10.3: Whole Genome Methods and Industrial Applications

Learning Objectives

- Explain the uses of genome-wide comparative analyses
- Summarize the advantages of genetically engineered pharmaceutical products

Advances in molecular biology have led to the creation of entirely new fields of science. Among these are fields that study aspects of whole genomes, collectively referred to as whole-genome methods. In this section, we’ll provide a brief overview of the whole-genome fields of genomics, transcriptomics, and proteomics.

Genomics, Transcriptomics, and Proteomics

The study and comparison of entire genomes, including the complete set of genes and their nucleotide sequence and organization, is called genomics. This field has great potential for future medical advances through the study of the human genome as well as the genomes of infectious organisms. Analysis of microbial genomes has contributed to the development of new antibiotics, diagnostic tools, vaccines, medical treatments, and environmental cleanup techniques.

The field of transcriptomics is the science of the entire collection of mRNA molecules produced by cells. Scientists compare gene expression patterns between infected and uninfected host cells, gaining important information about the cellular responses to infectious disease. Additionally, transcriptomics can be used to monitor the gene expression of virulence factors in microorganisms, aiding scientists in better understanding pathogenic processes from this viewpoint.

When genomics and transcriptomics are applied to entire microbial communities, we use the terms metagenomics and metatranscriptomics, respectively. Metagenomics and metatranscriptomics allow researchers to study genes and gene expression from a collection of multiple species, many of which may not be easily cultured or cultured at all in the laboratory. A DNA microarray (discussed in the previous section) can be used in metagenomics studies.
Another up-and-coming clinical application of genomics and transcriptomics is pharmacogenomics, also called toxicogenomics, which involves evaluating the effectiveness and safety of drugs on the basis of information from an individual’s genomic sequence. Genomic responses to drugs can be studied using experimental animals (such as laboratory rats or mice) or live cells in the laboratory before embarking on studies with humans. Changes in gene expression in the presence of a drug can sometimes be an early indicator of the potential for toxic effects. Personal genome sequence information may someday be used to prescribe medications that will be most effective and least toxic on the basis of the individual patient’s genotype.

The study of proteomics is an extension of genomics that allows scientists to study the entire complement of proteins in an organism, called the proteome. Even though all cells of a multicellular organism have the same set of genes, cells in various tissues produce different sets of proteins. Thus, the genome is constant, but the proteome varies and is dynamic within an organism. Proteomics may be used to study which proteins are expressed under various conditions within a single cell type or to compare protein expression patterns between different organisms.

The most prominent disease being studied with proteomic approaches is cancer, but this area of study is also being applied to infectious diseases. Research is currently underway to examine the feasibility of using proteomic approaches to diagnose various types of hepatitis, tuberculosis, and HIV infection, which are rather difficult to diagnose using currently available techniques.1

A recent and developing proteomic analysis relies on identifying proteins called biomarkers, whose expression is affected by the disease process. Biomarkers are currently being used to detect various forms of cancer as well as infections caused by pathogens such as *Yersinia pestis* and *Vaccinia virus*.2

Other “-omic” sciences related to genomics and proteomics include metabolomics, glycomics, and lipidomics, which focus on the complete set of small-molecule metabolites, sugars, and lipids, respectively, found within a cell. Through these various global approaches, scientists continue to collect, compile, and analyze large amounts of genetic information. This emerging field of bioinformatics can be used, among many other applications, for clues to treating diseases and understanding the workings of cells.

