Bioinformatics Chapter 6.

Genomic Mapping and Mapping Databases

지 성 욱
서울대 바이오정보기술연구센터
E-mail: swchi@cbit.snu.ac.kr

Sung-Wook Chi
Center for Bioinformation Technology (CBIT)
Seoul National University

장 병 탁
서울대 컴퓨터공학부 바이오지능연구실 &
바이오정보기술연구센터
E-mail: btzhang@cse.snu.ac.kr

Byoung-Tak Zhang
Center for Bioinformation Technology (CBIT) &
Biointelligence Laboratory
School of Computer Science and Engineering
Seoul National University

This material is available at http://bi.snu.ac.kr./ &
http://cbit.snu.ac.kr/

Outline

- Indroduction
- Genomic Mapping
- Types of Maps
- Data Repositories
  - GDB, NCBI, MGI/MGD
- Mapping Projects and Associated Resources
- Practical Uses of Mapping Resources

Introduction

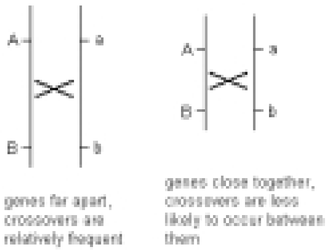
- “Map of Maps”
- The different types of markers and methods used for genomic mapping
- The inherent complexities in the construction and utilization of genome maps
- Several large community databases and method-specific mapping projects
- Practical examples of how these tools and resources can be used to aid in specific types of mapping studies

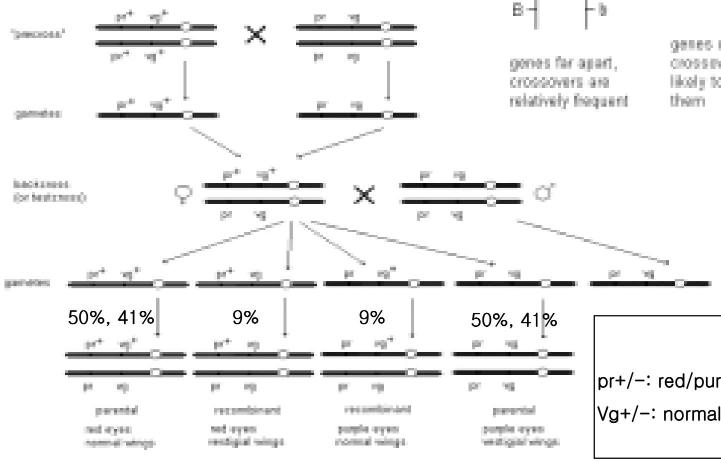
---

Genomic Mapping

- **Genetic Mapping**
  - Crossbreeding and pedigree
  - Calculation of recombination frequency by linkage analysis
- **Cytogenetic Mapping**
  - FISH(Fluorescent In Situ Hybridization)
- **Physical Mapping**
  - Molecular biology technique (hybridization, PCR)
  - Restriction Mapping
  - STS(Sequence Tagged Site) Mapping
    Radiation-hybrid method, Clone library based method.

Linkage Analysis

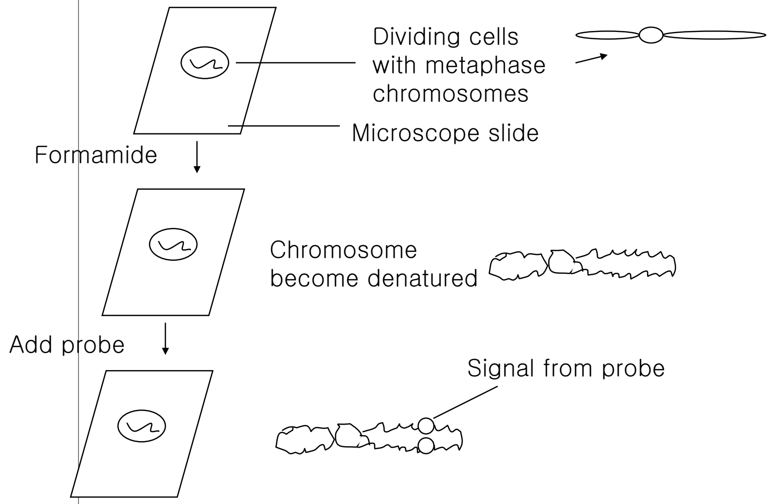




pr+/−: red/purple eyes

Vg+/−: normal/vestigial wings

FISH(Fluorescent In Situ Hybridization



Dividing cells with metaphase chromosomes

Microscope slide

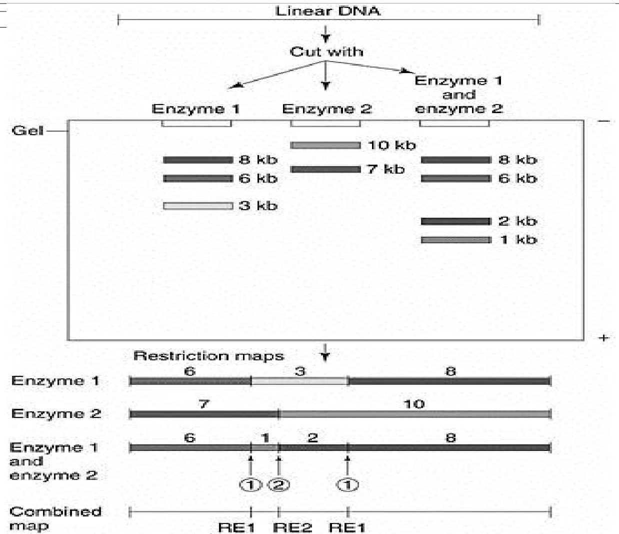
Formamide

Chromosome become denatured

Add probe

Signal from probe

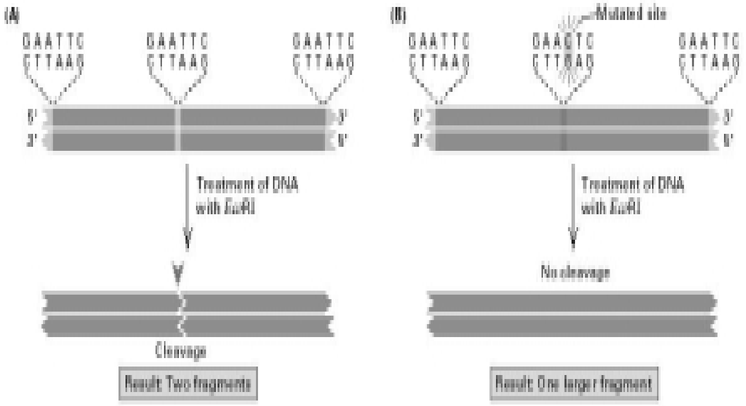
Restriction Mapping



Genomic Map Elements

- DNA Markers
  - A uniquely identifiable segment of DNA
  - Usually ranging in size from one to 300-400 nucleotide bases in size
  - Detection of markers may be either PCR based or hybridization based.
- Polymorphic Markers
  - Show sequence variation among individuals
  - To construct genetic linkage maps
  - RFLP (Restriction fragment length polymorphism)
- DNA Clones
  - Yeast artificial chromosome (YAC)
  - Bacterial artificial chromosome (BACs)
  - P1-artificial chromosome (PACs)

