



Genetics
Home
Reference

Help Me Understand Genetics
Genomic Research

Reprinted from <https://ghr.nlm.nih.gov/>

Lister Hill National Center for Biomedical Communications
U.S. National Library of Medicine
National Institutes of Health
Department of Health & Human Services

Published October 17, 2017

Genomic Research

Table of Contents

What are the next steps in genomic research?	3
What are single nucleotide polymorphisms (SNPs)?	5
What are genome-wide association studies?	6
What is the International HapMap Project?	7
What is the Encyclopedia of DNA Elements (ENCODE) Project?	9
What is pharmacogenomics?	11
What are whole exome sequencing and whole genome sequencing?	13
What are genome editing and CRISPR-Cas9?	15

What are the next steps in genomic research?

Discovering the sequence of the human genome was only the first step in understanding how the instructions coded in DNA lead to a functioning human being. The next stage of genomic research will begin to derive meaningful knowledge from the DNA sequence. Research studies that build on the work of the Human Genome Project are under way worldwide.

The objectives of continued genomic research include the following:

- Determine the function of genes and the elements that regulate genes throughout the genome.
- Find variations in the DNA sequence among people and determine their significance. The most common type of genetic variation is known as a single nucleotide polymorphism or SNP (pronounced “snip”). These small differences may help predict a person's risk of particular diseases and response to certain medications.
- Discover the 3-dimensional structures of proteins and identify their functions.
- Explore how DNA and proteins interact with one another and with the environment to create complex living systems.
- Develop and apply genome-based strategies for the early detection, diagnosis, and treatment of disease.
- Sequence the genomes of other organisms, such as the rat, cow, and chimpanzee, in order to compare similar genes between species.
- Develop new technologies to study genes and DNA on a large scale and store genomic data efficiently.
- Continue to explore the ethical, legal, and social issues raised by genomic research.

For more information about the genomic research following the Human Genome Project:

The National Human Genome Research Institute supports research in many of the areas described above. The Institute provides detailed information about its research initiatives at NIH (<https://www.genome.gov/researchatnhgri/>).

The Genome Institute at Washington University explains the 1000 Genomes Project (<http://genome.wustl.edu/projects/detail/1000-genomes-project/>), which furthers the work of the International HapMap Project (<https://www.genome.gov/10001688/>).

The Wellcome Trust Sanger Institute discusses the 1000 Genomes Project (<https://wellcome.ac.uk/press-release/1000-genomes-project-publishes-most->

comprehensive-map-date-human-genetic-variation) in a press release that describes the key objectives of the project.

What are single nucleotide polymorphisms (SNPs)?

Single nucleotide polymorphisms, frequently called SNPs (pronounced “snips”), are the most common type of genetic variation among people. Each SNP represents a difference in a single DNA building block, called a nucleotide. For example, a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain stretch of DNA.

SNPs occur normally throughout a person’s DNA. They occur once in every 300 nucleotides on average, which means there are roughly 10 million SNPs in the human genome. Most commonly, these variations are found in the DNA between genes. They can act as biological markers, helping scientists locate genes that are associated with disease. When SNPs occur within a gene or in a regulatory region near a gene, they may play a more direct role in disease by affecting the gene’s function.

Most SNPs have no effect on health or development. Some of these genetic differences, however, have proven to be very important in the study of human health. Researchers have found SNPs that may help predict an individual’s response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases. SNPs can also be used to track the inheritance of disease genes within families. Future studies will work to identify SNPs associated with complex diseases such as heart disease, diabetes, and cancer.

For more information about SNPs:

An audio definition of SNPs (<https://www.genome.gov/glossary/?id=185>) is available from the National Human Genome Research Institute’s Talking Glossary of Genetic Terms.

How scientists locate SNPs in the genome (<http://learn.genetics.utah.edu/content/precision/snips/>) is explained by the University of Utah Genetic Science Learning Center.

