid cross.** In this diagram seed color is determined by the Y/y allele, with the dominant Y allele resulting in yellow seeds. In the absence of the dominant Y allele, seeds will be green. All seeds of the first (*F1*) generation have yellow phenotypes, as all offspring will be Yy, with the yellow coloring driven by the dominant Y allele. In the second generation (*F2*) the ratio between yellow and green seeds becomes 3:1 as depicted in the Punnett square

unaffected individuals. Matings are illustrated by connecting horizontal lines, with their offspring depicted beneath. Each generation is assigned a Roman numeral (for example I, II, III) while individuals of a generation are assigned Arabic numbers (for example 1, 2, 3).

Using such trees we are able to determine the mode of inheritance for phenotypes of interest.

Mendel's Postulates

With his discovery of the basic principles of heredity, Mendel laid the foundations for the development of genetics. Following his experimentation, Mendel set down four postulates or laws of inheritance, some of which still hold true today.

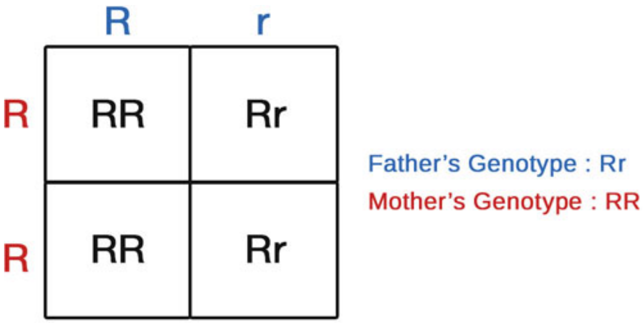


Fig. 1.5 **Example of a Punnett square.** The Punnett square shows an example of a cross between a heterozygous father (*R* and *r*) and a homozygous dominant mother (*R* and *R*). Typically an uppercase letter is used to denote a dominant allele, and a lowercase letter a recessive allele. As offspring 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

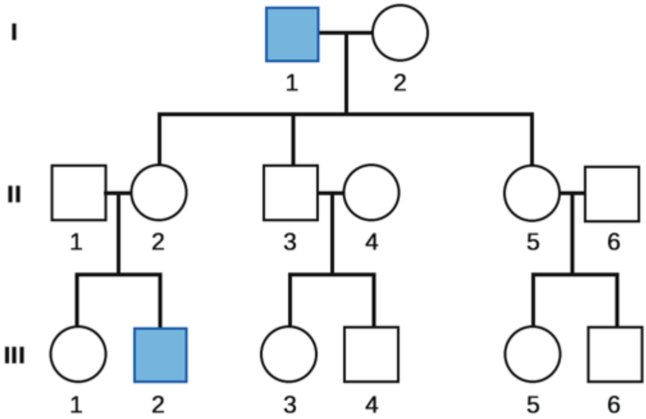


Fig. 1.6 **Example of a pedigree tree.** In this pedigree tree, one can see the inheritance pattern of a genetic disease across three generations. In generation I, the male (I-1) is affected and the female (I-2) unaffected. None of their offspring in generation II are affected, however, a male in generation III is (III-2). Key: *Square* = male, *circle* = female, *blue* = affected, *white* = unaffected

Principle of Paired Factors

A trait within an organism is represented by at least two factors. At its most basic we now know that this refers to the allelic variation of genes for example Yy as seen in Fig. 1.4. However, as our understanding of genetics develops we are finding that most traits are controlled by more than two factors.

Principle of Dominance

When two homozygous individuals with contrasting traits are crossed, the traits which appear in the first generation of offspring are always dominant, and those which do not appear are always recessive.

Law of Segregation

The law of segregation states that genes are distributed into gametes, so that each offspring inherits one allele from each parent.

Law of Independent Assortment

The law of independent assortment states that genes do not influence each other in regards to the sorting of alleles into gametes; every possible combination of alleles is equally likely to occur. Later research has identified that this is in fact incorrect. Genes or regions of DNA close to each other are more likely to segregate together. The technique of linkage analysis which we discuss in a later chapter shows how we can use this association to localize traits to specific genes.

Human Chromosomes

Almost every cell in our body contains 23 chromosome pairs. Of these 23, 22 are matching pairs known as autosomes and are numbered 1–22. The last pair are termed the sex chromosomes (or allosomes) consisting of an X-chromosome and the much smaller Y-chromosome. Women carry two copies of the X-chromosome, whereas men carry one X and one Y-chromosome, a single gene on the Y-chromosome *SRY* is thought to determine sex.

X-linked Inheritance

X-linked inheritance refers to the inheritance of genes that are present on X-, but not on Y-chromosomes (see example in Fig. 1.7). X-linked inheritance can be

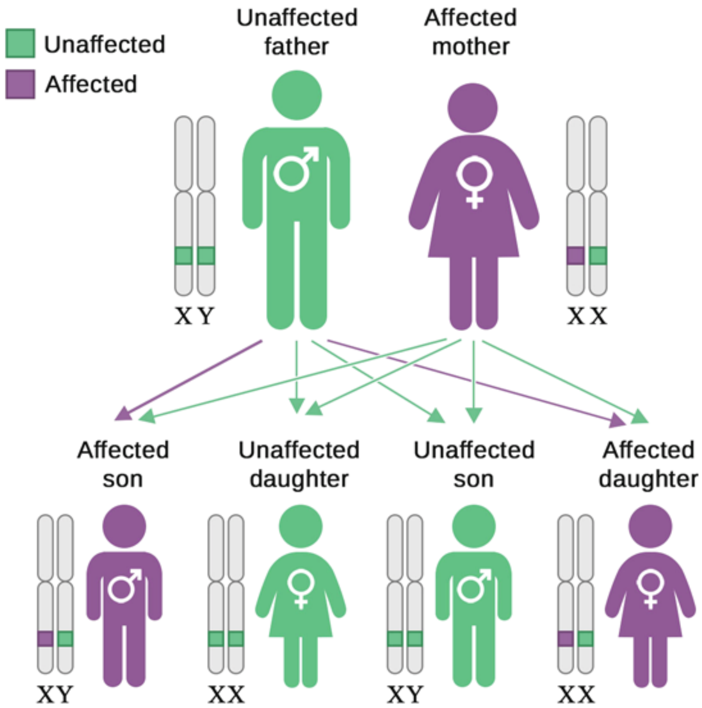


Fig. 1.7 **X-linked inheritance**. In this example, the mother is the carrier of a genetic mutation, located on the X-chromosome, which follows a dominant inheritance to her children

dominant or recessive; in the former instance, males and females are always affected if they carry a single X-chromosome with a dominant allele. In the latter, males will always be affected as they carry only a single X-chromosome, whereas women would have to carry two recessive alleles. As such recessive X-linked disorders are more frequently observed in males.

Those who carry a single allele for a recessive disorder, and are unaffected, are typically referred to as carriers.

Some examples of X-linked traits/conditions are: color blindness, hemophilia, and muscular dystrophy.

Color Blindness

Color blindness is a reduced or inexistent ability to see certain colors or differences in colors. In the most common type of color blindness, the individual is unable to differentiate between red and green. This type of color blindness is caused by recessive X-linked inheritance.

1.3 Let's Get Started

You're now equipped with all the knowledge you need to enter the Mendelian Inheritance simulation. Let's get started right away and test your understanding. Once you're sure you've mastered the basics you'll investigate if Joseph, a young boy with colorblindness, will pass on the condition to his children.

Techniques Used in the Virtual Lab

- Pedigree trees
- Punnett squares
- Breeding mice

Learning Objectives

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

- Explain how traits are passed on from parents to their offspring and what causes variation between siblings.
- Describe Mendel's Laws of Inheritance in color blindness.
- Compare and predict the phenotypes of offspring with given genotypes using Punnett squares and pedigree trees.
- Analyze dominant and recessive alleles, and how they contribute to an individual's phenotype.

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

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-c72e-48f5-b51e-f14f21b5eabd@11.2. Accessed 1 June 2018

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

Polymerase Chain Reaction

2

DAY 1 08:02 AM PROGRESS: 5%

Click on the detective to ask clarifying questions.

S. Stauffer et al., *Labster Virtual Lab Experiments: Basic Genetics*,
https://doi.org/10.1007/978-3-662-57999-2_2

2.1 Polymerase Chain Reaction Simulation

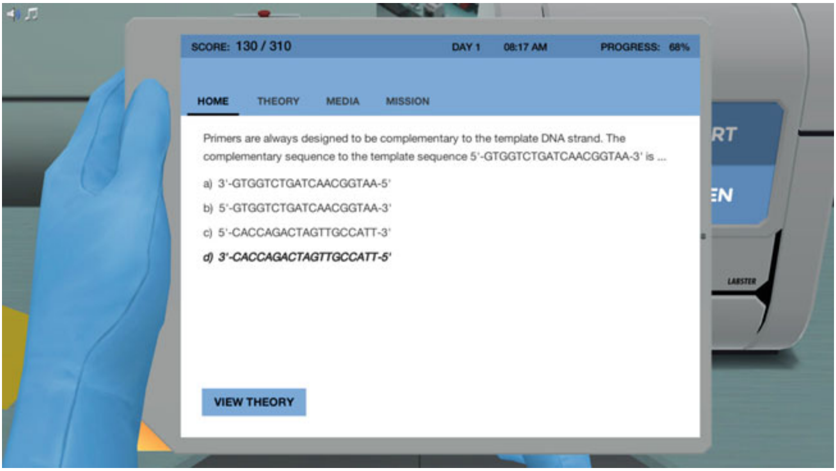
In the Polymerase Chain Reaction (PCR) simulation you will be thrown right into a crime scene where a murder has taken place. To investigate the crime scene your first task will be to collect blood samples in the hope that the murderer has left traces of their DNA.

Analyze the DNA from a Blood Sample

After sampling, you will go to the lab to isolate and analyze DNA from the sample you collected. By using a PCR kit, a thermocycler, and the purified DNA from the crime scene, it will be up to you to mix the correct reagents design the correct primers (Fig. 2.1) perform the PCR experiment.

See the Structure of DNA and Its Replication Up Close

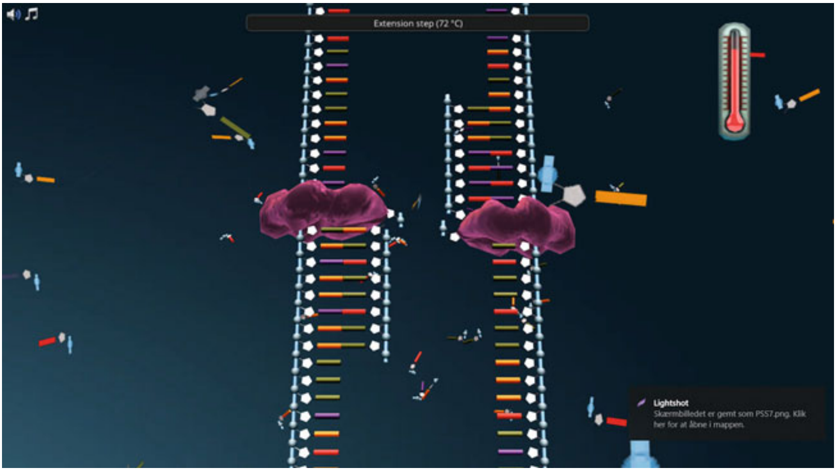
A 3D animation will show the PCR experiment at the molecular level (Fig. 2.2), illustrating the structure of DNA and its replication. Quiz questions will be asked throughout the experimental process, as well as at specific steps of the PCR itself.