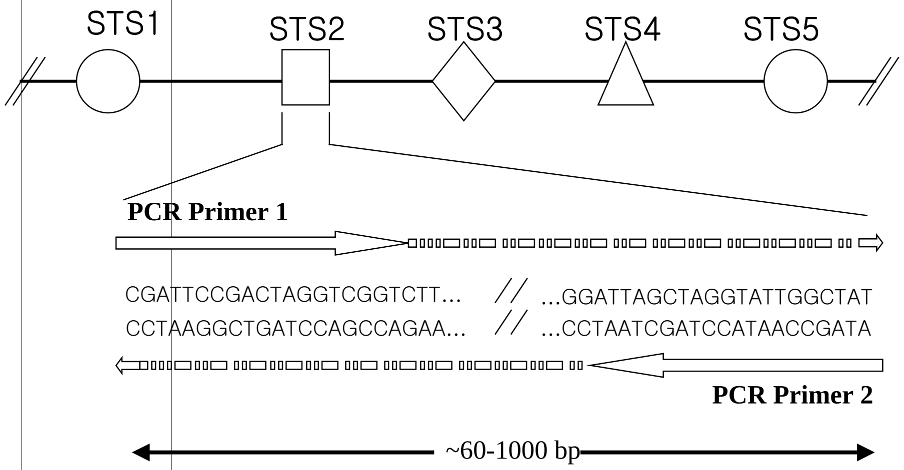
Basis for RFLP mapping



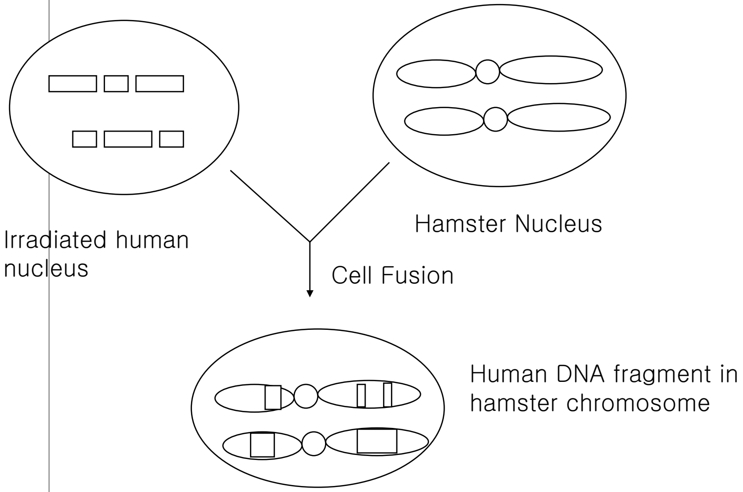
STS(Sequence Tagged Site)

- Simply a short sequence (100 ~ 500bp)
  - sequence must be known (PCR)
  - must have a unique location.
- Expressed sequence tags (ESTs)
  - comes from a unique gene
- SSLP (Simple sequence length polymorphism)
  - Minisatellites, VNTR(Variable number of tandem repeats)
  - Microsatelliyes, STRs(Simple tandem repeats)
- Random genomic sequences
  - obtained by sequence random pieced of cloned genomic DNA or from databases.

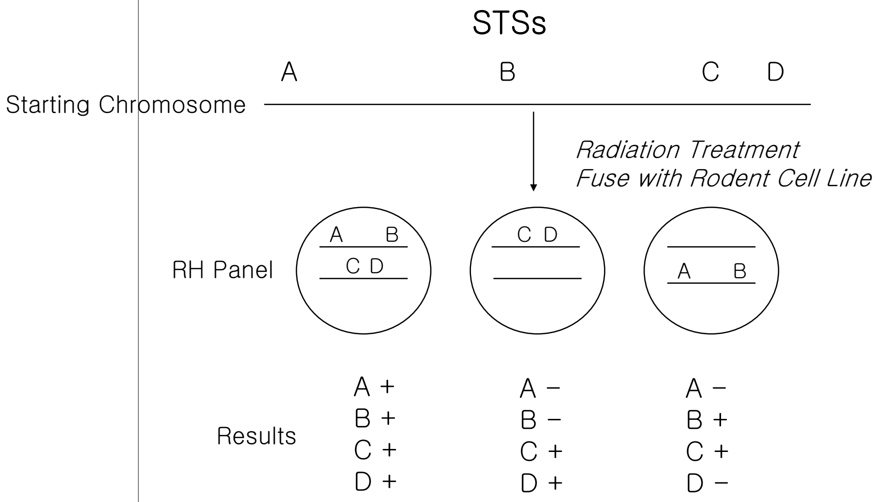
Sequence-Tagged Sites (STS) Mapping



Radiation Hybrid (RH) Method



Radiation Hybrid (RH) Mapping



STSs

Starting Chromosome: A B C D

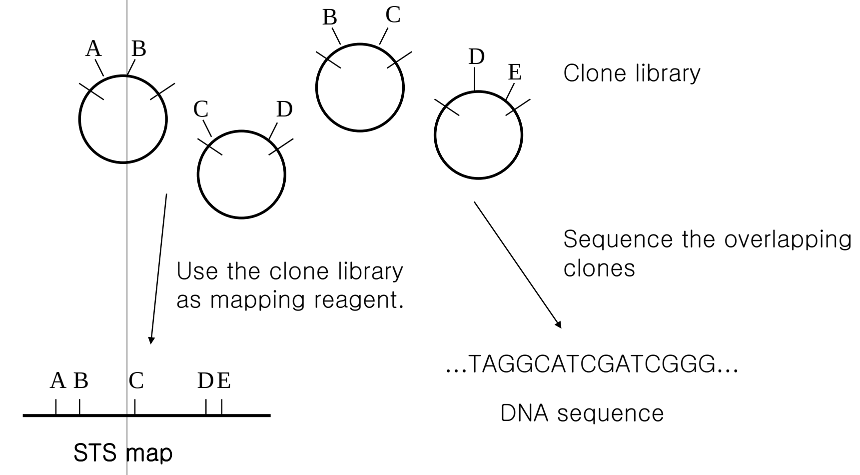
Radiation Treatment
Fuse with Rodent Cell Line

RH Panel

Results

A +	A −	A −
B +	B −	B +
C +	C +	C +
D +	D +	D −

Clone library based Mapping



Clone library

Use the clone library as mapping reagent.

