

系統生物學與高資料量篩檢 技術簡介

Chen, Shu-Hwa
Institute of Information Science
Academia Sinica
2008.08.13

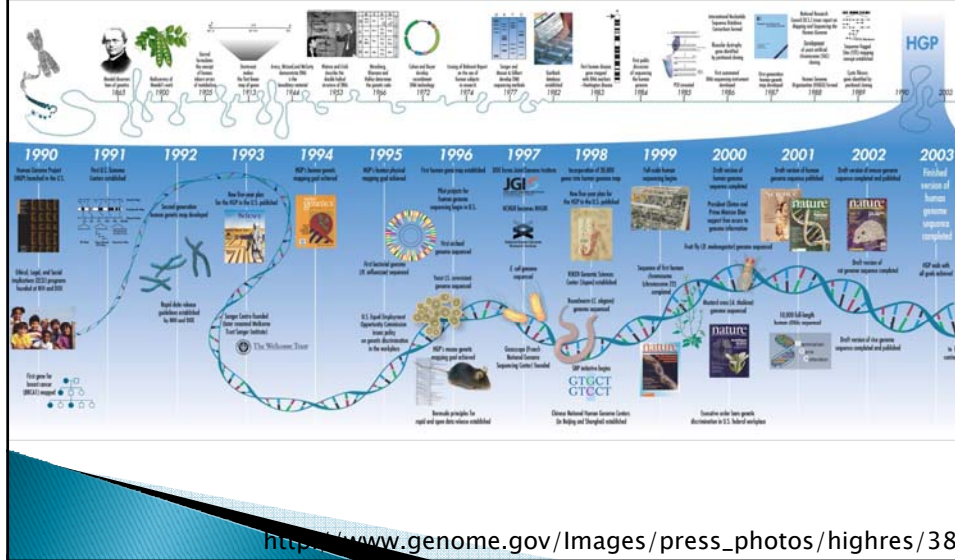
Bioinformatics

- ▶ Bioinformatics is the application of information technology to the **management and analysis** of biological data.
- ▶ Bioinformatics is an **interdisciplinary research** area that is the interface between the **biological and computational sciences**.
- ▶ **Bioinformatics** is the field of science in which biology, computer science, and information technology merge to form a single discipline.

Bioinformatics
Biology + Informatics + Statistics



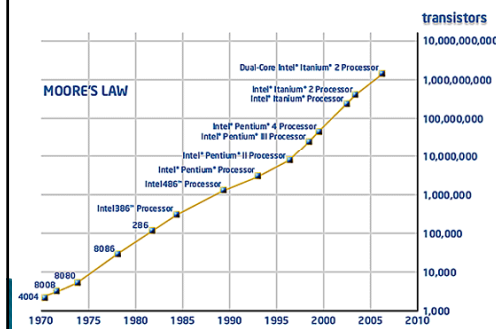
Milestones in Genetics



http://www.genome.gov/Images/press_photos/highres/38

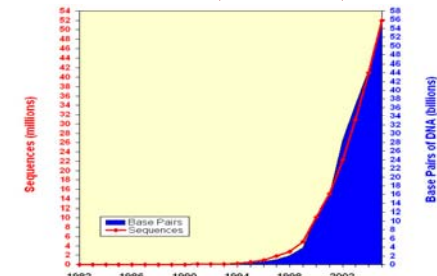
Rising of Bioinformatics

- ▶ Internet and WWW
- ▶ Computing Power
- ▶ Genome Projects
- ▶ High-Throughput Technology



<http://www.intel.com/technology/mooreslaw/index.htm>

Growth of GenBank(1982-2005)