For people interested in more technical data, several databases of known SNPs are available:

- NCBI database of single nucleotide polymorphisms (dbSNP) (<https://www.ncbi.nlm.nih.gov/SNP/>)
- National Bioscience Database Center of the Japan Science and Technology Agency (<https://biosciencedbc.jp/en/>)

What are genome-wide association studies?

Genome-wide association studies are a relatively new way for scientists to identify genes involved in human disease. This method searches the genome for small variations, called single nucleotide polymorphisms or SNPs (pronounced “snips”), that occur more frequently in people with a particular disease than in people without the disease. Each study can look at hundreds or thousands of SNPs at the same time. Researchers use data from this type of study to pinpoint genes that may contribute to a person’s risk of developing a certain disease.

Because genome-wide association studies examine SNPs across the genome, they represent a promising way to study complex, common diseases in which many genetic variations contribute to a person’s risk. This approach has already identified SNPs related to several complex conditions including diabetes, heart abnormalities, Parkinson disease, and Crohn disease. Researchers hope that future genome-wide association studies will identify more SNPs associated with chronic diseases, as well as variations that affect a person’s response to certain drugs and influence interactions between a person’s genes and the environment.

For more information about genome-wide association studies:

The National Human Genome Research Institute provides a detailed explanation of genome-wide association studies (<https://www.genome.gov/20019523>).

You can also search for clinical trials of genome-wide association studies online. ClinicalTrials.gov (<https://clinicaltrials.gov/>), a service of the National Institutes of Health, provides easy access to information about clinical trials. You can search for a specific clinical trial or browse by health condition or sponsor. You may wish to refer to a list of genome-wide association studies (<https://clinicaltrials.gov/search?term=GWAS+OR+%22Genome+Wide+Association%22>) that are accepting (or will accept) participants.

For people interested in more technical information, the NCBI’s Database of Genotypes and Phenotypes (dbGaP) (<https://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>) contains data from genome-wide association studies. An introduction to this database, as well as information about study results, is available from the dbGaP press release (https://www.nlm.nih.gov/archive//20120510/news/press_releases/dbgap_launchPR06.html). In addition, the National Human Genome Research Institute and the European Bioinformatics Institute jointly provide a Catalog of Published Genome-Wide Association Studies (<http://www.ebi.ac.uk/gwas/>).

What is the International HapMap Project?

The International HapMap Project is a scientific effort to identify common genetic variations among people. This project represents a collaboration of scientists from public and private organizations in six countries. Data from the project is freely available to researchers worldwide. Researchers can use the data to learn more about the relationship between genetic differences and human disease.

The HapMap (short for “haplotype map”) is a catalog of common genetic variants called single nucleotide polymorphisms or SNPs (pronounced “snips”). Each SNP represents a difference in a single DNA building block, called a nucleotide. These variations occur normally throughout a person’s DNA. When several SNPs cluster together on a chromosome, they are inherited as a block known as a haplotype. The HapMap describes haplotypes, including their locations in the genome and how common they are in different populations throughout the world.

The human genome contains roughly 10 million SNPs. It would be difficult, time-consuming, and expensive to look at each of these changes and determine whether it plays a role in human disease. Using haplotypes, researchers can sample a selection of these variants instead of studying each one. The HapMap will make carrying out large-scale studies of SNPs and human disease (called genome-wide association studies) cheaper, faster, and less complicated.

The main goal of the International HapMap Project is to describe common patterns of human genetic variation that are involved in human health and disease. Additionally, data from the project will help researchers find genetic differences that can help predict an individual’s response to particular medicines or environmental factors (such as toxins.)

For more information about the International HapMap Project:

The National Human Genome Research Institute defines haplotype (<https://www.genome.gov/glossary/index.cfm?id=99>) in their Talking Glossary of Genetic Terms. They also provide an overview of the project in their International HapMap Project fact sheet (<https://www.genome.gov/10001688>). The fact sheet includes a link to a more in-depth online tutorial on HapMap usage.