Sequence the overlapping clones

STS map

...TAGGCATCGATCGGG...

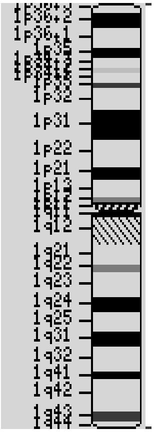
DNA sequence

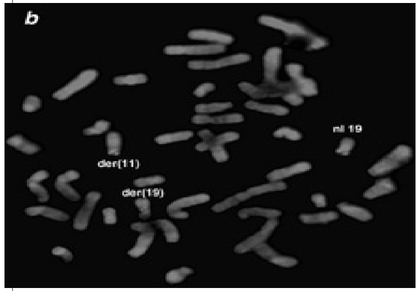
Type of Maps

- Cytogenetic maps
- Genetic Linkage maps
- Radiation hybrid maps
- Transcript maps
- Physical maps
- Comparative maps
- Integrated maps

Cytogenetic Maps

- Chromosome staining by Giemsa: G-bands
- Using FISH method
- Limited ordering range (<1-2mb)
- Not well-suited for high-throuput mapping.





Genetic Linkage Maps

- 1centiMorgans (cM) represents 1% probability of recombination.
- Not directly proportional to physical distance.
- MAP-O-MAT (map distance, statical support for order)

  \- http://compgen.rutgers.edu/mapomat
- CEPH(Centre d'Etude du Polymorphism Humain)

  \- Human DNAs from a set of reference pedigrees

  \- highly polymorphic STR markers , SNP

  \- http://www.cephb.fr/cephdb/

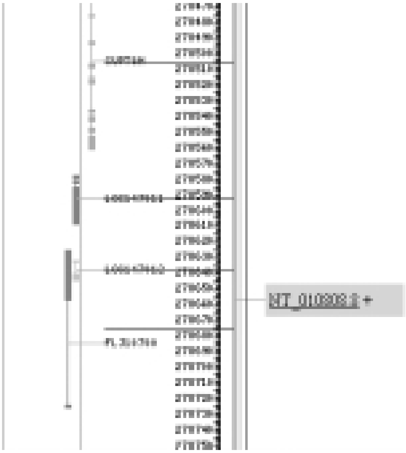
Radiation Hybrid Maps

- 1 Centirays (cR) representing a 1% probability of chromosome breakage.
- Radiation Hybrid Database (RHdb)

http://www.ebi.ac.uk/RHdb/

Transcript Maps

- Maps of transcribed sequences.
- Using Expressed Sequence Tag (EST) and known gene



Physical Maps

- STS content mapping (>1Mb)

  \- using PCR-based positional markers.

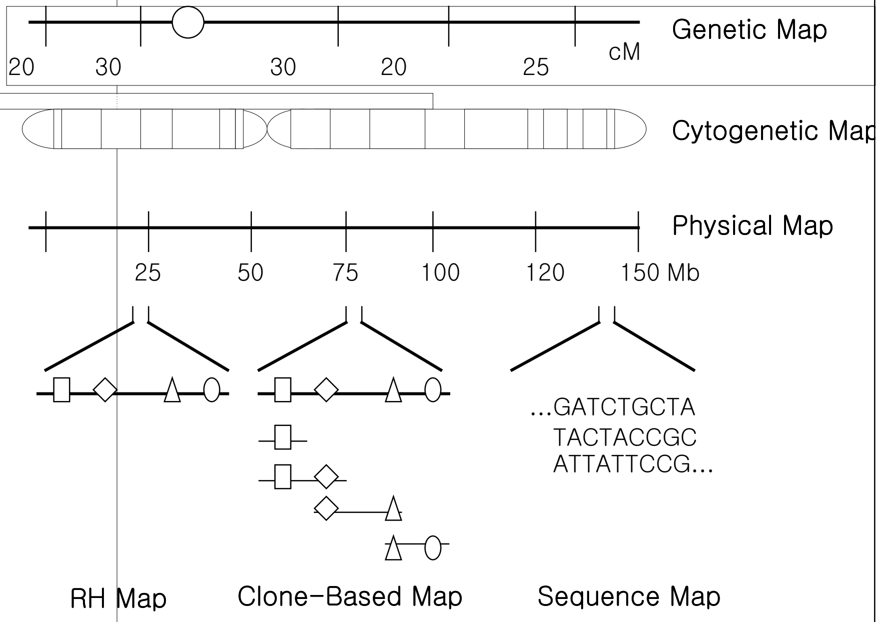
  \- Ordering of marker in clone library

  \- distance is measured by restriction mapping
- CEPHYAC map

  \- used a combination of fingerprinting.
- Optical Mapping

Comparative Maps

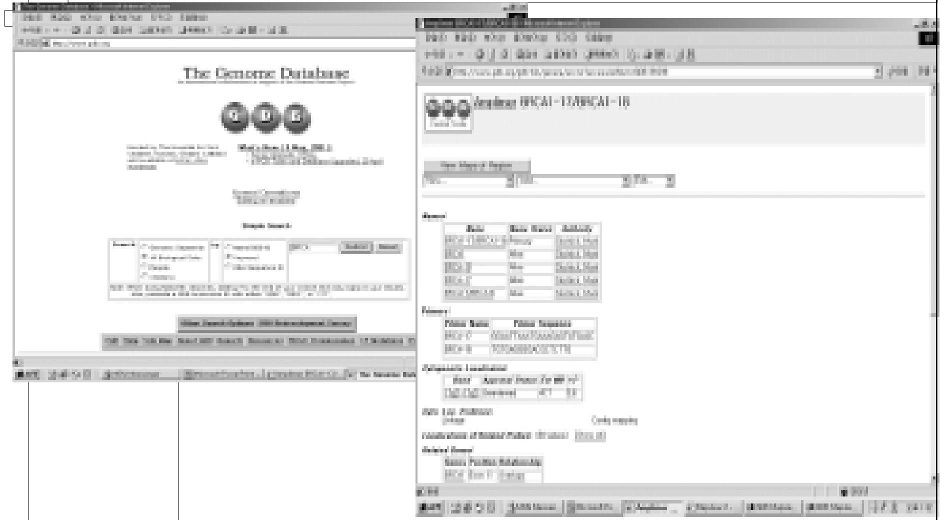
- Process of identifying conserved chromosome segments across different species.
- Orthologous genes sharing an identical linear order within a chromosome region.
- Identify conserved segments and ancient chromosome breakpoints.



