VI Spanish Meeting on Game Theory and Practice

Elche, 12-14 July 2004

Application of the Shapley value to microarray data analysis.

Stefano Moretti

DIMA: Mathematics Department, University of Genova and

IST: National Cancer Research Institute

Fioravante Patrone

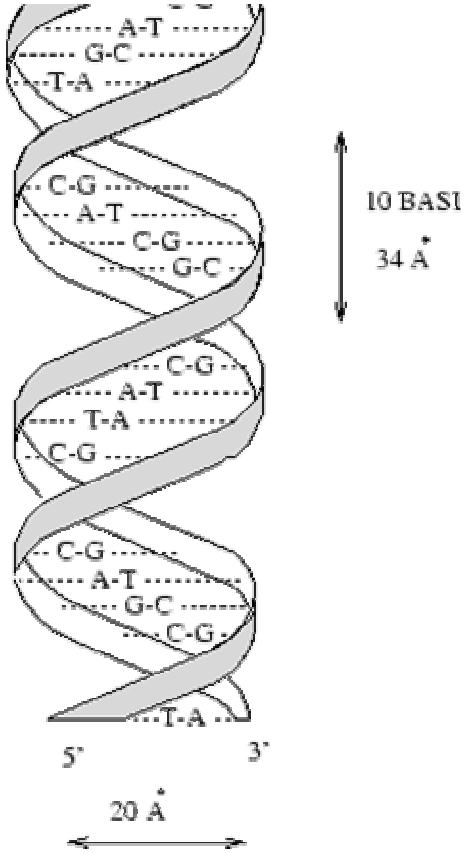
DIMA: Mathematics Department, University of Genova

N.B.

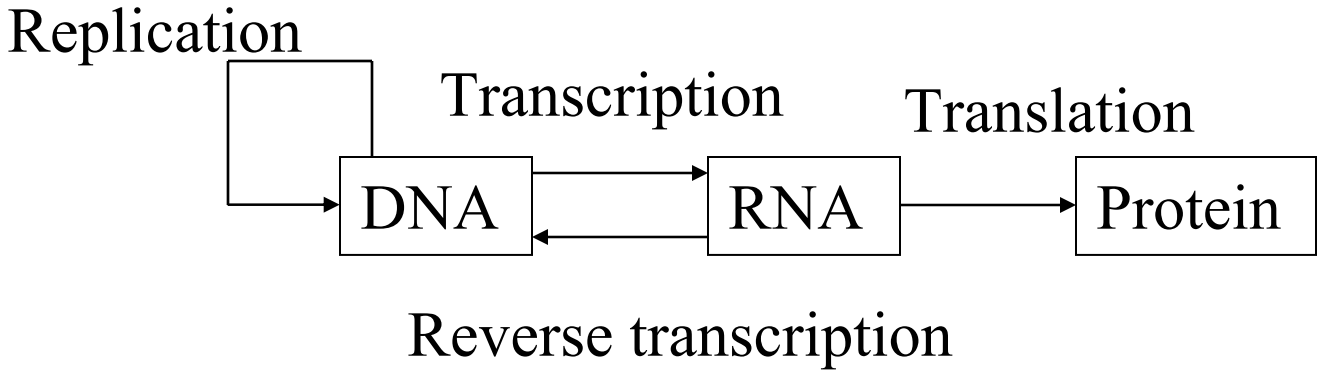
- E’ una versione “ridotta” della presentazione fatta ad Elche
- E’ stata omessa la parte tecnica in cui si individua una nuova caratterizzazione del valore Shapley, singificativa per lo specifico contesto applicativo

Plan:

- What is a “microarray” and why is it interesting?
- A game to play
- The Shapley value is of any help?
- Related developments and comments