Additionally, researchers can use reverse genetics, a technique related to classic mutational analysis, to determine the function of specific genes. Classic methods of studying gene function involved searching for the genes responsible for a given phenotype. Reverse genetics uses the opposite approach, starting with a specific DNA sequence and attempting to determine what phenotype it produces. Alternatively, scientists can attach known genes (called reporter genes) that encode easily observable characteristics to genes of interest, and the location of expression of such genes of interest can be easily monitored. This gives the researcher important information about what the gene product might be doing or where it is located in the organism. Common reporter genes include bacterial *lacZ*, which encodes beta-galactosidase and whose activity can be monitored by changes in colony color in the presence of X-gal as previously described, and the gene encoding the jellyfish protein green fluorescent protein (GFP) whose activity can be visualized in colonies under ultraviolet light exposure (Figure \(\PageIndex{1}\)).
Figure \(\PageIndex{1}\): (a) The gene encoding green fluorescence protein is a commonly used reporter gene for monitoring gene expression patterns in organisms. Under ultraviolet light, GFP fluoresces. Here, two mice are expressing GFP, while the middle mouse is not. (b) GFP can be used as a reporter gene in bacteria as well. Here, a plate containing bacterial colonies expressing GFP is shown. (c) Blue-white screening in bacteria is accomplished through the use of the lacZ reporter gene, followed by plating of bacteria onto medium containing X-gal. Cleavage of X-gal by the LacZ enzyme results in the formation of blue colonies. (credit a: modification of work by Ingrid Moen, Charlotte Jevne, Jian Wang, Karl-Henning Kalland, Martha Chekenya, Lars A Akslen, Linda Sleire, Per Ø Enger, Rolf K Reed, Anne M Øyan, Linda EB Stuhr; credit b: modification of work by “2.5JIGEN.com”/Flickr; credit c: modification of work by American Society for Microbiology)

Exercise \(\PageIndex{1}\)

1. How is genomics different from traditional genetics?
2. If you wanted to study how two different cells in the body respond to an infection, what –omics field would you apply?
3. What are the biomarkers uncovered in proteomics used for?

USE AND ABUSE OF GENOME DATA

Why can some humans harbor opportunistic pathogens like *Haemophilus influenzae*, *Staphylococcus aureus*, or *Streptococcus pyogenes*, in their upper respiratory tracts but remain asymptomatic carriers, while other individuals become seriously ill when infected? There is evidence suggesting that differences in susceptibility to infection between patients may be a result, at least in part, of genetic differences between human hosts. For example, genetic differences in human leukocyte antigens (HLAs) and red blood cell antigens among hosts have been implicated in different immune responses and resulting disease progression from infection with *H. influenzae*.

Because the genetic interplay between pathogen and host may contribute to disease outcomes, understanding differences in genetic makeup between individuals may be an important clinical tool. Ecological genomics is a relatively new field that seeks to understand how the genotypes of different organisms interact with each other in nature. The field answers questions about how gene expression of one organism affects gene expression of another. Medical applications of ecological genomics will focus on how pathogens interact with specific individuals, as opposed to humans in general. Such analyses would allow medical professionals to use knowledge of an individual’s genotype to apply more individualized plans for treatment and prevention of disease.

With the advent of next-generation sequencing, it is relatively easy to obtain the entire genomic sequences of pathogens; a bacterial genome can be sequenced in as little as a day.\(^{3}\) The speed and cost of sequencing the human genome has also been greatly reduced and, already, individuals can submit samples to receive extensive reports on their personal genetic traits, including ancestry and carrier status for various genetic diseases. As sequencing technologies progress further, such services will continue to become less expensive, more extensive, and quicker.

However, as this day quickly approaches, there are many ethical concerns with which society must grapple. For
example, should genome sequencing be a standard practice for everybody? Should it be required by law or by employers if it will lower health-care costs? If one refuses genome sequencing, does he or she forfeit his or her right to health insurance coverage? For what purposes should the data be used? Who should oversee proper use of these data? If genome sequencing reveals predisposition to a particular disease, do insurance companies have the right to increase rates? Will employers treat an employee differently? Knowing that environmental influences also affect disease development, how should the data on the presence of a particular disease-causing allele in an individual be used ethically? The Genetic Information Nondiscrimination Act of 2008 (GINA) currently prohibits discriminatory practices based on genetic information by both health insurance companies and employers. However, GINA does not cover life, disability, or long-term care insurance policies. Clearly, all members of society must continue to engage in conversations about these issues so that such genomic data can be used to improve health care while simultaneously protecting an individual’s rights.

Clinical Focus: Resolution

Because Kayla’s symptoms were persistent and serious enough to interfere with daily activities, Kayla’s physician decided to order some laboratory tests. The physician collected samples of Kayla’s blood, cerebrospinal fluid (CSF), and synovial fluid (from one of her swollen knees) and requested PCR analysis on all three samples. The PCR tests on the CSF and synovial fluid came back positive for the presence of *Borrelia burgdorferi*, the bacterium that causes Lyme disease.