<http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>

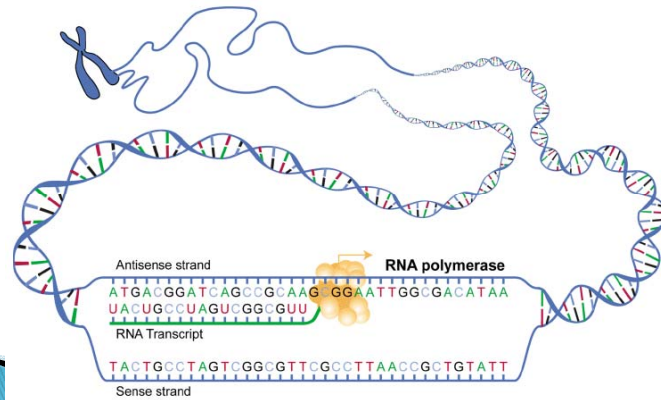
Biological information is coded

Binary Code

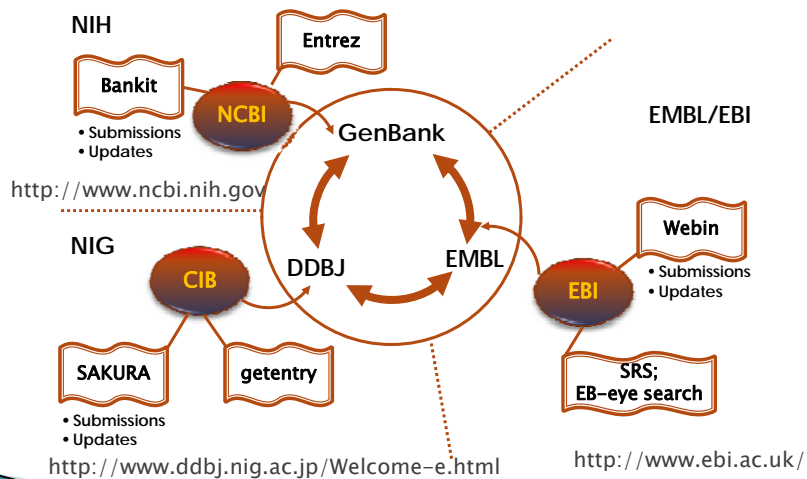
010001010100011101010
 100010101000111010101
 00010

Genetic Code

ATTCCATCGGAGTAATTCCATC
 GGAGTAATTCCATCGGAGTAAT



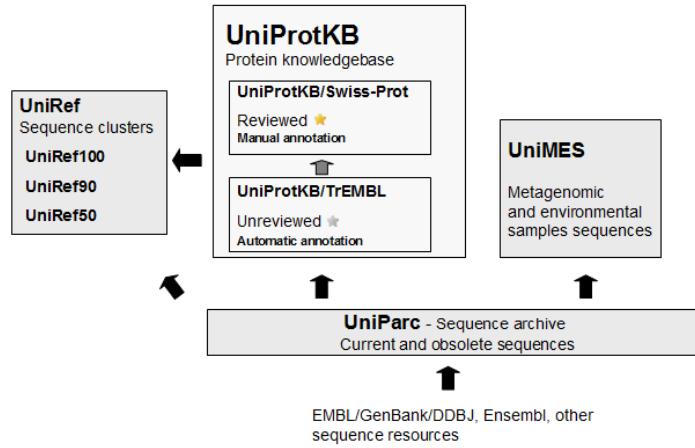
Data Exchanges among Three Major Nucleic Acid Public Databases



The dataflow for new submissions and updates between the three databases

Modified from Baxevanis AD & Ostell M. BFF. Bioinformatics. (2005) Wiley Inc. published. Figure 1.1 (p.5)

