#### Learning Outcomes

Upon completing this practical the student will be able to

- easily search for, find, and utilize online genetics sources;
- easily navigate through the National Center for Biotechnology Information (NCBI)'s web page;

1

• start answering biological questions through online sources.

### Background

Internet has become a fundamental tool in human genetics and genomics education and research. It provides access to numerous online sources that facilitate data storage, access, and analysis by end users including scientists, students, clinicians, patients, and public readers [1].

Development of online sources for human genetics and genomics began in the 1980s. Since then, online sources providing information and data on genetics and genomics research have been rapidly growing. Internet has become the most popular source for accommodating the most up-to-date research outcomes and innovations in the field [2].

A broad range of genetic and genomic data is available online for access and analysis. Examples include online databases for gene mapping, mapping of chromosomal regions, epigenome mapping, sequence data on DNA, RNA, and protein, species-specific sequence data and comparisons among species, disease-specific data, gene expression data from normal and diseased tissues, genome browsers, and literature information. Utilization of these sources could be tailored to address specific research questions by end users.

One of the most important advances in genetics has been the generation of whole genome sequence data. The availability of whole genome sequences offers a vast array of genetic data on coding regions, noncoding regions, chromosomal structure and gene organization, genes, gene locations on chromosomes, number of genes, number of base pairs, nucleotide sequences, and DNA variants. [3]

Human Genetics and Genomics: A Practical Guide, First Edition.

Bahar Taneri, Esra Asilmaz, Türem Delikurt, Pembe Savas, Seniye Targen, and Yagmur Esemen.

© 2020 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2020 by Wiley-VCH Verlag GmbH & Co. KGaA.

So far, diverse organisms including species from bacteria, eukarya, and archaea domains have been sequenced. Some of the earlier examples are *Haemophilus influenzae* (1995), *Saccharomyces cerevisiae* (1996), and *Caenorhabditis elegans* (1998) [4–6].

For the purposes of this book, only the human genome project will be detailed.

### Human Genome Project

Human genome project, launched in 1990, is an internationally renowned research project. It was constructed to define the genetic, physical, and sequence maps of the human genome by deciphering the entire nucleotide sequence. Revelation of the entire genomic sequence enabled identification of genes dispersed throughout the genome [7, 8].

Access to the complete genome sequence has revealed information on genes, regulatory elements, and chromosomal organization and structure and this has facilitated the start of a new era in the field of human genetics [8].

In the recent years, advancement of next generation sequencing technologies has enabled time-efficient and cost-effective sequencing of whole genomes. These high-throughput sequencing technologies define a milestone in human genetics and genomics research. Such technologies enable investigating genome integrity and genetic variations across the human genome and hold great promise for easing the diagnosis of Mendelian conditions as well as multifactorial disorders. Moreover, these technologies aid in translation of human genetic variants into clinics by targeting prediction, prevention, and variant-specific drug discovery [8].

Rapid accumulation of large volumes of sequencing data has made manual functional analysis impossible. Hence, computational tools and databases have been employed for data analysis of high-throughput sequencing. It should be noted that in most cases computational sequence analysis should be supplemented with further functional studies in the investigation of inherited human traits and disorders [8].

As an outcome of high-throughput sequence analysis large amounts of genomic and epigenomic data have become available in databases such as the National Center for Biotechnology (NCBI), University of California Santa Cruz (UCSC) Genome Browser, the encyclopedia of DNA elements (ENCODE), and the roadmap epigenomics project. These databases are detailed below.

### National Center for Biotechnology Information (NCBI)

NCBI is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health, legally established in 1988. Its initial aim was to develop computer-based systems for collection, interpretation, and presentation of accumulating data in human molecular biology, genetics, and biochemistry. It has since evolved to act as an international platform for data sharing and computer-based data analysis, comprehensively covering information from all organisms [9].

NCBI's homepage contains a wide range of biological information that can be accessed and downloaded. NCBI's Entrez interface provides the links for a vast array of databases, including **Gene**, a database with a specific focus on gene-specific data providing information on gene locus, gene product, nucleic acid sequence, annotated sequence variants, gene maps, and gene homologs with additional links to external databases. NCBI provides access to **OMIM (online Mendelian inheritance in man)** a well-established database providing information on heritable disorders and the disease-causing genes, their chromosomal locations and maps, phenotype–genotype relationships, and clinical features (molecular genetics and cytogenetic aspects are also involved). Another fundamental constituent of NCBI is the **PubMed** database, which provides access to the online archive of biomedical literature abstracts and articles through MEDLINE and scientific journals. The organization of NCBI allows interconnection among these various databases.

Several other databases are embedded within the NCBI, which are listed in NCBI's homepage and use of these databases is demonstrated through NCBI's training and tutorials. Table 1.1 lists some of these databases.

### UCSC Genome Browser

Upon completion of the first working draft of the human genome project in 2000s, the assembled genome has been accessible at the UCSC genome browser [10]. Today, the genome browser broadly serves to visualize and analyze genome assembly data that is gathered from UCSC or from other sources (user-supplied data). UCSC harbors the genome assembly of more than 93 organisms including humans, chimpanzees, mice, and zebra fish [10]. UCSC Genome Browser enables access to the following data: mapping and sequencing (i.e. base positions, GC contents, contigs, and scaffolds), genes and gene predictions (i.e. consensus CDS, noncoding RNA), phenotype and literature, mRNA and ESTs, expression, regulation (i.e. integrated regulation from ENCODE, CpG islands), comparative genomics, variation (i.e. single nucleotide polymorphisms [SNPs]), and repeats (i.e. microsatellite) [10, 11].