A-T
G-C
T-A
C-G
A-T
C-G
G-C
C-G
A-T
T-A
C-G
C-G
A-T
G-C
C-G
T-A
5'
3'
10 BASI
34 Å
20 Å

Fundamental principle: hybridization



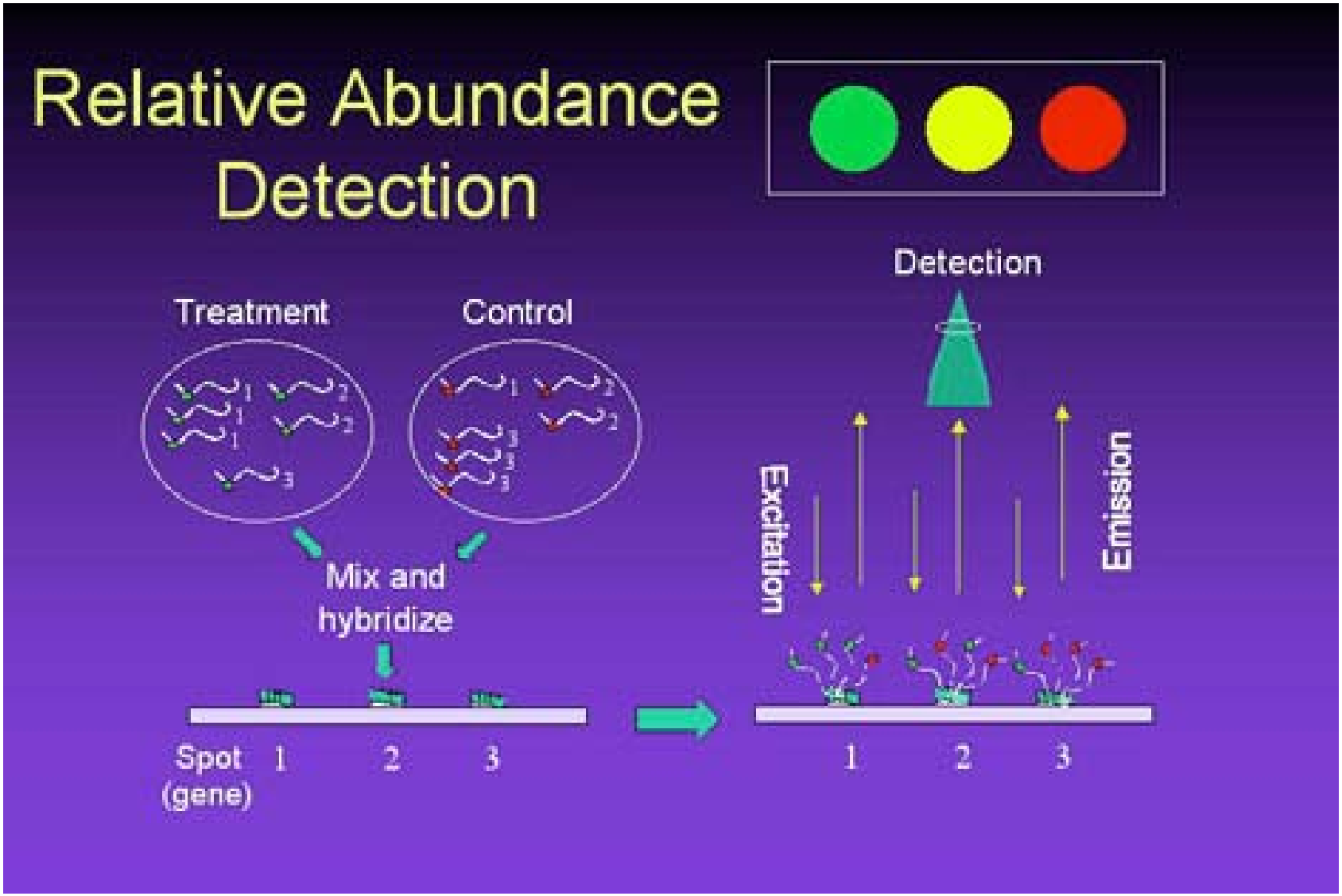
DNA ATATCGGCATCAGTCGATCGATCATCGATCGAT

↓

mRNA UAUAGCCGUAGUCAGCUAGCUAGUAGCUAGCUA

