15.1 Genomic Sequencing is an Extension of Genetic Mapping

- In Chp 13 and 14, transferring single genes from one organism to another was discussed. In the past, finding a gene of interest in an organism’s DNA took years. In 1990 the Human Genome Project set out to sequence the entire human genome and use that info to map all human genes. (The HGP also set out to map the genomes of other species commonly used in scientific research.)
- Mutant genes are the basis of genetic disorders—in humans and other organisms. Mutations were studied and cataloged, providing insight into genome organization (genomic maps).

- Mapping helps identify genes that cause disease
  - The first step in developing diagnostic tests and treatments for these disorders
Gene Linkage

• Genes located on the same chromosome tend to be inherited together and are said to show linkage

• When the degree of recombination between linked genes is measured (how frequently crossing-over takes place between two genes), the distance between them can be determined

Centimorgan (cM) (also called map unit, m.u.)

• First unit of distance between genes on chromosome

• One centimorgan equals a value of 1% crossing-over between two genes

• 1913-In the lab of T.H. Morgan first genetic map; 1933 won Nobel Prize in Physiology and Medicine
Crossing-Over Between Homologous Chromosomes in Meiosis

When two genes are linked, their chance of crossing-over increases as the distance between genes increases. This is called the recombination frequency.
First, two linked genes are identified. Crossover frequencies are used to construct genetic maps, giving the order and distance between genes:

- \( a \) and \( b \) are 12 cM or m.u. apart;
- \( a \) and \( c \) are 22 cM apart.

Note: These are *relative* gene locations not exact locations on a specific chromosome. It was not known exactly how many DNA bases were between genes.
Linkage and Recombination Can be Measured by Lod Scores

- **Lod method**
  - A statistical technique used to determine whether two genes are linked

- **Lod score** (*Lod stands for “log of the odds”)*
  - Calculated using computer software using a log scale
  - Genes are considered linked if the score is 3 or higher

- Research in humans using this technique was slow—only 5 cases of gene linkage had been discovered from 1933-1969!
1980 - marked the beginning of a new type of gene mapping

**Positional cloning**

- A recombinant DNA-based method of mapping and cloning genes with no prior information about the gene product or its function
- Inheritance of molecular markers is used to track the inheritance of genetic disorders in pedigrees and thus the gene responsible for the phenotype.
- Though this method was faster than gene-linkage studies, still only one gene could be identified at a time.
Some Genes Identified by Positional Cloning

<table>
<thead>
<tr>
<th>Chromosome 4</th>
<th>OMIM</th>
<th>Chromosome 17</th>
<th>OMIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huntington disease</td>
<td>143100</td>
<td>Breast cancer (BRCA1)</td>
<td>113705</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neurofibromatosis (NF1)</td>
<td>162200</td>
</tr>
<tr>
<td>Chromosome 5</td>
<td>OMIM</td>
<td>Chromosome 19</td>
<td>OMIM</td>
</tr>
<tr>
<td>Familial polyposis (APC)</td>
<td>175100</td>
<td>Myotonic dystrophy</td>
<td>160900</td>
</tr>
<tr>
<td>Chromosome 7</td>
<td>OMIM</td>
<td>Chromosome 21</td>
<td>OMIM</td>
</tr>
<tr>
<td>Cystic fibrosis</td>
<td>219700</td>
<td>Amyotrophic lateral sclerosis</td>
<td>105400</td>
</tr>
<tr>
<td>Chromosome 11</td>
<td>OMIM</td>
<td>X chromosome</td>
<td>OMIM</td>
</tr>
<tr>
<td>Wilms tumor</td>
<td>194070</td>
<td>Duchenne muscular dystrophy</td>
<td>310200</td>
</tr>
<tr>
<td>Ataxia-telangiectasia</td>
<td>208900</td>
<td>Fragile-X syndrome</td>
<td>309550</td>
</tr>
<tr>
<td>Chromosome 13</td>
<td>OMIM</td>
<td></td>
<td>OMIM</td>
</tr>
<tr>
<td>Retinoblastoma</td>
<td>180200</td>
<td>Adrenoleukodystrophy</td>
<td>300100</td>
</tr>
<tr>
<td>Chromosome 16</td>
<td>OMIM</td>
<td></td>
<td>OMIM</td>
</tr>
<tr>
<td>Polycystic kidney disease</td>
<td>173900</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mapping done by positional cloning technique. These types of chromosomal maps provided the framework for locating genes on each chromosome by the HGP. The red type represents real genes, the blue are genetic markers.
15.2 Genome Projects are an Outgrowth of Recombinant DNA technology