Primers are always designed to be complementary to the template DNA strand. The complementary sequence to the template sequence 5'-GTGGTCTGATCAACGGTAA-3' is ...

a) 3'-GTGGTCTGATCAACGGTAA-5'
b) 5'-GTGGTCTGATCAACGGTAA-3'
c) 5'-CACCAGACTAGTTGCCATT-3'
d) 3'-CACCAGACTAGTTGCCATT-5'

Fig. 2.1 Quiz question about primer design in the Polymerase Chain Reaction simulation



Lightshot
Skærmbilledet er gemt som PS57.png. Klik her for at åbne i mappen.

Fig. 2.2 3D animation in the Polymerase Chain Reaction simulation showing the steps of a PCR experiment on the molecular level

Contains:
Purified DNA from Murderer

DAY 1 06:07 AM PROGRESS: 37%

Mix primers, nucleotides, and purified DNA in the PCR tube.

Fig. 2.3 Experimental setup in the Polymerase Chain Reaction simulation to analyze the DNA samples of all suspects

Identify the Murderer

In the final experiment, you will load the sample from the crime scene on a DNA gel, together with samples from all the suspects, and then compare the patterns that emerge (Fig. 2.3). Will you be able to identify the murderer?

2.2 Polymerase Chain Reaction Theory Content

PCR is one of the most widely used molecular biology techniques in labs today. With this technique, a single copy (or a small number) of DNA fragments can be massively amplified for later analysis. PCR is used in all sorts of analyses but in the Polymerase Chain Reaction simulation you'll help investigators identify a murderer based on their DNA fingerprint. Before you get started we need to familiarize you with the PCR theory content. The following terms are important to successfully perform the experiments in the virtual lab simulation and to convict the murderer.

DNA

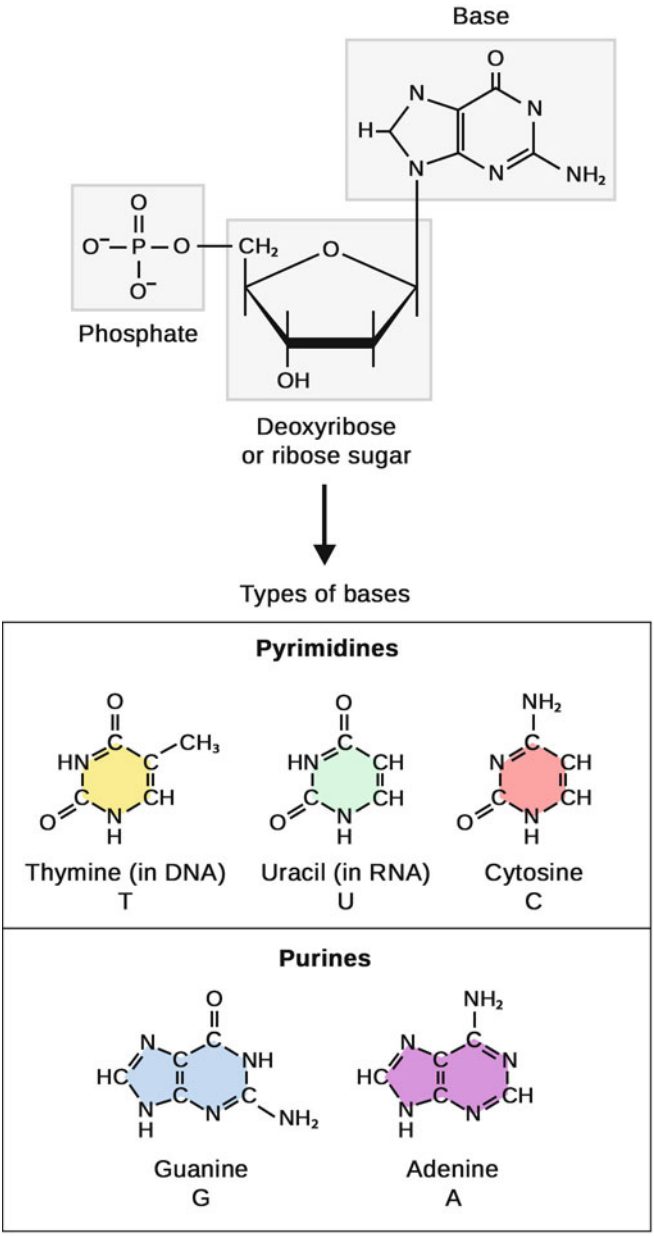
DNA is the hereditary material of all living organisms, including humans. It stores information using a code made up of four nucleotides: adenine (A), thymine (T), cytosine (C), and guanine (G) (Fig. 2.4). It is the order (or sequence) of these four nucleotides that determines the information provided by DNA. The information is used as an instruction for building and maintaining all the cells of an organism. The human genome consists of approximately 3.2 billion base pairs.

DNA Structure

The four bases of DNA are adenine, guanine, cytosine, and thymine. Uracil replaces thymine in RNA. Adenine and guanine are purines, a double-ringed class

Fig. 2.4 **Nucleotides in DNA and RNA.** A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and one or more phosphate groups (top). Bases can be divided into two categories: purines and pyrimidines (bottom). Purines have a double ring structure, and pyrimidines have a single ring

►


Base
Phosphate
Deoxyribose
or ribose sugar
OH
Types of bases
Pyrimidines
Thymine (in DNA)
T
Uracil (in RNA)
U
Cytosine
C
Purines
Guanine
G
Adenine
A

of molecules. Cytosine, thymine, and uracil are all pyrimidines (Fig. 2.4). A phosphodiester link connects the nucleotides with each other. The two strands are bound with weak hydrogen bonds between the complementary nucleotides. The nitrogenous bases are "inside" like rungs on a ladder (Fig. 2.5). Adenine (A) always pairs with thymine (T) and cytosine (C) pairs with guanine (G).

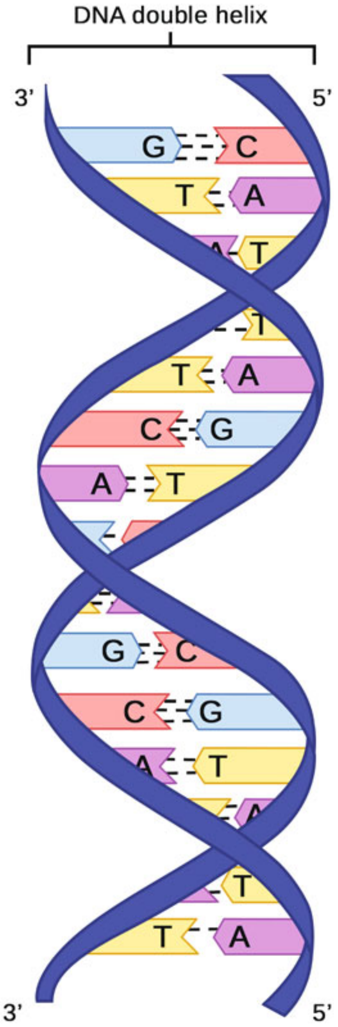


Fig. 2.5 DNA double-helix structure. DNA forms a double-helix composed of two antiparallel strands linked by weak hydrogen bonds between the nucleotides

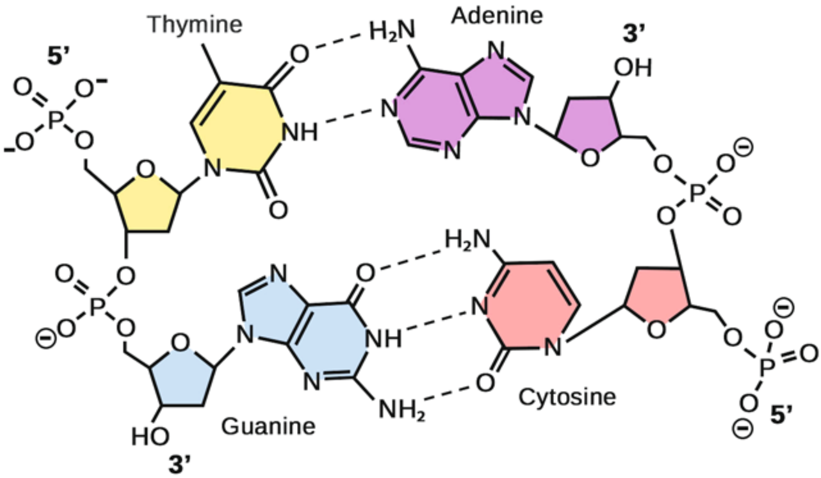


Fig. 2.6 Base-pairs in the DNA. The DNA double-helix consists of pairs of nucleotides, formed by thymine-adenine and guanine-cytosine. Nucleotides on the same strand bind to each-other through the phosphate group attached to the fifth carbon of the sugar molecule, and the hydroxyl group attached to the third carbon molecule forming a phosphodiester link. The complementary nucleotides on opposite strands bind using weak hydrogen bonds

Nucleotides bind to each-other through the phosphate group attached to the fifth carbon of the sugar molecule, and the hydroxyl group attached to the third carbon molecule forming a phosphodiester link (Fig. 2.6). This allows us to apply a direction when we are describing DNA strands, designated with 5′ and 3′, respectively.

The weak hydrogen bonds between complementary nucleotides give DNA its characteristic double helix structure. The double helix is said to be antiparallel because one strand runs in the 5′→3′ direction and the other runs in the 3′→5′ direction (Fig. 2.5).

Collecting a DNA Sample

DNA is stored in the nucleus of eukaryotic cells. This means that DNA can be harvested from any sample containing eukaryotic cells. Usually, a blood sample is collected, and DNA is then extracted from the white blood cells. Red blood cells cannot be used for DNA analysis as they do not contain a nucleus. If blood sampling is impossible or too risky, smaller amounts of DNA can also be extracted from a mouth swab (containing epithelial cells) or from a few hair follicles.

DNA Fingerprinting

DNA profiling is a way of analyzing differences among individuals at the DNA level. This method is also called DNA fingerprinting, because in the same way we have unique fingerprints, we also have unique DNA profiles. Large amounts of our DNA are the same in every person, but some regions, called variable number tandem repeats (VNTRs), consist of long stretches of small repeated sequences. The number of repeats in any of these sequences differs among individuals due to mutations accumulated through previous generations. The unique combination of these variable regions makes up a person's DNA profile. To identify the DNA profile of an individual, one must look at the DNA in these particular regions.

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).

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 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 a 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. 2.7). Therefore, after 30 cycles, at least one billion copies of the target sequence will be present in the tube.

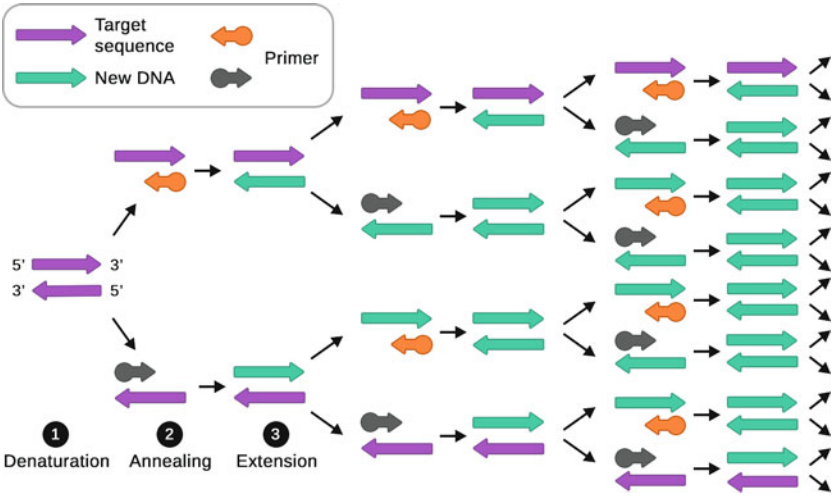


Fig. 2.7 **PCR experiment**: PCR consists of three steps: *1*. Denaturation, *2*. Annealing, and *3*. Extension. The steps are repeated many times (often up to 30 cycles), producing billions of DNA copies of specific regions

PCR

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. 2.7):

1. **Denaturation step (95 °C)**: 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 °C below the primer with the lowest melting temperature [T_m])**: At the annealing temperature the specific 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 °C):** At 72 °C, an enzyme called TAQ 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.

