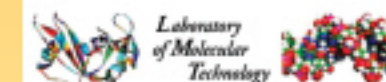




TMAP - Tissue Molecular Anatomy Project



Djamel Medjahed,^{1,2} Tawady Tontesh,² Gary W. Smithers,^{2,5} Brian Luke,^{2,5} David J. Munroe,^{1,2} and Peter F. Lemkin^{3,4}

¹Laboratory of Molecular Technology; ²SAIC-Frederick, Inc.; ³Laboratory of Experimental and Computational Biology; ⁴Center of Cancer Research;

⁵Advanced Biomedical Computing Center, National Cancer Institute at Frederick, P.O. Box B, Frederick, Maryland 21702 USA

Introduction

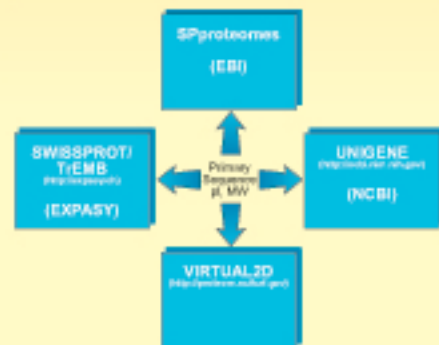
This project started initially as an effort to build interactive theoretical 2-dimensional pI/Mw charts (isoelectric focusing point vs molecular mass).

Ninety two proteome sets generated by the European Bioinformatics Institute were used to generate theoretical 2-Dimensional pI/Mw charts (isoelectric focusing point vs molecular mass).

The motivation was to build interactive, web-able tools to facilitate the putative assignment of protein to spots in 2D-PAGE maps. The collection of these maps known as VIRTUAL2D are accessible to the research community at <http://proteom.ncifcrf.gov>.

These two attributes are not the only available parameters. Tissue-specific expression maps can be computed on the basis of data harvested from various public databases. These maps can be thought of as snapshots of the profile at different states (drug-induced, diseased, etc.). Their potential role as predictive cancer proteomics is explored.

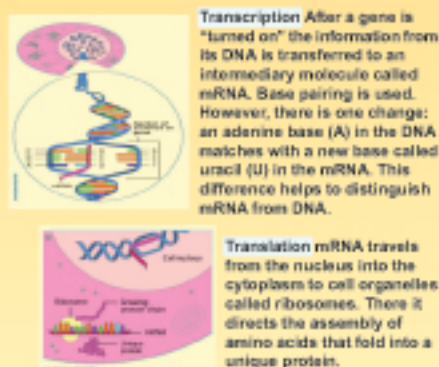
Public Databases Used to Build VIRTUAL2D



Values of amino acids pK's (determined [13] at high molar concentrations of urea) and masses used in pI/Mw computation

Ionizable Group	pK	Molecular Mass
C-terminal		
N-terminal	3.55	
Met	7.00	132.804
Thr	8.82	102.807
Ser	6.93	88.8000
Ala	7.59	72.8000
Val	7.44	100.834
Glu	7.70	130.817
Pro	8.36	96.918
Internal		
Asp	4.05	116.890
Glu	4.45	130.817
His	5.98	128.843
Cys	9.00	104.840
Tyr	10.0	164.878
Lys	10.0	144.861
Arg	12.0	157.889
C-terminal side chain groups		
Asp	4.55	116.890
Glu	4.75	130.817

From Gene to Protein: A Two-step Process



Transcription After a gene is "turned on" the information from its DNA is transferred to an intermediary molecule called mRNA. Base pairing is used. However, there is one change: an adenine base (A) in the DNA matches with a new base called uracil (U) in the mRNA. This difference helps to distinguish mRNA from DNA.

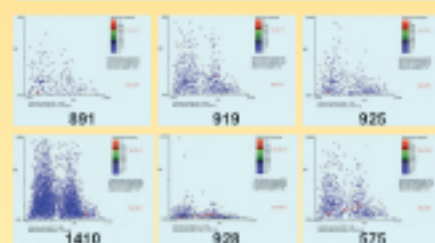
Translation mRNA travels from the nucleus into the cytoplasm to cell organelles called ribosomes. There it directs the assembly of amino acids that fold into a unique protein.