Organization of UniProt Databases

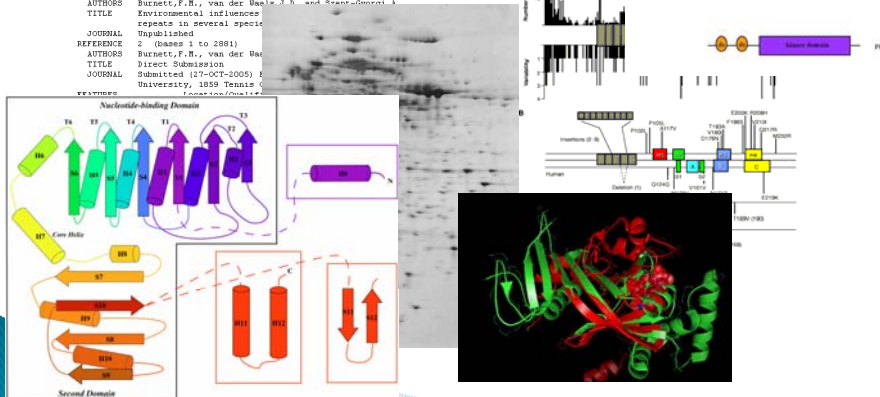


<http://beta.uniprot.org/help/about>

Various Formats for BioDB

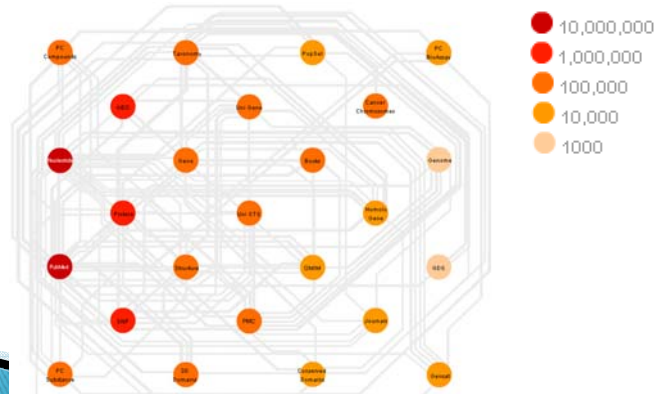
```

LOCUS eIF4E 3801 bp DNA linear INV 27-OCT-2005
DEFINITION Drosophila melanogaster eukaryotic initiation factor 4E [eIF4E]
gene, alternative splice products, complete cds.
ACCESSION
VERSION
KEYWORDS
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscophora;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 3801)
AUTHORS Burnett,F.H., van der Walde,J.H. and Scott-Gamer,J.
TITLE Environmental influences
repeats in several species
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 2881)
AUTHORS Burnett,F.H., van der Walde,J.H. and Scott-Gamer,J.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-2005)
UNIVERSITY, 1899 Tennis C
EXEMPTED
  
```



Databases and the Their Interconnections

- ▶ Most of DB records are linked to other records, both within a given database and between databases.
- ▶ Searches can begin anywhere within Entrez.



<http://www.ncbi.nlm.nih.gov/Database/datamodel/ind>

NAR Special Issues

<http://nar.oxfordjournals.org/>

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals • Life Sciences • Nucleic Acids Research

READ THIS JOURNAL

Nucleic Acids Research

[View Current Issue \(Volume 36 Issue 13 August 2008\)](#)
[Advance Access](#)
[Browse the Archive](#)

Nucleic Acids Research (NAR) is a fully Open Access journal, providing rapid publication of leading edge research into the nucleic acids under the following categories: chemistry, computational biology, genomics, molecular biology, nucleic acid enzymes, RNA and structural biology.

There is a Survey and Summary section, and methods papers are published in NAR Methods Online. Each year the first issue is devoted to biological databases, and a later issue to relevant web-based software resources.

SEARCH THIS JOURNAL [GO](#) [Advanced search](#)

Click here to view
PROTEIN FOLDING PAPERS FROM **peds**

RELATED JOURNALS

- ▶ [Bioinformatics](#)
- ▶ [Human Molecular Genetics](#)
- ▶ [Glycobiology](#)
- ▶ [Microbes](#)

READER SERVICES

- ▶ [Calendar of Events](#)
- ▶ [Specialties in the Nucleic Acids Area](#)
- ▶ [50 Most Frequently Cited Articles](#)
- ▶ [Library Recommendation Form](#)

THE JOURNAL

- ▶ [About this Journal](#)
- ▶ [NAR Methods online](#)
- ▶ [2008 Database Issue](#)
- ▶ [2008 Web Server Issue](#)
- ▶ [Referee Information](#)
- ▶ [Rights & Permissions](#)
- ▶ [Dispatch date of the next issue](#)
- ▶ [This journal is a member of the Committee on Publication Ethics \(COPE\)](#)

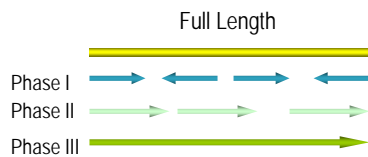
Impact factor: 6.954

Senior Executive Editors
R.J. Roberts, Ipswich, MA, USA
K.R. Fox, Southampton, UK
▶ [View full editorial board](#)