Cold Springs Harbor Laboratory provides a step-by-step guide (<http://cshlpress.com/pdf/sample/GenVa-06.pdf>) to using the International HapMap Project website.

You can also search for clinical trials involving haplotypes or associated with the International HapMap Project. ClinicalTrials.gov (<https://clinicaltrials.gov/>), a service of the National Institutes of Health, provides easy access to information about clinical trials. You can search for a specific clinical trial or browse by

health condition or sponsor. You may wish to refer to a list of haplotype-related studies (<https://clinicaltrials.gov/search?term=HAPMAP+OR+haplotype>) that are accepting (or will accept) participants.

What is the Encyclopedia of DNA Elements (ENCODE) Project?

The ENCODE Project was planned as a follow-up to the Human Genome Project. The Human Genome Project sequenced the DNA that makes up the human genome; the ENCODE Project seeks to interpret this sequence. Coinciding with the completion of the Human Genome Project in 2003, the ENCODE Project began as a worldwide effort involving more than 30 research groups and more than 400 scientists.

The approximately 20,000 genes that provide instructions for making proteins account for only about 1 percent of the human genome. Researchers embarked on the ENCODE Project to figure out the purpose of the remaining 99 percent of the genome. Scientists discovered that more than 80 percent of this non-gene component of the genome, which was once considered “junk DNA,” actually has a role in regulating the activity of particular genes (gene expression).

Researchers think that changes in the regulation of gene activity may disrupt protein production and cell processes and result in disease. A goal of the ENCODE Project is to link variations in the expression of certain genes to the development of disease.

The ENCODE Project has given researchers insight into how the human genome functions. As researchers learn more about the regulation of gene activity and how genes are expressed, the scientific community will be able to better understand how the entire genome can affect human health.

For more information about the ENCODE Project:

The University of California at Santa Cruz and Stanford University provide detailed information about the findings of the ENCODE Project (<https://www.encodeproject.org/>) as well as the Project’s experimental procedures and many other types of data.

Published research findings are available through Nature Magazine’s Nature Encode Explorer (<http://www.nature.com/encode/#/threads>), which gives the public access to scientific information collected from the ENCODE Project.

The Broad Institute of the Massachusetts Institute of Technology and Harvard University describes the purpose (<https://www.broadinstitute.org/news/mapping-genetic-world-beyond-genes>) of the ENCODE Project.

The National Human Genome Research Institute announces results of the ENCODE Project in a press release (<https://www.genome.gov/27549810>) and

provides an overview (<https://www.genome.gov/10005107>) of the ENCODE Project.

Some of the outcomes of the ENCODE Project (<http://www.cshl.edu/news-a-features/massive-genome-analysis-by-encode-redefines-the-gene-and-sheds-new-light-on-complex-disease.html>) are detailed by Cold Springs Harbor Laboratory.

What is pharmacogenomics?

Pharmacogenomics is the study of how genes affect a person's response to drugs. This relatively new field combines pharmacology (the science of drugs) and genomics (the study of genes and their functions) to develop effective, safe medications and doses that will be tailored to a person's genetic makeup.

Many drugs that are currently available are "one size fits all," but they don't work the same way for everyone. It can be difficult to predict who will benefit from a medication, who will not respond at all, and who will experience negative side effects (called adverse drug reactions). Adverse drug reactions are a significant cause of hospitalizations and deaths in the United States. With the knowledge gained from the Human Genome Project, researchers are learning how inherited differences in genes affect the body's response to medications. These genetic differences will be used to predict whether a medication will be effective for a particular person and to help prevent adverse drug reactions.

The field of pharmacogenomics is still in its infancy. Its use is currently quite limited, but new approaches are under study in clinical trials. In the future, pharmacogenomics will allow the development of tailored drugs to treat a wide range of health problems, including cardiovascular disease, Alzheimer disease, cancer, HIV/AIDS, and asthma.

