Bio-Computing

Young-Rae Cho, Ph.D.
Associate Professor
Division of Software / Division of Digital Healthcare
Yonsei University – Mirae Campus

Introduction to Bio-Computing

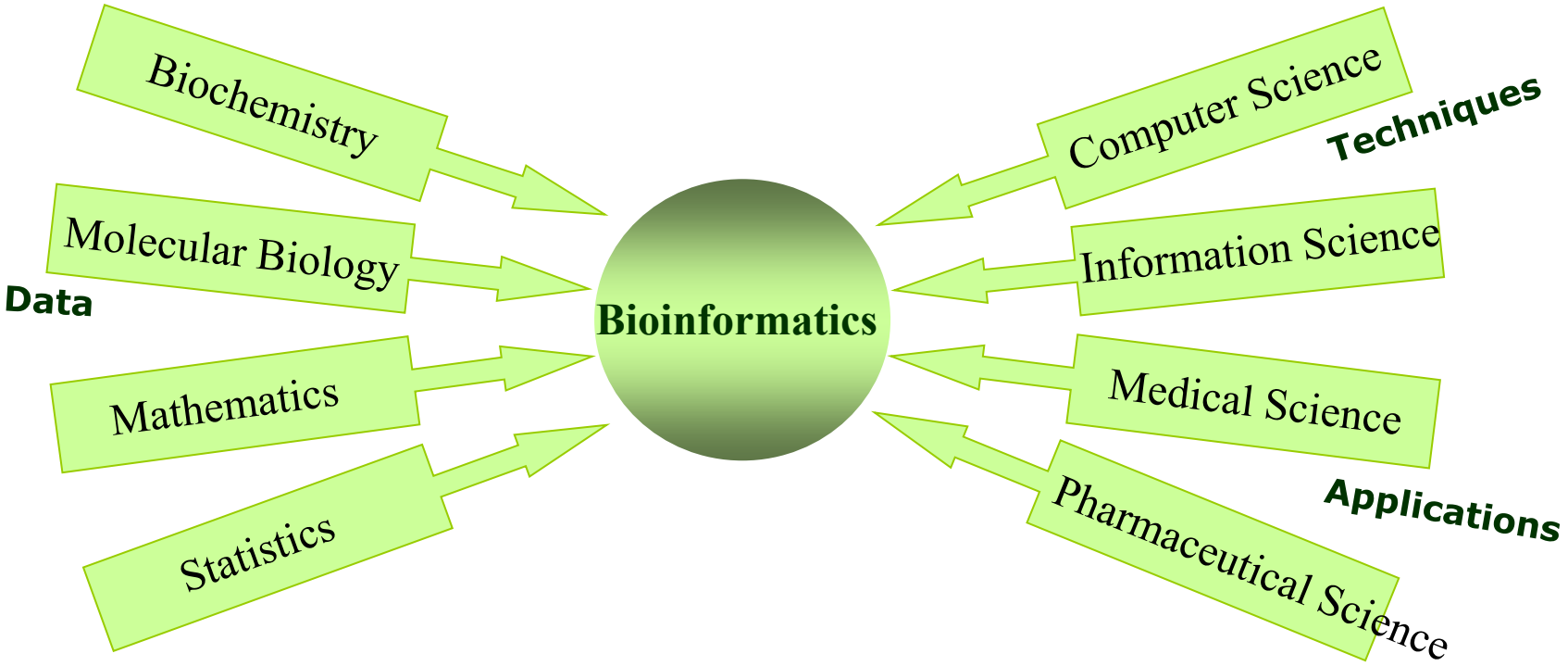
- Related Courses in US
  - Computational Biology (Undergraduate Junior or Senior Level)
  - Bioinformatics (Graduate Level)

- Courses at Yonsei University – Mirae Campus
  - SWE4016/DHC3003 “Bio-Computing” (Undergraduate Junior or Senior Level)
  - ITD6043 “Bio-Data Mining” (Graduate Level)

Bioinformatics ?