Kayla’s physician immediately prescribed a full course of the antibiotic doxycycline. Fortunately, Kayla recovered fully within a few weeks and did not suffer from the long-term symptoms of post-treatment Lyme disease syndrome (PTLDS), which affects 10–20% of Lyme disease patients. To prevent future infections, Kayla’s physician advised her to use insect repellant and wear protective clothing during her outdoor adventures. These measures can limit exposure to Lyme-bearing ticks, which are common in many regions of the United States during the warmer months of the year. Kayla was also advised to make a habit of examining herself for ticks after returning from outdoor activities, as prompt removal of a tick greatly reduces the chances of infection.

Lyme disease is often difficult to diagnose. *B. burgdorferi* is not easily cultured in the laboratory, and the initial symptoms can be very mild and resemble those of many other diseases. But left untreated, the symptoms can become quite severe and debilitating. In addition to two antibody tests, which were inconclusive in Kayla’s case, and the PCR test, a Southern blot could be used with *B. burgdorferi*-specific DNA probes to identify DNA from the pathogen. Sequencing of surface protein genes of *Borrelia* species is also being used to identify strains within the species that may be more readily transmitted to humans or cause more severe disease.

**Recombinant DNA Technology and Pharmaceutical Production**

Genetic engineering has provided a way to create new pharmaceutical products called recombinant DNA pharmaceuticals. Such products include antibiotic drugs, vaccines, and hormones used to treat various diseases. Table 1 lists examples of recombinant DNA products and their uses.

For example, the naturally occurring antibiotic synthesis pathways of various *Streptomyces* spp., long known for their antibiotic production capabilities, can be modified to improve yields or to create new antibiotics through the introduction
of genes encoding additional enzymes. More than 200 new antibiotics have been generated through the targeted inactivation of genes and the novel combination of antibiotic synthesis genes in antibiotic-producing *Streptomyces* hosts.

Genetic engineering is also used to manufacture subunit vaccines, which are safer than other vaccines because they contain only a single antigenic molecule and lack any part of the genome of the pathogen. For example, a vaccine for hepatitis B is created by inserting a gene encoding a hepatitis B surface protein into a yeast; the yeast then produces this protein, which the human immune system recognizes as an antigen. The hepatitis B antigen is purified from yeast cultures and administered to patients as a vaccine. Even though the vaccine does not contain the hepatitis B virus, the presence of the antigenic protein stimulates the immune system to produce antibodies that will protect the patient against the virus in the event of exposure.

Genetic engineering has also been important in the production of other therapeutic proteins, such as insulin, interferons, and human growth hormone, to treat a variety of human medical conditions. For example, at one time, it was possible to treat diabetes only by giving patients pig insulin, which caused allergic reactions due to small differences between the proteins expressed in human and pig insulin. However, since 1978, recombinant DNA technology has been used to produce large-scale quantities of human insulin using *E. coli* in a relatively inexpensive process that yields a more consistently effective pharmaceutical product. Scientists have also genetically engineered *E. coli* capable of producing human growth hormone (HGH), which is used to treat growth disorders in children and certain other disorders in adults. The HGH gene was cloned from a cDNA library and inserted into *E. coli* cells by cloning it into a bacterial vector. Eventually, genetic engineering will be used to produce DNA vaccines and various gene therapies, as well as customized medicines for fighting cancer and other diseases.