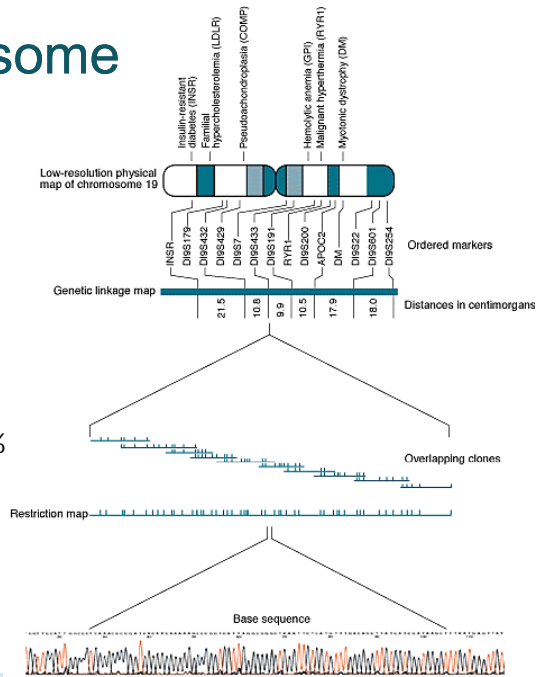
FOR AUTHORS

- ▶ [Instructions to authors](#)
- ▶ [Scope and Criteria for Consideration](#)
- ▶ [Online submission instructions](#)
- ▶ [Special Advertising Rates](#)

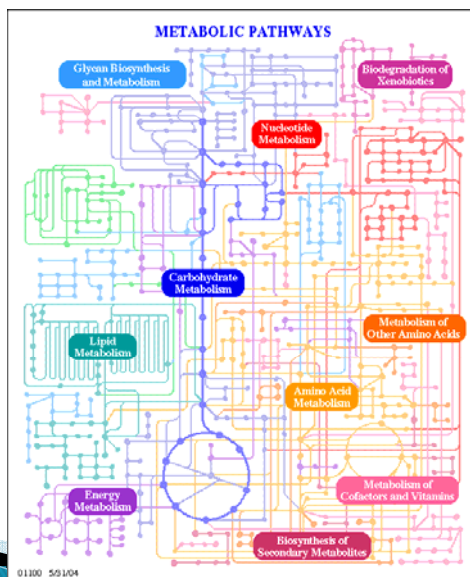
From Chromosome to Sequences



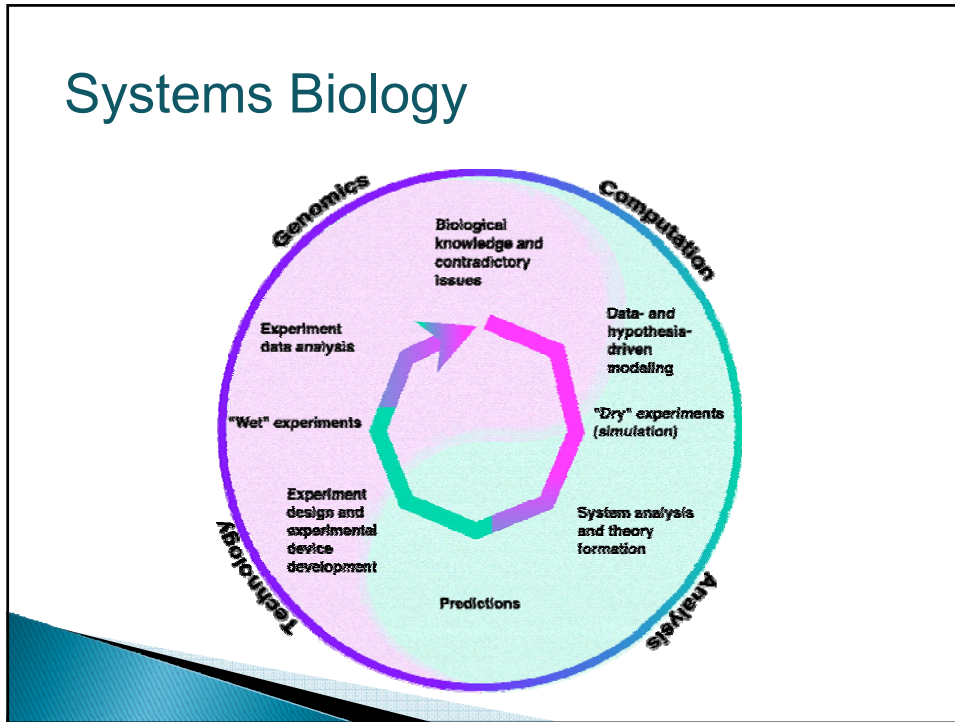
Coverage (8-10X) >> 99%
Error Rate << 0.01%