For more information about pharmacogenomics:

The National Institute of General Medical Sciences offers a list of Frequently Asked Questions about Pharmacogenomics (<https://www.nigms.nih.gov/education/Pages/factsheet-pharmacogenomics.aspx>).

A list of Frequently Asked Questions about Pharmacogenomics (<https://www.genome.gov/27530645>) is also offered by the National Human Genome Research Institute.

Additional information about pharmacogenetics is available from the Centre for Genetics Education (<http://www.genetics.edu.au/publications-and-resources/facts-sheets/fact-sheet-21-pharmacogenomics-pharmacogenetics>) as well as Genes In Life (<http://www.genesinlife.org/testing-services/testing-genetic-conditions/pharmacogenomic-testing>).

The Smithsonian National Museum of Natural History's exhibit 'Genome: Unlocking Life's Code' discusses the utility of pharmacogenomics (<http://unlockinglifescode.org/explore/genomic-medicine/pharmacogenomics>).

An interactive tutorial (<http://www.phgfoundation.org/tutorials/pharmacogenomics/>) about pharmacogenomics is available from the PHG Foundation.

PharmGKB (<https://www.pharmgkb.org/>) is a pharmacogenomics resource sponsored by the National Institutes of Health that collects information on human genetic variation and drug responses.

A list of clinical trials involving pharmacogenomics (<https://clinicaltrials.gov/search/?term=pharmacogenomics+OR+pharmacogenetics>) is available from ClinicalTrials.gov, a service of the National Institutes of Health.

What are whole exome sequencing and whole genome sequencing?

Determining the order of DNA building blocks (nucleotides) in an individual's genetic code, called DNA sequencing, has advanced the study of genetics and is one technique used to test for genetic disorders. Two methods, whole exome sequencing and whole genome sequencing, are increasingly used in healthcare and research to identify genetic variations; both methods rely on new technologies that allow rapid sequencing of large amounts of DNA. These approaches are known as next-generation sequencing (or next-gen sequencing).

The original sequencing technology, called Sanger sequencing (named after the scientist who developed it, Frederick Sanger), was a breakthrough that helped scientists determine the human genetic code, but it is time-consuming and expensive. The Sanger method has been automated to make it faster and is still used in laboratories today to sequence short pieces of DNA, but it would take years to sequence all of a person's DNA (known as the person's genome). Next-generation sequencing has sped up the process (taking only days to weeks to sequence a human genome) while reducing the cost.

With next-generation sequencing, it is now feasible to sequence large amounts of DNA, for instance all the pieces of an individual's DNA that provide instructions for making proteins. These pieces, called exons, are thought to make up 1 percent of a person's genome. Together, all the exons in a genome are known as the exome, and the method of sequencing them is known as whole exome sequencing. This method allows variations in the protein-coding region of any gene to be identified, rather than in only a select few genes. Because most known mutations that cause disease occur in exons, whole exome sequencing is thought to be an efficient method to identify possible disease-causing mutations.

However, researchers have found that DNA variations outside the exons can affect gene activity and protein production and lead to genetic disorders--variations that whole exome sequencing would miss. Another method, called whole genome sequencing, determines the order of all the nucleotides in an individual's DNA and can determine variations in any part of the genome.

While many more genetic changes can be identified with whole exome and whole genome sequencing than with select gene sequencing, the significance of much of this information is unknown. Because not all genetic changes affect health, it is difficult to know whether identified variants are involved in the condition of interest. Sometimes, an identified variant is associated with a different genetic disorder that has not yet been diagnosed (these are called incidental or secondary findings).

In addition to being used in the clinic, whole exome and whole genome sequencing are valuable methods for researchers. Continued study of exome and genome sequences can help determine whether new genetic variations are associated with health conditions, which will aid disease diagnosis in the future.