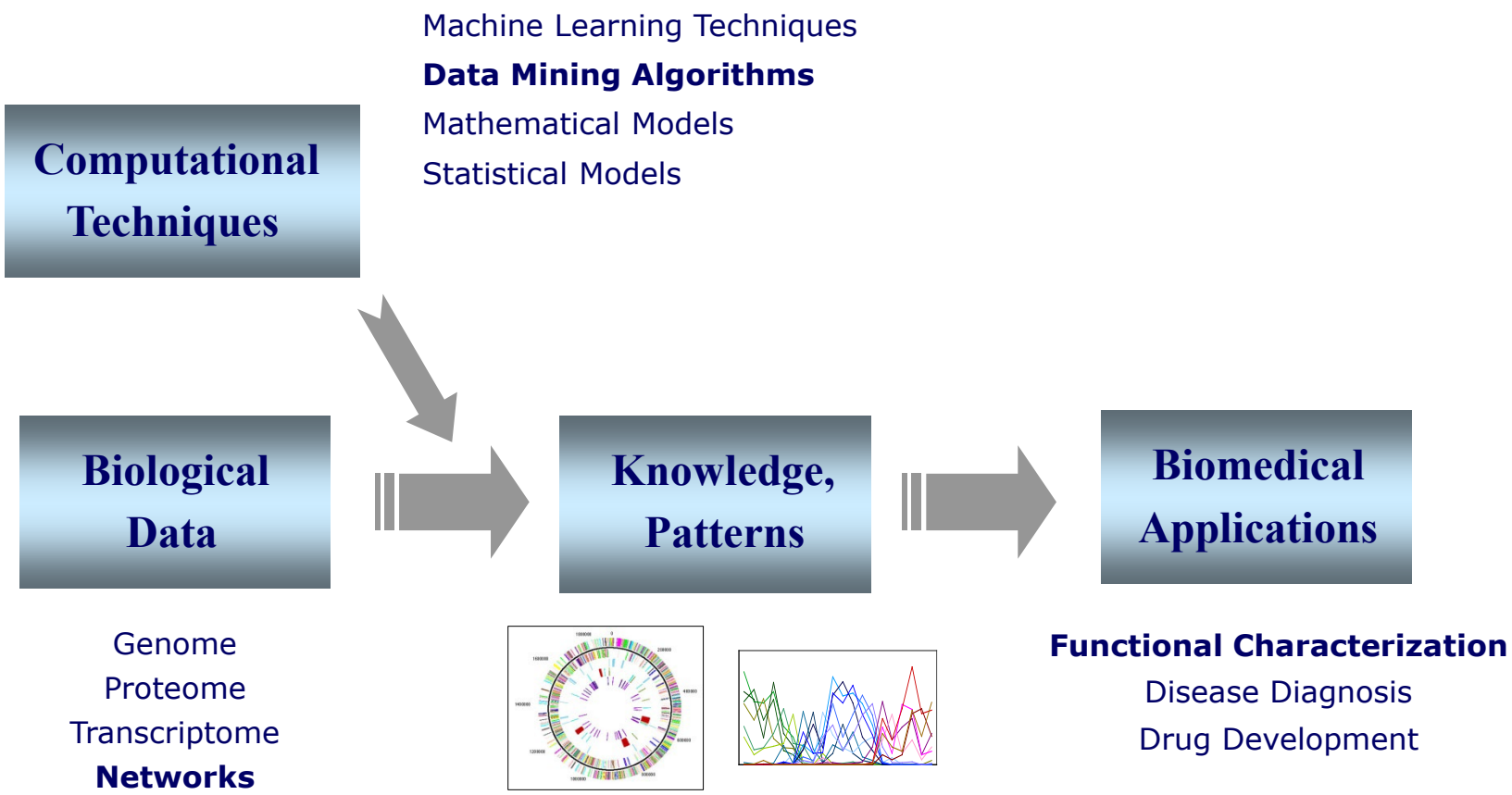
- Interdisciplinary Research Area

  to *manage* and *analyze* biological data using computational techniques



Bioinformatics ?