In addition, there are also various analysis tools available in the UCSC Genome Browser including the genome browser, which enables visualization of genomic data; BLAT, which is used for sequence alignment; *in-silico* PCR, which aligns primer sequences to genomes; and Gene Sorter, which enables search for similar genes by expression metrics.

## The Encyclopedia of DNA Elements (ENCODE)

ENCODE, a collaborative project of different nations funded by the National Human Genome Institute, serves as a follow-up of the human genome project. It aims to investigate the functional DNA sequences of the human genome.

| Database                                               | Description                                                                                                                                                                     | URL                                  |
|--------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|
| ClinVar                                                | Designates genomic variation and depicts how they relate to human health                                                                                                        | http://www.ncbi.nlm.nih.gov/clinvar/ |
| dbVar                                                  | Annotates genomic structural variations<br>including deletions, duplications,<br>insertions, inversions, translocations,<br>and other complex rearrangements                    | http://www.ncbi.nlm.nih.gov/dbvar/   |
| Gene                                                   | Provides descriptive and physical information about the gene of interest                                                                                                        | http://www.ncbi.nlm.nih.gov/gene/    |
| Genome                                                 | Organizes the genomic data of any<br>species and provides access to genome<br>resources of humans, microbe genome,<br>organelles, viruses, and prokaryotic<br>reference genomes | http://www.ncbi.nlm.nih.gov/genome/  |
| MedGen                                                 | Provides information on the medical aspects of the human genetics                                                                                                               | http://www.ncbi.nlm.nih.gov/medgen/  |
| Nucleotide                                             | Provides nucleotide sequences adopted<br>from several resources. Genome, gene,<br>and transcript sequences are available<br>through NCBI nucleotide database                    | http://www.ncbi.nlm.nih.gov/nuccore/ |
| OMIM<br>(online<br>Mendelian<br>inheritance<br>in man) | Provides descriptive information on<br>inherited disorders and their genetic<br>architecture                                                                                    | http://www.ncbi.nlm.nih.gov/omim/    |
| Protein                                                | Provides protein sequences from<br>numerous sources                                                                                                                             | http://www.ncbi.nlm.nih.gov/protein/ |
| РМС                                                    | Provides free access to biomedical and<br>life science journals at the U.S. National<br>Institutes of Health (NIH) and National<br>Library of Medicine (NLM)                    | http://www.ncbi.nlm.nih.gov/pmc/     |
| PubMed                                                 | Harbors comprehensive citations of<br>biomedical research and literature from<br>MEDLINE, life science journals, and<br>online books                                            | http://www.ncbi.nlm.nih.gov/pubmed/  |
| SNP                                                    | Harbors single nucleotide<br>polymorphisms (SNPs) and small-scale<br>variations of other types                                                                                  | http://www.ncbi.nlm.nih.gov/snp/     |

 Table 1.1
 List of selected databases accessible through NCBI's homepage.

ENCODE provides information on the reference human genome including coding and noncoding regions, i.e. protein-coding genes, noncoding transcripts, long noncoding RNAs, pseudogenes; subcellular localization of transcripts in humans by RNA sequencing; DNA methylation; chromatin accessibility indicator of DNA regulatory regions; transcription factor binding sites, three-dimensional space interaction; and integration of SNPs achieved from genome-wide association studies (GWAS) [12–14].

## **Roadmap Epigenomic Project**

The information gathered from the genome sequence alone is not sufficient to explain the physiological and pathological processes of the human body. When thinking about gene expression and its regulation throughout the genome, epigenetics must also be considered. Epigenetics studies the alteration of gene function through gene expression without changing the actual nucleotide sequence. Epigenome Project has been launched to reveal the epigenomic architecture of the human genome. This project studies epigenetic mechanisms including DNA methylation and histone modification. Understanding these mechanisms can provide better insight into gene expression and human diseases [15, 16].

It should be noted that this chapter does not exclusively cover all relevant online genetic sources.

