Application of Information Sciences to Analysis and Interpretation of Novel Genetic Data

NIA Alzheimer's Disease Centers Directors Meeting

Speaker:

Christopher Dubay

Division of Medical Informatics & Outcomes Research, OHSU

1

Introduction & Overview

- Speaker Introduction
- Description of the Talk & Bioinformatics
  - An Integrative Information Science
- Illustrative Exercise: Processing a 'Top 50' candidate genes list
- Future goals for Bioinformatics:
  - Systems Biology of Complex Phenotypes
  - The New Medicine

2

Bioinformatics

- General Tools
  - WP, Spreadsheets, Robotics, Instrumentation
- Communications
  - E-Mail, Networks, Internet & World Wide Web
- Databases
  - Storage, Organization
- Analysis Tools
  - Examination & Discovery
- **Informatics Has Changed How Science is Done**

3

Basic Point of Talk

- New Biology ->
- New Genetics ->
- New Medicine ->
- How Information Sciences & Technology can Support and Influence this Process

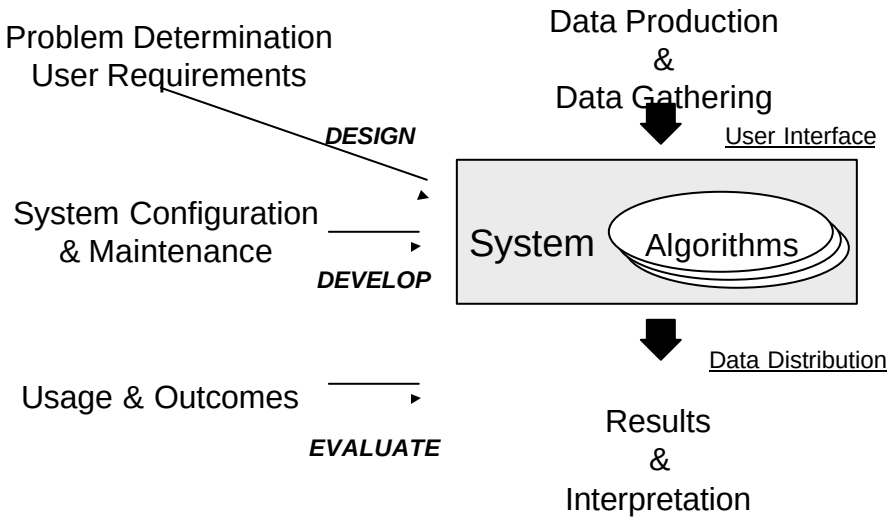
4

Bioinformatics Skill Set

- Practical Tools
- Cross Cultural Exchange
  - Language of Biomedical Research
  - Language of Informatics
- Solving Scientific Problems using Computers
  - Database Interoperation
  - Process Modeling & Data Visualization
- <u>Bioinformatics is an Information Science</u>

5

Informatics Research



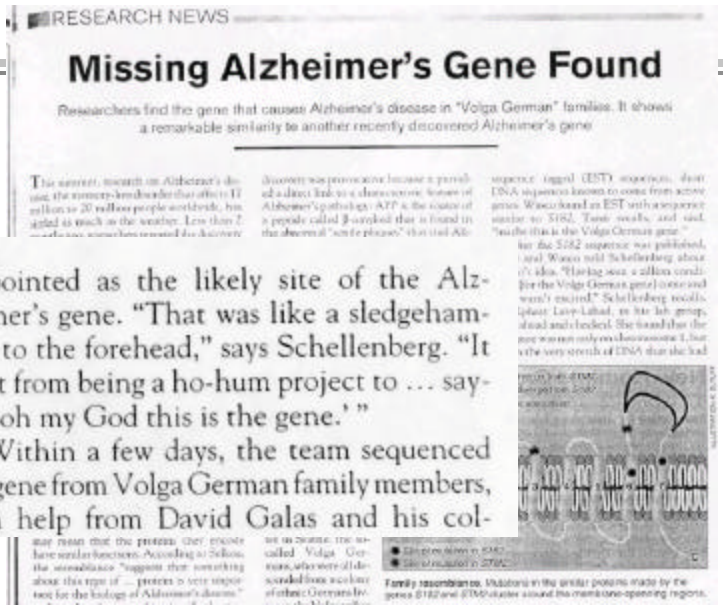
6

Bioinformatics Skill Set

- Practical Tools
- Cross Cultural Exchange
  - Language of Biomedical Research
  - Language of Informatics
- Solving Scientific Problems using Computers
  - Database Interoperation
  - Process Modeling & Data Visualization
- <u>Bioinformatics is an Information Science</u>

7

Bioinformatics Significance



RESEARCH NEWS

Missing Alzheimer's Gene Found

Researchers find the gene that causes Alzheimer's disease in "Volga German" families. It shows a remarkable similarity to another recently discovered Alzheimer's gene

pinpointed as the likely site of the Alzheimer's gene. "That was like a sledgehammer to the forehead," says Schellenberg. "It went from being a ho-hum project to … saying 'oh my God this is the gene.' "

Within a few days, the team sequenced the gene from Volga German family members, with help from David Galas and his col-

8

Illustrative Exercise

- What do you do with a 'Top 50' list of candidate genes?
  - Nomenclature
  - Similarity:
    - » Coding, non-Coding
    - » Structure & Function
    - » Cross-Species homology
  - Known Variation
  - Published Literature
    - » Systems Biology: pathways & interactions