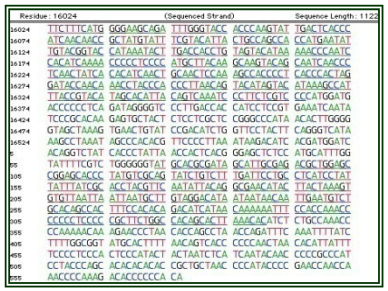
- Bioinformatics Research Process



Bioinformatics Milestone

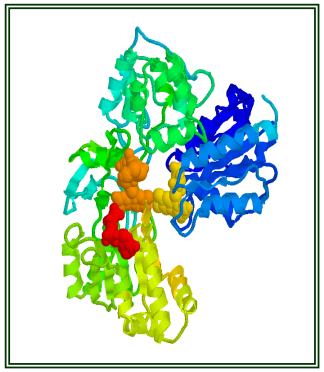
Computational Biology

Stage 1.
Sequence Analysis



- DNA sequencing
- Homolog search
- Motif finding

Stage 2.
Structure Analysis

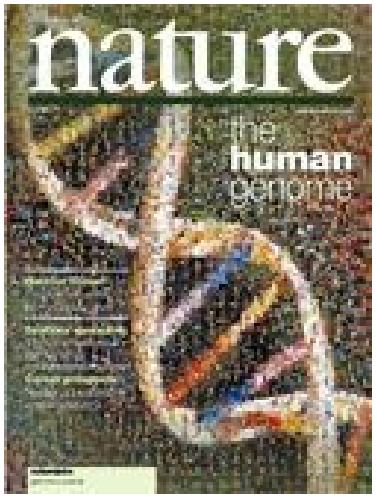


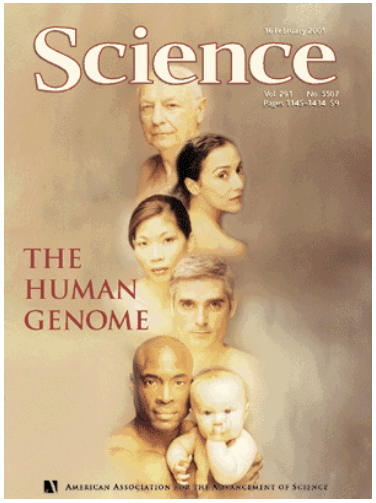
- Protein folding
- Homolog search
- Binding site prediction

Human Genome Project

- **Goal**
  - Identification of complete human genome
  - Mapping the gene from a functional standpoint

- **History**
  - Public project performed by Human Genome Project Consortium
  - Private project performed by Celera Genomics, Craig Venter
  - Both published final reports (called initial drafts) in 2001





The Number of Human Genes?

- **Human Genome Project Consortium**

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

*A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

30,000 ~ 40,000 genes

- **Celera Genomics**

THE HUMAN GENOME

The Sequence of the Human Genome

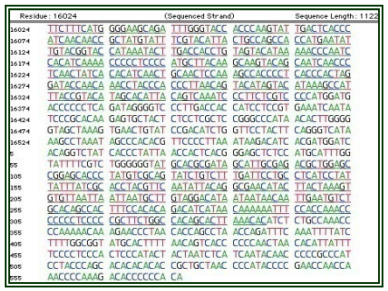
26,588 + 12,000 = 38,588 genes

- **Conclusion**
  - Nobody can confirm how many genes human has!

