Bioinformatics and Cancer

- Determine Proteins which are diagnostic of disease
- Determine Gene/Proteins which may impact prognosis
- Determine factors which may dictate therapeutic course
- Determine Gene products which are drug targets or vaccine candidates
- Determine natural variation which would correlate with disease risk and predisposition
Lung Cancer - a new disease

- First case described in the 18th Century by Giovanni Morgagni.
- By 1898 140 cases were on record.
- A 1912 pathology review by Isaac Adler claimed 374 cases had been recorded and noted “Lung cancer was amongst the rarest forms” but showed a “decided increase”.

www.bcgsc.ca
Modern Tobacco Usage

- 1839 “Bright” or “Virginia” tobacco invented.
- 1855 Safety match was invented.
- 1880 First cigarette machine was invented.
- 1921 Smoking tobacco remained illegal in 14 US States.
1929 Fritz Licknt, a Dresden physician, showed that lung cancer sufferers were more likely to be smokers.

- 1939 Franz Herman Muller produced the world’s first case controlled epidemiological study to show this link.
- 1996 Tobacco smoke was shown by Gerd Pfeifer and colleagues to cause mutations in the TP53 gene encoding the p53 protein.
Max Parkin et al. have estimated a global annual lung cancer toll of 1,240,000 new cases and 1,100,000 deaths for the year 2000.
Global Trends in Lung Cancer


info@bcgsc.ca  www.bcgsc.ca
The SAGE Technique

“Serial Analysis of Gene Expression”

[Diagram of SAGE technique]

Stage 1

[Expression chart for Gene product]

Stage 2

[Expression chart for Gene product]
Example DiscoveryDB Questions

• Show all the genes that are significantly up-regulated in a tumor and which are known cancer genes or are more than 50% identical to a known human cancer gene.

• Show all the proteins that are known to play a role in the process of apoptosis in human or are 70% identical to proteins known to be involved in apoptosis in any other organism.

• Show all the genes which are significantly down-regulated in the tumor and for which mutants are known in either the mouse, Drosophila, or C.elegans.
• Show me all the genes that have been implicated in a human disease, or genes that are known to be part of a biological pathway for which a disease gene has been determined.

• Show me all the Zn-finger proteins that are up-regulated, or any proteins known to bind to these Zn-finger proteins.

• Using literature data, show all the genes that are up-regulated and which are thought to bind or interact with telomeres in any organism.

• etc...
Natural Language Processing

Literature Sources
Most scientific information is in literature
- Pubmed 12M abstracts
- 400K per year

Information Overload
Filtering
- important concepts
- relevant facts

Relating Collected Facts
noun-verb-noun pattern

Eg.
- The XYZ gene is expressed in cancer.
- Gene A interacts with XYZ
A Gene Expression Karyotype for Lung Cancer
Waves of Transformation Events

- Wave 01
- Wave 02

www.bcgsc.ca
info@bcgsc.ca
Ho et al. Nature Jan 10 2002; 415:180-3

info@bcgsc.ca www.bcgsc.ca
The Bioinformatics of Mammalian Gene Expression

Building Regulatory Networks

• Genome Canada funded project
• Can we use bioinformatic tools to actually determine the regulatory elements controlling the expression of a gene?
• Collaboration with the European Bioinformatics Institute and ENSEMBL group
The Bioinformatics of Mammalian Gene Expression

• Genome Canada funded project
• Can we use bioinformatic tools to actually determine the regulatory elements controlling the expression of a gene?
• Collaboration with the European Bioinformatics Institute and ENSEMBL group

info@bcgsc.ca
www.bcgsc.ca
Comparative Genomics

Currently there will be a great resource in vertebrate genomic information

Mouse, Rat, Chicken, Pig, Fugu, Zebrafish

Orthologs between these species will share conserved upstream regions which often represent regulatory regions.
Potential Benefits

Derivation of a Regulatory network of Gene Expression

- **Gene Therapy**
  Being able to predict precise temporal and spatial expression of a gene will be crucial in developing this technology

- **Drug Discovery**
  Identify potential targets that could predictably alter the expression of particular genes

- **Diagnostic and Prognostic Tests**
  Determine new allelic variants for common diseases and Adverse Drug Reactions (ADRs)

- **Disease Predisposition**
  Effective determination of disease predisposition will be a major factor in future health care provision
MGE: 3D Concept Drawings

Isometric View

< Side Profile

GGAGATTCTGGGCCACTTTGGTTCCCCATGAGCCAAGACGGCACTTCTAATTTGCATTCCCTACCGGAGTCCCTGTCTGTAGCCAGCCTGGCTTTCAGCTGGTGCCCAAAGTGACAAATGTATCTGC AATGACAAAGGTACCC

From the diagram, it appears to be a representation of a genetic sequence or a part of a genome. The text seems to be a mix of nucleotide sequences and some sort of artistic rendering, possibly related to genetic engineering or molecular biology.
Sockeye Viewer

- Open Source
- Connects to any Ensembl database
- Java based, with the 3D API
- Coming Soon!

www.bcgsc.ca
info@bcgsc.ca