9

Nomenclature

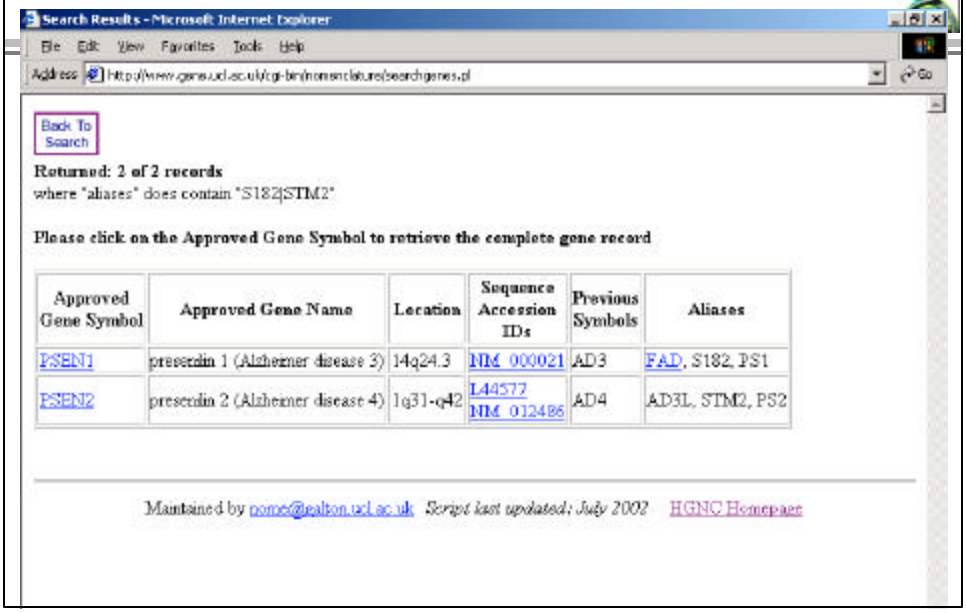
- Key to being able to find references
  - New and Old references
- Current Central Repository for Gene Names
  - HUman Genome Organization - Genew

10

HUGO - Genew

11

HUGO - Genew

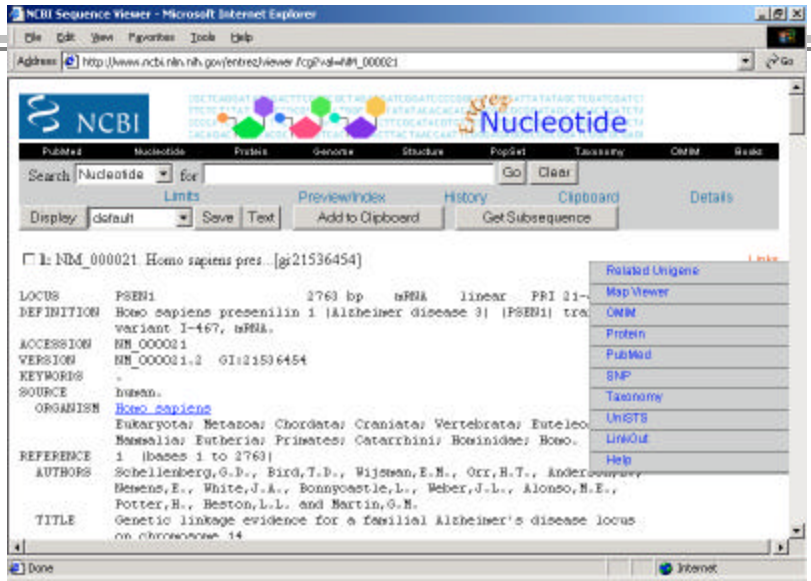


Similarity - Biosequence

- DNA Sequences
  - Get from GenBank
  - Find Genomic Location and Control
  - Gene Family?
- Protein Sequence
  - Function
  - Expression