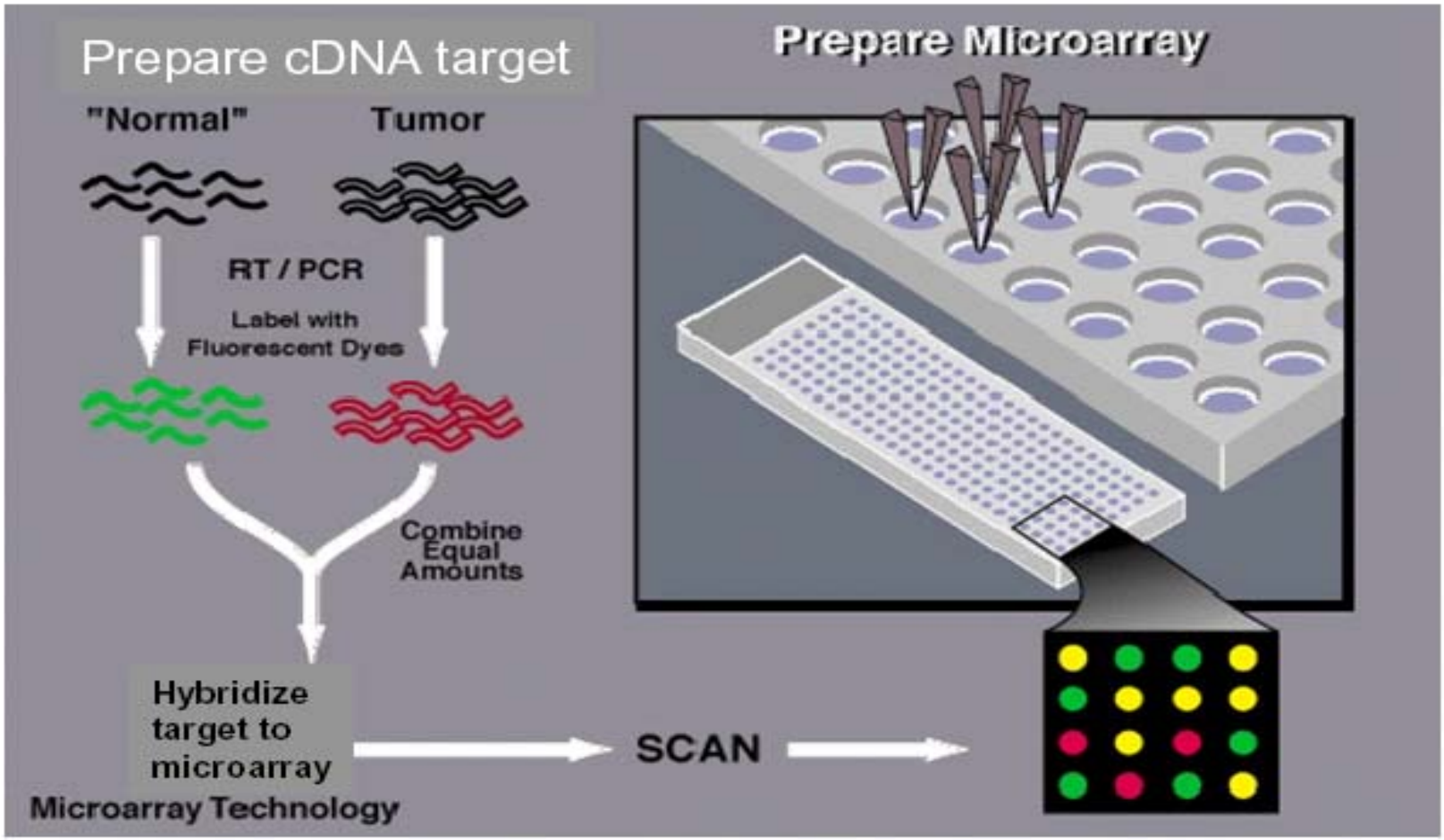
↓

cDNA ATATCGGCATCAGTCGATCGATCATCGATCGAT



(Slide source: http://www.bsi.vt.edu/)

Experiments with cDNA microarray



Institute for Systems Biology Protocol
Direct Incorporation of Cy3/Cy5 During Reverse Transcription