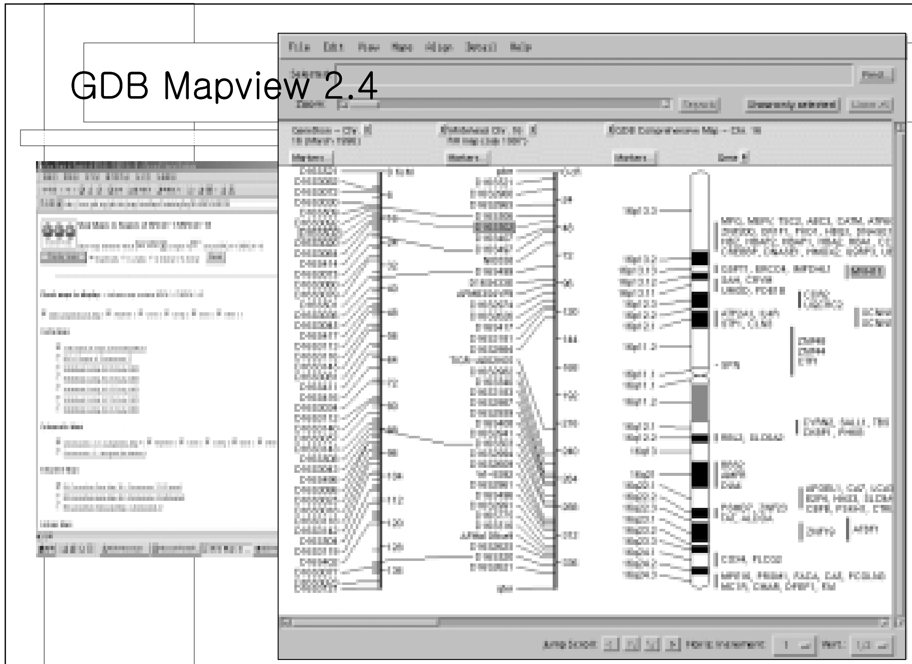
Genome Database (GDB)

- The official central repository for genomic mapping data created by the Human Genome Project
- Currently, GDB comprises descriptions of three types of objects from humans
  - Genomic Segments (genes, clones, amplimers, breakpoints, cytogenetic marakers, fragilesites, ESTs, syndromic regions, contigs, and repeats)
  - Maps (including cytogenetic, GL, RH, STS-content, and integrrated)
  - Variations (primarily relating to polymorphisms)

Genome Database (GDB)



GDB Mapview 2.4



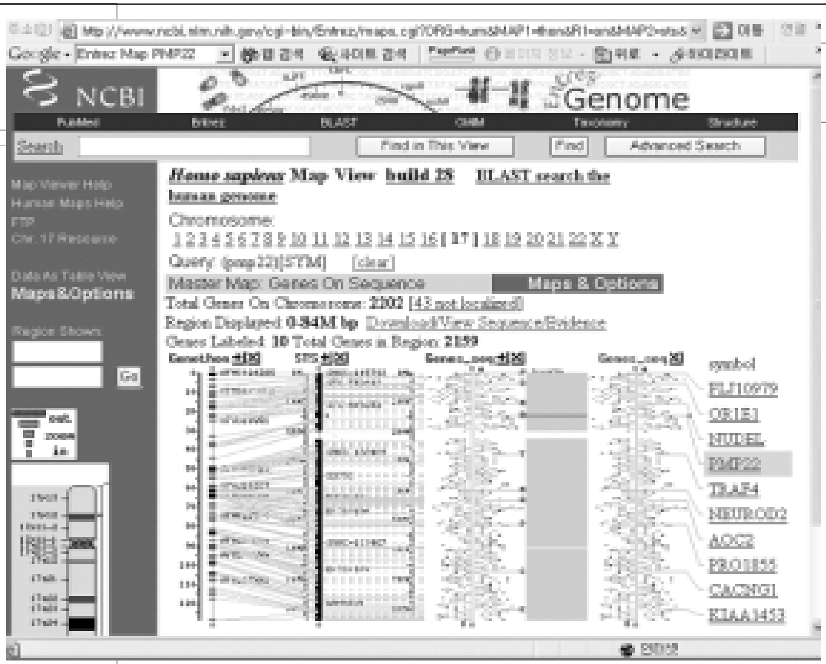
NCBI

- Entrez provides integrated access to several different types of data for over 600 organisms
  - ♦ Nucleotide sequences
  - ♦ Protein structures and sequences
  - ♦ PubMed/MEDLINE
  - ♦ Genomic mapping information
- NCBI Human Genome Map Viewer
  - ♦ Human genome sequence data as well as cytogenetic, genetic, physical, and radiation hybrid maps

---

Example: *Homo sampiens* Genome View page

- A genome-wide search for the term CMT* returns 33 hits. representing the loci for forms of Charcot-Marie-Tooth neuropathy on eight different chromosomes.
- Selecting the Genes_seq link for the PMP2 gene (the gene symbol for CMT1A, on chromosome 17) returns the view of the sequence map for the region surrounding this gene.