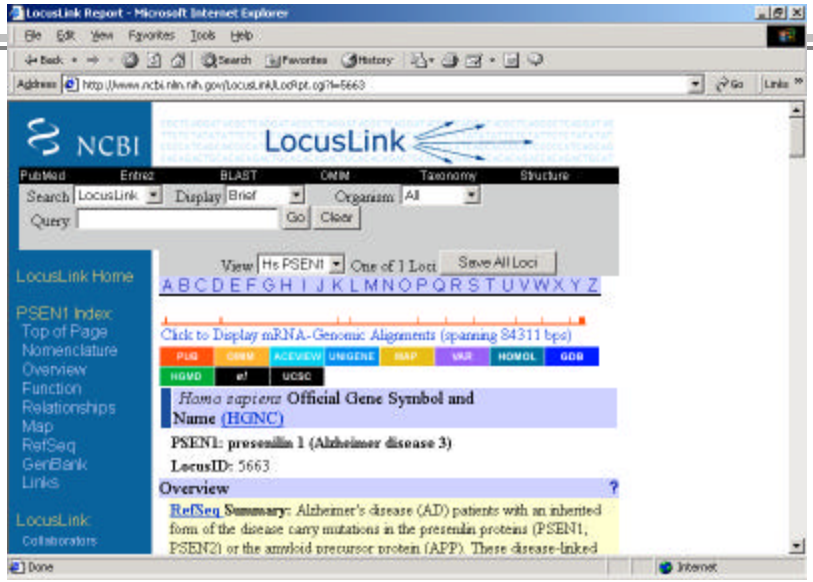
13

GenBank



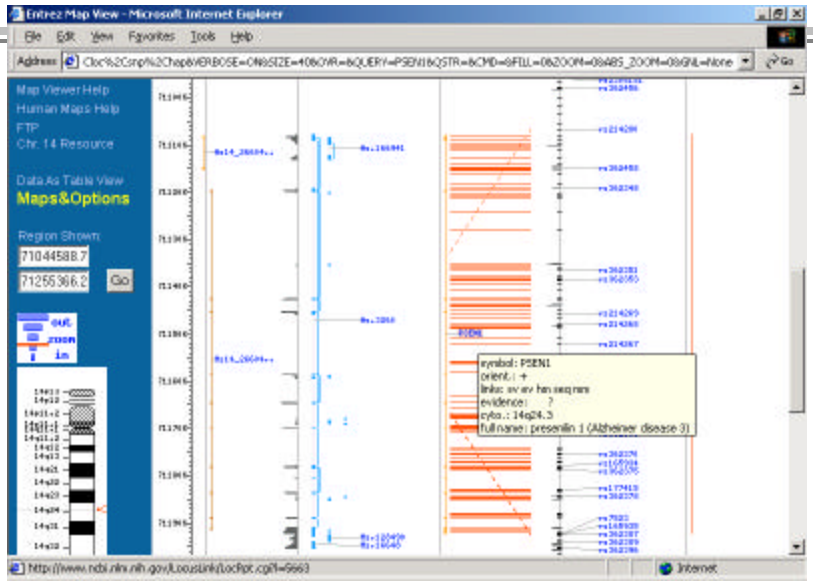
14

LocusLink



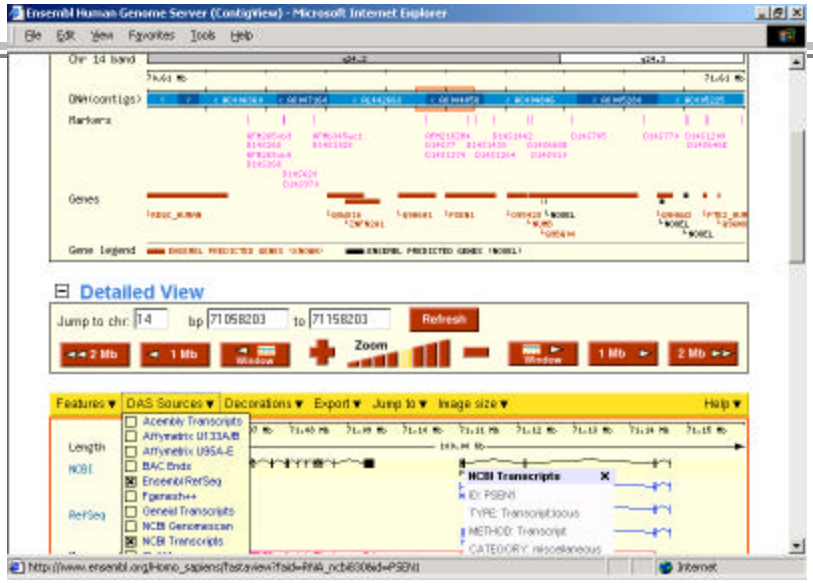
15

Genomic View - NCBI



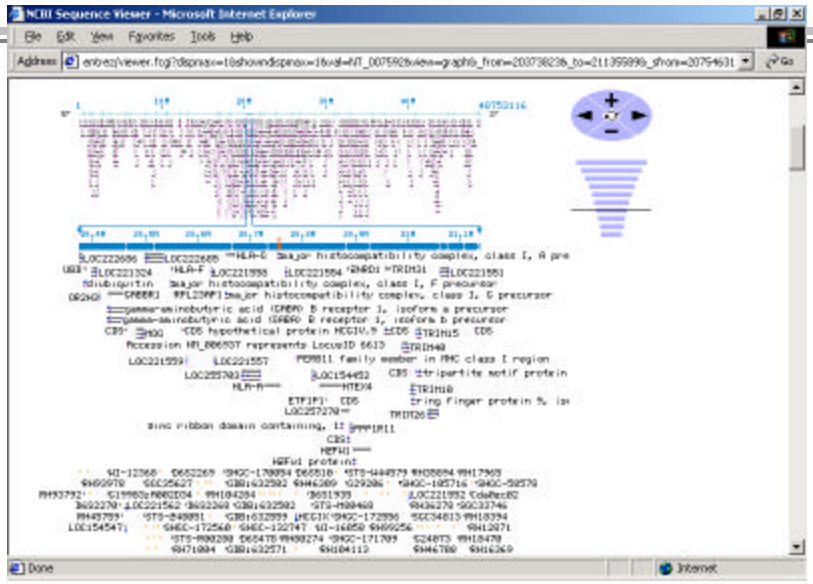
16

Genomic View - EMBL



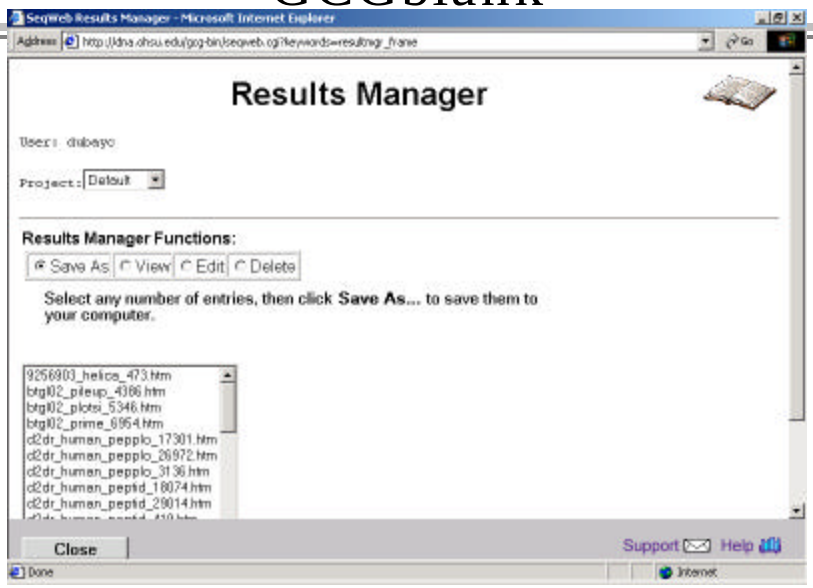
17

Genomic View - Entrez



18

Sequence & Results Manager: GCGblank



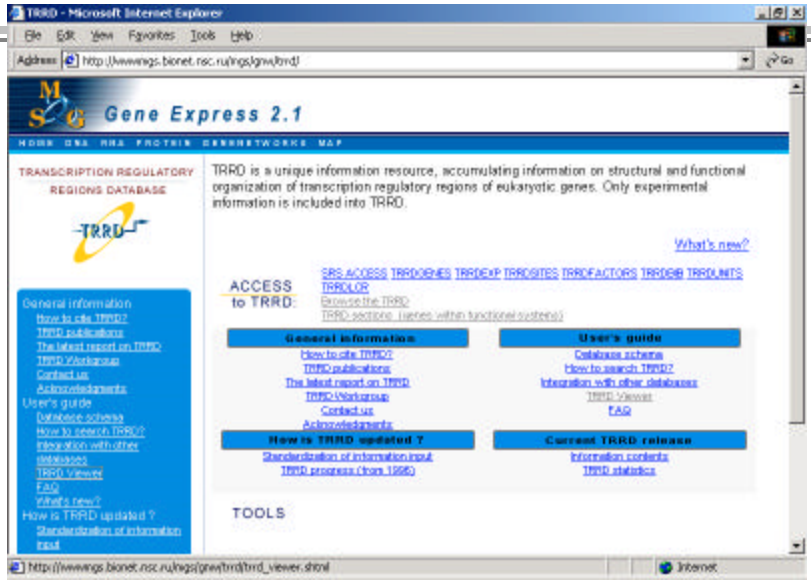