Written by Anne Rosenwald and Todd Eckdahl; August 2003)

Materials Needed:

Isolated and quantitated total RNA samples
Microarray slides (70-mer plus-strand oligomers)
RNase-free water
oligo dT primer (16- to 18-mer) at 1 μg/ul
Coverslips, 22 x 40mm size from Corning
100 mM DTT (dithiothreitol)
low dTTP dNTP mix (10 mM each dATP, DCTP, dGTP, **1 mM dTTP**)
Cy3-dUTP and Cy5-dUTP (1 mM each [separately])
3 M Ammonium Acetate, pH 5.2
100% Ethanol, 70% Ethanol

Superscript II Reverse Transcriptase, 5X first strand buffer
RNase A (4 mg/ml)
RNase H (2 unit/μl)

Reverse Transcription and Cy-dye Incorporation

1. In duplicate, aliquot 50 μg total RNA (one for each treatment)
    a. already checked with denaturing agarose gel
    b. quantitated with UV spectrophotometer
    c. precipitated (eg. 1/10 volume 3M NaOAc pH 5.2, 2 volumes EtOH)
2. To each tube, add 2.5 μg oligo dT
3. Adjust volume to 11 μl with DEPC-treated H_2O
4. Heat to 75°C for 10 min
5. Cool slowly to room temperature and spin down (Note: keep at RT from this point on)
6. Add the following in order:
    a. 4 μl Superscript first strand 5X buffer
    b. 2 μl DTT (100 mM)
    c. 1 μl dNTPs (10 mM each dATP, dCTP, cGTP and **1 mM dTTP**)
    d. 1 μl Cy-dye labeled dUTP (1 mM) (One gets Cy-3 dUTP and one gets Cy-5 dUTP)
    e. 1 μl Superscript Reverse Transcriptase II (200 units/ ul; make sure this is Exonuclease-free)
7. Mix gently and incubate at room temperature for 10 min
8. Incubate at 42°C for 2-3 hours (no more than ~5 hours; do not do this step overnight)
9. Heat sample to 95°C for 2 min
10. Place samples on ice, spin down (can store at –20°C at this point if necessary)

Degrade RNA

1. Make sure contents of tubes are spun down
2. Add 0.5 µl of RNase A (4 mg/ml) at room temperature (Promega)
3. Add 0.5 µl of RNase H (2 U/µl) (Fermentas) (Note: Not clear if really necessary to use both enzymes; the RNase H is fairly expensive)
4. Incubate at 37°C for 15-30 min

(Alternative RNA Degradation)

1. Add 3.5 µl 0.5 M NaOH/50 mM EDTA, then
2. Heat at 65 C, 10 minutes
3. Add 5 µl 1 M Tris 7.5

Purification (Using Qiagen PCR CleanUp Kit)

1. Add 25 µl high-quality H_2O to samples (add only 17.5 ul if alternative RNA degradation was done)
2. Add 2.7 µl 3 M Sodium Acetate, pH 5.2
3. Add 250 µl QIAquick buffer PB
4. Apply each sample to a QIAquick column (the DNA should stick to the column here)
5. Centrifuge for 30 sec at full speed
6. Take the column flow-through and replace back onto the top of the column and spin a 2^{nd} time
7. Place the flow-through back in the original tube and save in case of problems with the purification.
8. Wash with 400 µl QIAquick buffer PE, spin 30 sec at full speed and discard flow-through (your DNA remains on the column)
9. Repeat step 8, discarding flow-through
10. Spin the column briefly once more to get rid of remainder of wash solution
11. Place column in a clean, well-labeled 1.5 ml elution tube
12. Apply 30 µl buffer EB to center of column without touching the membrane
13. Wait one min, then centrifuge 1 min at full speed (gradually increase from 0 to full speed to avoid shearing off the eppendorf tube lids). Your DNA is in the flow-through this time!
14. Again apply 30 µl buffer EB to center of column without touching the membrane. DNA should be in a volume of 60 µl.