KEGG Kyoto Encyclopedia of Genes and Genomes



Systems Biology



An Example of a Systems Biology Approach

articles

Global analysis of protein localization in budding yeast

Won-Ki Huh^{1*}, James V. Falvo^{1*}, Luke C. Gerke¹, Adam S. Carroll¹, Russell W. Howson¹, Jonathan S. Weissman^{1,2} & Erin K. O'Shea¹

¹Howard Hughes Medical Institute, University of California–San Francisco, Department of Biochemistry and Biophysics, and ²Department of Cellular and Molecular Pharmacology, 600 16th Street, San Francisco, California 94143-2240, USA

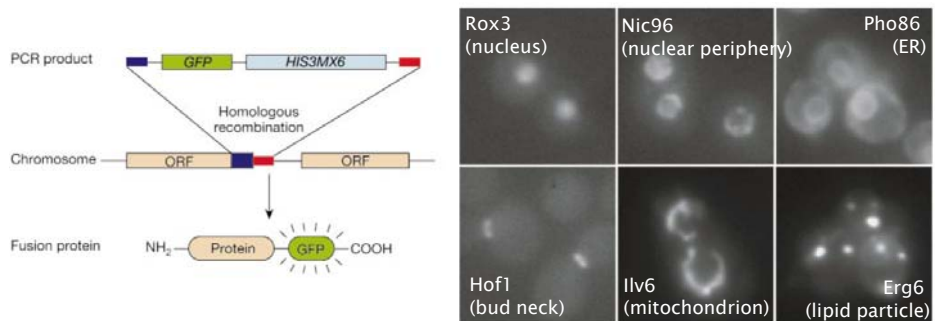
*These authors contributed equally to this work.

A fundamental goal of cell biology is to define the functions of proteins in the context of compartments that organize them in the cellular environment. Here we describe the construction and analysis of a collection of yeast strains expressing full-length, chromosomally tagged green fluorescent protein fusion proteins. We classify these proteins, representing 75% of the yeast proteome, into 22 distinct subcellular localization categories, and provide localization information for 70% of previously unlocalized proteins. Analysis of this high-resolution, high-coverage localization data set in the context of transcriptional, genetic, and protein–protein interaction data helps reveal the logic of transcriptional co-regulation, and provides a comprehensive view of interactions within and between organelles in eukaryotic cells.

NATURE (2003), 425:686

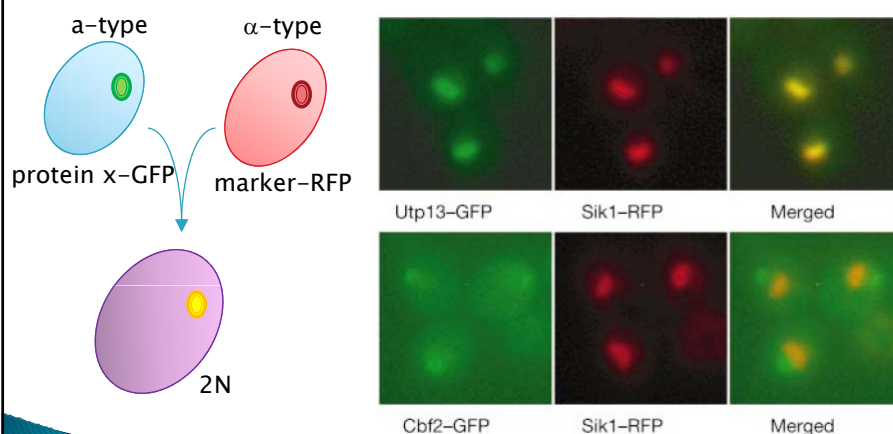
Microscopic analysis of yeast strains expressing GFP-tagged proteins