19

Promoters & Enhancers

- Transcriptional Elements
- For Genes & Genomic Region
  - Enhancers can be distant

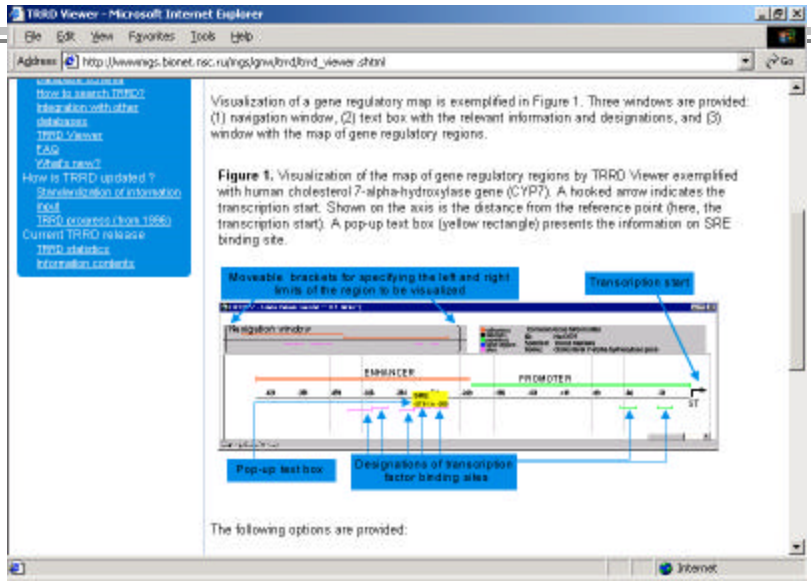
20

Transcription Factor Database



21

TRRD - Viewer



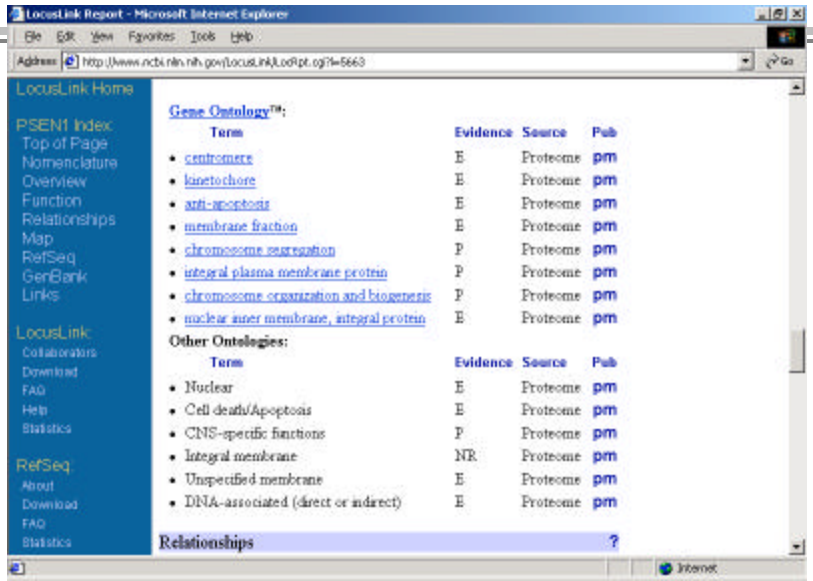
22

Similarity - Function

- Group genes of similar function
  - Spatially
  - Temporally
  - Action
- Need for common Vocabulary

