Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma

Iizuka, N., et al. *FEBS Letters*, 2005. **579**(5): p. 1089.

Erin Bridgeford

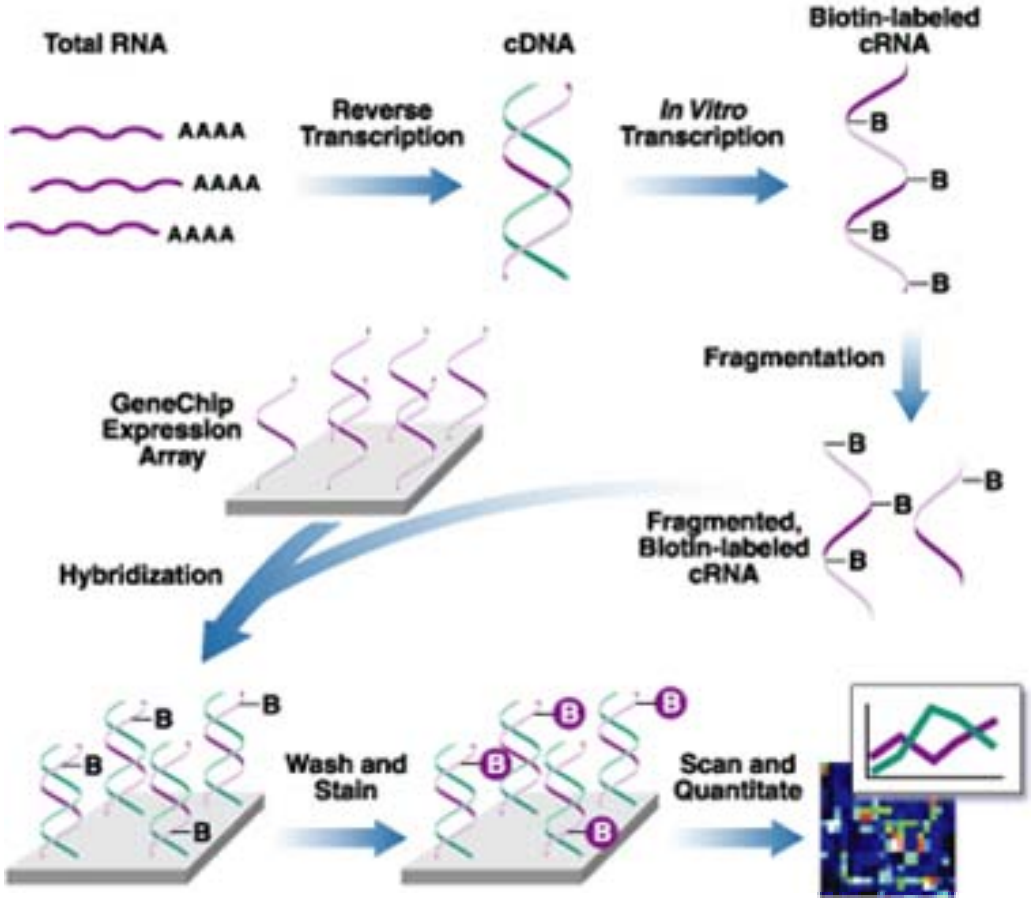
Nancy Guillén

Microarrays to investigate problems in cell biology

- Data from transcription state of the cell under certain conditions
- Each experiment produces lots of data
- Finding single change in gene expression
- Look at overall patterns of gene expression
- Hypothesis driven vs. fishing expedition

Image removed due to copyright reasons.

GeneChip® Microarray



Courtesy of Affymetrix. Used with permission.

- Probe is 25-mer oligonucleotide for high specificity
- Multiple probes for each expression or genotype measurement
- Optimized probe set

Analysis of gene expression profiling data

- Due to high volume of data, systematic methods for organization are required to convert data into a manageable set
- Strategies grouped in two categories:
  - Discrimination or supervised learning
  - Clustering or unsupervised learning (k-means, self-organizing-maps)
- Underlying biological phenomena might get lost in abstraction

Self-organizing maps for clustering of expression data

- SOM is a similarity graph, and a clustering diagram
- Converts complex, nonlinear statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display.
- SOM has a series of partitions with a predefined geometrical configuration and, initially, their reference vectors are random
- Genes or samples are mapped to the relevant partitions, depending on which reference vector they are most similar to
- Demo