For more information about DNA sequencing technologies and their use:

Genetics Home Reference discusses whether all genetic changes affect health and development (<https://ghr.nlm.nih.gov/primer/mutationsanddisorders/neutralmutations>).

A scientist at the Genome Institute at Washington University in St. Louis describes the different sequencing technologies (<http://genome.wustl.edu/articles/detail/dna-sequencing-technology-a-perspective-from-dr-elaine-mardis/>) and what the new technologies have meant for the study of the genetic code.

An illustration of the decline in the cost of DNA sequencing (<https://www.genome.gov/sequencingcosts/>), including that caused by the introduction of new technologies, is provided by the National Human Genome Research Institute.

The American College of Medical Genetics and Genomics has laid out their policies regarding whole exome and whole genome sequencing (https://www.acmg.net/StaticContent/PPG/Clinical_Application_of_Genomic_Sequencing.pdf), including when these methods should be used, what results may arise, and what the results might indicate.

GeneReviews compares whole genome sequencing, whole exome sequencing, and the sequencing of a selection of individual genes (<https://www.ncbi.nlm.nih.gov/books/NBK279899/>), particularly their use in the diagnosis of genetic conditions.

The PHG Foundation provides an overview of whole genome sequencing (<http://www.phgfoundation.org/file/10365/>) and how it can be used in healthcare.

The Mount Sinai School of Medicine Genomics Core Facility describes the techniques used in whole exome sequencing (<http://icahn.mssm.edu/research/genomics/core-facility/whole-exome>).

The Smithsonian National Museum of Natural History's exhibit 'Genome: Unlocking Life's Code' discusses the advancements (<http://unlockinglifescode.org/explore/genome-within-us/reading-lifes-code>) made in DNA sequencing.

What are genome editing and CRISPR-Cas9?

Genome editing (also called gene editing) is a group of technologies that give scientists the ability to change an organism's DNA. These technologies allow genetic material to be added, removed, or altered at particular locations in the genome. Several approaches to genome editing have been developed. A recent one is known as CRISPR-Cas9, which is short for clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9. The CRISPR-Cas9 system has generated a lot of excitement in the scientific community because it is faster, cheaper, more accurate, and more efficient than other existing genome editing methods.

CRISPR-Cas9 was adapted from a naturally occurring genome editing system in bacteria. The bacteria capture snippets of DNA from invading viruses and use them to create DNA segments known as CRISPR arrays. The CRISPR arrays allow the bacteria to "remember" the viruses (or closely related ones). If the viruses attack again, the bacteria produce RNA segments from the CRISPR arrays to target the viruses' DNA. The bacteria then use Cas9 or a similar enzyme to cut the DNA apart, which disables the virus.

The CRISPR-Cas9 system works similarly in the lab. Researchers create a small piece of RNA with a short "guide" sequence that attaches (binds) to a specific target sequence of DNA in a genome. The RNA also binds to the Cas9 enzyme. As in bacteria, the modified RNA is used to recognize the DNA sequence, and the Cas9 enzyme cuts the DNA at the targeted location. Although Cas9 is the enzyme that is used most often, other enzymes (for example Cpf1) can also be used. Once the DNA is cut, researchers use the cell's own DNA repair machinery to add or delete pieces of genetic material, or to make changes to the DNA by replacing an existing segment with a customized DNA sequence.

Genome editing is of great interest in the prevention and treatment of human diseases. Currently, most research on genome editing is done to understand diseases using cells and animal models. Scientists are still working to determine whether this approach is safe and effective for use in people. It is being explored in research on a wide variety of diseases, including single-gene disorders such as cystic fibrosis, hemophilia, and sickle cell disease. It also holds promise for the treatment and prevention of more complex diseases (<https://ghr.nlm.nih.gov/primer/mutationsanddisorders/complexdisorders>), such as cancer, heart disease, mental illness, and human immunodeficiency virus (HIV) infection.