Data Contained in a Single (Colon Cancer) Library File in CGAP

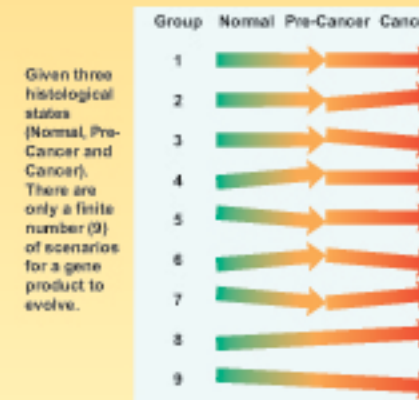
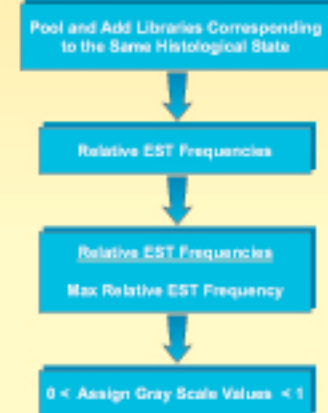


The number and frequency of detection of each EST provides the third dimension for computing relative expression and TMAPs.

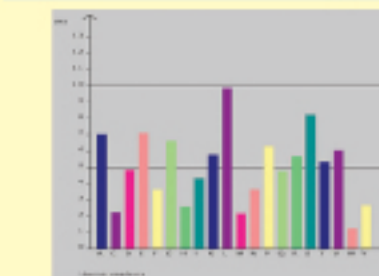
Expression Maps for Individual CGAP Prostate Libraries



Computing Expression Levels for Multiple Similar Libraries



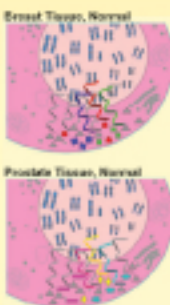
Amino Acid Distribution in Homo Sapiens



On the fly interaction and identification. By using the controls, one can zoom in on a particular area. Simply moving the mouse over or clicking on any spot will either display a short description or bring up comprehensive information from the hyper-linked web server of choice (Protoplot uses java code modified from MicroArray Explorer).

Expression Profiles (<http://proteom.ncifcrf.gov>)

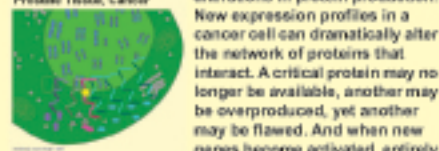
In any type of tissue, only a limited set of the genes are switched on at any given time. Also, from one tissue type to another, the limited set of genes involved will vary. Thus, each tissue can be identified by its unique pattern of gene expression. This pattern is often called an "expression profile" or a "molecular signature".



Here is an example of a normal breast cell and a normal prostate cell. Although both of these cells have many mRNAs and proteins in common (grey), they also have unique differences.

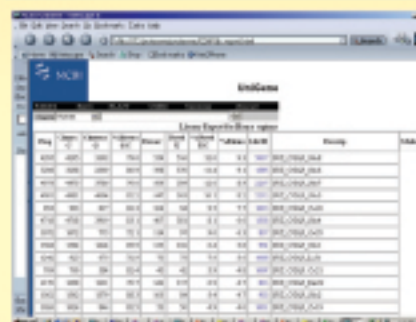
Expression Profiles in Cancer (<http://proteom.ncifcrf.gov>)

It is possible to measure differences between a normal and a cancer tissue of the same type - for example, normal and cancerous prostate. Samples of normal prostate tissue will have identical expression profiles. But when a normal prostate tissue is transformed into cancerous prostate tissue, the expression profile changes. Any changes in gene expression ultimately cause alterations in protein production. New expression profiles in a cancer cell can dramatically alter the network of proteins that interact. A critical protein may no longer be available, another may be overproduced, yet another may be flawed. And when new genes become activated, entirely new proteins may be introduced.



Many different combinations of gene changes and protein interactions are seen in cancerous tissue.