- Instead of finding and mapping disease genes one by one, the Human Genome Project (HGP) planned to sequence all the DNA in the human genome and to use this information to identify all the genes in the genome.

- Included in the HGP plan- the genomes of model organisms will also be sequenced.
Organisms Included in the Human Genome Project

**Table 15.2** Model Organisms Included in the Human Genome Project

<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome Size Million Base Pairs (Mb)</th>
<th>Estimated Number of Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em> (bacterium)</td>
<td>4.6</td>
<td>4,300</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em> (yeast)</td>
<td>12</td>
<td>6,000</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em> (roundworm)</td>
<td>97</td>
<td>20,000</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em> (plant)</td>
<td>120</td>
<td>25,000</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em> (fruit fly)</td>
<td>165</td>
<td>13,600</td>
</tr>
<tr>
<td><em>Mus musculus</em> (mouse)</td>
<td>3,000</td>
<td>30,000</td>
</tr>
<tr>
<td><em>Homo sapiens</em> (human)</td>
<td>3,200</td>
<td>20,000–25,000</td>
</tr>
<tr>
<td>Year</td>
<td>Event</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>Human Genome Project (HGP) begins on October 1</td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>First genetic map of genome</td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>Revised goals call for sequencing genome by 2005</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>High-resolution genetic map</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>First physical map of genome</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>16,000 human genes catalogued</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>National Human Genome Research Institute (NHGRI) created</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>Celera Corporation announces plans to sequence the human genome</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>Full-scale sequencing begins in HGP</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>HGP and Celera jointly announce draft sequence of genome</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>Working draft of genome published</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>Mouse genome sequenced</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>Sequence of gene-coding portion of human genome finished</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>Rat and chicken genomes sequenced</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>Chimpanzee genome sequenced</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>Rhesus monkey genome sequenced</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>Orangutan genome sequenced</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>Neanderthal genome sequenced</td>
<td></td>
</tr>
</tbody>
</table>
15.3 Genome Projects Have Created New Scientific Fields

- The size of the human genome required development of new technologies

- **Genomics** - The study of the organization, function, and evolution of genomes

- **Bioinformatics** - The use of computers and software to acquire, store, analyze, and visualize genomic information and share that info on the WWW.

- **Annotation**
  - Analysis of genomic nucleotide sequence data to identify protein-coding genes, non-protein-coding genes, their regulatory sequences, and functions
Gene Sequencing Computers
Goals of Genomics

- Create genetic and physical maps of genomics
- Catalog all non-gene sequences and determine their copy number
- Find the chromosomal location of all genes in a genome and annotate each gene
- Compile lists of expressed genes and non-expressed sequences
- Elucidate gene function and gene regulation
- Identify all proteins encoded by a genome and their functions
- Compare genes and proteins between species
- Characterize DNA variations between species
- Implement and manage Web-based databases
Some Sub-Fields of Genomics

- **Comparative genomics**
  - Compares genomes of different species for clues to the evolutionary history of genes or a species

- **Structural genomics**
  - Derives three-dimensional structures for proteins

- **Pharmacogenomics**
  - Analyzes genes and proteins to identify targets for therapeutic drugs
15.5 What Have We Learned So Far About the Human Genome? (1)

- *We will not discuss section 15.4*

- Only about 5% of our 3 billion nucleotides of DNA encode proteins

- Genes are distributed unequally on chromosomes
  - On each chromosome a cluster of genes is separated from the next cluster by stretches of DNA that do not encode genes.

