

SAGE Visualization Tool for Gene Expression Analysis

Presented by:
Timothy Chan and Zsuzsanna Hollander

1

Gene Expression and Motivation

1. All living things are made up of cells.
2. All cells contain genes which have the information to create all sorts of proteins in our bodies including our nails, hair, enzymes etc.
3. Different cell types contain the same DNA, but are different because different proteins are synthesized and produced.
4. A cell can change the expression level of its genes in response to various signals (ie. Stress, heat, damage, etc).
5. Gene expression levels are different in diseased cells and normal cells.

2

SAGE

1. Advent of large-scaled gene expression technologies have allowed simultaneous analysis of 10's of thousands of genes.
2. SAGE (Serial Analysis of Gene Expression) is a sequenced based method to quantify gene expression levels in cells.
3. Method based on taking a small sequence (called a TAG) of an mRNA to represent a gene.

3

Sample Data

TAG	COUNT	TAG	COUNT	TAG	COUNT
CCCATGCTC	1286	CACACTCAC	245	FTCACTGTGA	150
CCTCAGCTA	715	ACTAACAACC	229	ACGAGGGGAG	142
CTAAGACTC	559	AGCCCFACAA	222	TGCTCCFACC	140
GCOCAGTCA	519	ACTTTTCAA	217	CAACCTATCC	140
CACCTAATTG	469	GCAGGGTGGG	207	CCCCCTGGAT	136
CCGTATATC	448	GACATCAAGT	198	ATTTGAGTGC	136
TTCATAGACC	400	ATCGTGGGGG	193	GCAGGGCCTC	128
ACATTGGGTG	377	GACCCAGGAT	190	GGGTGCAGT	127
GTGAAACCC	359	GTGAACCCCT	188	GGAAACAGA	119
CCACTGGACT	359	CTGGCCCTGG	186	TCACCGGTCA	118
TGATTGACT	358	GCTTTATTGG	185	GTGCAGTGG	118
AGCCTTGGCC	344	CTAGCCTCAC	172	CCTCAGGATA	114
ATTTGAGAAG	320	GCRAAACCT	167	CTCATAGGA	113
GTGAPACGG	294	AAAAATTCT	161	ATCATGGGGA	110

4

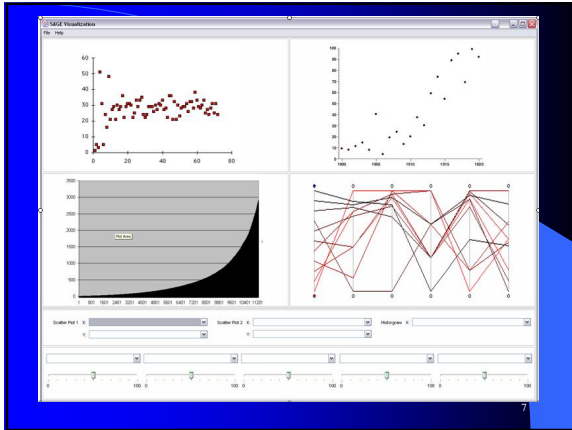
Problems

1. A typical experiment requires ~30,000 gene expression comparisons where normal and a diseased cell is compared.
2. Statistical measures are used to filter out candidate genes to reduce the dimensionality of the data but it is tedious and time consuming to play with these measures until a good set is found.
3. Finding significant genes would be much easier with some sort of visualization tool.

5

Proposed Solution

6



Milestones

ID	Task Name	Duration	Start	Finish	% Complete
1	Project Proposal	24 days	Thu 05/02/04	Mon 01/03/04	100%
2	Research	31 days	Thu 05/02/04	Mon 08/03/04	75%
3	Design	15 days	Wed 25/02/04	Wed 10/03/04	90%
4	Implementation	36 days	Wed 10/03/04	Thu 15/04/04	10%
5	Paper Writing	29 days	Mon 22/03/04	Tue 20/04/04	0%

Milestones

ID	Task Name	Duration	Start	Finish	% Complete
1	Project Proposal	24 days	Thu 05/02/04	Mon 01/03/04	100%
2	Research	31 days	Thu 05/02/04	Mon 08/03/04	75%
3	Design	15 days	Wed 25/02/04	Wed 10/03/04	90%
4	Implementation	36 days	Wed 10/03/04	Thu 15/04/04	10%
5	Paper Writing	29 days	Mon 22/03/04	Tue 20/04/04	0%

- ## Research
- Research existing SAGE Software Visualization tools
 - Read up on papers on sliders, scatter plots, parallel coordinates
 - Research and Review Swing and find appropriate Swing IDE to work with

- ## Difficulties
- Java IDE
 - JBuilder
 - NetBeans
 - Sun Java Studio

Milestones

ID	Task Name	Duration	Start	Finish	% Complete
1	Project Proposal	24 days	Thu 05/02/04	Mon 01/03/04	100%
2	Research	31 days	Thu 05/02/04	Mon 08/03/04	75%
3	Design	15 days	Wed 25/02/04	Wed 10/03/04	90%
4	Implementation	36 days	Wed 10/03/04	Thu 15/04/04	10%
5	Paper Writing	29 days	Mon 22/03/04	Tue 20/04/04	0%

Milestones

ID	Task Name	Duration	Start	Finish	% Complete
1	Project Proposal	24 days	Thu 05/02/04	Mon 01/03/04	100%
2	Research	31 days	Thu 05/02/04	Mon 08/03/04	75%
3	Design	15 days	Wed 25/02/04	Wed 10/03/04	90%
4	Implementation	36 days	Wed 10/03/04	Thu 15/04/04	10%
5	Paper Writing	29 days	Mon 22/03/04	Tue 20/04/04	0%

13

Implementation

- GUI Implementation
- Parser/Loader
- Integration of Scatter Plot, Histogram, Parallel Coordinate Modules

14

Difficulties

- Integrating graphing modules
 - parallel coordinate
 - scatter plot
 - histogram

15