Bioinformatics Milestone

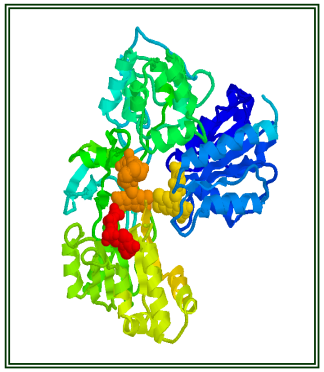
Computational Biology → Functional Genomics → Systems Biology

Stage 1. Sequence Analysis



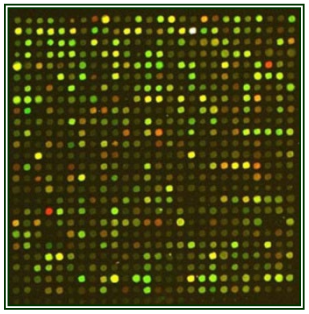
- DNA sequencing
- Homolog search
- Motif finding

Stage 2. Structure Analysis



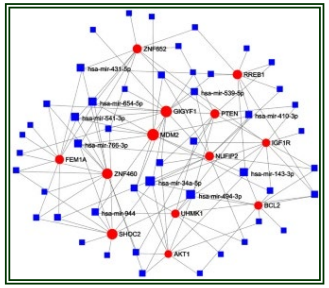
- Protein folding
- Homolog search
- Binding site prediction

Stage 3. Genome Analysis



- Function prediction
- Gene clustering
- Gene classification

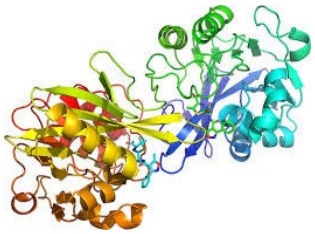
Stage 4. Network Analysis



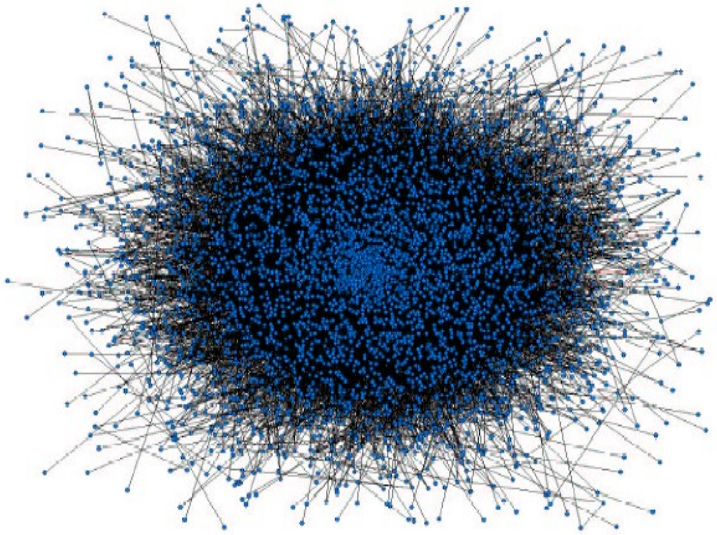
- Network modeling
- Module finding
- Pathway prediction

Computational Network Biology - Data

- **Interactome**
  - Genome-wide PPIs (Protein-Protein Interactions)
  - Problem?
    - Large scale & Unreliability
  - Demand?
    - Computational, integrative approaches

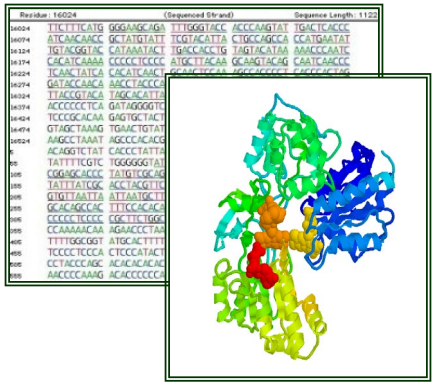


- **PPI Networks**
  - Undirected unweighed graph, G(V,E)
  - Problem?
    - Complex connectivity
  - Demand?
    - Computational, systematic approaches



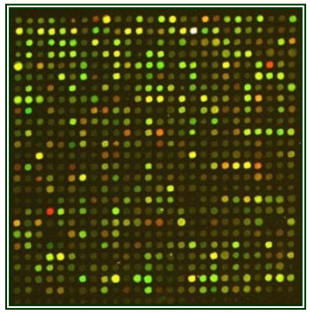
From Bioinformatics To Biomedical Informatics