► Constructing recombinant strains

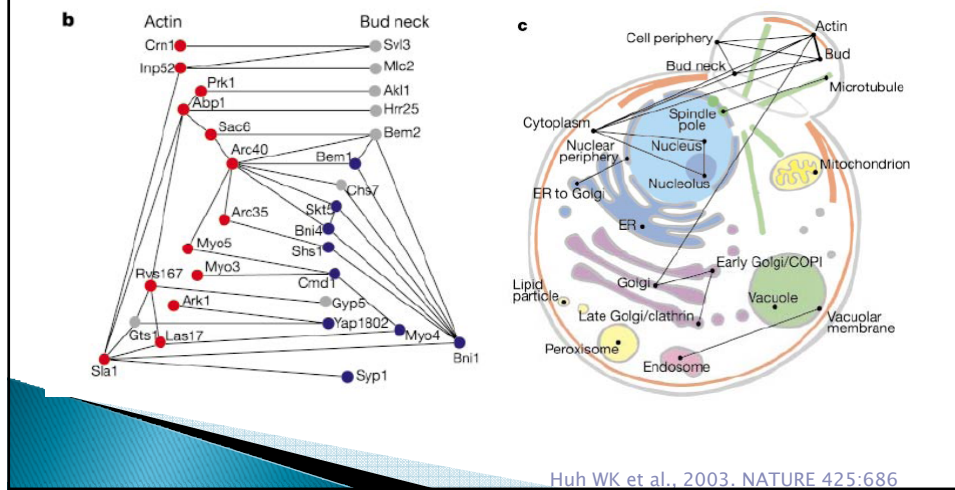


Huh WK et al., 2003. NATURE 425:686

Representative co-localization experiment



Relationship between genetic and physical interactions and subcellular localization



High-throughput *omics

Subject	Inclusive set (For an individual)	Statistical Study (many individuals)
Genes	Genome	Genomics
Transcripts	Transcriptome	Transcriptomics
Proteins	Proteome	Proteomics
Metabolites	Metabolome	Metabolomics
Phenotype	Phenome	Phenomics

High-Throughput Technology

- ▶ Using robotics, data processing and control software, liquid handling devices, and sensitive detectors to quickly conduct millions of biochemical, genetic or pharmacological tests.

AUTOMATION

Brute-Force Approach → Fast and Huge Amount of Data

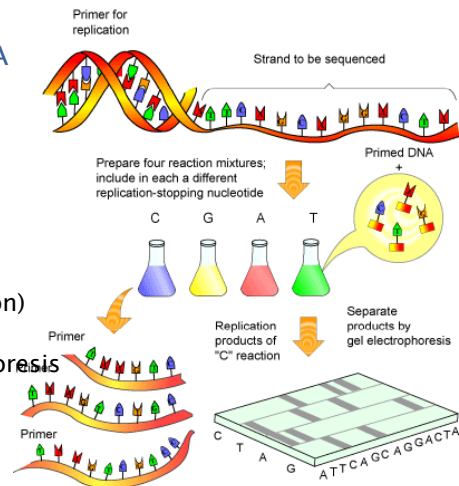
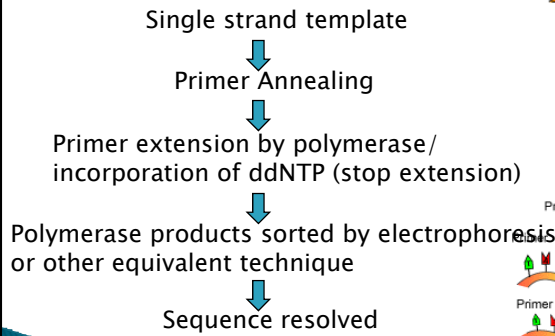
High-Throughput Technology

- ▶ Genomics
 - HTP Sequencing
 - Accessing varied version of alleles
- ▶ Transcriptomics
 - Microarray for gene expression data
- ▶ Proteomics
 - Isotope-Coded Affinity Tags
- ▶ Anatomical and Histological Images

DNA Sequencing and Resequencing

Sanger Method

Sequencing Method based on DNA polymerase and chain terminator



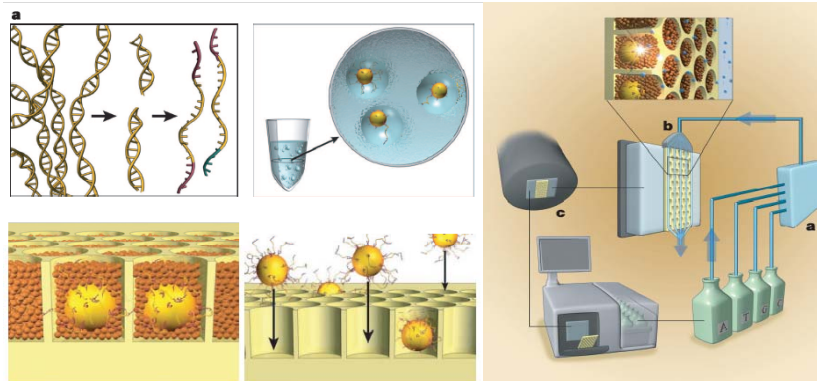
Animation of automated DNA sequencing using Sanger Method