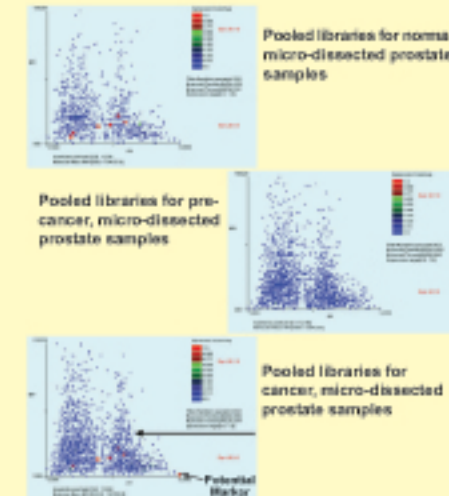
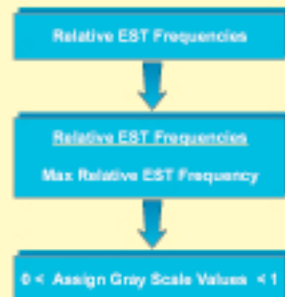
CGAP Libraries within UNIGENE



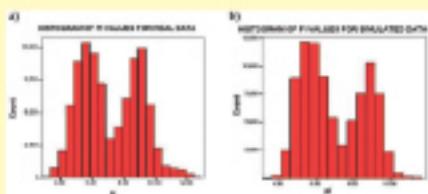
CGAP Prostate Libraries from Micro-dissected Samples

Normal CGAP_ID (UNIGENE_ID)	Pre-Cancer CGAP_ID (UNIGENE_ID)	Cancer CGAP_ID (UNIGENE_ID)
NCL_CGAP_P12(1011)	NCL_CGAP_P12(282)	NCL_CGAP(378)
NCL_CGAP_P12(1083)	NCL_CGAP_P12(311)	NCL_CGAP(382)
NCL_CGAP_P12(1419)	NCL_CGAP_P12(338)	NCL_CGAP(318)
NCL_CGAP_P1(1071)	NCL_CGAP_P1(538)	NCL_CGAP(336)
NCL_CGAP_P1(981)	NCL_CGAP_P1(1048)	NCL_CGAP(386)
NCL_CGAP_P12(319)		NCL_CGAP(1016)
NCL_CGAP_P12(324)		
NCL_CGAP_P1(910)		
NCL_CGAP_P1(1008)		

Computing Expression Levels for Single Library



Each of these spots is hyper-linked to a Pre-selected server allowing on the fly identification

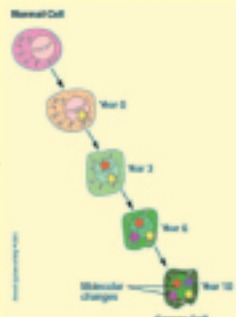


a) Histogram of pI "real" values extracted from the latest combined release of SWISS-PROT/EMBL, for Homo-sapiens. The data points are grouped in bins that are 0.25 pI unit-wide. b) Histogram of pI values, simulated data for each group, 1543 sequences are randomly simulated ranging from 50 to 600 amino acids in length, in increments of 10.

Cancer Genome Anatomy Project (CGAP)

Cancer develops when a cell undergoes several molecular changes. These changes may take place over an extended period of time, and only after the cell has accumulated a critical number of these changes does it become cancerous.

In 1996, the Cancer Genome Anatomy Project (CGAP) began. This program of the U.S. National Cancer Institute (NCI) focuses on cataloging the gradual molecular changes that occur throughout this transformation process.



Workflow for Building Expression Maps



TMAP (Tissue Molecular Anatomy Project)

- BLOOD
- BRAIN
- BREAST
- COLON
- HEAD & NECK
- HEART
- KIDNEY
- LIVER
- LUNG
- OVARY
- PANCREAS
- PROSTATE
- SKIN
- UTERUS

Conclusions

Tissue specific maps have been constructed from public databases.

Varying expression profiles may reveal (Cancer) histology specific panels of markers.

Statistical analysis is currently underway to assert the significance of these markers.

This sort of predictive study is meant to be the starting point for further investigations.

Correspondence: Dr. Djamel Medjahed, National Cancer Institute at Frederick, P.O. Box B, Frederick, Maryland 21702-1201 USA. E-mail: medjahed@ncifcrf.gov

Djamel Medjahed
SAIC-Frederick, Inc.
Contract NO1-CO-1260
Funded by the National Cancer Institute