## References

- Waggoner, D.J. (2014). Internet resources in medical genetics. *Current Pro*tocols in Human Genetics 80: 9.12.1. Available from: https://doi.org/10.1002/ 0471142905.hg0912s62 (accessed 11 August 2017).
- **2** Web resources: internet resources featured in this guide. *Nature Genetics* 32 (1): 77–79. Available from: http://www.nature.com/doifinder/10.1038/ng981 (accessed 11 August 2017).
- 3 Vucic, E.A., Thu, K.L., Williams, A.C. et al. (2010) Copy number variations in the human genome and strategies for analysis. In: *Genetic Variation, Methods and Protocols* (eds. Barnes, M.R. and Breen, G.) Springer. 103–117. Available from: https://www.springer.com/gp/book/9781603273664 (accessed 11 August 2017).
- 4 Fleischmann, R.D., Adams, M.D., White, O. et al. (1995). Whole-genome random sequencing and assembly *Haemophilus influenzae* of Rd. *Science* 269 (5223): 496–512. Available from: http://science.sciencemag.org/content/269/ 5223/496.long (accessed 11 August 2017)\ignorespaceshttps://doi.org/10.1126/ science.7542800.
- 5 Goffeau, A., Barrell, B.G., Bussey, H. et al. (1996). Life with 6000 genes. *Science* 274 (5287): 546. Available from: https://pdfs.semanticscholar.org/0635/ d3f7ac72619a11f5b5d1ad55577588344852.pdf (accessed 11 August 2017).
- 6 C. elegans Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282 (5396): 2012–2018. Available from: https://www.ncbi.nlm.nih.gov/pubmed/9851916 (accessed 11 August 2017).
- 7 Olson, M.V. (1993). The human genome project. *Proceedings of the National Academy of Sciences of the United States of America* 90 (10): 4338–4344. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8506271 (accessed 11 August 2017).
- 8 Lander, E.S., Linton, L.M., Birren, B. et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409 (6822): 860–921. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/11237011; http://www.nature.com/ nature/journal/v409/n6822/pdf/409860a0.pdf (accessed 11 August 2017).

- **9** National Center for Biotechnology Information. Available from: http://www .ncbi.nlm.nih.gov/ (accessed 11 August 2017).
- 10 University of California Santa Cruz UCSC Genome Browser. Available from: https://genome.ucsc.edu/ (accessed 16 August 2017).
- 11 Karolchik, D., Hinrichs, A.S., and Kent, W.J. (2007). The UCSC Genome Browser. *Current Protocols in Bioinformatics*; Chapter 1: Unit 1.4. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18428780 (accessed 11 August 2017) https://doi.org/10.1002/0471250953.bi0104s17.
- 12 ENCODE: Encyclopedia of DNA Elements. Available from: https://www .encodeproject.org/ (accessed 16 August 2017).
- 13 Qu, H. and Fang, X. (2013). A brief review on the human encyclopedia of DNA elements (ENCODE) project. *Genomics, Proteomics & Bioinformatics* 11 (3): 135–141. Available from: https://www.ncbi.nlm.nih.gov/pubmed/ 23722115 (accessed 11 August 2017)https://doi.org/10.1016/j.gpb.2013.05.001.
- 14 Diehl, A.G. and Boyle, A.P. (2016). Deciphering ENCODE. Trends in Genetics 32 (4): 238–249. Available from: https://www.ncbi.nlm.nih.gov/pubmed/ 26962025 (accessed 11 August 2017).
- **15** Roadmap Epigenomics Project . Available from: http://www .roadmapepigenomics.org/ (accessed 16 August 2017).
- 16 Bernstein, B.E., Stamatoyannopoulus, J.A., Costello, J.F. et al. (2010). The NIH roadmap epigenomics mapping consortium. *Nature Biotechnology* 28 (10): 1045–1048. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20944595 (accessed 11 August 2017).

## **Exercise Questions**

- 1 In this exercise, you will be using online genetics sources in order to gain a broader understanding of genetics databases, which are commonly used in education and scientific research as bioinformatics tools. For this task, you have been provided with the screenshots of NCBI. In order to familiarize yourself with this tool, you are guided through the basic processes of using NCBI in a stepwise manner.
  - **Step 1**: Enter the following link to your browser: http://www.ncbi.nlm.nih .gov. You will see the NCBI homepage appear on your screen as shown in Figure 1.1.
  - **Step 2**: Click on the arrow next to the box labeled **All Databases**. Here, you will find a list of different databases, each of which is tailored to the purpose of your search (Figure 1.2).
  - **Step 3**: Scroll the mouse down to find and select **PubMed**. PubMed is an electronic database that catalogs biomedical journals and articles of your interest (Figure 1.3).
  - **Step 4**: In this task, you are asked to search the PubMed for articles including information on the *BRCA* genes. Figure 1.4 illustrates the search for the *BRCA 1 gene*. As you type the name of the gene into the browser, PubMed