<http://www.dnalc.org/dnalc/resources/shockwave/cycseq.html>

DNA Sequencing and Resequencing

454 Sequencing Technology:

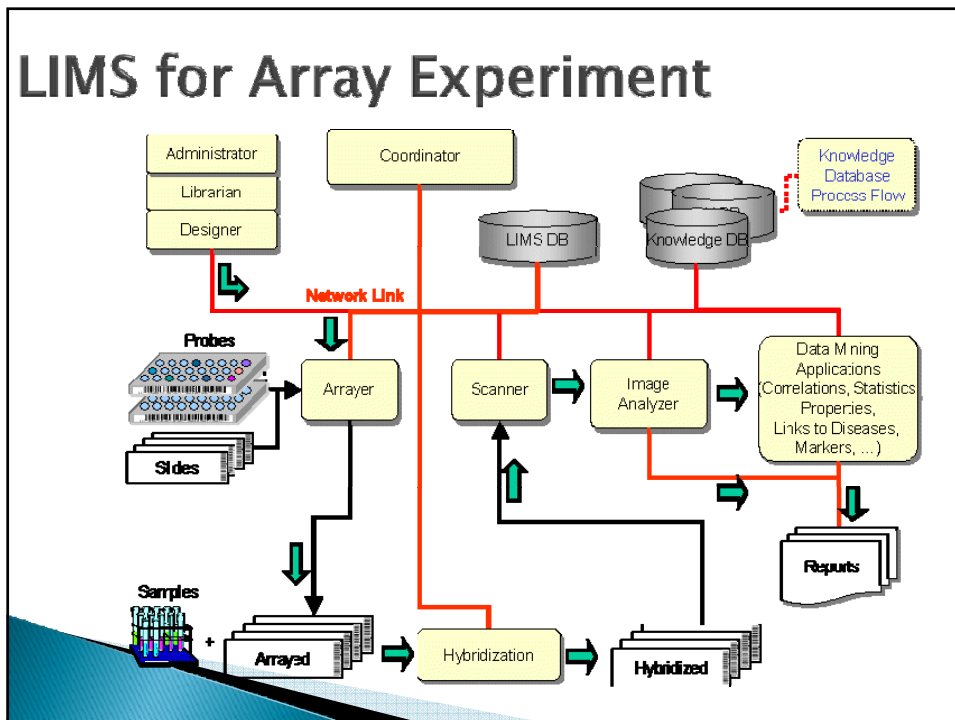
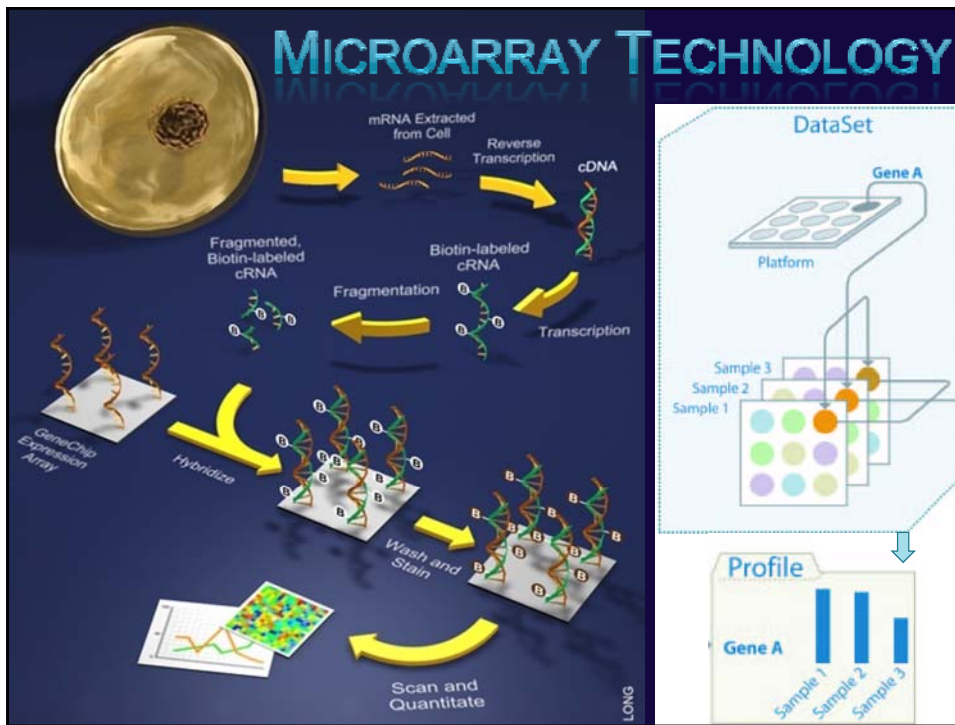
emulsion-based template/ picoliter reaction well/ pyrophosphate sequencing