<table>
<thead>
<tr>
<th>Recombinant DNA Product</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial natriuretic peptide</td>
<td>Treatment of heart disease (e.g., congestive heart failure), kidney disease, high blood pressure</td>
</tr>
<tr>
<td>DNase</td>
<td>Treatment of viscous lung secretions in cystic fibrosis</td>
</tr>
<tr>
<td>Erythropoietin</td>
<td>Treatment of severe anemia with kidney damage</td>
</tr>
<tr>
<td>Factor VIII</td>
<td>Treatment of hemophilia</td>
</tr>
<tr>
<td>Hepatitis B vaccine</td>
<td>Prevention of hepatitis B infection</td>
</tr>
<tr>
<td>Human growth hormone</td>
<td>Treatment of growth hormone deficiency, Turner's syndrome, burns</td>
</tr>
<tr>
<td>Human insulin</td>
<td>Treatment of diabetes</td>
</tr>
<tr>
<td>Interferons</td>
<td>Treatment of multiple sclerosis, various cancers (e.g., melanoma), viral infections (e.g., Hepatitis B and C)</td>
</tr>
<tr>
<td>Tetracenomycins</td>
<td>Used as antibiotics</td>
</tr>
</tbody>
</table>
Recombinant DNA Product | Application
---|---
Tissue plasminogen activator | Treatment of pulmonary embolism in ischemic stroke, myocardial infarction

Exercise 2

1. What bacterium has been genetically engineered to produce human insulin for the treatment of diabetes?
2. Explain how microorganisms can be engineered to produce vaccines.

**RNA Interference Technology**

In *Structure and Function of RNA*, we described the function of mRNA, rRNA, and tRNA. In addition to these types of RNA, cells also produce several types of small noncoding RNA molecules that are involved in the regulation of gene expression. These include antisense RNA molecules, which are complementary to regions of specific mRNA molecules found in both prokaryotes and eukaryotic cells. Non-coding RNA molecules play a major role in RNA interference (RNAi), a natural regulatory mechanism by which mRNA molecules are prevented from guiding the synthesis of proteins. RNA interference of specific genes results from the base pairing of short, single-stranded antisense RNA molecules to regions within complementary mRNA molecules, preventing protein synthesis. Cells use RNA interference to protect themselves from viral invasion, which may introduce double-stranded RNA molecules as part of the viral replication process (Figure 2).

![RNA Interference Diagram](https://bio.libretexts.org/Courses/Manchester_Community_College_(MCC)/The_Introduction_to_Microbiology_(Cases)/10%3…)

Researchers are currently developing techniques to mimic the natural process of RNA interference as a way to treat viral infections in eukaryotic cells. RNA interference technology involves using small interfering RNAs (siRNAs) or microRNAs (miRNAs) (Figure 3). siRNAs are completely complementary to the mRNA transcript of a specific gene of interest while miRNAs are mostly complementary. These double-stranded RNAs are bound to DICER, an endonuclease that cleaves the RNA into short molecules (approximately 20 nucleotides long). The RNAs are then bound to RNA-induced silencing complex (RISC), a ribonucleoprotein. The siRNA-RISC complex binds to mRNA and cleaves it. For miRNA, only one of the two strands binds to RISC. The miRNA-RISC complex then binds to mRNA, inhibiting translation. If the miRNA is completely complementary to the target gene, then the mRNA can be cleaved.
Taken together, these mechanisms are known as gene silencing.

![Diagram illustrating gene silencing](https://bio.libretexts.org/Courses/Manchester_Community_College_(MCC)/The_Introduction_to_Microbiology_(Cases)/10%3…)

Figure 4: This diagram illustrates the process of using siRNA or miRNA in a eukaryotic cell to silence genes involved in the pathogenesis of various diseases. (credit: modification of work by National Center for Biotechnology Information)

### Key Concepts and Summary

- **The science of genomics** allows researchers to study organisms on a holistic level and has many applications of medical relevance.
- **Transcriptomics** and **proteomics** allow researchers to compare gene expression patterns between different cells and shows great promise in better understanding global responses to various conditions.
- The various –omics technologies complement each other and together provide a more complete picture of an organism’s or microbial community’s (metagenomics) state.
- The analysis required for large data sets produced through genomics, transcriptomics, and proteomics has led to the emergence of **bioinformatics**.
- **Reporter genes** encoding easily observable characteristics are commonly used to track gene expression patterns of genes of unknown function.
- The use of recombinant DNA technology has revolutionized the pharmaceutical industry, allowing for the rapid production of high-quality **recombinant DNA pharmaceuticals** used to treat a wide variety of human conditions.
- RNA interference technology has great promise as a method of treating viral infections by silencing the expression of specific genes

### Footnotes


Contributors

- Template:ContribOpenSTAXMicrobiology