|                                                  | v To ♥                         |                                |                                    | Sign in to NC                                                                                                        |
|--------------------------------------------------|--------------------------------|--------------------------------|------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| National Center for<br>Biotechnology Information |                                |                                |                                    |                                                                                                                      |
| NCBI Home                                        | Welcome to NCBI                |                                |                                    | Popular Resources                                                                                                    |
| Resource List (A-Z)                              | The National Center for Biotec | hnology Information advances   | science and health by providing    | PubMed                                                                                                               |
| All Resources                                    | access to biomedical and gene  | omic information.              |                                    | Bookshelf                                                                                                            |
| Chemicals & Bioassays                            | About the NCBI   Mission   C   | Organization   NCBI News & Blo | 2g                                 | PubMed Central                                                                                                       |
| Data & Software                                  |                                |                                |                                    | PubMed Health                                                                                                        |
| DNA & RNA                                        | Submit                         | Download                       | Learn                              | BLAST                                                                                                                |
| Domains & Structures                             | Deposit data or                | Transfer NCBI data to          | Find help documents,               | Nucleotide                                                                                                           |
| Genes & Expression                               | manuscripts into NCBI          | your computer                  | attend a class or watch a tutorial | Genome                                                                                                               |
| Genetics & Medicine                              | databases                      |                                |                                    | SNP                                                                                                                  |
| Genomes & Maps                                   | - I A I                        |                                |                                    | Gene                                                                                                                 |
| Homology                                         | - T                            |                                |                                    | Protein                                                                                                              |
| Literature                                       |                                |                                |                                    | PubChem                                                                                                              |
| Proteins                                         |                                |                                |                                    |                                                                                                                      |
| Sequence Analysis                                | Develop                        | Analyze                        | Research                           | NCBI News & Blog                                                                                                     |
| Taxonomy                                         | Use NCBI APIs and code         | Identify an NCBI tool for      | Explore NCBI research              | NCBI Replacing Obsolete NCBI                                                                                         |
| Training & Tutorials                             | libraries to build             | your data analysis task        | and collaborative projects         | Genomes (chromosome) and Removing                                                                                    |
| Variation                                        | applications                   |                                | 1                                  | (alu_repeats) BLAST databases<br>09 Aug 201                                                                          |
|                                                  |                                | 38                             | . <u>(*)</u>                       | Sequence Viewer 3.22 now available<br>03 Aug 201                                                                     |
|                                                  |                                |                                |                                    | Sequence Viewer 3.22 has several new<br>features, improvements and bug fixes,<br>including improved rendering on BAM |
|                                                  |                                |                                |                                    | Former NCRI Director Dr. David 1                                                                                     |

**Figure 1.1** NCBI Homepage (Image on Internet; accessed 12 November 2018). Source: Available from: https://www.ncbi.nlm.nih.gov. Courtesy of the U.S. National Library of Medicine.

| S NCBI Resources 🖸                                                                                                                                                                                                         | How To 🖂                                                                                                                                                                                   |                                                                     |                                                                                                                                           |                                                                                                                 | Sign in to NCBI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SNCBI<br>National Center for<br>Biotechnology Information                                                                                                                                                                  | <ul> <li>All Databases</li> <li>Assembly</li> <li>Biocollections</li> </ul>                                                                                                                |                                                                     |                                                                                                                                           |                                                                                                                 | Search                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| NCBI Home<br>Resource List (A-Z)<br>All Resources<br>Chemicals & Bioassays<br>Data & Schware<br>DNA & RNA<br>Domaina & Structures<br>Genes & Expression<br>Genetics & Medicine<br>Genomes & Maps<br>Homology<br>Literature | BioProject<br>BioSampie<br>BioSayne<br>Clonvar<br>Clone<br>Conserved Domains<br>dbGaP<br>dbVar<br>EST<br>Gene<br>Genome<br>Genome<br>GEO DataSets<br>GEO Profiles<br>GSS<br>GTR            | ICBI<br>for Biotec<br>al and gen<br><u>Mission   (</u><br>t<br>NCBI | hnology Information advances<br>smic Information.<br>Dransization   NCBI News & Bic<br>Download<br>Transfer NCBI data to<br>your computer | science and health by providing<br>the<br>Learn<br>Find help documents,<br>attend a class or watch a<br>tutonal | Popular Resources<br>PubMed<br>Bockholf<br>PubMed Central<br>PubMed Health<br>BLAST<br>Nucleotide<br>Genome<br>SNP<br>Gene<br>Protein<br>PubChem                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Proteins<br>Sequence Analysis<br>Taxonomy<br>Training & Tutoriats<br>Variation                                                                                                                                             | Identical Protein Groups<br>MedGen<br>MeSH<br>NCBI Web Site<br>NLM Catalog<br>NLL Catalog<br>NLL Catalog<br>NLL Catalog<br>NLL Catalog<br>NLM Catalog<br>PMC<br>PopSet<br>Probe<br>Protein | p<br>nd code                                                        | Analyze<br>Identify an NCBI tool for<br>your data analysis task                                                                           | Research<br>Explore NCBI research<br>and collaborative projects                                                 | NCBI News & Blog<br>NCBI Reptacing Obsolete NCBI<br>Genemas (chromosome) and Removing<br>Human ALU repeat elements<br>(Gui_repeat) BLAST database<br>(Gui_repeat) BLAST database<br>(Gui_repeat)<br>(Gui_repeat) BLAST database<br>(Gui_repeat)<br>(Gui_repeat) BLAST database<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui_repeat)<br>(Gui<br>repeat)<br>(Gui<br>repea |

**Figure 1.2** NCBI homepage (All Databases; accessed 12 November 2018). Source: Available from: https://www.ncbi.nlm.nih.gov. Courtesy of the U.S. National Library of Medicine.

will show additional keywords that might aid you further in your search. Refer to the image below and try and perform this search.