Gene expression profiles in hepatocellular carcinoma

- Microarray studies aimed at translating molecular information into clinical practice
- Studies for breast cancer and large-B-cell lymphoma
- Studies generally include cohort of patients followed for years after treatment
- Link gene clusters with good or poor prognosis (survival, recurrence)
- HCC outcome complicated by the fact that cirrhosis (pre-neoplastic) compromises liver functionality
- Heterogeneous nature of human HCC
- Results have to provide rationale for a molecular classification of the tumor to be able to predict outcomes and guide treatments

Hepatocellular carcinoma and hepatitis B (HBV) and C (HCV) viruses

- Mutagenic effect of virus
- Chronic inflammation and disease leads to malignant neoplastic event
- Molecular basis not well understood

What are they trying to do with this?

- Goal

  Understand the relation between development and dedifferentiation of HCC

- Hypothesis

  Disease progression: chronic HCV infection → well differentiated HCC → moderately differentiated HCC → poorly differentiated HCC

- Approach

  Perform a comprehensive analysis of gene expression levels and identify discriminatory genes for each stage to elucidate the molecular basis of HCC using a global picture of expression patterns

Materials and Methods: Sample Selection

Samples taken from 76 HCC patients

- 50 seropositive for HCVAb
- 26 seronegative for HCV
- All seronegative for HBV surface antigen
- Histopathology on HCV+ samples
  - 7 well-differentiated HCC (group G1)
  - 35 moderately-differentiated HCC (G2)
  - 8 poorly-differentiated (G3)

Control Groups

- Two control groups:
- Group L0 comprised of 6 nontumorous, histologically normal liver samples from patients with benign or metastatic liver tumors
  - 1 focal nodular hyperplasia
  - 2 hemangiomas
  - 3 metastatic tumors (2 from colon cancer, 1 gastric)
  - All seronegative for HCV Ab and HBVsAg
- Group L1: Five HCV-infected nontumorous samples from 5 HCC patients
  - Two chronic hepatitis
  - 3 liver cirrhosis
- Concerns in sample selection:
  - No normal samples of liver as baseline
  - No samples from HCV+ patients without HCC

Materials and Methods: DNA Microanalysis

- Resected specimens divided in two groups:
  - One frozen immediately after surgery for later RNA extraction
  - One preserved in 10% formaldehyde and embedded in paraffin
    - Used to demonstrate that non-necrotic tissues were source of RNA
- RNA extraction performed
- Quality control of RNA:
  - Look for genomic DNA contamination
  - Check for RNA decay by agarose gel electrophoresis
    - If ratio of 28S/18S rRNA is around 2.0, suggests RNA had not decayed before or during extraction
    - Reduced 28S/18S ratios indicate poor quality RNA

Materials and Methods: Microarray Analysis

- Synthesis of cDNA and cRNA (see Iizuka et al, Cancer Research 62, 2002)
- Oligonucleotide microarray screening
  - huU95A DNA Chips (12,600 probes that correspond to 8900 named genes) for initial screen

Image removed due to copyright reasons.

Materials and Methods: Gene Selection

- At first pass, 3559 genes selected
  - Expression levels were greater than 40 arbitrary units (arbitrary units = intensity/brightness of probed spot over brightness of local background)
- Fisher ratio applied to evaluate which genes could help discriminate among the groups:
  - Measures the difference between two means normalized by the average variance (ie, estimates signal-to-noise ratio).
  - Larger Fisher ratio suggests a stronger likelihood for a gene's ability to discriminate between groups.

Gene selection cont'd

- Random permutation test perfomed to validate Fisher ratio :
  - Looks to find undesired structure in random data.
  - If original result is due to chance, then randomly relabelling data should achieve similar ratios
  - Genes with $P<0.005$ were selected
- Different numbers of genes for each group deemed discriminatory:
  - L0 to L1: 152 genes
  - L1 to G1: 191 genes
  - G1 to G2: 54 genes
  - G2 to G3: 40 genes

Materials and Methods: Identifying Discriminatory Genes

- Percentage of genes identified by chance (false discovery rate) calculated
- Ratio of false positives/total positives
- A high FDR value can still be meaningful