23

GO: Gene Ontology



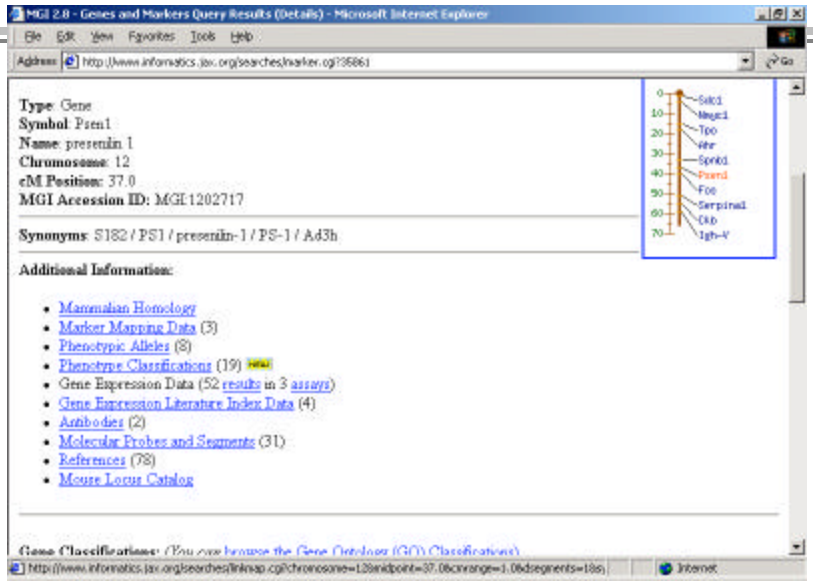
24

Similarity - Orthologs & Paralogs

- Function Clues from:
  - What is know in other organisms
  - Evolution of gene in organisms