**Step 5**: Now go back to NCBI homepage and choose the database **Gene** as shown previously in *Step 2*. Using this database, perform a new search for the *BRCA1* gene. With this search, you will find gene-specific information

| S NCBI Resources                                                           | Gene                                                                             |                                          |                                                                                     |                                 | Sign in to NCB                                                                                                       |        |
|----------------------------------------------------------------------------|----------------------------------------------------------------------------------|------------------------------------------|-------------------------------------------------------------------------------------|---------------------------------|----------------------------------------------------------------------------------------------------------------------|--------|
| SNCBI<br>National Center for<br>Biotechnology Information                  | Genome<br>GEO DataSets<br>GEO Profiles<br>GSS                                    |                                          |                                                                                     |                                 | Search                                                                                                               |        |
| NCBI Home<br>Resource List (A-Z)<br>All Resources<br>Chemicals & Bioassays | GTR<br>HomoloGene<br>Identical Protein Groups<br>MedGen<br>MeSH<br>NCBI Web Site | for Biotec<br>al and gene<br>Mission   C | hnology Information advances<br>omic information.<br>Organization   NCBI News & Bic | science and health by providing | Popular Resources<br>PubMed<br>Bookshelf<br>PubMed Central<br>PubMed Hoelth                                          |        |
| Data & Software                                                            | NLM Catalog                                                                      |                                          | Download                                                                            | Loorn                           | BLAST                                                                                                                |        |
| DNA & RNA                                                                  | OMIM                                                                             |                                          | Transfer NORL data to                                                               | Learn Fied hele desuments       | Nucleotide                                                                                                           |        |
| Domains & Structures                                                       | PMC<br>PopSet                                                                    | PMC                                      | NCBI                                                                                | CBI your computer               | attend a class or watch a                                                                                            | Genome |
| Genes & Expression                                                         |                                                                                  | pSet                                     | tu                                                                                  | tutorial                        | SNP                                                                                                                  |        |
| Genetics & Medicine                                                        | Protein                                                                          | 1                                        |                                                                                     | le.                             | Gene                                                                                                                 |        |
| Genomes & Maps                                                             | Protein Clusters                                                                 |                                          | 1.1.1                                                                               |                                 | Protein                                                                                                              |        |
| Homology                                                                   | PubChem BioAssay                                                                 |                                          |                                                                                     |                                 | PubChem                                                                                                              |        |
| Literature                                                                 | PubChem Compound                                                                 |                                          |                                                                                     |                                 | Publik                                                                                                               |        |
| Proteins                                                                   | PubMed                                                                           |                                          |                                                                                     |                                 |                                                                                                                      |        |
| Sequence Analysis                                                          | PubMed Health                                                                    | p                                        | Analyze                                                                             | Research                        | NCBI News & Blog                                                                                                     |        |
| Taxonomy                                                                   | SNP                                                                              | nd code                                  | Identify an NCBI tool for                                                           | Explore NCBI research           | NCBI Replacing Obsolete NCBI                                                                                         |        |
| Training & Tutorials                                                       | Sparcle                                                                          |                                          | your data analysis task                                                             | and collaborative projects      | Genomes (chromosome) and Removing<br>Human Al U repeat elements                                                      |        |
| Variation                                                                  | SnA<br>Structure<br>Taxonomy<br>ToolKit<br>ToolKitAll                            |                                          | 888                                                                                 | <u>s</u>                        | (alu_repeats) BLAST databases<br>09 Aug 2017<br>Sequence Viewer 3.22 now available<br>03 Aug 2017                    |        |
|                                                                            | ToolKitBook<br>ToolKitBookgh<br>UniGene                                          |                                          |                                                                                     |                                 | Sequence Viewer 3.22 has several new<br>features, improvements and bug fixes,<br>includino improved rendering on BAM |        |

Figure 1.3 NCBI homepage (PubMed; accessed 12 November 2018). Source: Available from: https://www.ncbi.nlm.nih.gov. Courtesy of the U.S. National Library of Medicine.



**Figure 1.4** NCBI-PubMed database searching the *BRCA1* gene (Image on internet; accessed 11 August 2017). Source: Available from: http://www.ncbi.nlm.nih.gov/pubmed/?term=BRCA1+ gene. Courtesy of the U.S. National Library of Medicine.

including the gene name, gene symbol, gene location, associated pathways, variations, links to related phenotypes, and locus-specific resources (Figure 1.5).

Step 6: In this step, you are asked to perform a search for BRCA2 gene in Homo sapiens (human) and its homologs in Mus musculus (mouse) and