PCR Preparation

To perform a PCR experiment you need:

- Primers
- Nucleotides
- Taq polymerase
- DNA template
- To take extra care to avoid contamination
- A thermocycler

If any of these components is missing no reaction will occur and no DNA will be synthesized.

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 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.

Meticulous primer design is a critical issue to avoid non-specific PCR products. There are several types of software available to assist in primer design that help you to take the main characteristics of the primers into consideration:

- **Location in the genome**: primers have to be complementary to the beginning of the DNA sequence that you are planning to amplify. If you aim to amplify one specific gene it is important to avoid primers potentially annealing to other targets.
- **Melting temperature (T_m)**: primers should anneal to the DNA template at 60 °C. Usually a T_m of 65 °C will be sufficient. The guanine-cytosine (GC) content is very important in defining this temperature.
- **Structure**: secondary structures and complementary pairing between the primers should be avoided to prevent primer-dimers during the experiment.

Nucleotides

As nucleotides are the building blocks of DNA. They are made up of a sugar molecule (deoxyribose in DNA, ribose in RNA), they are needed in a PCR to supply the synthesis of new DNA strands.

Taq Polymerase

DNA polymerases are the enzymes responsible for DNA synthesis in living organisms, and therefore they are designed to be functional at physiological conditions. Hence, the DNA polymerase from most organisms would not work well in a PCR because most polymerases degrade at 90 °C. If such a polymerase were to be used for PCR, it would be necessary to add new DNA polymerase after each cycle. Luckily, Thomas Brock discovered a thermostable DNA polymerase: the Taq polymerase. This DNA polymerase was isolated from the bacterium *Thermus aquaticus* found in the hot springs at Yellowstone National Park. The Taq polymerase is active at temperatures up to 95 °C. This ensures that multiple cycles of PCR can be performed in a single continuous process, and no additional polymerase is required.

Avoid Contamination

When preparing for a PCR experiment, one must be extra careful to avoid potential contamination. PCR is a very powerful and sensitive technique to amplify DNA. This means that the presence of even a tiny contaminant, for example, DNA coming from other samples or even yourself, will also be amplified, competing with the original template, and confusing the results. To prevent contamination, one must always use gloves and work in a very clean environment.

PCR 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 will then be determined by the primers (Fig. 2.8). 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 separated according to their length in a matrix gel, such as agarose. This technique is often used to separate DNA or RNA molecules, for example, in the case of DNA profiling or to study RNA integrity.

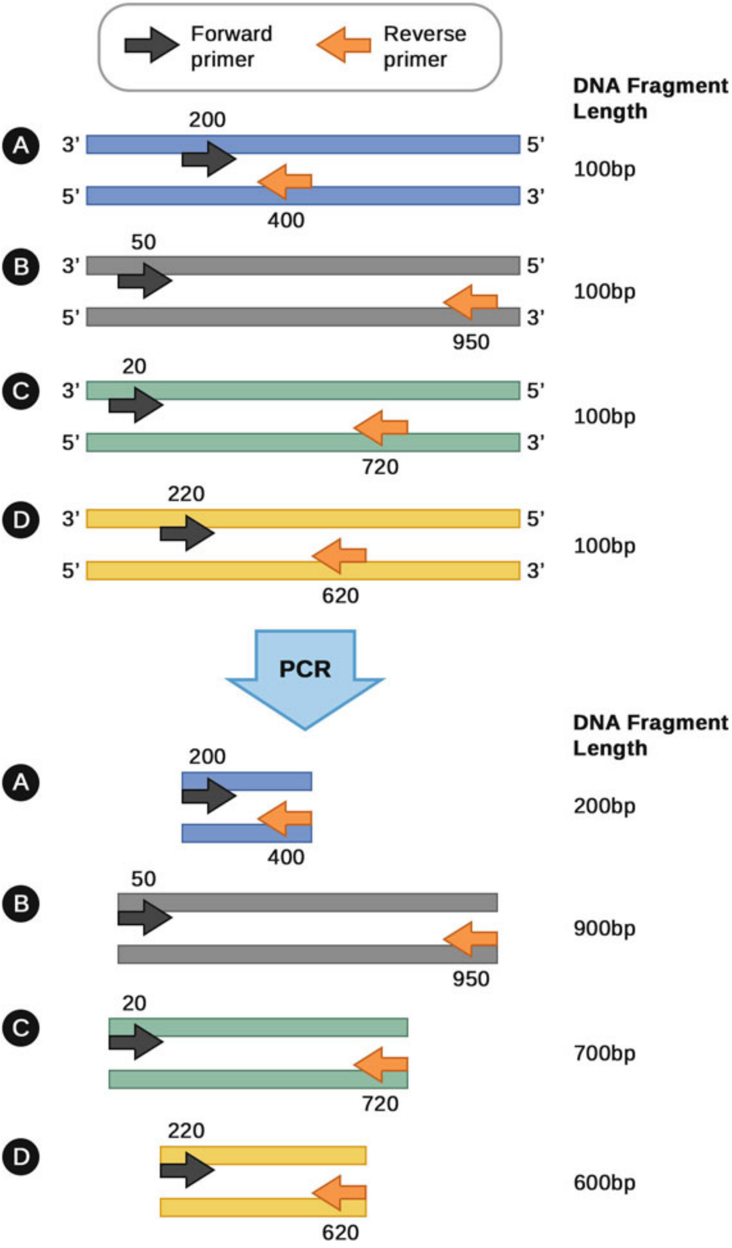

Forward primer
Reverse primer
DNA Fragment Length
A
200
3'
5'
5'
3'
400
100bp
B
50
950
100bp
C
20
720
100bp
D
220
620
100bp
PCR
DNA Fragment Length
A
200
400
200bp
B
50
950
900bp
C
20
720
700bp
D
220
620
600bp

Fig. 2.8 **Examples of 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

◀

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

DNA Fragment Analysis

After the gel electrophoresis of DNA samples is complete, the DNA needs to be stained, for example using a chemical called ethidium bromide, in order to visualize the molecules. Typically, different DNA fragments are visible as bands at specific distances from the top of the gel on the basis of their molecular size (Fig. 2.9a). The smaller the molecular size of a fragment, the faster it travels in the gel, and thus, the longer the distance it migrates (Fig. 2.9b). The exact sizes of the

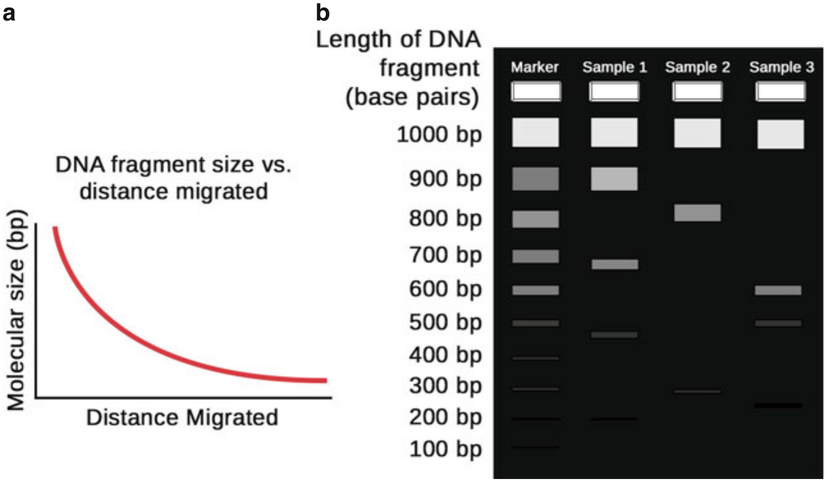


Fig. 2.9 **Agarose gel electrophoresis.** The relationship between DNA fragment size and distance (**a**) and schematic of DNA bands after a gel electrophoresis run (**b**)

DNA fragments in the sample can be estimated by comparing the distance with the molecular weight standard samples (also called a DNA ladder or marker, which is described below).

Besides the molecular size of a DNA molecule, the conformation also has an influence on the migration behavior during gel electrophoresis. Circular plasmids move at different speeds depending on their conformation. A supercoiled plasmid (which is an under- or overwound form of a plasmid) travels faster because it has less friction against the agarose matrix than a linearized plasmid. A mixture of fragments of varying sizes appears as a long smear. If the fragments are too large, as occurs in the case of uncut genomic DNA, they will form a single large band at the top of the gel.

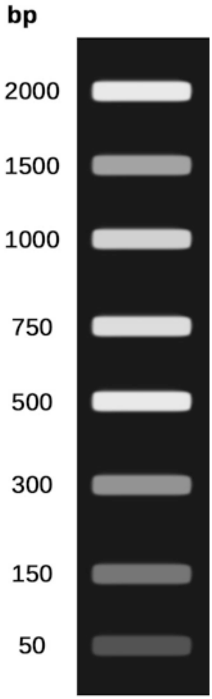


Fig. 2.10 Example of a DNA ladder. A gel electrophoresis experiment is usually run with a adder in one of the wells. Fragment lengths are given in base pairs

DNA Ladder

A DNA "ladder" or molecular weight standard sample is a mixture of DNA fragments with known lengths. It is used as a scale for determining the lengths of unknown nucleic acid fragments when performing gel electrophoresis experiments. The size of the fragments is determined by running a gel with the ladder in a well next to the samples with unknown lengths. The bands showing from the DNA ladder have predetermined lengths as shown in Fig. 2.10.

2.3 Let's Get Started

You're now equipped with all the knowledge you need to enter the Polymerase Chain Reaction simulation. Let's put your new knowledge of PCR to the test. Go to the crime scene and collect the evidence, then identify the DNA fingerprint of the murderer and compare it with those of the suspects. Will you be able to solve the puzzle and help identify the murderer?

Techniques Used in the Lab
- Polymerase chain reaction (PCR)
- Gel electrophoresis
- DNA profiling

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

- Explain the function of DNA polymerase in DNA replication and synthesis
- Perform a PCR experiment using DNA from a blood sample as the template
- Carry out a DNA gel electrophoresis
- Interpret the unique signature of the human genome and the use of tandem repeated regions (TRR) in DNA profiling

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-c72e-48f5-b51e-f14f21b5eabd@11.2. Accessed 1 June 2018
Nicholl DST (2008) An introduction to genetic engineering, 3rd edn. Cambridge University Press, Cambridge
Urry LA et al (2014) Campbell biology, 10th edn. Pearson, Boston

Animal Genetics

3

Collect blood from animal III-3 using the syringe.