Determining the incorporation of labeled nucleotides

1. Use a 384-well spec plate (if available) so that small volumes can be used
2. Use undiluted labeled sample and buffer as the blank (~40 µl of each). Read absorbances at 260 nm, 280 nm, 550 nm for Cy3 and 650 for Cy5.
3. Calculations:

a. Extinction coefficients are 150,000 for Cy3 and 250,000 for Cy5
b. Purity: Corrected A260/A280 (want this to be ~1.8)
c. Pmol dye/ µl of sample = A(dye)/(extinction coefficient *10^6)
d. dsDNA = 50 µg/ml
e. ssDNA = 33 µg/ml
f. ssRNA = 40 µg/ml

Example:
Total dsDNA (µg) = 50 µg/ml * 1 ml/ 1000 µl * A260 * Volume (µl) * DF

4. If labeling efficiencies are fairly close for each dye and the nucleic acid yield is also similar, then about 30 pmol of dye of the sample per hybridization is a good starting point to product good intensity on most arrays. It is better to control the amounts of starting RNA/cells and have a good internal control and normalization scheme to deal with incorporation discrepancies.

Institute for Systems Biology
Direct Incorporation of Cy3/Cy5 During Reverse Transcription

Cost Analysis

Reagent	Supplier	Startup Cost	Number of Two Label Expts	Cost per Two Label Expt
oligo dT	Operon	$75 for 100 ug	20	$3.75
Superscript II	Invitrogen Life Technologies	$220 for 10,000 units	25	$8.80
Cy3–dUTP	Amersham	25 nmoles for $295	25	$11.80
Cy5-dUTP	Biosciences	25 nmoles for $304	25	$12.16
Rnase A	Promega	$212 - 1 ml, 4 mg/ml	1000	$0.22
Rnase H	Promega	$146 for 50 units	25	$5.84
QIAquick	Qiagen	50 columns for $76	25	$3.04
Totals		**$1328 for 25 Experiments**		**$45.61 per Expt**

Matrix:

Arrays

Genes		Array1	Array2	Array3	Array4	Array5	...
	Gene1	0.46	0.30	0.80	1.51	0.90	...
	Gene2	-0.10	0.49	0.24	0.06	0.46	...
	Gene3	0.15	0.74	0.04	0.10	0.20	...
	Gene4	-0.45	-1.03	-0.79	-0.56	-0.32	...
	Gene5	-0.06	1.06	1.35	1.09	-1.09	...
	...	...	...	...	...	...	

M = $\log_2$(Red intensity / Green intensity)

expression measure, e.g, RMA

Intensity rate

- **M<0**, the gene is more "*expressed*" in the control sample (marked in green)
- **M=0**, the gene is equally expressed
- **M>0**, the gene is more "*expressed*" in the tumor (or treated...) sample (marked in green)

Microarray expression data from desease samples

	s1	s2	s3
g1	4.2	20	12
g2	1.1	9.8	1.6
g3	7	2.4	6.1

Microarray expression data from normal samples

	s1	s2	s3
g1	4.1	6.3	2.7
g2	4.2	7.8	2.1
g3	5	3.5	0.5

Discretized matrix

	s1	s2	s3
g1	0	1	1
g2	1	1	1
g3	1	0	1

cutoffs

<	>
2.7	6.3
2.1	7.8
0.5	5

Microarray TU game:

- Players are genes;
- games with [0-1] characteristic function;
- on each sample:
  - -If a coalition has value **1** then that coalitions activates the disease;
  - -If a coalition has value **0** then that coalition does not activate the disease.