Ethical concerns arise when genome editing, using technologies such as CRISPR-Cas9, is used to alter human genomes. Most of the changes introduced with genome editing are limited to somatic cells, which are cells other than egg and sperm cells. These changes affect only certain tissues and are not passed

from one generation to the next. However, changes made to genes in egg or sperm cells (germline cells) or in the genes of an embryo could be passed to future generations. Germline cell and embryo genome editing bring up a number of ethical challenges, including whether it would be permissible to use this technology to enhance normal human traits (such as height or intelligence). Based on concerns about ethics and safety, germline cell and embryo genome editing are currently illegal in many countries.

Scientific journal articles for further reading

Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest.* 2014 Oct;124(10):4154-61. doi: 10.1172/JCI72992. Epub 2014 Oct 1. Review. PubMed: 25271723 (<https://www.ncbi.nlm.nih.gov/pubmed/25271723>). Free full-text available from PubMed Central: PMC4191047 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4191047/>).

Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell.* 2014 Jun 5;157(6):1262-78. doi:10.1016/j.cell.2014.05.010. Review. PubMed: 24906146 (<https://www.ncbi.nlm.nih.gov/pubmed/24906146>). Free full-text available from PubMed Central: PMC4343198 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4343198/>).

Komor AC, Badran AH, Liu DR. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell.* 2017 Apr 20;169(3):559. doi:10.1016/j.cell.2017.04.005. PubMed: 28431253 (<https://www.ncbi.nlm.nih.gov/pubmed/28431253>).

Lander ES. The Heroes of CRISPR. *Cell.* 2016 Jan 14;164(1-2):18-28. doi:10.1016/j.cell.2015.12.041. Review. PubMed: 26771483 (<https://www.ncbi.nlm.nih.gov/pubmed/26771483>).

For more information about CRISPR-Cas9 and other genome editing technologies:

The National Human Genome Research Institute has a series of fact sheets about genome editing:

- Overview of genome editing (<https://www.genome.gov/27569222/genome-editing/>)
- How does genome editing work? (<https://www.genome.gov/27569223/how-does-genome-editing-work/>)
- How is genome editing used? (<https://www.genome.gov/27569224/how-is-genome-editing-used/>)

- What are the ethical concerns about genome editing? (<https://www.genome.gov/27569225/what-are-the-ethical-concerns-about-genome-editing/>)
- What do people think about genome editing? (<https://www.genome.gov/27569226/what-do-people-think-about-genome-editing/>)
- What's happening in genome editing right now? (<https://www.genome.gov/27569227/whats-happening-in-genome-editing-right-now/>)

Questions and answers about CRISPR (<https://www.broadinstitute.org/what-broad/areas-focus/project-spotlight/questions-and-answers-about-crispr>) are available from the Broad Institute.

The Personal Genetics Education Project has a fact sheet, Genetic Modification, Genome Editing, and CRISPR (<https://pged.org/genetic-modification-genome-editing-and-crispr/>), that provides an introduction to genome editing.

Yourgenome.org (from the Wellcome Genome Campus) provides information for the public about CRISPR-Cas9 (<https://www.yourgenome.org/facts/what-is-crispr-cas9>).

A video illustrating how CRISPR-Cas9 works (<https://www.youtube.com/watch?v=2pp17E4E-O8>) is available from the McGovern Institute for Brain Research at MIT.

The American Society of Human Genetics has published a position statement on human germline genome editing ([http://www.cell.com/ajhg/fulltext/S0002-9297\(17\)30247-1](http://www.cell.com/ajhg/fulltext/S0002-9297(17)30247-1)).

ClinicalTrials.gov has a list of human studies using genome editing (<https://clinicaltrials.gov/ct2/results?cond=&term=CRISPR+OR+genome+editing+OR+gene+editing>) related to various diseases.



<https://ghr.nlm.nih.gov/>

Lister Hill National Center for Biomedical Communications
U.S. National Library of Medicine
National Institutes of Health
Department of Health & Human Services

Published on October 17, 2017