Homo sapiens Map View build 28 BLAST search the human genome

Chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 [17] 18 19 20 21 22 X Y

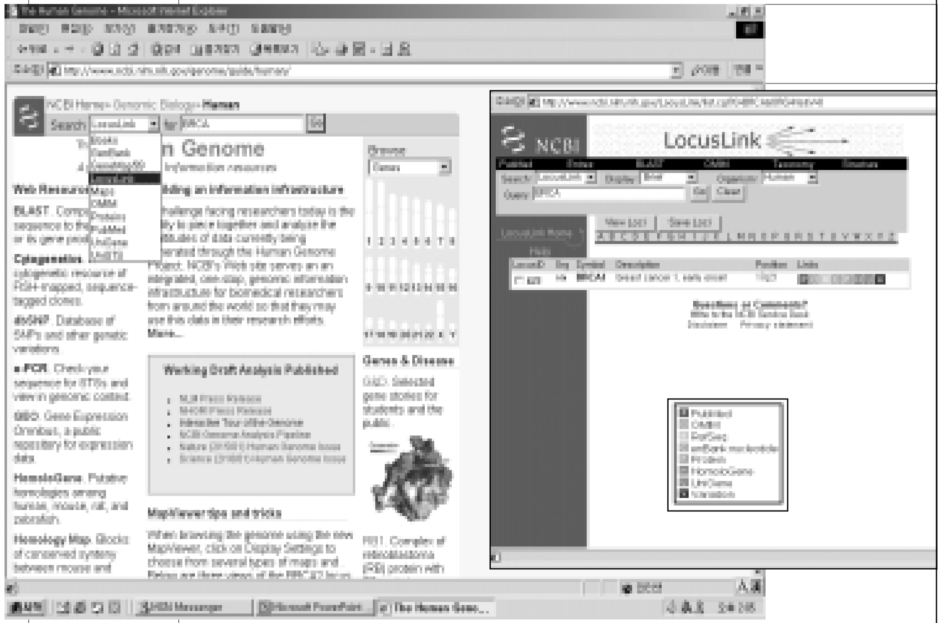
Query: (pmp22)[STM] [clear]

Master Map: Genes On Sequence Maps & Options

Total Genes On Chromosome: 2202 [43 not localized]

Region Displayed: 0-84M bp Download/View Sequence/Evidence

Genes Labeled: 10 Total Genes in Region: 2159



NCBI Home > Genomic Biology > Human

Human Genome

Building an information infrastructure

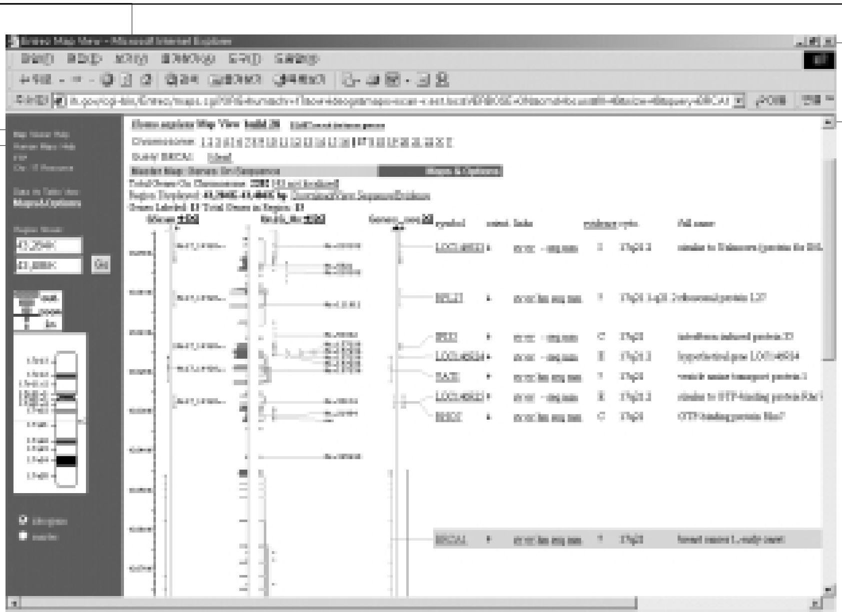
Working Draft Analysis Published

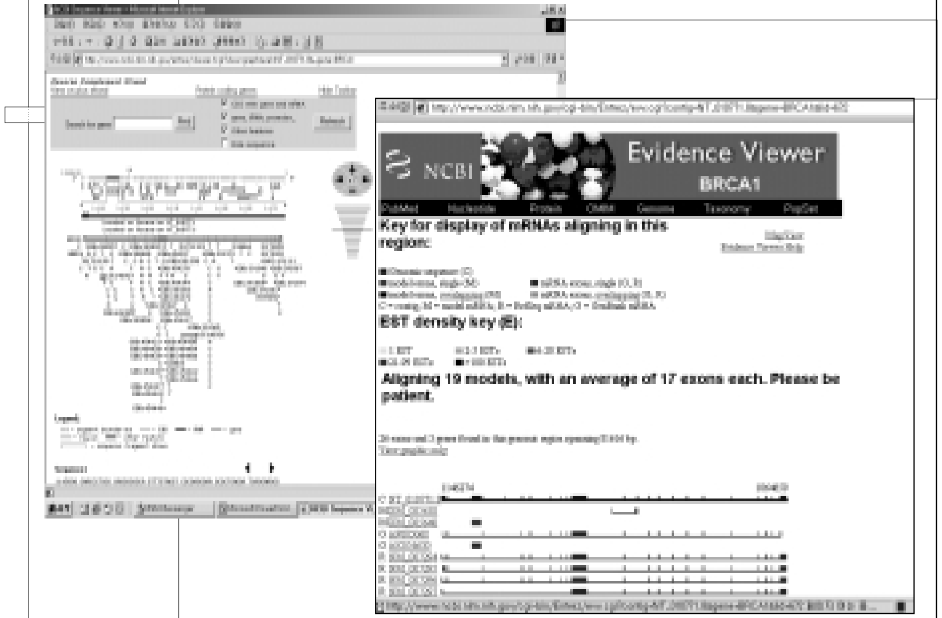
MapViewer tips and tricks

Genes & Disease

LocusLink

Questions or Comments?





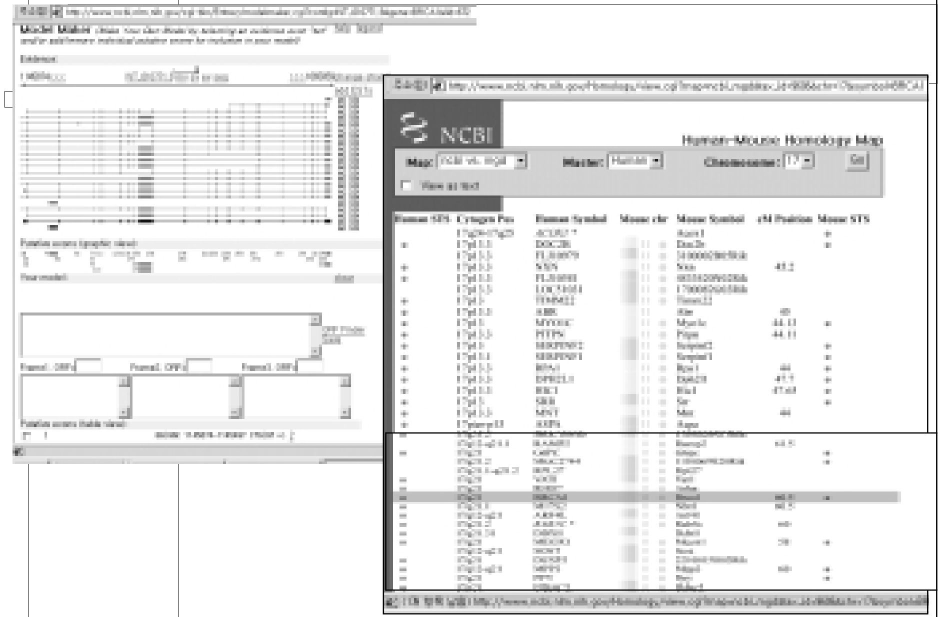
Evidence Viewer

BRCA1

Key for display of mRNAs aligning in this region:

EST density key (E):

Aligning 19 models, with an average of 17 exons each. Please be patient.