EXAMPLE

Microarray discr. data

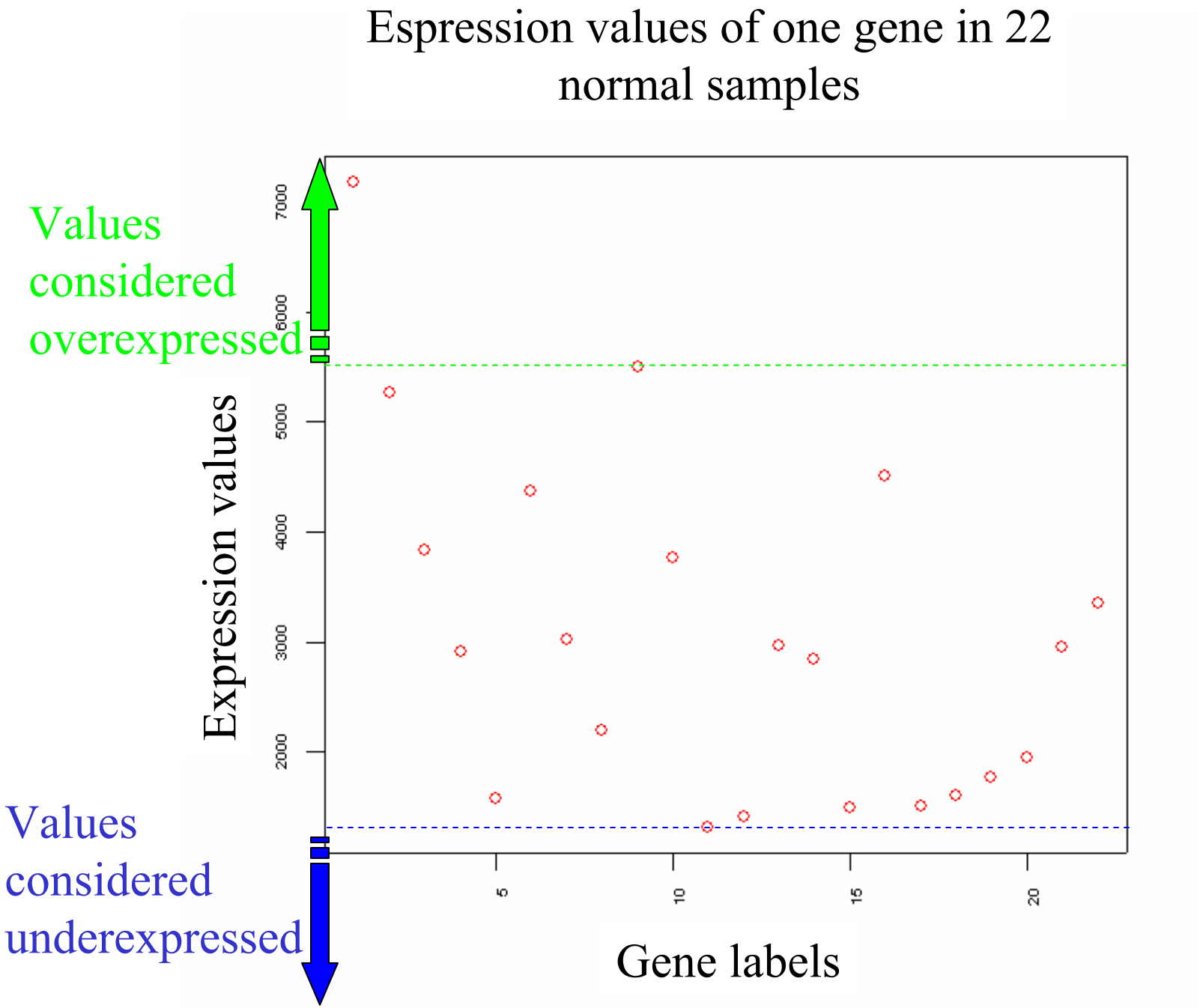
	s1	s2	s3
g1	0	1	1
g2	1	1	1
g3	1	0	1

The corresponding *[0,1]-game* <{g1,g2,g3},v>:
v({g1,g2})=v({g3,g2})=1/3
v({g1,g2,g3})=1 and
v(S)=0 for each other different coalition S.