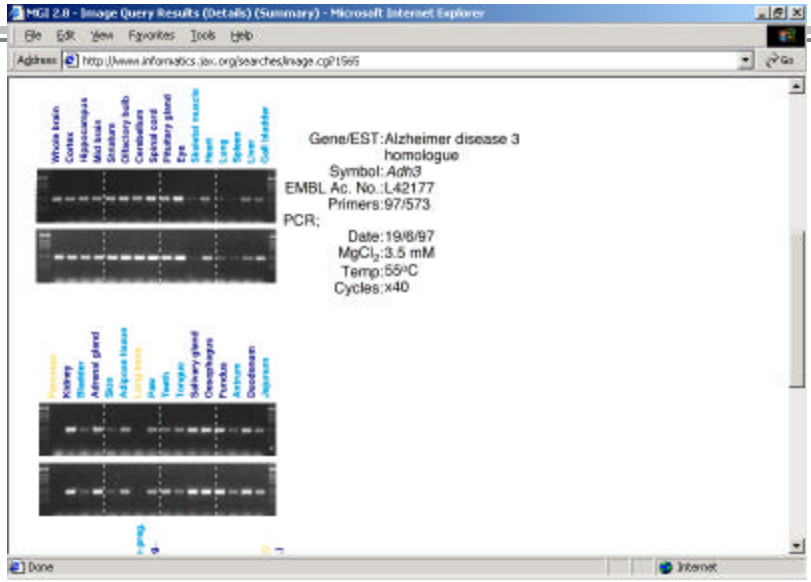
25

Jackson Labs: MGI



26

Jackson Labs: MGI



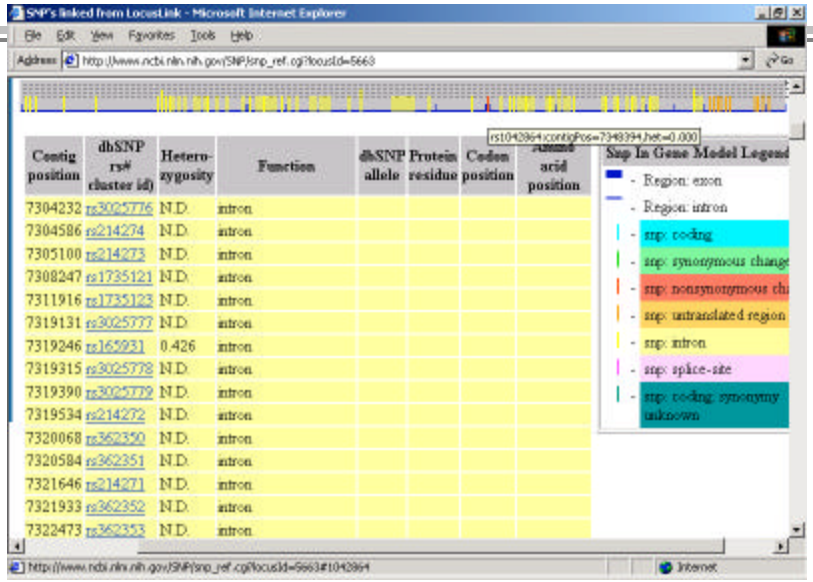
27

Variation

- Today: SNPs
  - Functional & Non-Functional
- Yesterday: RFLPs & SSLPs

28

dbSNP

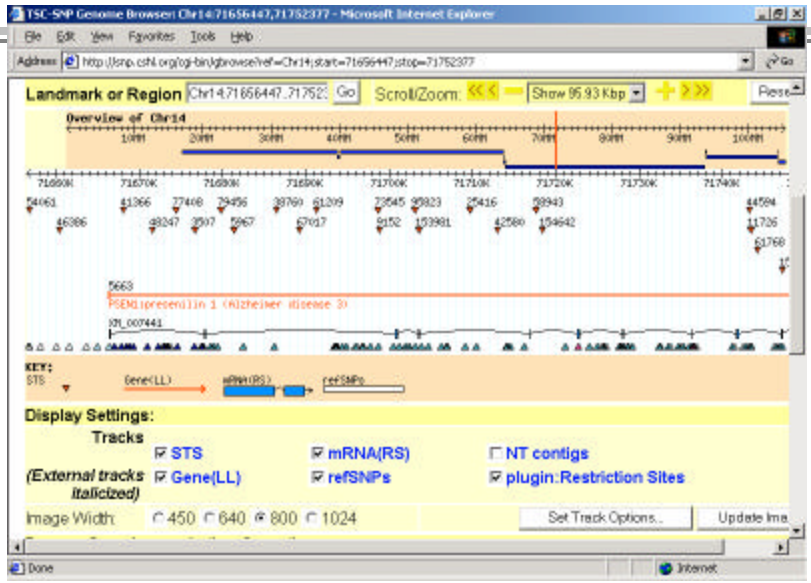


29

SNP Consortium

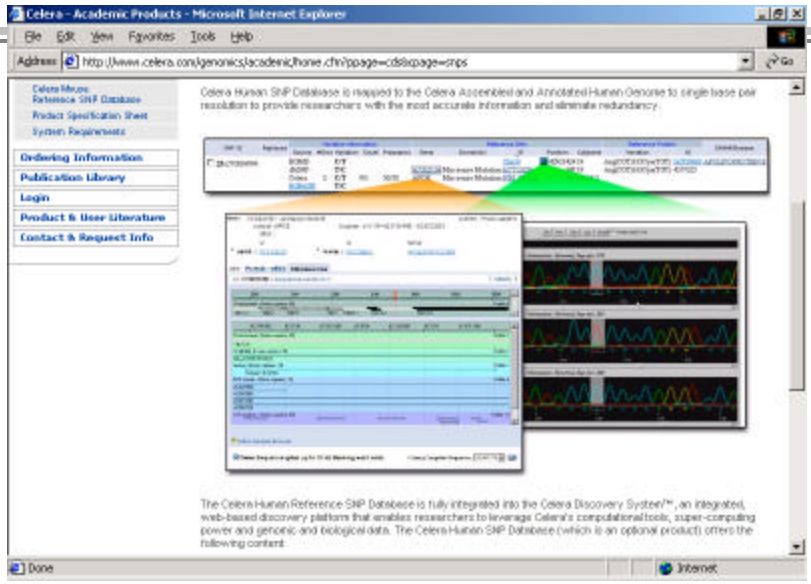
30

SNP Consortium



31

SNP (Celera) View



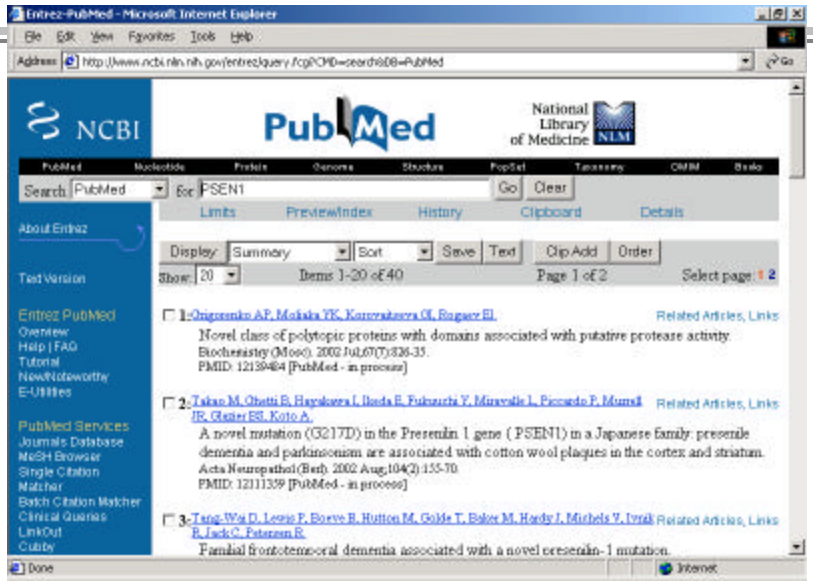
32

Published Literature

- Searching the Medline Abstracts
  - The MESH Vocabulary
- Now: Full Text & Web supplements
- Future: MIAME & other standards
- 3D Protein Structures