S. Stauffer et al., *Labster Virtual Lab Experiments: Basic Genetics*,
https://doi.org/10.1007/978-3-662-57999-2_3

3.1 Animal Genetics Simulation

In the Animal Genetics simulation, you will learn about Mendelian inheritance and how a mutation in a DNA sequence can give rise to an altered phenotype. With the aim of identifying the gene that causes double-muscling in cattle, you will use polymerase chain reaction (PCR) and gel electrophoresis to search for differences in the animals' genomes.

Fill in the Pedigree

As a researcher you will be called to a farm to investigate some double-muscled cattle. For your first task you will fill in a pedigree to help identify the double-muscled cattle (Fig. 3.1). By studying the pedigree, you will learn about heritable traits and whether the inheritance of double-muscling is autosomal or sex-linked and whether it is dominant or recessive. You will then draw a blood sample, and from that, extract DNA and perform further experiments in the lab.

Identify the Candidate Gene

Your next task will be to perform genome scanning, which involves the analysis of the entire genome to identify the candidate gene using genetic markers (Fig. 3.2). By detecting marker patterns and similar gene sequences in the DNA of double-

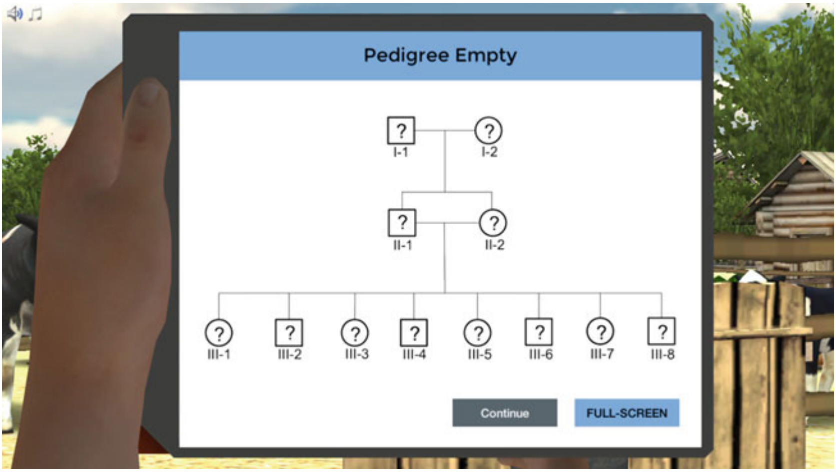


Fig. 3.1 The pedigree task in the first mission of the Animal Genetics simulation

DAY 1 08:25 AM PROGRESS: 81%

Use the scalpels to collect meat from 3 different samples.

Fig. 3.2 Experimental setup in the Animal Genetics simulation to identify the candidate gene responsible for double-muscling in cattle

DAY 1 08:24 AM PROGRESS: 81%

Use the scalpels to collect meat from 3 different samples.

Fig. 3.3 Gel-electrophoresis experiment in the Animal Genetics simulation muscled cattle, your goal will be to narrow down the genomic location of the candidate gene. Experimentally, this will be achieved by amplifying the markers using PCR and analyzing the PCR product size using gel electrophoresis (Fig. 3.3).

Develop a DNA Test
After identifying the gene responsible for double-muscling, your last task will be to develop a DNA test. This test will be used to determine the origin of organic meat; as it is prohibited to label meat as organic if it is derived from double-muscled cattle. You will apply your newly developed DNA test on three meat packages from an organic farm.

Will you be able to use your understanding of animal genetics to find the gene which causes double-muscling and develop a reliable commercial test for organic meat?

3.2 Animal Genetics Theory Content

Mendelian and molecular biology techniques, such as PCR, are often used together to investigate populations, and identify genes linked with specific traits. In the Animal Genetics simulation you will investigate a mutation on a cattle farm which leads to double-muscling. But, before you can identify the gene in which the mutation resides and design a rapid test to detect it you will need to familiarize yourself with the Animal Genetics theory content.

Hereditary Traits

Offspring can inherit traits if they have a genetic basis. If a physical trait is more frequently observed in some families/breeds than others, it is a sign the trait is genetic and hereditary.

To rule out other causes of the trait, environmental factors (for example a shared diet or similar housing) should be examined and excluded. The next step is to collect family and phenotype data to construct a pedigree tree. This will give an overview of the inheritance pattern.

Mode of Inheritance

When studying specific hereditary traits, four major modes of inheritance have to be considered:

- Autosomal recessive
- Autosomal dominant
- X-linked recessive
- X-linked dominant

The mode of inheritance depends on whether the genetic trait is dominant or recessive, and if it is carried on the autosomes or the X chromosome.

Autosomal Recessive

An autosomal recessive trait is a recessive trait linked to an autosome (a chromosome that is not a sex chromosome). The main characteristics observed for these kinds of traits are:

- The trait may jump generations
- Affected individuals are often inbred
- Males and females are equally affected

Autosomal Dominant

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

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

X-linked Recessive

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

- The trait may jump generations
- Males are more frequently affected, as they only have one X chromosome

X-linked Dominant

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

- All generations have affected individuals
- Affected males pass the trait on to all their daughters

Genome Scanning

Easily detectable DNA markers that are linked to an inherited trait can be identified using a technique known as genome scanning. The location of these markers indicates the chromosomal region where the gene responsible for the trait lies. This narrows the search down to a fraction of the genome.

The process of identifying gene markers that co-segregate with an inherited trait is called linkage analysis.

Linkage Analysis

Linkage analysis identifies loci that co-segregate in a pedigree, typically because they are closely located in the genome. It can be performed using two traits, or between DNA markers and a locus involved in an inherited disease.

Today, linkage analysis is typically performed to identify a candidate region and thereby focuses the search for a candidate gene to a limited chromosomal region.

Short Tandem Repeat Markers

Short tandem repeats (STRs) are polymorphic sequences found throughout the mammalian genomes and are often used as DNA markers to compare loci between different DNA samples.

The core repeats in an STR can be di- (two bases, CA) tri- (three bases, CAG) or tetra-nucleotides (four bases, ATGC). These core sequences are repeated a number of times after each other (for example, CACACACACACA or (CA)6).

Different alleles frequently contain different numbers of repeats in the individual STR loci as these STR regions are more susceptible to mutation. In a population, multiple alleles can be present, but in an individual (diploid), there is a maximum of two different alleles; one on each homologous chromosome.

The STR-loci are flanked by unique sequences, and therefore the individual STR-loci can be amplified by PCR and the sizes of the alleles estimated by gel electrophoresis. In this manner, the genetic makeup of two or more different DNA samples can be compared; a process that is also called genotyping.

Types of Mutations

Mutations are changes in the nucleotide sequence as compared to a normal (also called wild-type) sequence. Depending on where a mutation is located it in the genome it may have no, or varying effects on the phenotype of an organism.

Mutations can be categorized according to how they change the order of bases:

- **Substitution**: one base is replaced by another
- **Deletion**: one or more bases are deleted
- **Insertions**: one or more bases are inserted

In addition, mutations located in coding sequences can be categorized according to their effect on the corresponding protein:

- **Synonymous/silent mutation**: a substitution that does not change the amino acid
- **Non-synonymous/missense mutation**: a substitution that changes the amino acid
- **Nonsense mutation**: generates a stop codon
- **Frameshift mutation**: insertion or deletion not divisible by three that alters the reading frame

DNA Sequencing

DNA sequencing is a technique used for “reading” the precise order of nucleotides of a DNA strand. Small DNA fragments, whole genes, or even genomes can be sequenced. DNA sequencing can be performed, 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.

The most widely used DNA sequencing technique is based on chain-termination developed by Frederick Sanger. This method is very similar to a PCR, but it involves only one primer, which anneals close to the region of interest at the 3′

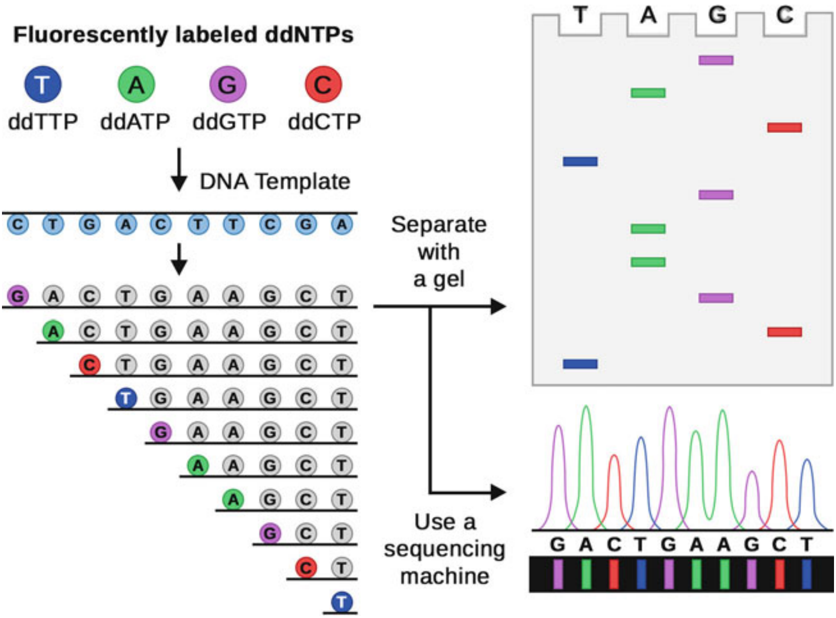


Fig. 3.4 Schematic overview of Sanger (chain-termination) method for DNA sequencing. DNA fragments of all possible lengths are produced during the elongation cycles, each terminated with a fluorescently-labeled stop nucleotide. When the fragments are separated according to their size on a gel, the order of colors detected will indicate the precise DNA sequence

end of the DNA template. During the sequencing reaction, a mixture of normal nucleotides (dATP, dTTP, dCTP, and, dGTP) and "stop"-nucleotides (ddATP, ddTTP, ddCTP, and ddGTP) are added to the DNA template (Fig. 3.4). The stop nucleotides are each labeled with a specific fluorescent color. In every cycle the target DNA is replicated by a DNA polymerase until a stop-nucleotide is added, which prevents 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.

DNA Test

A DNA test for a known disease-causing mutation is used to determine the genetic status (sick/healthy/carrier) of an individual. A DNA test is usually expected to be used for testing many DNA samples, and it should therefore be cheap, reliable, and fast. PCR in combination with gel electrophoresis is often used as diagnostic tests. If the mutation is a deletion or an insertion, the mutated DNA region is amplified by PCR. The size of the product is then estimated by gel electrophoresis revealing whether the product is longer/shorter as compared with the wild-type. Two known control samples (sick and healthy) should be included in all tests for comparison.

3.3 Let's Get Started

Combining Mendelian genetics with PCR? You're really becoming an expert in genetics and are now equipped with all the knowledge you need to enter the Animal Genetics simulation. Do you think you'll be able to work all the way through the experiment? From constructing a pedigree to scanning a genome; will you be able to detect the double-muscling mutation?

Techniques Used in the Lab
- PCR
- Gel electrophoresis
- Pedigree/linkage analysis

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

- Explain different hereditary traits and modes of inheritance
- Construct a pedigree analysis based on observed phenotypes
- Perform genome scanning to identify candidate genes for double-muscling in cattle
- Develop a DNA test for double-muscling in cattle

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

OpenStax (2018) Biology. OpenStax CNX. http://cnx.org/contents/185cbf87-c72e-48f5-b51e-f14f21b5eabd@11.2. Accessed 1 June 2018

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

4

Gene Expression

DAY 1 08:04 AM PROGRESS: 14%

Go to Workbench 1 to begin with the experiments

S. Stauffer et al., *Labster Virtual Lab Experiments: Basic Genetics*,
https://doi.org/10.1007/978-3-662-57999-2_4

4.1 Gene Expression Simulation

In the Gene Expression simulation, you will take part in an exciting project about identifying new genetic factors to explain why obesity affects some individuals more than others.