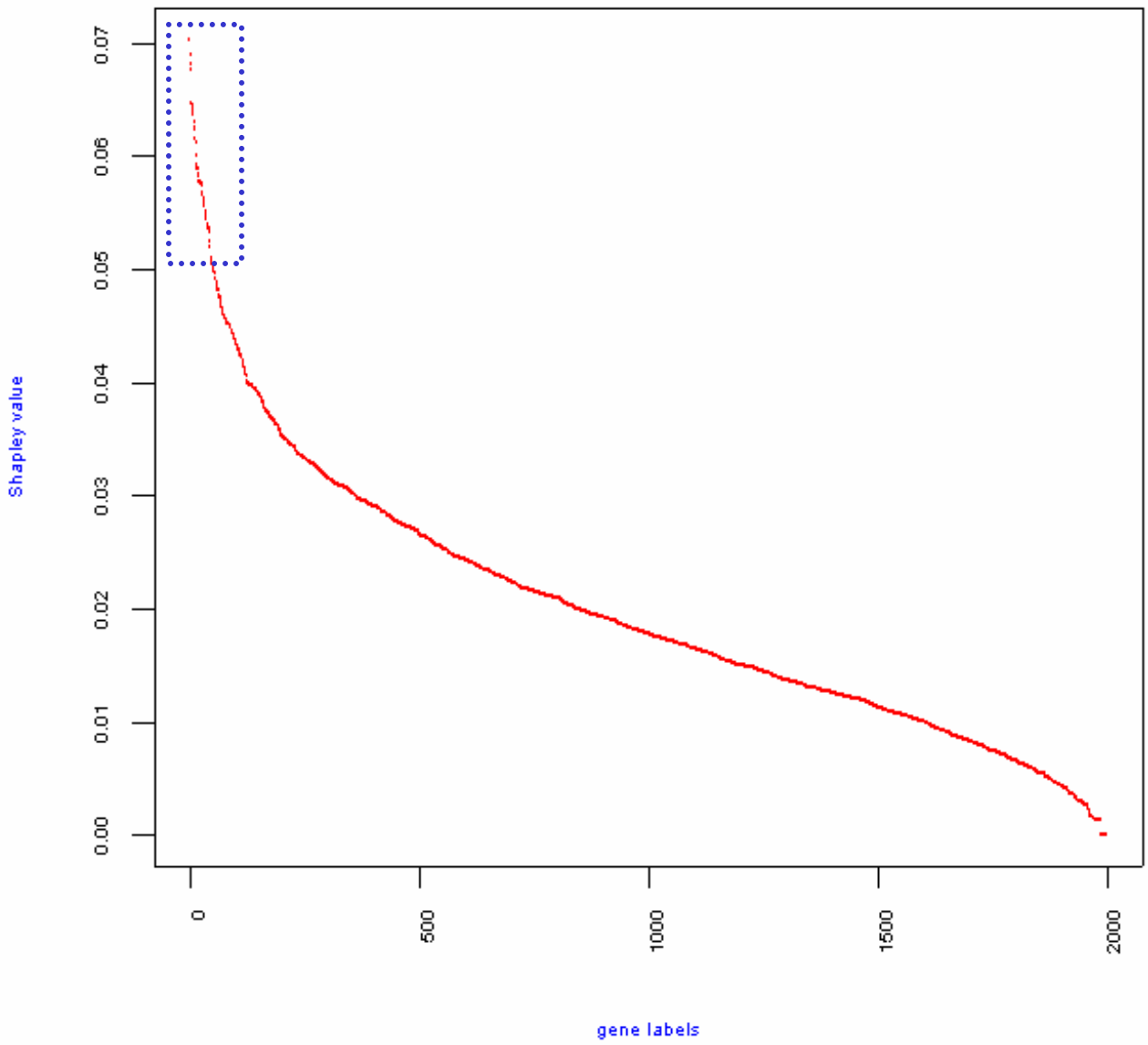
The Shapley value is: (5/18,8/18,5/18).