Genome sequencing in microfabricated high-density picolitre reactors

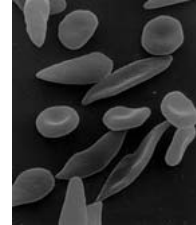
Nature. 2005 Sep 15; 437(7057):326-7

Demo <http://www.454.com/flash/roche-gene.asp>



SNP (Single Nucleotide Polymorphism)

	codon	5	6	7
Hb β (normal)	CCT	GAG	GAG	
	Pro	Glu	Glu	
Hb β (Sickle)	CCT	GTG	GAG	
	Pro	Val	Glu	

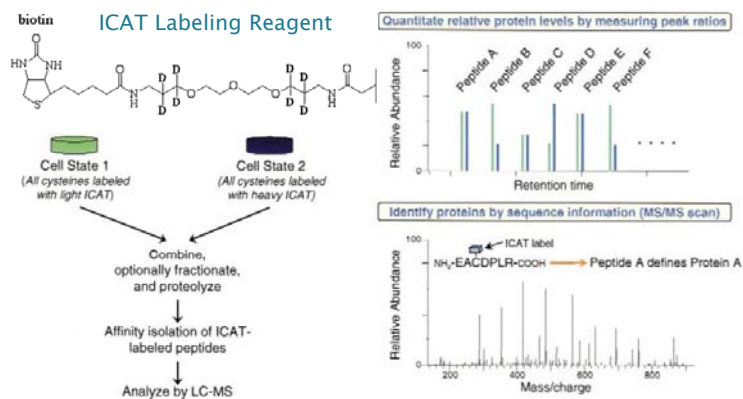


http://en.wikipedia.org/wiki/Red_blood_cells

- ▶ SNP Density in human genome (estimated):
 - One SNP site / 500~1000 bases
- ▶ How to detect SNPs?
 - Sequencing-based method
 - PCR-based method
 - Array-based method

Isotope-Coded Affinity Tags

<http://www.proteomecenter.org/PDFs/Gygi.NatBiotech.99.pdf>



Animation Demo of ICAT Technology

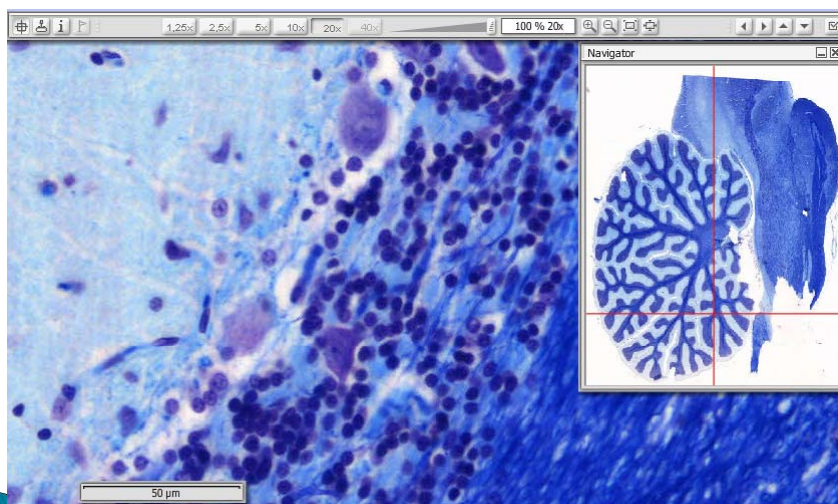
<http://www.bio.davidson.edu/Courses/genomics/ICAT/ICAT.html>

High-Throughput Microscopy



Adopted from a speech of Urban Liebel, Mitochek project group, EMBL Heidelberg
harvester.embl.de/media/2006-01-13-dresden.pdf

A whole-view on a slice of body



<http://dotslide.olympus-sis.com/>