Prepare Samples for a Gene Expression Analysis

In your first mission, you will collect fat-tissue samples from lean and obese pigs (Fig. 4.1) and isolate messenger RNA (mRNA), the direct product of gene expression (Fig. 4.2). In the next step, you will synthesize DNA that is complementary to the isolated mRNA, also called cDNA, for further analysis by next-generation sequencing (NGS).

Perform an NGS Experiment

With your cDNA samples at hand, you will perform an NGS experiment and learn all the important steps of this high-tech molecular biology procedure. Immersive 3D animations will show what is happening at the molecular level inside the NGS machine to better understand the technology (Fig. 4.3). Finally, you will learn how to analyze the sequencing data produced and evaluate the quality of your samples.

DAY 1 08:03 AM PROGRESS: 9%

Get samples from both a lean and an obese pig.

Fig. 4.1 Sample collection in the first mission of the Gene Expression simulation

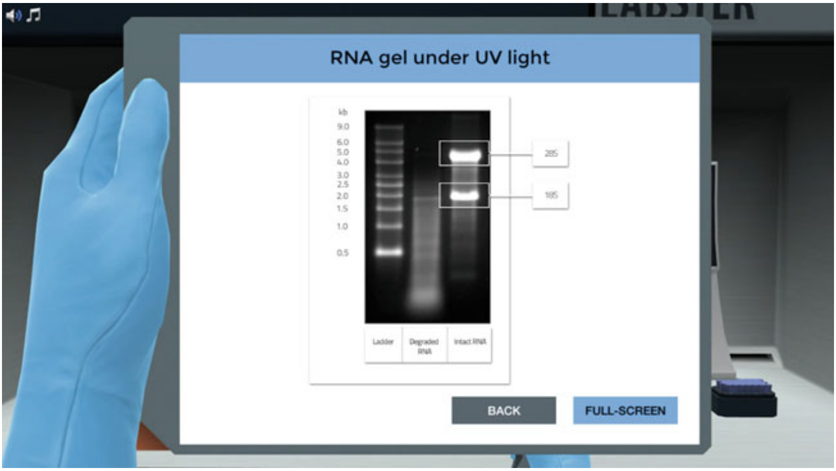
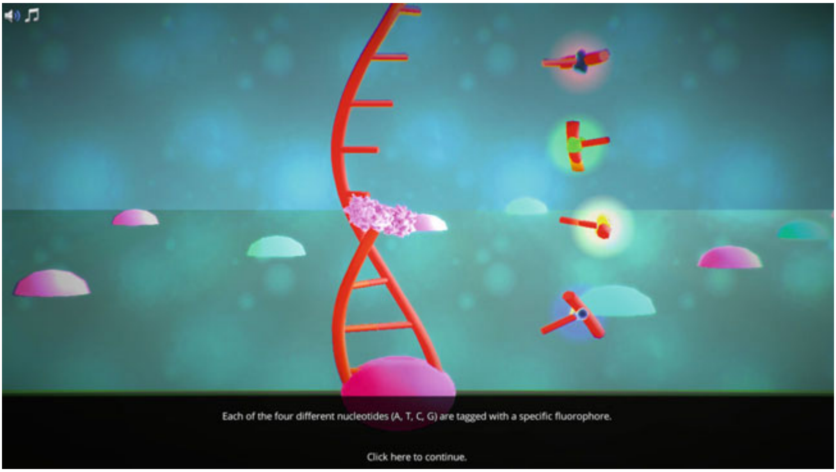


Fig. 4.2 RNA gel in the Gene Expression simulation


Each of the four different nucleotides (A, T, C, G) are tagged with a specific fluorophore.

Click here to continue.

Fig. 4.3 3D animation in the Gene Expression simulation, visualizing NGS at the molecular level

This will give you an understanding of how mRNA relates to the gene expression levels and how this can affect obesity.

Validate Your Results
After identifying a candidate gene involved in obesity, you will be tasked to set up a quantitative PCR (qPCR) experiment to validate your results. You will prepare the qPCRs, obtain the amplification and melting curves, and normalize the output to compare gene expression levels correctly. Will you be able to find a gene that is responsible for obesity in pigs?

4.2 Gene Expression Theory Content

Almost every cell in our body holds our full genetic code, which contains approximately 20,000 genes. However, not all of these genes are expressed equally; this variation in expression gives rise to the different tissues of our body, determines variation between individuals and can even lead to the development of certain diseases or disorders. In the Gene Expression simulation, you will perform a next-generation sequencing experiment to identify changes in the level of expression of a gene which may be linked to obesity. Before we get started you will need to understand the Gene Expression theory content which underpins these cutting-edge techniques.

Obesity and Genetics

Obesity has become a worldwide epidemic, and dramatic changes in lifestyle (increased caloric intake and decreased exercise) are without a doubt the most important drivers behind the trend. However, over the past few decades it has become evident that genes can also influence the development of obesity, and increasing effort is being made to identify genes and genetic variants of relevance.

It has been found that mutations in specific genes may cause an increased tendency for obesity. In addition, an alteration in the expression level of genes may affect the development of obesity. Therefore, both mutation studies and expression studies are relevant when investigating the genetic background of obesity. In particular, searching for genes that are differentially expressed in obese and lean individuals can give valuable data because these genes may play an important role in obesity.

The Pig as a Model

When studying obesity, the pig is preferred as a model as it resembles humans in several ways: body and organ size, diet composition, eating habits, physiology, and genome. Moreover, the use of an animal model provides the possibility of controlled mating, diet, and unlimited sampling; which are the usual limitations in human studies.

Gene Expression

Every cell contains a complete genome, but only a subset of genes is expressed at a specific time. The set of genes that are expressed gives the cell its characteristics and functionalities. This is a dynamic process, with the expression of specific genes being turned on and off in response to internal and external stimuli.

A gene is considered expressed when it is transcribed and a functional product produced. The direct product of transcription is an RNA molecule: messenger RNA (mRNA), regulatory RNAs such as micro-RNA (miRNA), or structural RNAs, such as ribosomal RNA (rRNA). Regulatory and structural RNAs are already functional products while mRNAs require translation into a protein. In the case of protein-coding genes, the quantity of mRNA transcribed reflects the level of gene expression.

To measure gene expression levels, different techniques can be used, for example, next-generation sequencing or quantitative PCR.

RNA Isolation

Typically, before extracting mRNA as a direct product of gene expression, the total RNA needs to be purified from a sample. The first step in an RNA isolation protocol is to separate the RNA from the rest of the cellular contents.

A common method used to isolate total RNA is guanidium thiocyanate-phenol-chloroform extraction. This method is based on guanidinium isothiocyanate, a powerful protein denaturant, and on the organic solvent phenol. In addition, a buffer is used that maintains RNA integrity, disrupts cells, and denatures proteins and is based on the research of Piotr Chomczynski and Nicoletta Sacchi. It takes slightly more time than column-based approaches, but it can handle larger amounts of tissue or cells and therefore yields more RNA.

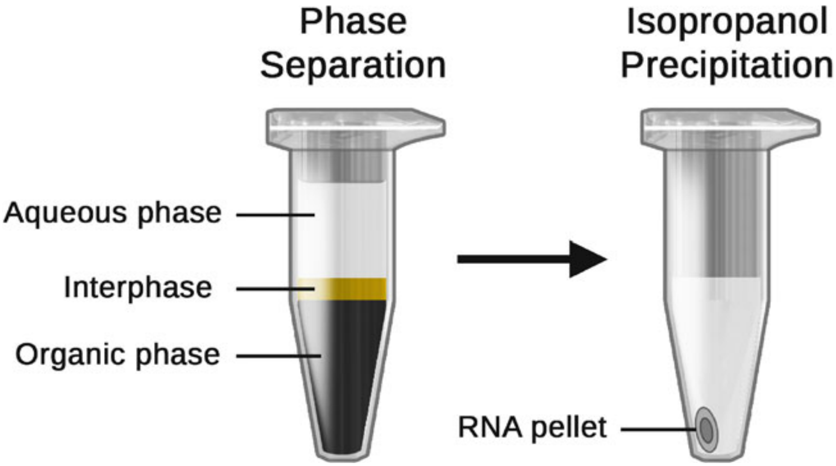


Fig. 4.4 **RNA isolation.** The polar RNA will remain in the aqueous phase upon chloroform addition and phase separation (*left*). Following precipitation with isopropanol and centrifugation, an RNA pellet will form (*right*)

The Main Steps of RNA Isolation Are as Follows:

- Cell lysis and disruption of cellular structures
- Separation of RNA from cell debris
- Purification of RNA from DNA and proteins
- Precipitation of RNA
- Wash and re-suspension of pure RNA

The cell lysate consists of RNA, DNA, proteins, and cellular debris. Centrifugation can be used to separate the macromolecules (proteins, RNA, and DNA) from the cell debris. Phase separation of the sample will occur when chloroform is added to the sample, based on the different densities of the cellular debris. The polar RNA will remain in the upper aqueous phase, DNA will accumulate in the interphase, and the denatured proteins will dissolve in the lower organic phase (Fig. 4.4, left).

The RNA from the aqueous phase is precipitated by adding isopropanol, which precipitates polar molecules like RNAs and salts (Fig. 4.4, right). The salts can be washed away with 75% ethanol to obtain pure RNA. Finally, the RNA pellet is dried and re-suspended using RNase free water as a buffer.

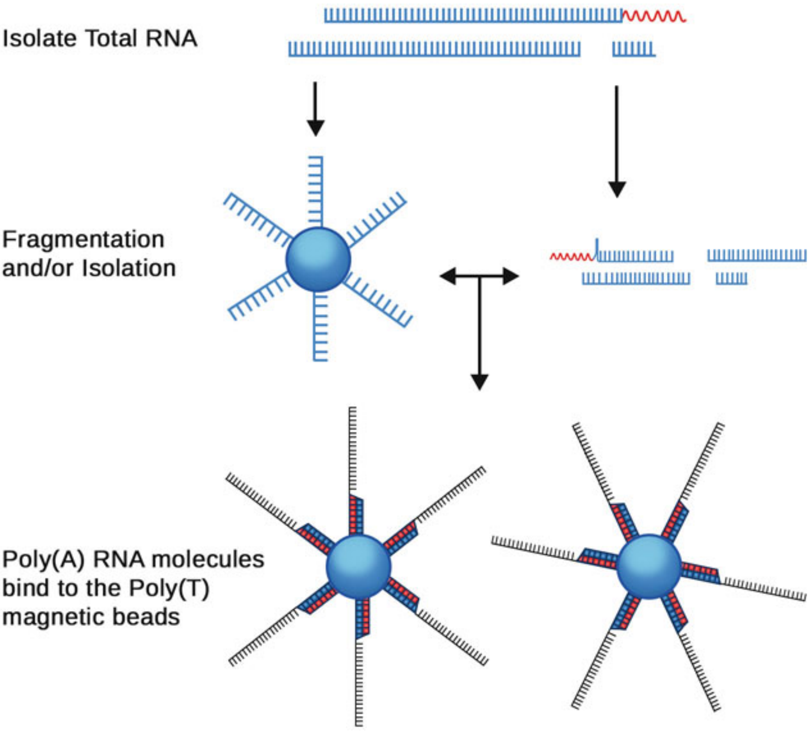


Fig. 4.5 **mRNA purification using magnetic beads.** By using poly(T) coated magnetic beads, mRNA containing a polyA tail can be specifically captured from an RNA mixture

mRNA Purification

The pellet obtained during RNA isolation, as described here, contains the entire RNA content of the cells: mRNA, but also miRNA, siRNA and rRNA for example. For gene expression studies, we need to specifically isolate the mRNA, it being the direct product of protein-coding gene transcription. There are various methods that can be used to purify mRNA from the total RNA in an extract. For example, using column separation or magnetic beads (Fig. 4.5).