| S NCBI Resources 🕑 H                                                                                 | łow To 🗹                                          |                                                                                    |                                                                                 |                                                                                  |               | Sign in to NCBI                                                                                                                                               |
|------------------------------------------------------------------------------------------------------|---------------------------------------------------|------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Gene                                                                                                 | Gene ᅌ Bl<br>Cr                                   | RCA1                                                                               |                                                                                 |                                                                                  |               | Search Help                                                                                                                                                   |
| Gene sources<br>Genomic<br>Mitochondria<br>Organelles<br>Plasmids                                    | Tabular - 20 per<br>Did you mear<br>Search Gene t | page - Sort by Relevance<br>BRCA1 as a gene symbol<br>or <u>BRCA1</u> as a symbol. | •<br>>!?                                                                        | ٤                                                                                | Send to: 👻    | Filters: Manage Filters Results by taxon                                                                                                                      |
| Categories<br>Alternatively spliced<br>Annotated genes<br>Non-ooding<br>Protein-ooding<br>Pseudogene | Search result<br>Items: 1 to 20 c                 | S<br>f 14210<br>discontinued or replaced ite                                       | << First                                                                        | < Prev Page 1 of 711 Next 2                                                      | • Last >>     | lop Organisms [Iree]<br>Horos sapiens (784)<br>Mus musculus (109)<br>Camelina sativa (82)<br>Salmo salar (69)<br>Brassica napus (57)<br>Al other taxa (13109) |
| Sequence content<br>CCDS<br>Ensembl<br>RefSeq<br>RefSeqGene                                          | Name/Gene ID BRCA1 ID: 672                        | Description<br>BRCA1, DNA repair<br>associated [Homo<br>sapiens (human)]           | Location<br>Chromosome 17,<br>NC_000017.11<br>(4304429543125483,<br>complement) | Aliases<br>BRCAI, BRCC1, BROVCA1,<br>FANCS, IRIS, PNCA4,<br>PPP1R53, PSCP, RNF53 | MIM<br>113705 | Find related data                                                                                                                                             |
| Status clear<br>✓ Current<br>Chromosome                                                              | r D <u>Brca1</u><br>ID: 12189                     | breast cancer 1, early<br>onset [ <i>Mus musculus</i><br>(house mouse)]            | Chromosome 11,<br>NC_000077.6<br>(101488761101551957,<br>complement)            |                                                                                  |               | Find items                                                                                                                                                    |
| more                                                                                                 | D: 497672                                         | BRCA1, DNA repair<br>associated [Rattus<br>norvegicus (Norway rat)]                | Chromosome 10,<br>NC_005109.4<br>(8939482189455009,<br>complement)              |                                                                                  |               | Search details           BRCA1[All Fields] AND alive[prop]                                                                                                    |
| Show additional filters                                                                              | D: 403437                                         | breast cancer 1, early<br>onset [Canis lupus<br>familiaris (dog)]                  | Chromosome 9,<br>NC_006591.3<br>(1995198120025494)                              |                                                                                  |               | Search See more                                                                                                                                               |
|                                                                                                      | D: 827854                                         | protein BREAST<br>CANCER                                                           | Chromosome 4,<br>NC_003075.7                                                    | AT4G21070, ARABIDOPSIS<br>THALIANA BREAST                                        |               |                                                                                                                                                               |

**Figure 1.5** NCBI-gene database searching for *BRCA1* gene (Image on Internet; accessed 17 July 2016). Source: Available from: http://www.ncbi.nlm.nih.gov/gene/?term=BRCA1. Courtesy of the U.S. National Library of Medicine.

| S NCBI Resources                                                                                                                                                                | How To 🗹                                                                                                      |                                                                                                                                      |                                                                                                              |                                                                                                   |                          | Sign in to NCBI                                                                                                                                                                                 |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Gene                                                                                                                                                                            | Gene \$                                                                                                       | BRCA2  <br>Create RSS Create a                                                                                                       | alert Advanced                                                                                               |                                                                                                   |                          | Soarch Help                                                                                                                                                                                     |
| Gene sources<br>Ganomic<br>Organelles<br>Categories<br>Alternatively sploed<br>Annotalied genes<br>Non-coding<br>Protein-coding<br>Preter-coding<br>Sequence<br>content<br>CCDS | Tabular + 20 pe<br>See BRCA2<br>brca2 in Hom<br>Search resul<br>Items: 1 to 20<br>See also 30<br>Name/Gene ID | r page - Sort by Relev.<br>BRCA2, DNA repair a:<br>o sapiens Mus muscult<br>ts<br>of 1887<br>discontinued or replaced<br>Description | ance +<br>ssociated<br>is Ratius norregicus Al 22<br><first <="" pr<br="">Licens.</first>                    | 77 Gene records                                                                                   | Send to: + > Last >> MIM | Hide sidebar>>> Filters : Managa Filters Results by taxon To Organisms [Iree] Mus musculus (31) Droophils melanopaster (22) Ratus norvegics (15) Tupak chirensis (12) Al other taxa (1588) More |
| Ensembl<br>RefSeq<br>RefSeqGene<br>Status de<br>Vourrent<br>Chromosome<br>locations<br>more                                                                                     | BRCA2<br>ID: 675<br>ar<br>D: 12190                                                                            | BRCA2, DNA repair<br>associated [Homo<br>sapiens (human)]<br>breast cancer 2,<br>early onset [Mus<br>musculus (house<br>mouse)]      | Chromosome 13,<br>NC_000013.11<br>(3231548032399672)<br>Chromosome 5,<br>NC_000071.6<br>(150522021150570147) | BRCC2, BROVCA2,<br>FACD, FAD, FAD1,<br>FANCD, FANCD1,<br>GLM3, PNCA2,<br>XRCC11<br>Fancd1, RAB163 | 600185                   | Find related data Database: Select  Find forms                                                                                                                                                  |
| Clear all<br>Show additional filters                                                                                                                                            | <ul> <li>Brca2</li> <li>ID: 360254</li> <li>Brca2</li> <li>ID: 37916</li> </ul>                               | BRCA2, DNA repair<br>associated [ <i>Rattus</i><br>norvegicus (Norway<br>rat)]<br>Breast cancer 2,<br>early onset homolog            | Chromosome 12,<br>NC_005111.4<br>(503660544754)<br>Chromosome 2R,<br>NT_033778.4                             | Dmel_CG30169,<br>30169, BBC42                                                                     |                          | Search details<br>BRCA2[All Fields] AND<br>alive[prop]                                                                                                                                          |
|                                                                                                                                                                                 |                                                                                                               | [Drosophila<br>melanogaster (fruit                                                                                                   | (2452617224529581)                                                                                           | BcDNA:SD25109,<br>CG13583, CG13584,                                                               |                          | Search See more                                                                                                                                                                                 |

**Figure 1.6** NCBI-gene database searching for *BRCA2* gene (Image on Internet; accessed 11 August 2017). Source: Available from: http://www.ncbi.nlm.nih.gov/gene/?term=BRCA2. Courtesy of the U.S. National Library of Medicine.