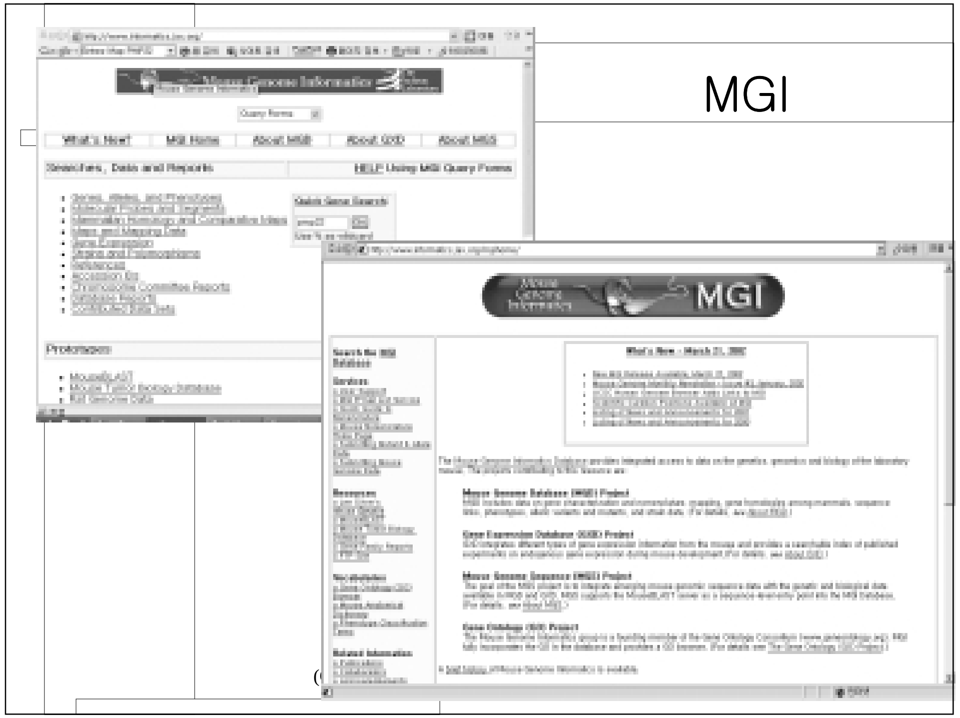
Model Maker

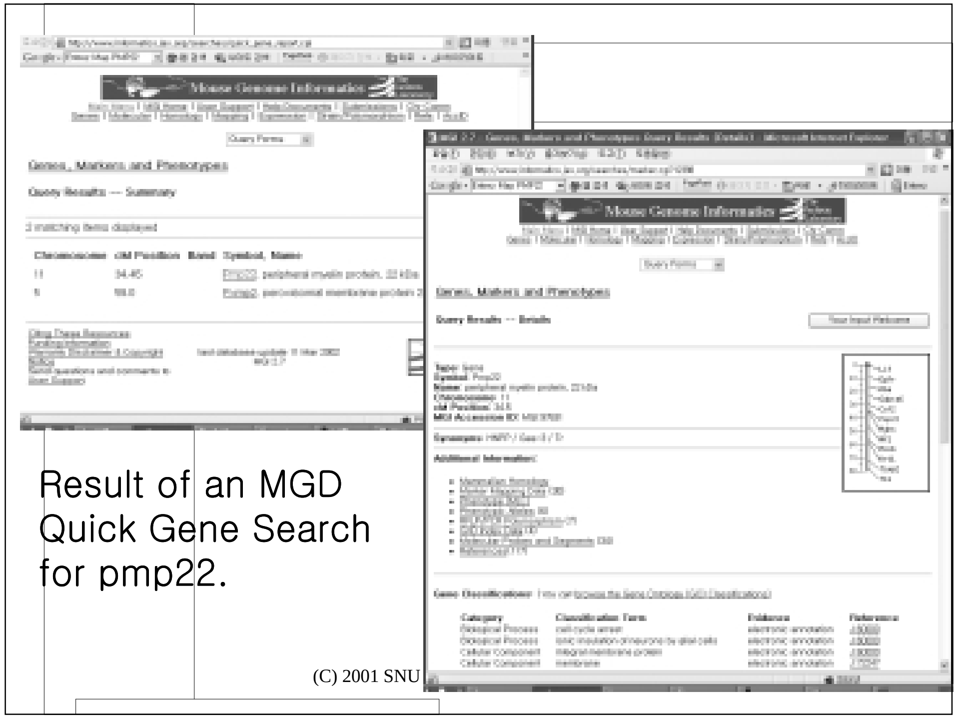
Human-Mouse Homology Map

MGI/MGD

- The Mouse Genome Initiative Database (MGI): the primary public mouse genomic catalogue resource.
- Located at The Jackson Laboratory, the MGI currently encompasses three cross-linked topic-specific database
    - ♦ The Mouse Genome Database (MGD)
    - ♦ The mouse Gene Expression Database (GXD)
    - ♦ The Mouse Genome Sequence project (MGS)