Magnetic beads coated with oligo(dT) primers bind to the mRNA's polyA tail, a sequence of adenine nucleotides at the 3′ end of every mRNA molecule. When placed in a magnetic rack, the beads with the attached mRNA will stick to the tube walls. One can then easily remove the supernatant containing all the unwanted material (including the other subtypes of RNA), and wash the beads using 80% ethanol. After several rounds of washing the mRNA molecules can be eluted from the magnetic beads and transferred to a new tube.

Complementary DNA (cDNA) Synthesis

When investigating protein-coding regions of the genome, working with mRNA presents some challenges. Most importantly, RNA is easily degraded due to ubiquitous RNases. Therefore, reverse transcription from mRNA to the complementary DNA (cDNA) creates a more stable and reliable sample to analyze, for example by PCR. cDNA is synthesized by the enzyme reverse transcriptase (RT) from the mRNA template. Because it comes from mRNA, cDNA only includes the expressed sections of the gene (exons), compared to genomic DNA which contains the entire gene sequence.

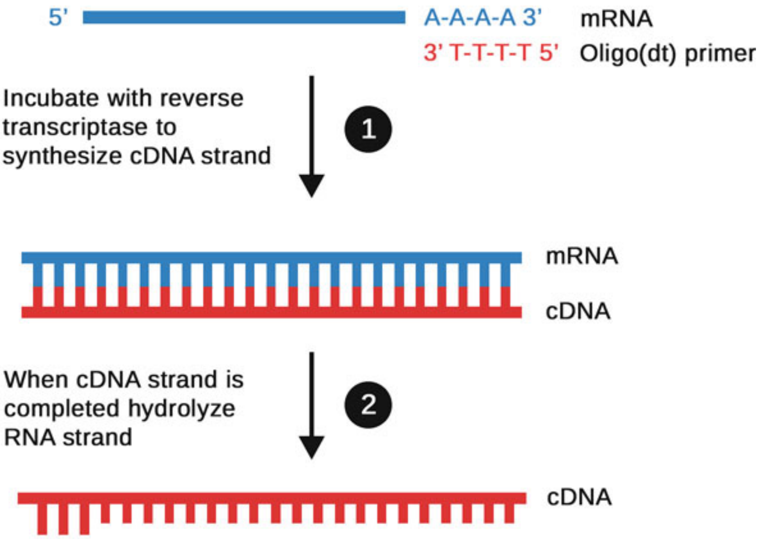


Fig. 4.6 **Steps to create cDNA from a mRNA template.** Using an oligo(dT) primer, mRNA is reverse transcribed into cDNA, followed by hydrolyzation of the RNA strand

The primer used for cDNA synthesis from mRNAs is usually an oligo(dT) primer, which is able to anneal to the polyA tail of mRNA molecules (Fig. 4.6). Random hexamer primers are also sometimes used, which bind randomly to any RNA in the sample depending on the hexamer's sequence, and thus all RNA molecules are reverse transcribed, not just mRNA. After cDNA synthesis, a DNA-RNA hybrid is formed. Finally, the RNA is digested by RNaseH resulting in a single-stranded cDNA molecule that can be directly used in any PCR-based technique.

In theory, each RNA strand is reverse-transcribed to cDNA (with a 1:1 ratio), so the final amount of cDNA still reflects the quantity of mRNA in the initial sample, and therefore the level of gene expression. The cDNA is then ready to be sequenced, for example using NGS techniques.

Next-generation Sequencing (NGS)

NGS is an advanced DNA-sequencing technology. In contrast to the "first-generation" sequencing methods, such as Sanger sequencing, NGS enables high-throughput sequencing of entire genomes at once.

In NGS, many short DNA molecules are sequenced at the same time in an automated process. These short DNA sequences are then assembled by aligning their sequence to a reference sequence (reference genome), thereby revealing the complete DNA sequence of a genome. Follow the key steps of NGS close up in a series of 3D molecular animations.

NGS is a very powerful technique that can be used for many different applications, such as:

- Single nucleotide polymorphism (SNP) discovery
- RNA-sequencing
- Whole genome sequencing
- Detection of genetic aberrations (such as mutations or chromosomal rearrangements)
- Detection of DNA methylation patterns

NGS Experimental Procedure

There are many different NGS platforms, and each of them utilizes a different technique to sequence DNA. Here, we will focus on the "reversible dye termination technology". In this method, the DNA is first fragmented into small sizes, and two short DNA molecules called "adapters" are ligated to each end of the sample. These adapters will function as primer-docking sites to amplify the DNA during PCR and to bind to the sequencing chip. Before analyzing the data, the adapters need to be removed because they are not biological sequences.

Because of the need for the "adaptors", the extracted cDNA needs to be processed in preparation for sequencing. The different sample preparation steps are outlined below:

- Fragmentation
- End-repair
- A-tailing
- Adapter ligation
- PCR amplification

The sequencing process itself begins with the single-strand cDNA molecules capped with adapters and primers being attached to the surface of a slide (called the flow cell) and amplified with the polymerase enzyme creating local clonal DNA colonies. These DNA colonies are also referred to as DNA clusters. Each of the clusters contains exactly the same DNA sequence, hence the term clonal DNA colonies. In a similar fashion to first-generation sequencing, the complementary strand to the single-stranded DNA is then built, using nucleotides individually labeled with a different a fluorescent dye and $3'$ blocker. After the addition of the corresponding nucleotide onto the complementary strand, the elongation stops due to the blocker and an image of the fluorescence is captured, thus identifying the nucleotide. Following this, the blocker sitting at the $3'$ terminal is chemically removed from the DNA, allowing the next cycle of nucleotide addition to start.

NGS Data Analysis

When analyzing NGS data, you first need to ascertain the reliability of your sequencing experiment, which means checking the probability that the nucleotide at each single position is the correct one. The overall reliability of sequenced DNA fragment is usually evaluated with a scoring system. Second, you need to remove all the potential artifacts: for example the adapters added during NGS preparation need to be removed because they have no biological meaning and were inserted purely for the sequencing process.

Finally, you need to align the small DNA fragments together (their sequence would usually overlap, so that you can reconstruct longer fragments) and map these fragments back to a reference genome, thus identifying their gene of origin. The quantity of fragments mapped back to a single gene will provide information on the level of expression of that specific gene. You can thus compare the level of expression of a specific gene across several samples, for example between the fat and lean pigs in the case of obesity covered in the Gene Expression simulation.

Sequencing is a great way to measure gene expression, but solid data should be validated using a different technique, for example by quantitative PCR (qPCR).

qPCR

qPCR is a technique that monitors the accumulation of a DNA product from a PCR after each cycle using fluorescent dyes. It allows quantification of the initial number of copies of a particular type of DNA in a sample. This technique benefits from high sensitivity, specificity, reproducibility, repeatability, and accuracy.

qPCR is often used to quantify RNA transcripts and thus the expression of a gene of interest. This method is called reverse transcription (RT)-qPCR.

Primer Design for qPCR

qPCR requires accurate preparation beforehand. The primers need to be carefully designed and effective controls incorporated in the experiment. It is very important to avoid primer-dimers and other non-specific PCR products, because they will incorporate the fluorophore, leading to an inaccurate quantification in qPCR experiments. In addition to the usual considerations when designing primers, there are some important guidelines for designing primers for qPCR:

- Design primers that give a product (amplicon) of 100–200 bp (optimally 100 bp). Primers should be around 20 bp long.
- If possible, design primers to span exon-exon junctions (in two different exons) in order to prevent issues with possible genomic DNA contamination. These primers may detect some splice variants, but not others. If this design is not possible, it is very important to treat the RNA sample with DNase prior to reverse transcription.

Reference Genes for qPCR

Reference genes, also called housekeeping genes, are usually genes essential for the cell because they are required for basic cellular functions. Reference genes are expected to be expressed at relatively constant levels in most of the cells of an organism under normal and disease conditions, and thus can be used as controls in qPCR experiments.

Differences in RNA expression level between two samples can be the result of both true biological variation or technical variation. It is therefore very important to minimize the technical variation by using proper normalization against stable reference genes. Because reference genes are expected to provide the same results across qPCR experiments, you can evaluate if a difference of expression level for your target gene is due to biological or technical variation by testing if your reference gene data falls into the expected range. This procedure will allow a better appreciation of the true biological variation.

In fact, many factors can contribute to variability in the sample analysis, but it is usually related to the reverse transcription efficiency for each sample. The more contaminants remaining in your RNA sample, the less efficient is the reverse transcription reaction, and this effect varies between the samples.

Typical reference genes are *GAPDH* (*glyceraldehyde 3-phosphate dehydrogenase*), and *ACTB* (*beta-actin*).

qPCR Analysis

The qPCR results consist of an amplification plot and melting curve.

- The **amplification plot** depicts the change in fluorescence as log (ΔRn) in each PCR cycle (Fig. 4.7).
- The **melting curve** shows at what temperature the DNA strands melt/separate. This is very useful to evaluate if the amplification of a single transcript is measured, or if there are contaminants in your starting sample or issues with the primer design.

As the number of PCR cycles increases, the quantity of DNA incorporating the dye and thus the fluorescence level increases exponentially. After reaching a certain number of cycles, the fluorescence will become detectable and rise above the background noise. It will cross an arbitrary fluorescence threshold of your choice, which should correspond to the beginning of the log phase (Fig. 4.7). The

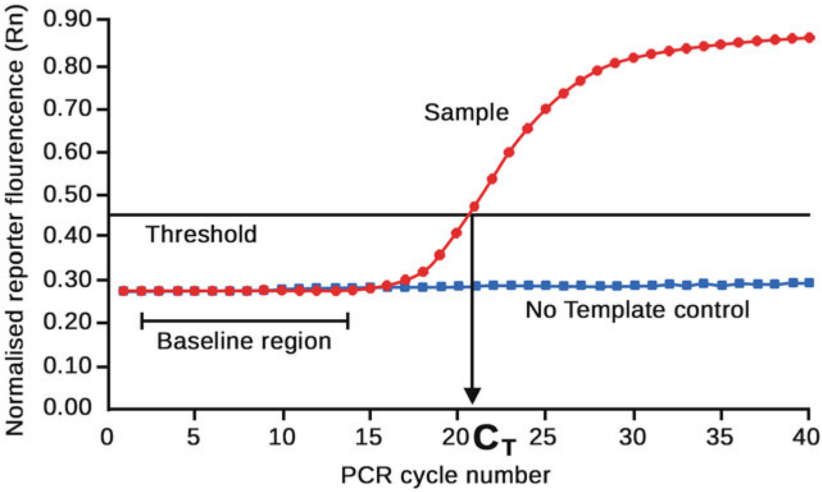


Fig. 4.7 qPCR analysis. Example of an amplification plot of a qPCR for one sample and the control

cycle at which the fluorescence crosses this specified threshold is called the C_t value (also called C_q for quantification cycle). A low C_t indicates a high amount of initial transcript since the fluorescence threshold was crossed early, corresponding to a high expression of the investigated gene. A high C_t value will indicate small amounts of transcript, and therefore low expression of the gene.