Stage 1/2.
Sequence &
Structure Analysis



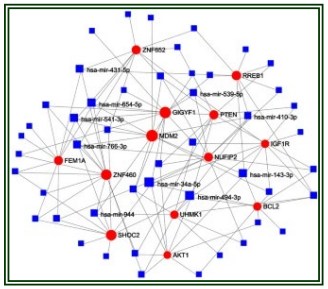
- DNA sequencing
- Homolog search
- Motif finding
- Protein folding
- Binding site prediction

Stage 3.
Genome Analysis



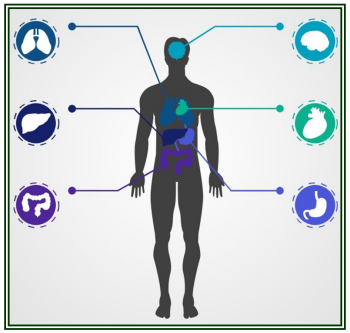
- Function prediction
- Gene clustering
- Gene classification

Stage 4.
Network Analysis



- Network modeling
- Module finding
- Pathway prediction

Stage 5.
Medical Data
Analysis



- Disease Diagnosis
- Drug Development

Personalized Medicine / Precision Medicine

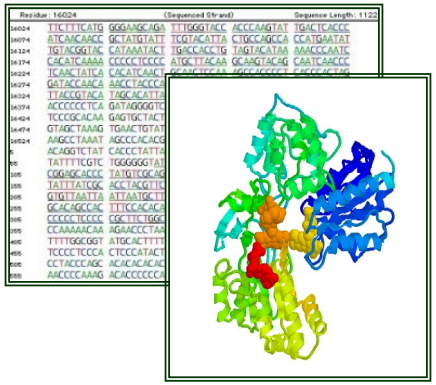
- **Generalized Concept**
  - Customized healthcare including medical treatment and prevention
  - Background: All individuals have genetic variation

- **Use of Big Data**
  - Genomic data and patients' medical record data

- **Support by Advanced Technology**
  - Machine learning, deep learning, and artificial intelligence (AI) techniques

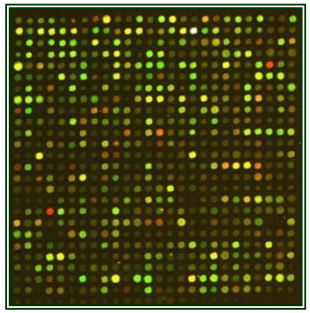
Scope of Bio-Computing

Stage 1/2. Sequence & Structure Analysis



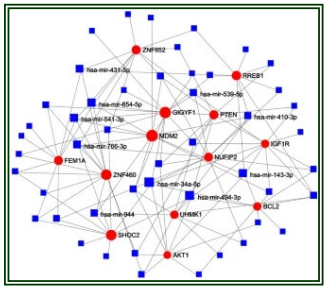
- DNA sequencing
- Homolog search
- Motif finding
- Protein folding
- Binding site prediction

Stage 3. Genome Analysis



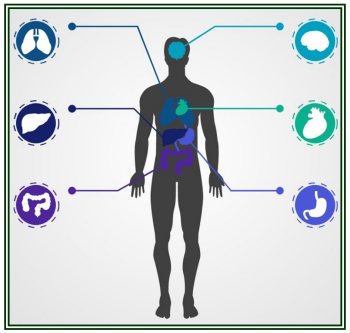
- Function prediction
- Gene clustering
- Gene classification

Stage 4. Network Analysis



- Network modeling
- Module finding
- Pathway prediction

Stage 5. Medical Data Analysis



- Disease Diagnosis
- Drug Development

Subjects of Bio-Computing

- **Sequence Analysis**
  - Pairwise Sequence Alignment
  - Multiple Sequence Alignment
  - Pattern Matching and Finding
  - Phylogenetic Analysis

- **Genome Analysis**
  - Gene Clustering
  - Gene Classification

- **Network Analysis**
  - Network Modelling
  - Module Finding