# Biology 644

**Old Title:** Bioinformatics for Molecular Biologists

**Potential New Title:** Integrated Bioinformatics Using R for Both Wet and Dry Scientists

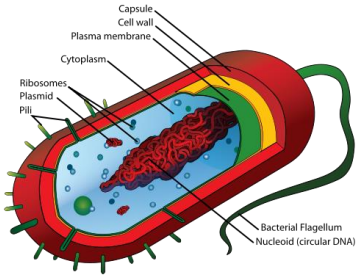


## Outline

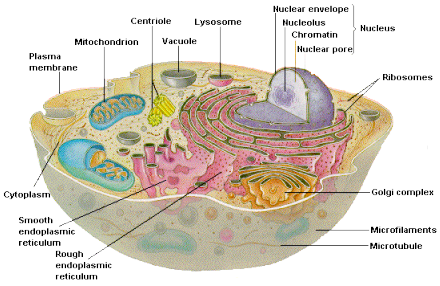
- Temporary Seating Chart
- Review Syllabus
- Brief Molecular Biology Review Lecture
- R Lab
- - Install Packages
- - Chapter 1: Sections 1.1 – 1.5
- - Chapter 1 Supplemental



## Bacterial Cell (Prokaryotic)

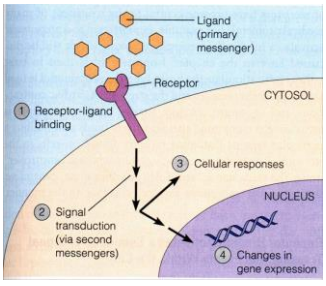


## Eukaryotic Cell



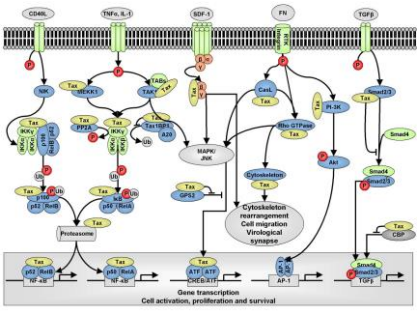



## Simplest Cell Signaling Diagram You Will Ever See

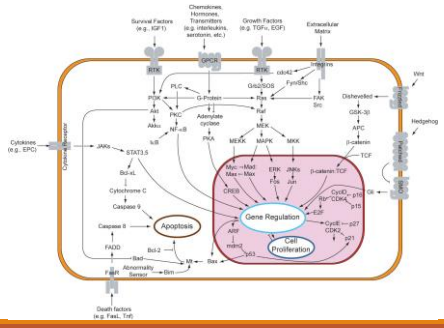





## Slightly More Complicated Cell Signaling Diagram

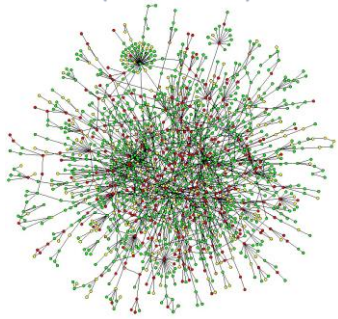


## Even More Complicated Cell Signaling Diagram



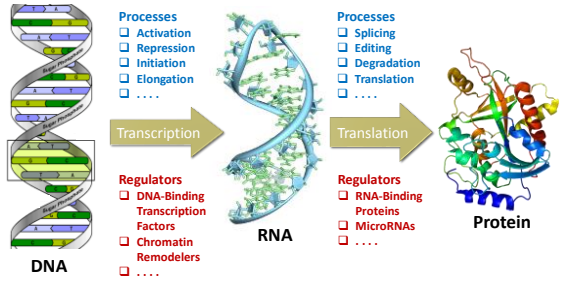



## Much More Complicated Cell Signaling Diagram (a.k.a. Hairball)



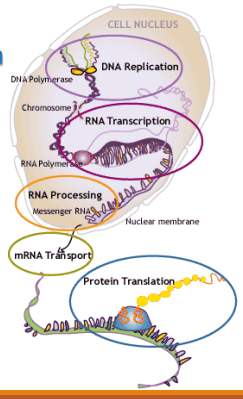


## The Central Dogma of Molecular Biology



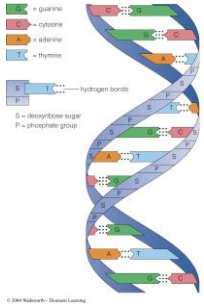
## DNA -> RNA -> Protein

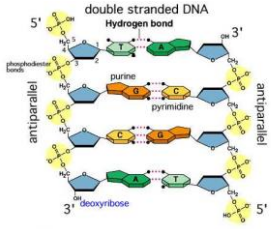
## In the Cell





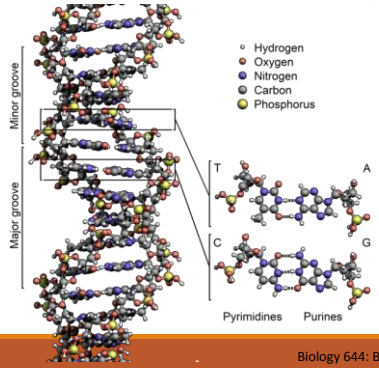

## DNA Sequence



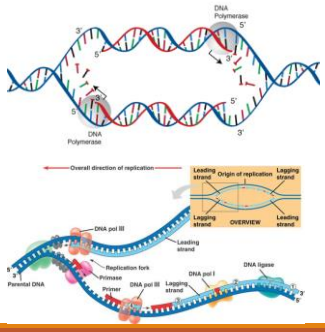






## DNA Structure

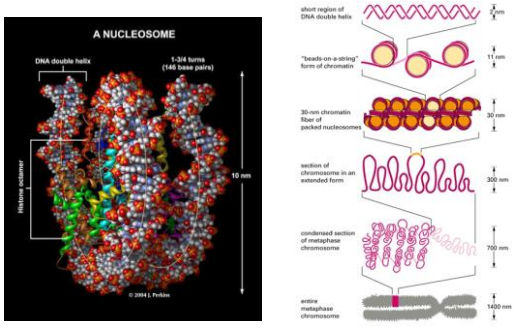


## DNA Replication



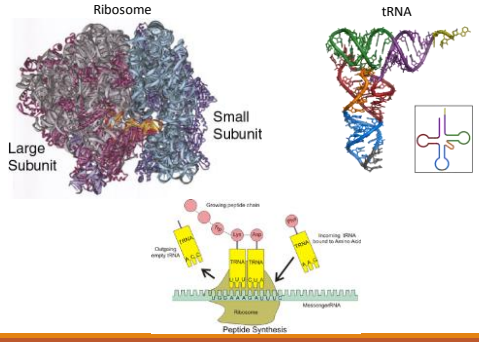



## DNA Wrapping and Chromosomal Packing

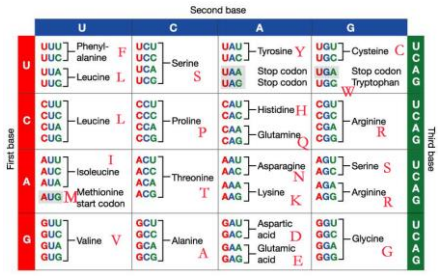





## mRNA Translation with Ribosomes and tRNAs



## Codon Alphabet






## Bioinformatics: the analysis, organization, integration, and annotation of biological data

- **Sequence Bioinformatics**: patterns about the sequence can reveal insight into transcription, translation, and function of synthesized proteins.
  - **Genome analysis**: Genome assembly, genome annotation, gene finding, alternative splicing, EST analysis and comparative genomics.
  - **Sequence analysis**: Multiple sequence alignment, sequence search and clustering, function prediction, motif discovery, functional site recognition in protein, RNA and DNA sequences.
  - **Phylogenetics**: Phylogeny estimation, models of evolution, comparative biological methods, population genetics.