MGI





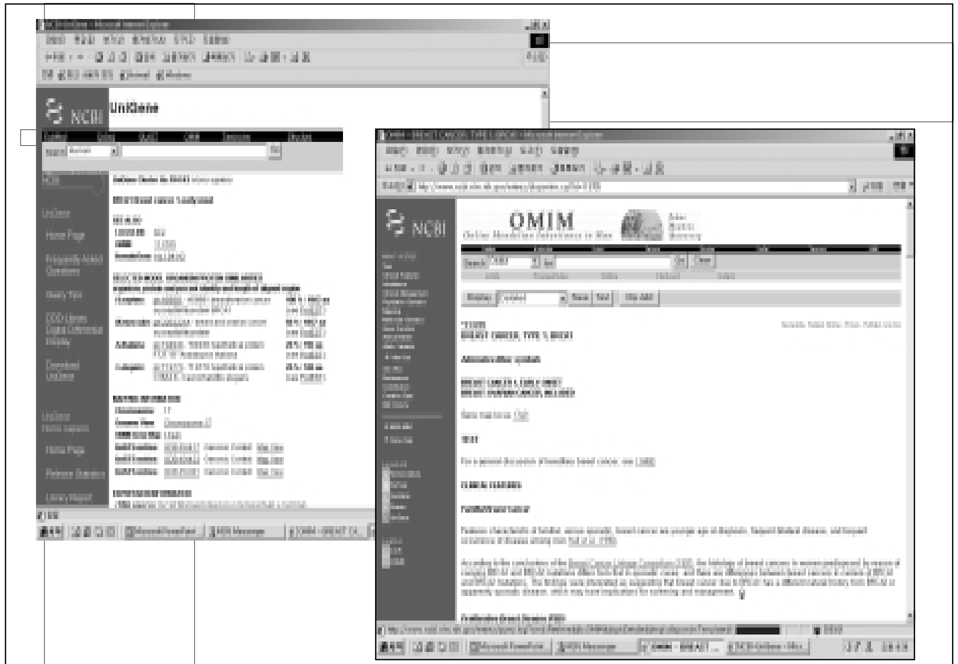
Result of an MGD Quick Gene Search for pmp22.

Mapping Projects and Associated Resources

- Cytogenetic resources
- Genetic linkage map resources
- Radiation hybrid map resources
- STS content maps and resources
- DNA sequence
- Integrated maps and genomic cataloguing
- Comparative resources
- Single-chromosome and regional map resources

Cytogenetic Resources

- The central repository for human cytogenetic information is GDB
    - Query for marker map and map information using cytogenetic coordinates
    - Cross-referencing cytogenetic positions with genes or regions of interest
- NCBI’s LocusLink, UniGene, OMIM catalogues
    - Valuable repositories of cytogenetic positions
- NCI’s Cancer Chromosome Aberratio Project
- Southeastern Regional Genetic Group
- Coriell Cell Repositories



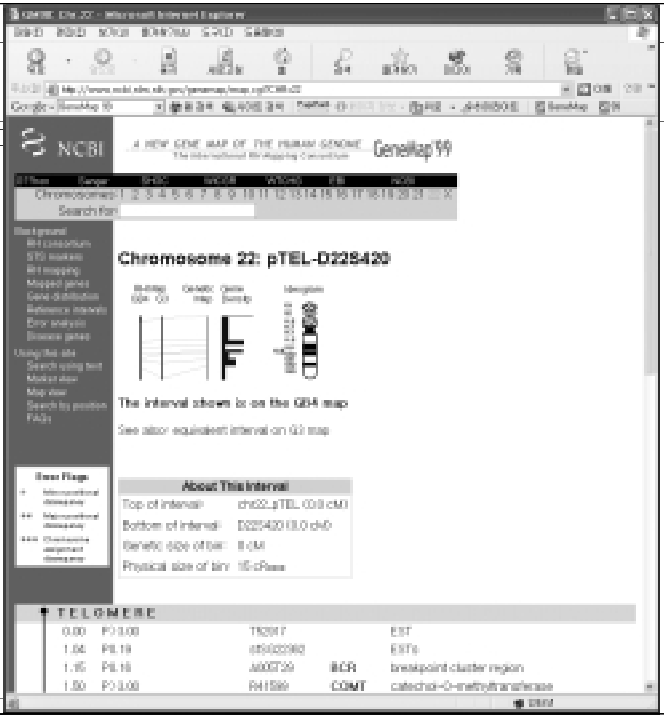
Genetic Linkage Map Resources

- The starting point for many disease-gene mapping projects
- The backbone of many physical mapping efforts
- Genome-wide maps
    - The Cooperative Human Linkage Center
        - The CHLC has identified, genotyped, and/or mapped over 3,300 microsatellite repeat markers
    - The group at Généthon
        - Identified and genotyped over 7,800 dinucleotide repeat markers
    - The Center for Medical Genetics
        - Identified over 300 dinucleotide repeats
        - Constructed high-desity maps using over 8,000 markers
- The ABI PRISM linkage mapping
    - Dinucleotide repeat markers derived from the Généthon linkage map
- The Map-O-MAT Web site
    - A marker-based linkage map server that provides several map-specific queries

---

Radiation Hybrid Map Resources

- Intermediate level of resolution between linkage and physical maps
- They are helpful for sequence alignment and will aid in completion of the human genome sequencing project
- The Radiation Hybrid Database (RHdb)
- The lowest-resolution human RH panel: the Genebridge4 (GB4)
- An intermediate level panel: The Stanford Generation 3 (G3)
- The height resolution panel (“The Nest Generation”, or TNG): Stanford



Chromosome 22: pTEL-D22S420

The interval shown is on the GB4 map

See also: equivalent interval on G3 map

About This Interval

Top of interval:	chr22_pTEL (0.0 cM)
Bottom of interval:	D22S420 (0.0 cM)
Genetic size of bin:	0 cM
Physical size of bin:	15 cR3000

1.15 BCR breakpoint cluster region

1.50 COMT catechol-O-methyltransferase

STS Content Maps and Resources

- The WICGR physical map
  - STS content based and contains more than 10,000 markers for witch YAC clones have benne identified.
  - Integrated with the Généthon GL and the WICGR RH maps
- CEPH/Généthon YAC project
  - Centered around screening of the CEPH MegaYAC library with a large set of STSs

---

DNA Sequence

- The existing human and forthcoming mouse draft genomic sequences are excellent sources for confirming mapping information, positioning and orienting localized markers, and bottom-up mapping of interesting genomic regions
- NCBI tools like BLAST can be very powerful in finding marker/sequence links.
- As the mammalian sequencing projects progress, a “sequence first” approach to mapping becomes more feasible

---

Integrated Maps and Genomic Cataloguing

- GDB’s Comprehensive Maps provide an estimated position of all genes, markers, and clones in GDB on a megabase scale.
- The estimated Mb position of the marker on each map:

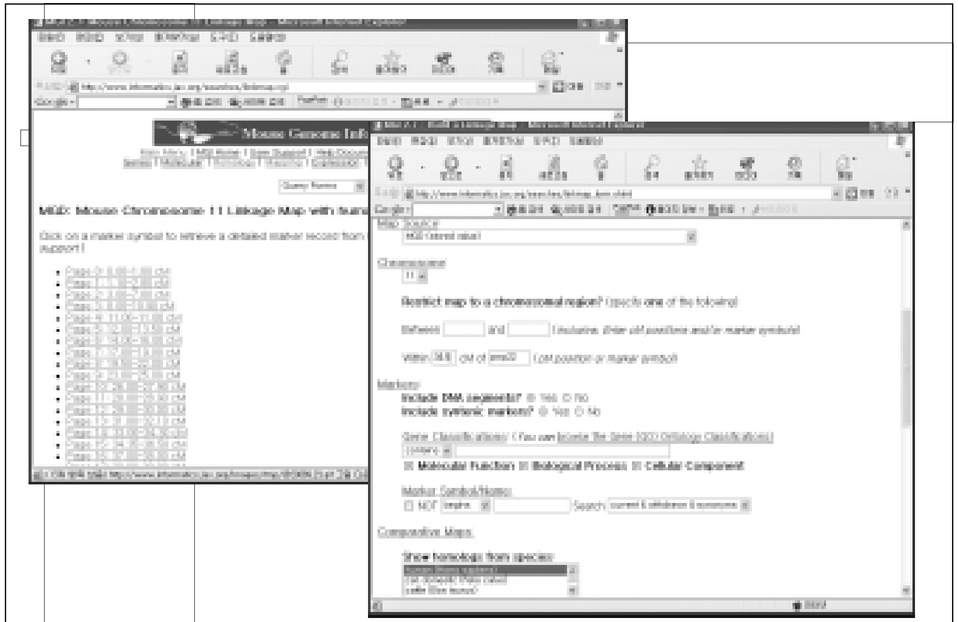
Element	Chromosome	Map	Coordinate	Units	EST MB^{-}	+/−
D1S228	1	CeneMap ‘99	782.0000	cR	32.2	0.0

- LDB and UDB are two additional sites that infer physical position of a large, heterogeneous set of markers from existing maps using algorithm analogous to GDB’s
- The eGenome project uses a slightly different approach for creating integrated maps of the human genome.
- Genomic catalogues help both to provide a single initial source containing most of the publicly available genomic information for a region and to make the task of monitoring new information easier.

Comparative Resources

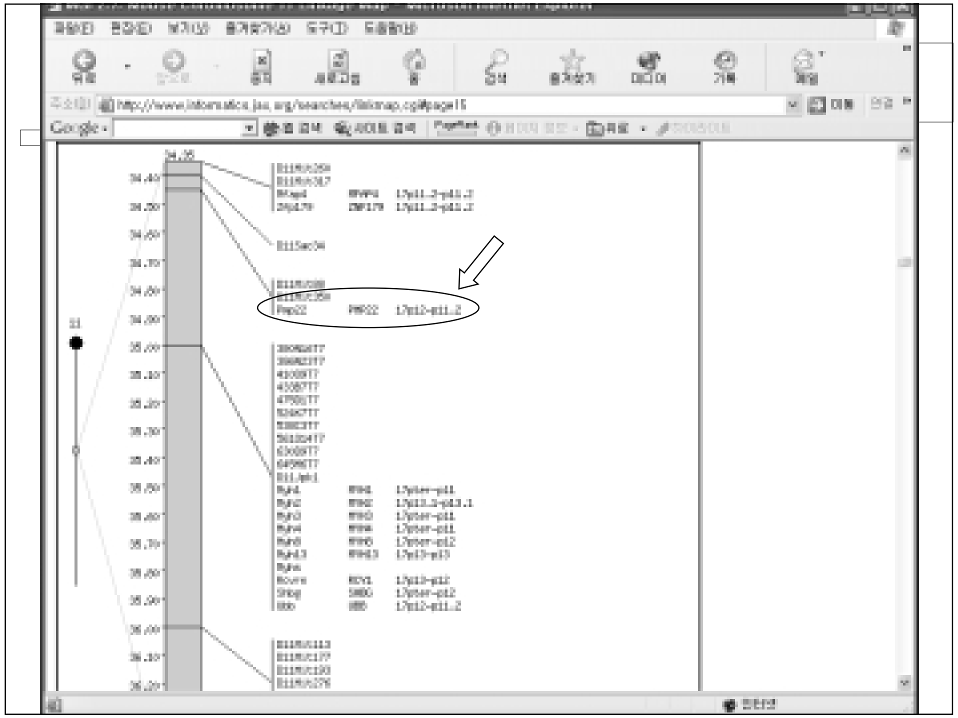
- Studying the evolution and relatedness of genes between species and finding disease genes through position based orthology.
- NCBI’s LocusLink database: Links to HomoloGene, a resource of curated and computed cross-species gene homologies
- The GDB does provide homology maps that simplify the reported literature for mouse, human, rat, and 17 other species.
- The Comparative Mapping by Annotation and Sequence Similarity (COMPASS) approach has been by researchers studying the cattle genome to construct cattle-human comparative maps with 638 identified human orthologs

---



URL: http://www.informatics.jax.org/searches/linkmap_form.shtml

---



Single-Chromosome and Regional Map Resources

- Data generated for only a single chromosome or a subchromosomal region are often important for fine mapping
- Most published human chromosome maps are listed and be viewed at GDB’s Web site.
- Human Genome Organization (HUGO) has developed individual human chromosome Web pages.
- The Sanger Centre and the WUGSC have two of the most advanced collections of chromosome-specific genomic data, informatics tools, and resources.

Practical USES of Mapping Resources

- Defining a genomic region
- Determining and ordering the contents of a defined region
- Defining a map position from a clone or DNA sequence

Defining a Genomic Region

- A genomic region of interest is best defined by two flanking markers that are commonly used mapping purposes
    - Ploymorphic Généthon markers in humans
    - MIT microsatellites in mice
- Commonly used markers are often present on multiple, independently derived maps, so their “position” on the chromosome provides greater confidence for anchoring a regional endpoint
- In contrast, the exact location of less commonly used markers is often locally ambiguous.

Determining and Ordering the Contents of a Defined Region

- A good way to start is to identify a map that contains both flanking markers, either from a chromosome-wide or genome-wide map from the sources
- The map or maps containing the flanking markers can be used to create a consensus integrated map of the region.
- The most reliable tool for marker ordering is a DNA sequence or sequence contig.

Defining a Map Position from a Clone or DNA Sequence

- To determine whether the element of interest has already been localized
- If gene-specific or closely linked markers have been used previously for mapping, a position can usually be described in terms specific to the mapping method that was employed.
- If no previously localization exists for a genomic element, some experimental work must be undertaken.