- Humans have 20,000 to 30,000 genes
  - Far fewer than the predicted 80,000 to 100,000
What Have We Learned So Far about the Human Genome? (2)

- There are more proteins in the body than genes
  - mRNAs are processed in many ways so 20,000 to 30,000 genes can produce 300,000 proteins

- Genomes of humans and other higher organisms are similar
  - We share half our genes with the fruit fly and more than 90% with mice
What Have We Learned So Far about the Human Genome? (3)

- New types of mutation have been discovered that are related to specific diseases
  - Different mutations in a single gene can give rise to different genetic disorders
  - Some mutations in DNA repair genes can destabilize the genome

- Nucleotide variation is common
  - The human genome contains several million locations where single nucleotide differences occur in humans (SNPs)
Functional Assignments for Human Genes

- Cell adhesion (577, 1.9%)
- Chaperone (159, 0.5%)
- Cytoskeletal structural protein (876, 2.8%)
- Extracellular matrix (437, 1.4%)
- Immunoglobulin (264, 0.9%)
- Ion channel (406, 1.3%)
- Motor (376, 1.2%)
- Structural protein of muscle (296, 1.0%)
- Proto-oncogene (902, 2.9%)
- Select calcium binding protein (34, 0.1%)
- Intracellular transporter (350, 1.1%)
- Transporter (533, 1.7%)
- Molecular function unknown (12809, 41.7%)

- Nucleic acid enzyme (2308, 7.5%)
- Signaling molecule (376, 1.2%)
- Receptor (1543, 5.0%)
- Kinase (868, 2.8%)
- Select regulatory molecule (988, 3.2%)
- Transferase (610, 2.0%)
- Synthase and synthetase (313, 1.0%)
- Oxidoreductase (656, 2.1%)
- Lyase (117, 0.4%)
- Ligase (56, 0.2%)
- Isomerase (163, 0.5%)
- Hydrolase (1227, 4.0%)
### Table 15.4  Comparison of Selected Genomes

<table>
<thead>
<tr>
<th>Organism</th>
<th>Approximate Size of Genome (Date Completed)</th>
<th>Number of Genes</th>
<th>Approximate Percentage of Genes Shared with Humans</th>
<th>Web Access to Genome Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humans (Homo sapiens)</td>
<td>3,200 million bp (February 2003)</td>
<td>20,000–25,000</td>
<td>100%</td>
<td><a href="http://www.ornl.gov/hgmis/">http://www.ornl.gov/hgmis/</a></td>
</tr>
<tr>
<td>Yeast (Saccharomyces cerevisiae)</td>
<td>12 million bp (1996)</td>
<td>~6,000</td>
<td>31%</td>
<td><a href="http://www.yeastgenomes.org">http://www.yeastgenomes.org</a></td>
</tr>
</tbody>
</table>
To better understand a genetic disorder several questions must be answered:

- Where is the gene located?
- What is the normal function of the protein encoded by this gene?
- How does the mutant gene or protein produce the disease phenotype?
15.7 Proteomics is an Extension of Genomics

- Proteomics is the study of the structure and function of proteins, which is important in development of new diagnostic tests and drugs

- **Proteomics**
  - Study of expressed proteins in a cell at a specific time under a particular set of circumstances
Role of Proteomics

- Understanding gene function and its changing role in development and aging
- Identifying proteins that are biomarkers for diseases; used to develop diagnostic tests
- Finding proteins for development of drugs to treat diseases and genetic disorders
Proteins Expressed in a Cell

- Proteins can be separated by size and electric charge and displayed on a gel
- Size separation is down the X-axis
- Charge separation is from left to right

Fig. 15-14, p. 346
15.8 Ethical Concerns about Human Genomics

- To deal with the impact of genomic information on society, the HGP set up the ELSI (Ethical, Legal, and Social Implications) program to ensure that genetic information would be safeguarded, not used in discriminatory ways.

- Privacy and confidentiality of genetic information

- Fairness in the use of genetic information by insurers, employers and others

- Discrimination caused by someone’s genetic status

- Use of genetic information in reproductive decisions
ELSI Focuses on several Issues

- Privacy and confidentiality of genetic information
- Fairness in the use of genetic information by insurers, employers and others
- Discrimination caused by someone’s genetic status
- Use of genetic information in reproductive decisions