  - **Analysis of high-throughput biological data**: Microarrays (nucleic acid, protein, array CGH, genome tiling, and other arrays), EST, SAGE, MPSS, proteomics, mass spectrometry.
  - **Genetics and population analysis**: Linkage analysis, association analysis, population simulation, haplotyping, marker discovery, genotype calling.
  - **Systems biology**: Systems approaches to molecular biology, multiscale modeling, pathways,gene networks.
  - **Computational Proteomics**: Filtering and indexing sequence databases, Peptide quantification and identification, Genome annotations via mass spectrometry, Identification of post-translational modifications, Protein-protein interactions, Computational approaches to analysis of large scale Mass spectrometry data, Exploration and visualization of proteomic data, Data models and integration for proteomics and genomics, Querying and retrieval of proteomics and genomics data etc.
- **Structural Bioinformatics** – 3D structure reveals further information about DNA, protein- DNA binding, RNA, protein-RNA binding, Protein-Protein interactions, etc.
  - Structural genomics via **mass spectrometry**



# R

- A free software programming language and software environment for statistical computing and graphics
- Implementation of the S programming language
- Also contains lexical scoping semantics inspired by Scheme
- A GNU project that is freely available under the GNU General Public License
- Written primarily in C, Fortran, and R
- Pre-compiled binary versions are provided for many popular operating systems
- Contains thousands of packages to analyze many different types of biological data (and data from statistics, geology, finance…)


- A free, open source and open development software project for the analysis and comprehension of genomic data generated by wet lab experiments in molecular biology
- Provides widespread access to a broad range of powerful statistical and graphical methods for the analysis of genomic data.
- Based primarily on the statistical R programming language, but also contains contributions in other programming languages
- A large number of genome annotation packages
- Most Bioconductor components are distributed as R packages
- Used for the analysis of:
  - Single channel Affymetrix
  - Two or more channel cDNA/Oligo microarrays
  - SAGE, sequence, and SNP data
  - Much more….