One should always check if the negative controls produce a signal. If there is a signal for the no-template controls (just water and the PCR mix without any cDNA), the experiment is likely contaminated and the results are not valid.

There are different methods to calculate the relative amounts of each DNA from the measured fluorescence and the corresponding observed C_t. The most common method is the "delta delta C_t" method, which allows you to assess the biological variation in the expression level of your target gene.

4.3 Let's Get Started

You're now equipped with all the knowledge you need to enter the Gene Expression simulation. Follow the process all the way through; from isolating the RNA to performing a cutting-edge NGS experiment and validating your findings. Will you identify a gene associated with obesity to help healthcare providers deal with the epidemic?

Techniques Used in the Lab

- RNA isolation
- cDNA synthesis
- Next-generation sequencing (NGS)
- Data analysis of sequenced fragments
- Validation using qPCR

Learning Objectives

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

- Explain the central dogma of molecular biology
- Explain and perform RNA extraction and mRNA purification
- Perform PCR, RT-PCR, qPCR, and NGS techniques
- Relate cell functions with gene expression

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

Chomczynski P, Sacchi N (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. Nat Protoc 1:581

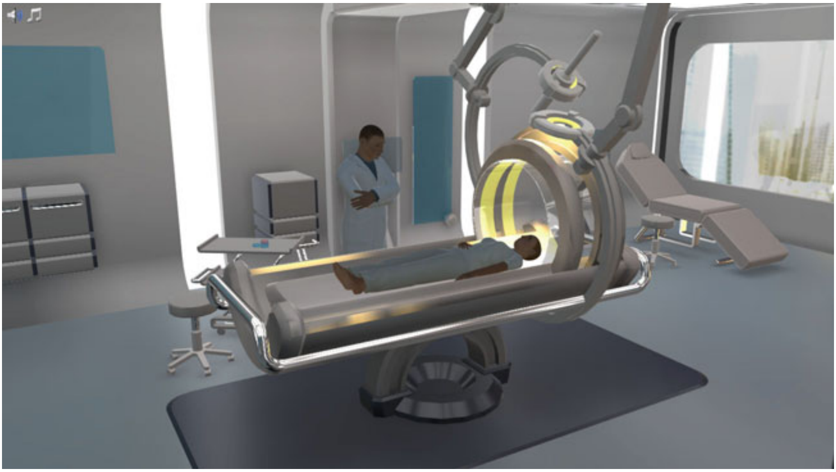
Hartwell L et al (2015) Genetics: From Genes to Genomes, 5th edn. McGraw-Hill, Boston

OpenStax (2018) Biology. OpenStax CNX. http://cnx.org/contents/185cbf87-c72e-48f5-b51e-f14f21b5eabd@11.2. Accessed 1 June 2018

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

Gene Regulation

5



S. Stauffer et al., *Labster Virtual Lab Experiments: Basic Genetics*,
https://doi.org/10.1007/978-3-662-57999-2_5

5.1 Gene Regulation Simulation

The Gene Regulation simulation is based on the study which won the Nobel Prize for physiology in 2012. In this simulation, you will learn about the regulation of gene expression and its real-life applications. Your mission will be to help a visually impaired child by generating a type of stem cell that can be differentiated to retinal-pigmented epithelial cells for transplantation.

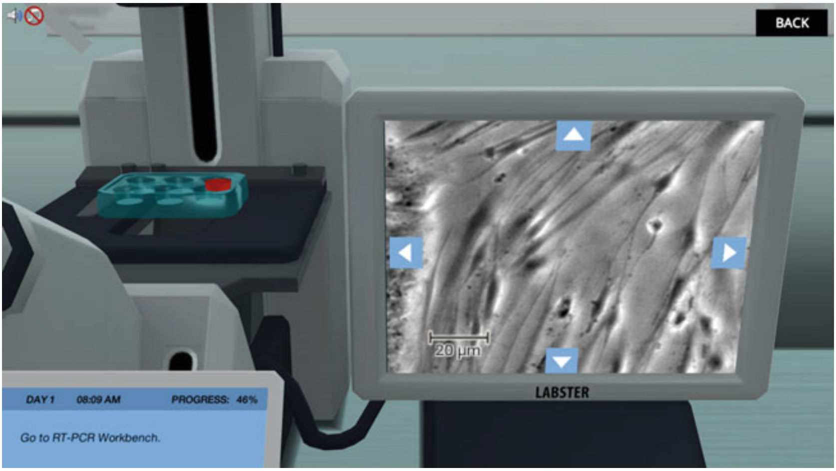
Transforming Fibroblasts into Induced Pluripotent Stem Cells

Your first task will be to determine which transcription factors are important in maintaining stem cell-like features. By analyzing differential gene expression between fibroblasts and stem cells, you will be able to identify promising transcription factors for this task (Figs. 5.1 and 5.2). You will use these transcription factors and a viral vector system to transform mature fibroblast cells into induced pluripotent stem (iPS) cells.

Gene Expression Analysis

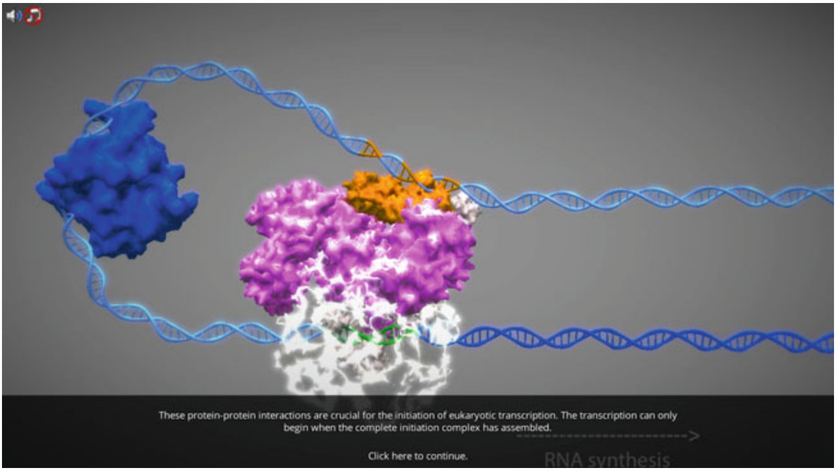
Following the transformation experiment, your task will be to confirm that the transcription factor encoding genes you selected are up-regulated in fibroblasts. You will check the mRNA transcripts using reverse transcription PCR (RT-PCR) and

Fig. 5.1 Introduction scene of the Gene Regulation simulation presenting the background information for your mission



Go to RT-PCR Workbench.

Fig. 5.2 Microscopy analysis of cells in the Gene Regulation simulation



These protein-protein interactions are crucial for the initiation of eukaryotic transcription. The transcription can only begin when the complete initiation complex has assembled.

Fig. 5.3 3D molecular animation in the Gene Regulation simulation protein levels using western blotting. Finally, you will experience gene regulation on a molecular level by following 3D animations (Fig. 5.3).

After completing the Gene Regulation simulation, you will have a deeper understanding of RT-PCR and western blotting, two standard techniques commonly used to analyze gene expression.

Will you be able to use your knowledge of iPS cells to help a visually impaired child?

5.2 Gene Regulation Theory Content

Regulation of gene expression is incredibly complicated. However, an understanding of this regulation can lead to some amazing discoveries, such as those of Shinya Yamanaka whose study of stem cell biology won him the Nobel Prize for physiology in 2012. Before replicating Yamanaka's work in the Gene Regulation simulation you will require a good understanding of the topics covered in the gene regulation theory content.

Gene Regulation

Gene regulation is important in all organisms. In both prokaryotic and eukaryotic organisms gene expression events are constantly turned on and off in response to internal and external environmental conditions.

Early gene regulation conserves more energy compared to regulation in later stages. For example, selective blocking of transcription is much more energy efficient than waiting for the transcription and translation process to finish before finally degrading or inhibiting the protein. In prokaryotic organisms, gene expression is generally regulated during the transcription stage using operons (a cluster of genes with a single promoter that can be regulated by positive and negative controls). In this chapter, we will focus on eukaryotic gene expression.

Eukaryotic Gene Regulation

Most eukaryotic organisms are multicellular and are comprised of a variety of cells with different functions, regardless of having the same genome. To create different cell types, eukaryotic organisms rely on gene regulation. Gene regulation is im-

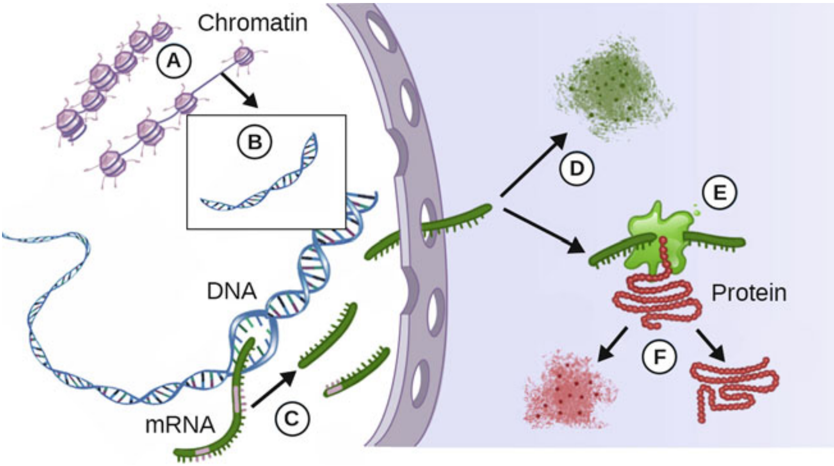


Fig. 5.4 **Different steps of gene regulation in a eukaryotic cell.** In eukaryotic cells, gene regulation can occur at different stages: (*A*, *B*) before transcription via chromatin modifications, (*C*) during transcription through transcription factors and RNA processing, (*D*) after transcription by mRNA degradation, and (*E*, *F*) during and after translation

portant for maintaining the proper specialized functions of each cell. In eukaryotic organisms, gene regulation can occur at different stages (see Fig. 5.4).

Chromatin Modification

Chromatin modification is a form of epigenetics. Epigenetics refers to changes in the expression of a gene or set of genes without changing the DNA sequence. Unlike mutations in DNA, which are mostly permanent, chromatin modifications can be reversed. Chromatin remodeling provides initial control of gene expression by regulating DNA accessibility to its transcription initiation apparatus. Chromatin modification can be achieved by histone acetylation and DNA methylation.

Histone Acetylation

Eukaryotic cells package their DNA by tightly wrapping it around proteins called histones. Histones are a family of small, positively charged proteins. The human histones are named H1, H2A, H2B, H3, and H4. Having a negative charge due to the phosphate-sugar backbone, DNA tightly binds to the positively charged histones. The resulting DNA-protein complex is called chromatin. The basic unit of chromatin is a nucleosome, which is comprised of an octameric complex of the core histone proteins and 146 base pairs of DNA.

The structure of chromatin not only serves as a packaging mechanism, but also as a regulator of gene expression by affecting the accessibility of the DNA transcription apparatus including transcription factors and RNA polymerase. When DNA is located in highly condensed chromatin (heterochromatin), it becomes physically inaccessible to the DNA transcription apparatus, and so transcription is blocked.