33

MedLine via PubMed



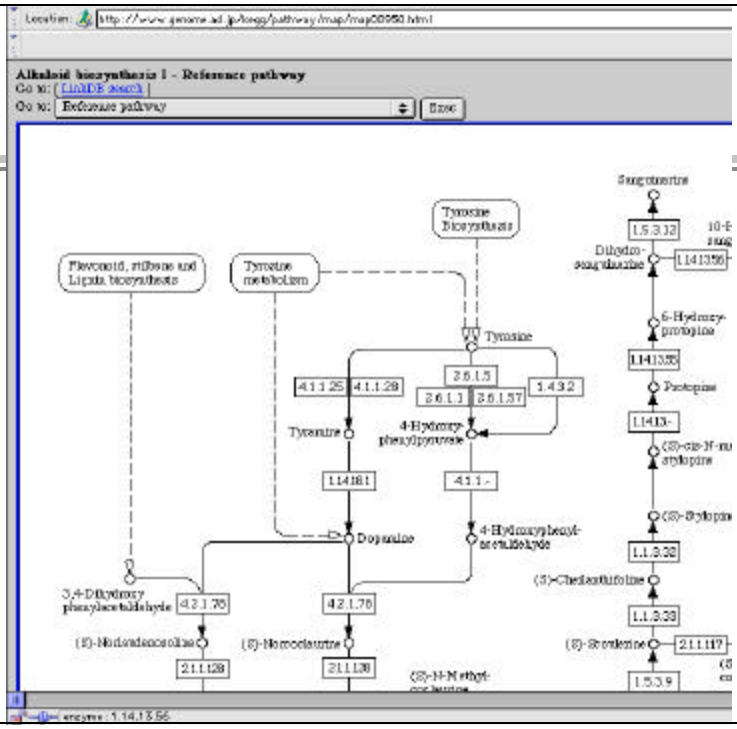
34

Systems Biology - Pathways & Interactions

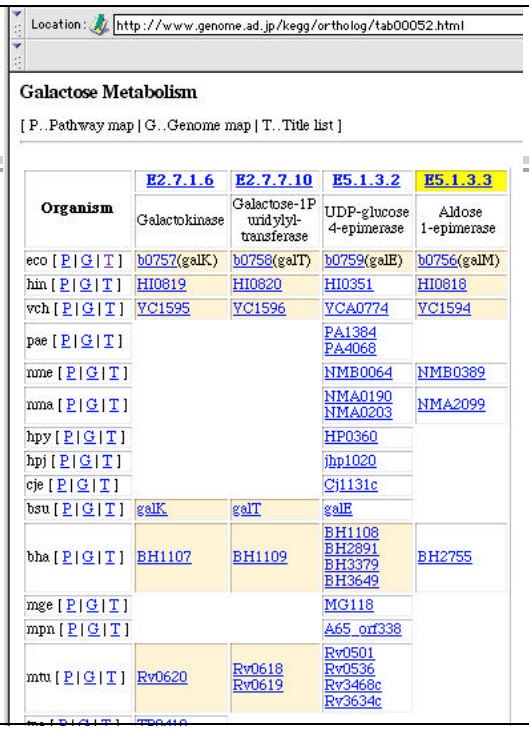
- Use our growing knowledge of:
  - Pathways
  - Gene & Protein Interactions

35

36



37



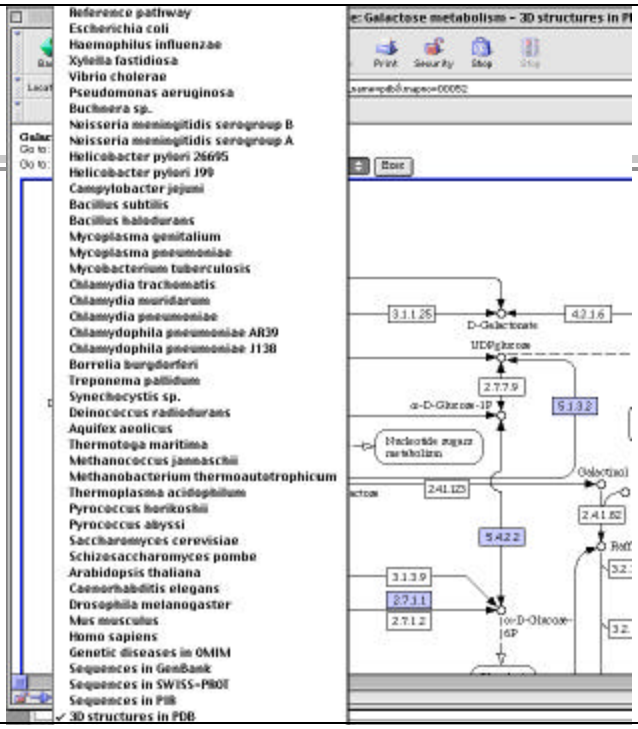
Location: http://www.genome.ad.jp/kegg/ortholog/tab00052.html

Galactose Metabolism

[P..Pathway map | G..Genome map | T..Title list]