*Rattus norvegicus* (rat) (Figures 1.6 and 1.7). Explore the Results by taxon shown on upper right corner of Figure 1.6.

**Step** 7: Now, you are asked to perform a search using OMIM (online Mendelian inheritance in man). Select **OMIM** from the **All Databases** box through the NCBI homepage (Figure 1.8).

| S NCBI Resources 🗹 I                  | How To 🗹                                                                                                                                                                                             |                      | Sign in to NCBI                                              |
|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|--------------------------------------------------------------|
| Gene                                  | Gene CAdvanced                                                                                                                                                                                       |                      | Search                                                       |
| Full Report +                         |                                                                                                                                                                                                      | Send to: +           | Hide sidebar >>                                              |
| Brca2 breast canc                     | er 2, early onset [ <i>Mus musculus</i> (house mouse) ]                                                                                                                                              |                      | Table of contents Summary                                    |
| Gene ID: 12190, updated on            | 6-Aug-2017                                                                                                                                                                                           |                      | Genomic context                                              |
| Summary                               |                                                                                                                                                                                                      | ≈ ?                  | Genomic regions, transcripts, and products                   |
| Official Symbol<br>Official Full Name | Brca2 provided by MGI<br>breast cancer 2, early onset provided by MGI                                                                                                                                |                      | Expression<br>Bibliography                                   |
| See related                           | Ensembl:ENSMUSG00000041147 Vega:OTTMUSG00000055512                                                                                                                                                   |                      | vanation                                                     |
| Gene type                             | protein coding                                                                                                                                                                                       |                      | Pathways from BioSystems                                     |
| RefSeq status                         | VALIDATED                                                                                                                                                                                            |                      | Interactions                                                 |
| Organism<br>Lineage                   | Organism <u>Mus musculus</u><br>Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria;<br>Euarzhontodires: Glines: Rodentia: Myomomba: Murcidea: Murinae: Mus |                      | General gene information<br>Markers, Homology, Gene Ontology |
| Also known as                         | Fancd1; RAB163                                                                                                                                                                                       |                      | General protein information                                  |
| Orthologs                             | human all                                                                                                                                                                                            |                      | NCBI Reference Sequences (RefSeq)                            |
| - Genomic context                     |                                                                                                                                                                                                      | ▲ ?                  | Related sequences                                            |
| Location: 5 G3; 5 89.52 d             | See Brca2 in <u>Genome Data View</u>                                                                                                                                                                 | er <u>Map Viewer</u> | Additional links                                             |
| Exon count: 27                        |                                                                                                                                                                                                      |                      | Genome Browsers                                              |

**Figure 1.7** NCBI-gene database searching for *BRCA2* gene mouse homolog (Image on Internet; accessed 11 August 2017). Source: Available from: http://www.ncbi.nlm.nih.gov/gene/12190. Courtesy of the U.S. National Library of Medicine.

| 🗧 NCBI 🛛 Resources 🗹          | How To 🗹                                 |                                                                                                                            |                                                                                                                                                                            | Sign in to NCBI                                                       |
|-------------------------------|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| OMIM                          | OMIM ¢                                   | vanced                                                                                                                     |                                                                                                                                                                            | Bearch Help                                                           |
| OMII                          | M                                        | OMIM<br>OMIM is a comprehensive, au<br>available and updated daily. C<br>Medicine, Johns Hopkins Univ<br>home is omim.org. | thoritative compendium of human genes and genetic ph<br>MIM is authored and edited at the McKusick-Nathans in<br>ersity School of Medicine, under the direction of Dr. Ada | enotypes that is freely<br>stitute of Genetic<br>Hamosh. Its official |
| Using OMIM                    |                                          | OMIM tools                                                                                                                 | Related Resources                                                                                                                                                          |                                                                       |
| Getting Started               |                                          | OMIM API                                                                                                                   | ClinVar                                                                                                                                                                    |                                                                       |
| FAQ                           |                                          |                                                                                                                            | Gene                                                                                                                                                                       |                                                                       |
|                               |                                          |                                                                                                                            | GTR                                                                                                                                                                        |                                                                       |
|                               |                                          |                                                                                                                            | MedGen                                                                                                                                                                     |                                                                       |
| Last updated on: 09 Aug 2017  |                                          |                                                                                                                            |                                                                                                                                                                            |                                                                       |
| You are here: NCBI > Genetics | s & Medicine > Online Mendelian Inherita | nce in Man (OMIM)                                                                                                          |                                                                                                                                                                            | Support Center                                                        |

**Figure 1.8** NCBI-OMIM database homepage (Image on Internet; accessed 11 August 2017). Source: Available from: http://www.ncbi.nlm.nih.gov/omim/. Courtesy of the U.S. National Library of Medicine.