Each histone has a 20 amino acid-long "tail" at its N-terminus that protrudes outward. These histone tails are accessible to various modifying enzymes that remove or catalyze additional specific chemical groups such as acetyl, methyl, and phosphate groups. Histone acetyltransferase can add acetyl groups to the positively charged histone tail, thus changing its charge. By reducing the positive charge of a histone tail, the affinity of DNA to histones decreases; therefore opening up the compacted chromatin and promoting gene expression. Another enzyme called histone deacetylase does the opposite of histone acetyltransferase: this enzyme removes the acetyl group from histone tails, thus repressing transcription.

You will be able to follow how histones are acetylated in one of the 3D molecular animations of the Gene Regulation simulation.

DNA Methylation

Methylation is the addition of a methyl group (–CH3). A methylation reaction is catalyzed by an enzyme called DNA methyltransferase. In mammals, DNA methylation primarily occurs at the 5′-carbon cytosine (C) bases that are adjacent to guanine (G) bases, resulting in 5′-methylcytosine (5mC). These C and G-rich regions are called CpG islands and are abundant in promoter regions.

DNA methylation can repress the expression of genes by blocking the promoter from its DNA transcription apparatus. DNA methylation is reversible; an enzyme called demethylase catalyzes the removal of methyl groups from cytosine.

Transcription Factors

Transcription factors are proteins that mediate the binding of RNA polymerase and the initiation of transcription. Transcription only starts when the appropriate transcription factors are attached and the RNA polymerase is bound in the correct orientation. The transcription factors and RNA polymerase on the promoter form the transcription initiation complex.

The rate of transcription can be strongly increased or decreased by the binding of specific transcription factors. Enhancers and silencers can affect the expression of genes located hundreds, or even thousands of nucleotides away. This is made possible by DNA bends that bring the enhancers/silencers and promoters in close proximity.

Post-transcriptional Regulation

Gene regulation can also occur immediately after transcription by processing the RNA. In eukaryotic organisms, RNA processing is performed by alternative RNA splicing and mRNA degradation. Short interfering RNAs (siRNAs) are part of a complex mechanism that regulates gene expression by binding to specific mRNAs and inhibiting translation.

Stem Cells

A stem cell is a single cell that can replicate itself and differentiate into other cell types. Stem cells do not have any tissue-specific structures that allow them to perform specialized functions. They are divided into groups depending on their abilities to differentiate into different tissues:

- **Totipotent**: cells capable of giving rise to all other types of cells. This includes the fertilized egg and the subsequent new cells formed until approximately day 4 of embryonic development.
- **Pluripotent**: this term refers both to embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs). These cells can give rise to all types of cells, except those that form the amniotic sac and the placenta. ESCs come from the inner cell mass in the blastocyst, while iPSCs are generated from somatic cells.
- **Multipotent**: also called adult stem cells, are cells that can develop into closely related cell types.

Induced Pluripotent Stem (iPS) Cells

IPS cells are stem cells have revolutionized the field of stem cell research. Somatic cells, or differentiated cells, are reprogrammed into iPS cells by introducing four genes that encode four transcription factors: *OCT4*, *SOX2*, *KLF4* and *MYC*. These

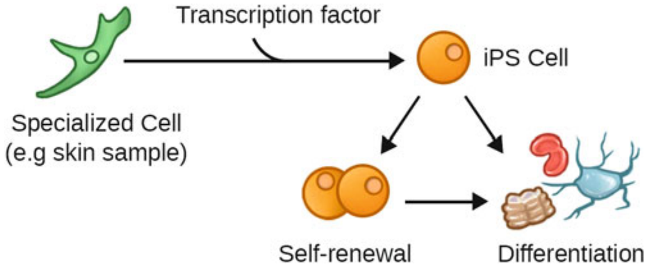


Fig. 5.5 **Generation of iPS cells.** Reprogramming of somatic cells (e.g. skin cells) into iPS cells can be achieved by introducing genes encoding for specific transcription factors. These cells can differentiate into other cell types and have the ability for self-renewal

genes are active in pluripotent stem cells but are largely dormant in differentiated cells.

To generate iPS cells, the four transcription factor genes are introduced into somatic cells, typically using retroviral vectors (Fig. 5.5). Once these genes are integrated into the somatic cell genome, they are expressed as transcription factor proteins. The transcription factors will induce the expression of endogenous genes such as *NANOG*, *SOX2* and *OCT4*, which are important for inducing and maintaining pluripotency characteristics.

Application of iPS Cells

IPS cells have the unique characteristics of self-renewal and the capability of differentiating into other cell types. In contrast to ESCs, iPS cells are not derived from human embryos thus raising fewer ethical concerns. In addition, since iPS are cultured from the patient's own cells, their application decreases the chance of immunological reactions that often lead to rejection of transplanted cells or tissues. These characteristics make iPS cells appealing for therapeutic uses such as cell therapy for leukemia, autoimmune disease and, recently, even as a cure for blindness. However, residual expression of the four transcription factors may induce tumors. *MYC,* for example, is a well-known oncogene.

iPS Cells for the Treatment of Retinal Diseases

Eyes are complex organs, comprised of many types of cells with specific roles. Together, these cells form a system to detect light from the environment and transfer this information to the brain. The retinal pigment epithelium (RPE) is the cell layer responsible for the conversion of light waves into a cell signal. The RPE cells are paramount in maintaining the structure and function of the retina and photoreceptors. They are also important for processing nutrients that supplement the retina. RPE cells can be damaged by a variety of diseases, such as, age-related macular degeneration (AMD), retinitis pigmentosa, Stargardt's disease and Leber's congenital neurosis. Progressive damage of the RPE can lead to eye dysfunction and even blindness.

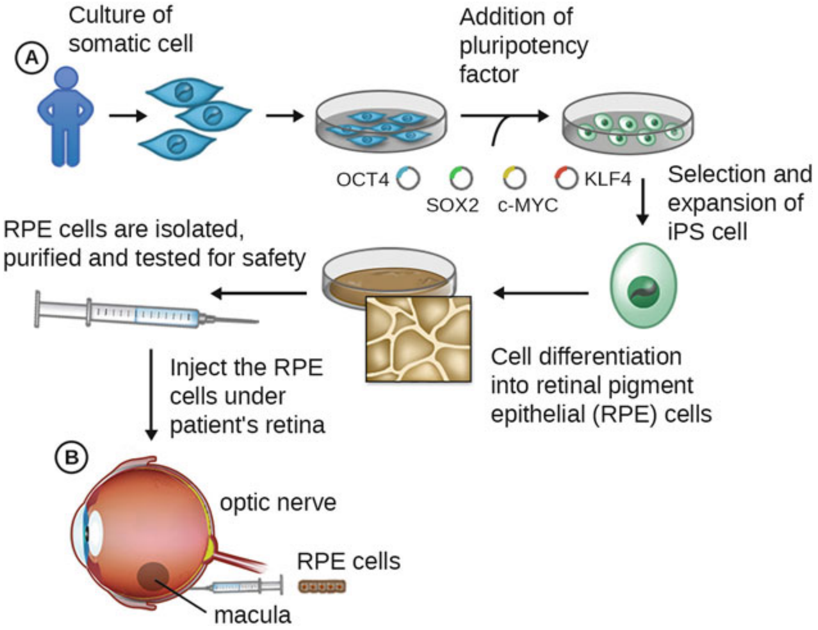


Fig. 5.6 **Treatment of retinal diseases with iPS cells.** Somatic cells are cultured and reprogrammed into iPS cells by the introduction of pluripotency factors (*A*). Once those cells are differentiated into RPE cells, they are tested and injected under the patient's retina (*B*)

RPE donors are rare and difficult to find; current studies are attempting to use iPS to generate RPE cells for eye transplants (Fig. 5.6). Treatment is being performed by injecting RPE cells under a patient's retina.

Retroviral Vectors for Gene Delivery

Retroviruses are lipid-enveloped viruses comprised of linear single-stranded RNA genomes of 7 to 11-kilobases. This type of RNA viruses is characterized by an RNA-dependent DNA polymerase that reverse transcribes the viral RNA genome into DNA, which is then integrated into the host DNA.

Due to their ability to stably integrate the viral genome into the host DNA, retroviruses are commonly used vectors for gene delivery, as for example in the case of iPS cell generation. Typically, gamma-retroviral and lentiviral vectors are used for gene delivery.

Reverse Transcription PCR (RT-PCR)

In the Gene Regulation simulation, you will perform a reverse transcription PCR (RT-PCR) experiment. This method is commonly used to qualitatively detect mRNA expression. Since mRNA is more fragile than DNA and cannot be amplified by PCR it first needs to be converted into complementary DNA (cDNA), before being amplified using PCR. These processes have previously been described in this book (Chaps. 2.2 and 4.2).

Western Blotting

Besides measuring the transcribed mRNA levels, gene expression can be analyzed on the protein level for example by western blotting—an important method in protein analysis and identification. A western blot can detect one specific protein in a solution that contains numerous other proteins. The term western blot specifically refers to the transfer of protein from a protein gel to a membrane and their subsequent detection using antibodies. Fig. 5.7 shows an overview of the steps needed to perform a western blot experiment.

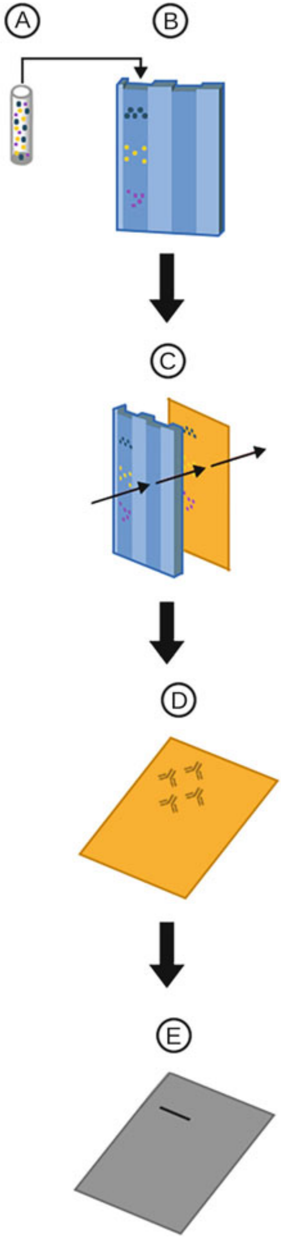


Fig. 5.7 **Western blot experimental steps.** Following sample preparation (*A*), the protein samples are loaded onto an SDS polyacrylamide gel (*B*) and separated by gel electrophoresis. In the next step, the separated proteins are transferred onto a membrane (*C*) and detected by antibodies targeting the protein of interest (*D*). Finally, bands representing the amount of specific proteins are visualized on a membrane (*E*)

5.3 Let's Get Started

Wow! You're a genetics master, you've covered all the theory topics and are now ready to face your final challenge in the Gene Regulation simulation. Can you repeat the Nobel Prize-winning work of Shinya Yamanaka and create some iPS, then differentiate them into RPE as a therapeutic treatment for a visually impaired child?

Techniques Used in the Lab

- Microscopy
- RT-PCR
- Western blotting

Learning Objectives

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

- Explain how gene expression can be regulated
- Describe the different levels of gene regulation (mRNA and protein)
- Detect mRNA (RT-PCR) and protein expression (western blotting)
- Describe the concept of stem cells and their potential use

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-c72e-48f5-b51e-f14f21b5eabd@11.2. Accessed 1 June 2018

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