Organism	E2.7.1.6 Galactokinase	E2.7.7.10 Galactose-1P uridylyl-transferase	E5.1.3.2 UDP-glucose 4-epimerase	E5.1.3.3 Aldose 1-epimerase
eco [P | G | T]	b0757(galK)	b0758(galT)	b0759(galE)	b0756(galM)
hin [P | G | T]	HI0819	HI0820	HI0351	HI0818
vch [P | G | T]	VC1595	VC1596	VCA0774	VC1594
pae [P | G | T]			PA1384 PA4068	
nme [P | G | T]			NMB0064	NMB0389
nma [P | G | T]			NMA0190 NMA0203	NMA2099
hpy [P | G | T]			HP0360	
hpj [P | G | T]			jhp1020	
cje [P | G | T]			Cj1131c	
bsu [P | G | T]	galK	galT	galE	
bha [P | G | T]	BH1107	BH1109	BH1108 BH2891 BH3379 BH3649	BH2755
mge [P | G | T]			MG118	
mpn [P | G | T]			A65_orf338	
mtu [P | G | T]	Rv0620	Rv0618 Rv0619	Rv0501 Rv0536 Rv3468c Rv3634c	

38



39

Bioinformatics Future

- Application of our discovered knowledge to heath care
- Pharmacogenomics
- Delivering Tools to the Clinician

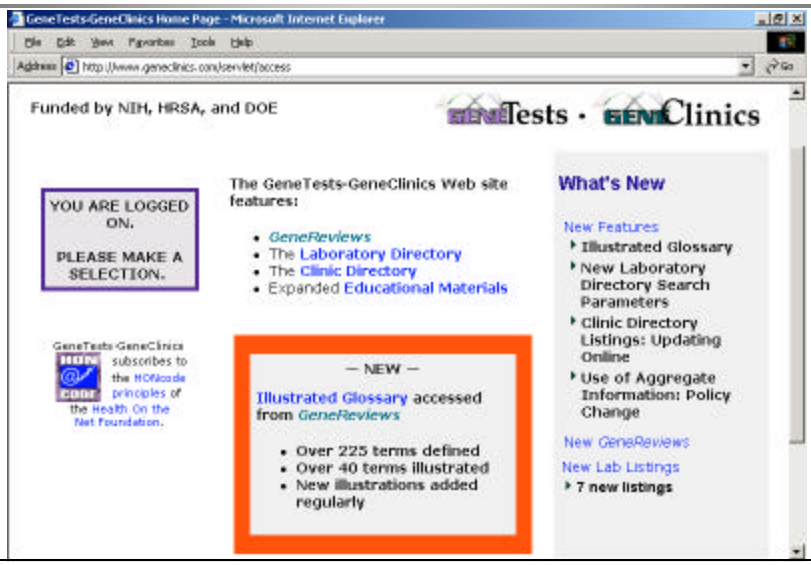
40

GeneClinics: Knowledge Base

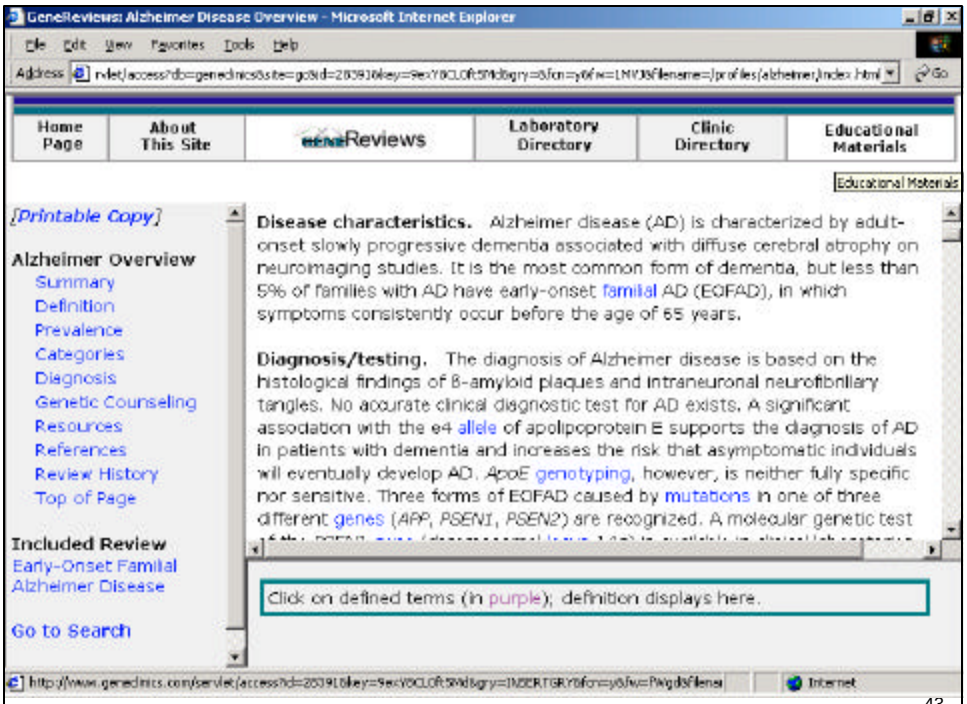
- Expert-Authored, Up-to-Date
  - Rapid Information Growth
  - Genetics Health Paradigm
- Meet Need of Health Professionals
- OO-DBMS via XML & WWW
- Develop Electronic Peer-Review
- Evaluate Utility & Methods
- Now: GeneSeek -> Integration

41

GeneClinics: Knowledge Base



42



43

Questions:

- Are these the best possible interfaces to the data?
- How can Bioinformatics Support Medical Informatics?
- Web links for sites in the talk available at:

http://medir.ohsu.edu/~bioinf/acdmurls.htm

44