Materials and Methods: Comparing Classes

- For class comparison, minimum distance classifier designed with top 40 genes from each class:
  - Finds centers of classes and measures between those centers and a test image's center
- Self-organizing map
  - Algorithm used for clustering data
  - Provides visualization of multi-dimensional data

Materials and Methods: Some Concerns

- No indication that laser capture microdissection (LCM) or any more precise method of tissue selection was used: analyzing stroma and vasculature as well
- Technology does not always generate reproducible or consistent results even with optimized samples.
  - Per Stearns, “The current state of the art provides 5−10% variation in signal intensities among replicate array elements on the same microarray, and 10−30% variation among corresponding array elements on different microarrays.[1]”

[1]Stears et al, Trends in Microarray Analysis, Nature Medicine, 9 (140-145), 2003

Materials and Methods: Concerns, cont'd

- Several manipulations of data required to estimate and select genes of interest
- Each step can introduce assumptions/bias of authors in selection
- May select out biologically relevant data
- Assumption of certain percentage of false positives

Table 1: Clinicopathologic characteristics of 50 HCV-positive HCCs

Factors	Well (G1)*	Moderately (G2)*	Poorly (G3)*	P-value
Sex				P = 0.8007
Male	4	24	6	
Female	3	11	2	
Age (year)**	65.3 ± 7.0	65.4 ± 7.1	67.2 ± 9.5	P = 0.9612 (G1 vs. G2) P = 0.6595 (G1 vs. G3) P = 0.5406 (G2 vs. G3)
Primary lesion				P = 0.0568
Single tumor	6	15	2	
Multiple tumors	1	20	6	
Capsule formation				P = 0.3339
Present	4	29	6	
Absent	3	6	2	
Tumor size (cm)**	2.0 ± 0.8	5.0 ± 3.2	6.0 ± 7.0	P = 0.0007 (G1 vs. G2) P = 0.0279 (G1 vs. G3) P = 0.6397 (G2 vs. G3)
Stage*				P = 0.0656
I	6	10	2	
II	1	17	3	
IIIA/IV	0	8	3	
Microscopic venous invasion*				P = 0.0381
(-)	7	21	3	
(+)	0	14	5	
Alpha-feto protein (ng/ml)				P = 0.1504
< or = 100	6	24	3	
> 100	1	11	5	
Non-tumorous liver				P = 0.7569
Normal or chronic hepatitis	2	15	2	
Liver cirrhosis	5	20	6	

Fisher's exact test, Student's *t* test and Mann-Whitney's *U* test were used to elucidate differences in backgrounds between each group.
* Tumor differentiation, stage, and microscopic venous invasion were determined on the basis of TNM classification of UICC. G1-G3 tumors are equal to types I-III of Edmondson and Steiner classification, respectively.
** Mean ± S.D.

Table 1: Clinicopathologic characteristics

Tumor size

- Significantly larger in groups G2 and G3 compared with G1
- Significance determined by Mann-Whitney *U* test

Tumor invasiveness

- No vessel involvement in group G1
- Significantly more frequent vessel involvement in G2 and G3

Clinicopathologic characteristics, cont'

- Tumor stage
  - Tended to be more advanced from G1 to G3
    - P = 0.066 by Fisher's exact test: (borderline value?)
- Based on clinicopathologic characteristics, authors posit that HCC develops sequentially from L0 to L1 up to G3

Figure 1: Discriminatory Genes in Development of HCC

Image removed due to copyright reasons.

Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Figure 1a: Discriminatory Genes

Reading the array:

- Each row is a gene
- Each column is a group/sample
- Red is upregulated, green downregulated

Figure 1a:

- 152 differentially expressed genes
- Criteria for selection:
  - Downregulated genes: fold change of L1 vs L0 <1
    - 85 downregulated
  - Upregulated genes: fold change of L1 vs L0 >1
    - 67 upregulated

Figure 1a: L1 vs L0

Image removed due to copyright reasons.

Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Table 2: Downregulated genes from L1 compared with L0

- Dystrophin: anchors cytoskeleton to cell membrane.
  - Absence increases cell permeability
  - May get lysis, more inflammation
- Fibronectin:
  - Tissue repair
  - Embryogenesis
  - Blood clotting
  - Cell migration/adhesion
- Many genes with unknown functions

Fisher Ratio	GB Number	Description	Symbol	Locus	Function
Eighteen genes downregulated in L1 in comparison with L0					
50.45	M18533	Dystrophin	DMD	Xp21.2	Cytoskeleton
23.02	AF035316	Homolog to tubulin beta chain		6p24.3	Unknown
20.65	AL049942	Zinc finger protein 337	ZNF337	20p11.1	Unknown
18.34	L27479	Friedreich ataxia region gene X123	X123	9q13-q21	Unknown
16.63		Fibronectin (Alt. Splice 1)			Extracellular matrix
16.13	U19765	Zinc finger protein 9	ZNF9	3q21	Transcription/retroviral nucleic acid binding protein
14.91	X55503	Metallothionein IV	MTIV	16q13	Detoxification
13.71	AL046394	Poly(rC) binding protein 3	PCBP3	21q22.3	RNA-binding protein/ post-transcriptional control
12.56	AB007886	KIAA0426 gene product	KIAA0426	6p22.2-p21.3	Unknown
12.41	AL050139	Hypothetical protein FLJ13910	FLJ13910	2p11.1	Unknown

Top-40 Discriminatory Genes in L0 and L1

22 genes upregulated in L1 compared with L0

- Many are inflammatory in nature
  - Secondary to HCV infection?
  - Possible relationship to other oncogenic process?
    - Upregulation of Ras suspcious?
  - Normal liver baseline or HCV-infected liver only would be useful comparison

Table 2

40.49	AI362017	Cystatin C	CST3	20p11.21	Cysteine protease inhibitor
21.66	L13977	Prolylcarboxypeptidase (angiotensinase C)	PRCP	11q14	Metabolism/lysosome-related protein
20.59	D32053	Lysyl-tRNA synthetase	KARS	16q23-q24	Protein biosynthesis
13.70	AF038962	Voltage-dependent anion channel 3	VDAC3	8p11.2	Transport of adenine nucleotides
11.90	AL008726	Protective protein for beta-galactosidase (cathepsin A)	PPGB	20q13.1	Lysosomal protein/ enzyme activator
11.71	J03909	Interferon, gamma-inducible protein 30	IFI30	19p13.1	Lysosomal thiol reductase/IFN-inducible
11.32	Z69043	Signal sequence receptor, delta	SSR4	Xq28	Translocatation of newly synthesized polypeptides
11.17	AL080080	Thioredoxin-related transmembrane protein	TXNDC	14q21.3	Redox reaction
11.15	M63138	Cathepsin D	CTSD	11p15.5	Lysosomal aspartyl protease/proteolysis
11.12	L09159	Ras homolog gene family, member A	ARHA	3p21.3	Oncogenesis/actin cytoskeleton

Twenty-two Genes Upregulated in L1 in Comparison with L0

Figure 1b: Differential Expression of G1 compared with L1

Figure 1b: L1 to G1

L1 to G1: 191 genes differentially expressed

- 95 upregulated in G1
  - Types include signal transduction, transcription, and RNA processing,
  - ATOX1 increase noted in previous study with HCV-related HCC
- 96 downregulated in G1
  - Includes tumor suppressor/apoptotic genes (BCL2, IGFB3),
  - Cell proliferation genes (FOS and IGFBP4 – may also be associated with apoptosis)

Image removed due to copyright reasons.

Please see:

Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Figure 1c: G1 to G2

G1 considered well-differentiated, G2 moderately differentiated

- 54 genes differentially-expressed
- 25 genes upregulated in G2
  - Many related to protein modification, transcription, and translation

- 15 genes downregulated in G2
  - Many IFN-related genes
    - OAS2 (antiviral protein)
    - STAT1 (transcription pathway)
    - PSME1 (proteolysis)
    - Suggestive of decreased immune response
    - Earlier paper noted IFN-inducible genes in HCV-related HCC but not HBV-related HCC

Image removed due to copyright reasons.

Please see:

Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Figure 1d: G2 to G3

G3 considered poorly differentiated

- More vascular invasion, larger tumor size than G1

- 40 genes differentially expressed
- 10 genes upregulated in G3
  - LGALS9 (galectin; associated with cell adhesion, growth regulation, apoptosis, metastasis)
  - TGFB1 (may trigger invasiveness of HCC cells via integrin)
- 30 genes downregulated in G3
  - SDCI (cell adhesion, metastasis)
    - Another study found decreased levels found in HCC with high metastatic potential)

Image removed due to copyright reasons.
Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Figures 1e-h: 40 most discriminatory genes for each transition

Image removed due to copyright reasons.

Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Most discriminatory genes determined by looking at greatest differential expression (highest Fisher ratios) from one transition to another (L0 to L1, etc)

Figure 1e-h

- Almost no overlap between discriminatory genes in all groups: 17 out of 437 (0.39%)
- False discovery rate (FDR) (i.e., genes identified by chance) of all groups were all extremely low
  - L0 vs L1: FDR of 0%
  - L1 vs G1: 0%
  - G1 vs G2: 0.24%
  - G2 vs G3: 0.29%
- Overall trend towards smaller numbers of significant genes as identified by this technology as the cancer becomes more advanced (i.e., dedifferentiation continues)
  - Possible significance of this?

Significance of selected genes

To verify significance of selected genes, authors constructed the Minimum Distance classifier

- Quick recap: The **minimum distance classifier** finds centers of classes and measures between those centers and a test image's center. The distance is defined as an index of similarity so that the minimum distance is identical to the maximum similarity.

Figure 2

Authors' classification results:

a) 92% accuracy

b) 98% accuracy

c) 84% accurad) 100% accuracy

Image removed due to copyright reasons.

Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

- High accuracy reported between classes
- Are results/authors' conclusions of pre- and post-transition discrimination and grouping of molecular signatures reasonable?

Arrangements of samples by SOM

- 61 samples mapped according to expression levels of the top 40 genes for each transition (total = 160 genes)
- G2 tumors classified into two subtypes:
  - Without venous invasion
  - With venous invasion
- Tumor size assigned to samples
- p53 abnormality data applied to 22 of the HCC samples

Figure 3: Visualization of sample arrangement by SOM: development and classification

Image removed due to copyright reasons.

Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Figure 3 results

- 8x5 cells on hexagonal grid – 40 clusters
- Clusters showed a sigmoidal curve in the order L0, L1, G1, G2, G3
- G2 w/o venous invasion closer to G1
- G2 w/ venous invasion closer to G3

Figure 4: Tumor size and p53 status

Image removed due to copyright reasons.

Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Figure 4 results

- Tumor size not always consistent with differentiation state
- HCCs become progressively less differentiated as they enlarge
- HCCs with WT-p53 located within or close to G1 clusters
- Most HCCs with mutant p53 located at most distant points from L0, L1 and G1 clusters
- Genetic abnormality of p53 is a feature of late stage HCC

Arrangement of HCV - / HCC samples by SOM

Image removed due to copyright reasons.

Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Figure S3 results

- SOM for HCV –/ HCCs failed to arrange samples sequentially according to differentiation state
- Changes in identified discriminatory genes are specific for HCV +/ HCCs

Validation of microarray data by quantitative RT-PCR

- One discriminatory gene for each transition selected at random to validate microarray data by analysis with real time RT-PCR
  - CD74 for L0→L1
  - IGFBP3 for L1→G1
  - STAT1 for G1→G2
  - TGFB1 for G2→G3
- Abundance of each transcript calculated the as the mean copy number per 100 ng RNA for each tissue
- Data compared by Student's t test or Mann–Whitney U test and Pearson's correlation coefficient.

Validation of microarray data by quantitative RT-PCR

Image removed due to copyright reasons.

Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

Comparison of expression patterns as measured by microarray and RT-PCR

Image removed due to copyright reasons.

Please see:
Iizuka, N., et al. "Self-organizing-map-based molecular signature representing the development of hepatocellular carcinoma." *FEBS Letters* 579, no. 5 (February 14, 2005): 1089-100.

RT-PCR validation results

- Expression patterns of CD74, IGFBP3, STAT1 and TGFB1 reproduced by real time quantitative RT-PCR
- Is the data accurately reproduced ?

Conclusions

- Differential genetic expression for each stage of development with characteristic molecular signature
- No overlap for discriminatory genes for each transition
- Patterns valid for HCV+/HCC only
- Provide additional biomarkers
  - Diagnosis and treatment

Questions and concerns about the paper

- Sample and tissue concerns
  - Isolated cancer cells?
  - Lack of HCV infected tissue from non cancerous patients
- Introduced bias from statistical manipulations
- Variability and lack of reproducibility
- Choosing SOM to present data
- Occult HBV infection because of lack of data for core antigen and viral DNA

Further discussions