**Step 8**: In this step, you will search for a neurodegenerative disorder, muscular dystrophy. Type in muscular dystrophy in the search window and begin your search. Record the first two entries and explore their associated links titled "Gene summaries" and "Genetic tests".

#### Additional Exercise Questions 11

| S NCBI Resou                                                                                                | irces 🗹 How To 🗹                                            |                             |                                             |                                                                | Sign in to NCBI |
|-------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|-----------------------------|---------------------------------------------|----------------------------------------------------------------|-----------------|
| OMIM                                                                                                        | OMIM                                                        | Create alert Limits Advant  | ced                                         | Search                                                         | Help            |
| Summary + 20 pe                                                                                             | r page <del>-</del>                                         |                             | Send to: +                                  | Filter your results:<br>All (431)                              |                 |
| Items: 1 to 20 of                                                                                           | 431                                                         |                             | << First < Prev Page 1 of 22 Next > Last >> | OMIM OHISTS (102)                                              | Manage Filters  |
| 204730 - AN     OSTEOPOR     OMIM: 204730     Gene summari                                                  | INO ACIDURIA WIT<br>OSIS, AND ACIDOS<br>es Genetic tests Mi | TH MENTAL DEFICIENCY, DWARF | ISM. MUSCULAR DYSTROPHY.                    | Find related data<br>Database: Select                          | \$              |
| 310000 - MUSCULAR DYSTROPHY, MABRY TYPE     Send Items     Gene summaries. Genetic tests. Medical Iterature |                                                             |                             |                                             |                                                                |                 |
| 254000 - ML<br>3. OMIM: 254000<br>Gene summari                                                              | ISCULAR DYSTRO                                              | PHY, CONGENITAL, WITH INFAN | TILE CATARACT AND HYPOGONADISM              | Search details<br>muscular[All Fields]<br>dystrophy[All Fields | AND             |
| 254100 - ML     OMIM: 254100     Gene summari                                                               | ISCULAR DYSTRO                                              | PHY, CONGENITAL, WITH RAPID | PROGRESSION                                 | Search                                                         | See more        |

**Figure 1.9** NCBI-OMIM database searching for muscular dystrophy (Image on Internet; accessed 11 August 2017). Source: Available from: https://www.ncbi.nlm.nih.gov/omim/? term=muscular+dystrophy. Courtesy of the U.S. National Library of Medicine.

## **Additional Exercise Questions**

- 1 Search online and provide the names for each of the following:
  - a. Two conferences in the field of human genetics.
  - b. Two societies or associations about human genetics.
  - c. Two scientific journals in the field of human genetics.
  - d. Two books in the field of human genomics.
- **2** Using NCBI, answer the following questions:
  - a. When you search for the *BRCA1* gene in PubMed, what is the name of the first article that appears in the list of results?
  - b. State the name of the scientific journal.
  - c. List the date of publication.
  - d. State the name of the first author in the article.
  - e. List the PubMed ID of this publication.
- **3** Using NCBI gene database, search for the *CFTR* gene in humans (*Homo sapiens*).
  - a. What is the full name of this gene?
  - b. What is the function of this gene?
  - c. On which chromosome is this gene located?
  - d. Is this gene also known by other aliases or designations? If so, provide an example.
  - e. Is this gene found in other organisms? If so, provide the name of one such organism.
- **4** In the Exercise Questions section, you have searched for muscular dystrophy (MD) through OMIM. As shown in Figure 1.9, several types of muscular dystrophy exist. Select a specific type of MD from the search results and

click on the link provided to answer the following questions accordingly.

- a. What is the OMIM number for this disease?
- b. Provide the full name of the disease.
- c. What is the cytogenetic location of the mutation causing the disease?
- d. Write down two clinical features of the disease.
- **5** Now, you are asked to repeat the same steps in Question 4 for color blindness. When you search OMIM, you will see that there are several types of color blindness. Select a specific type of color blindness from the search results and click on the link provided to answer the following questions accordingly.
  - a. What is the OMIM number for this disease?
  - b. Provide the full name of the disease.
  - c. What is the cytogenetic location of the mutation causing the disease?
  - d. Write down two clinical features of the disease.
- **6** You are given a list of gene symbols below. By using the online tools you explored earlier, find the full gene name, gene function, and one of the associated disorders for each of the genes listed below.

| Gene Symbol | Gene name | Gene function | Associated disease |
|-------------|-----------|---------------|--------------------|
| P53         |           |               |                    |
| RB1         |           |               |                    |
| HTT         |           |               |                    |
| APC         |           |               |                    |
| PTEN        |           |               |                    |

7 You are given a list of genetic disorders. By using the online tools you explored earlier, find the gene locations and known mutations for each of the syndromes and diseases listed below.

| Genetic disorder                                       | Associated genes/<br>genomic locations | Type of mutations causing the disease |
|--------------------------------------------------------|----------------------------------------|---------------------------------------|
| Cri-du-chat syndrome                                   |                                        |                                       |
| Prader–Willi<br>syndrome<br>Angelman syndrome          |                                        |                                       |
| Beckwith–Wiedemann<br>syndrome<br>Huntington's disease |                                        |                                       |
| Sickle cell anemia                                     |                                        |                                       |