Application 1:
tumor versus normal

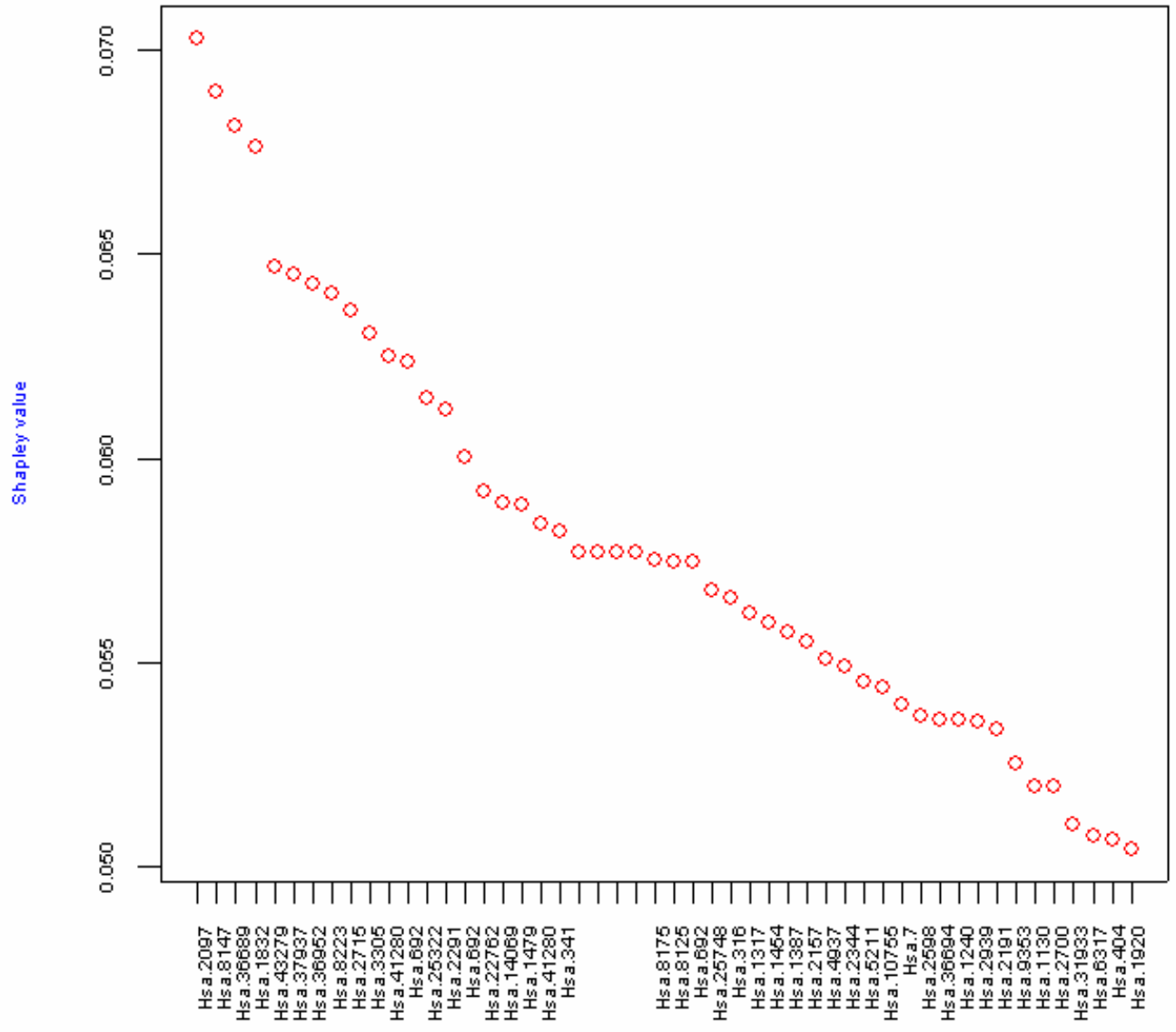
- Alon et al. (1999) were interested in identifying coregulated families of genes in *tumor* and *normal* colon tissues.
- They studied 6500 human genes.
- Genes were collected on 62 samples,
  - 40 tumor colon tissues
  - 22 normal colon tissues


Espression values of one gene in 22 normal samples
Values considered overexpressed
Values considered underexpressed
Expression values
Gene labels
7000
6000
5000
4000
3000
2000
5
10
15
20

Shapley value of 2000 genes



Shapley value of 100 genes



Group of four genes with the highest Shapley values (in decreasing order from top to down).

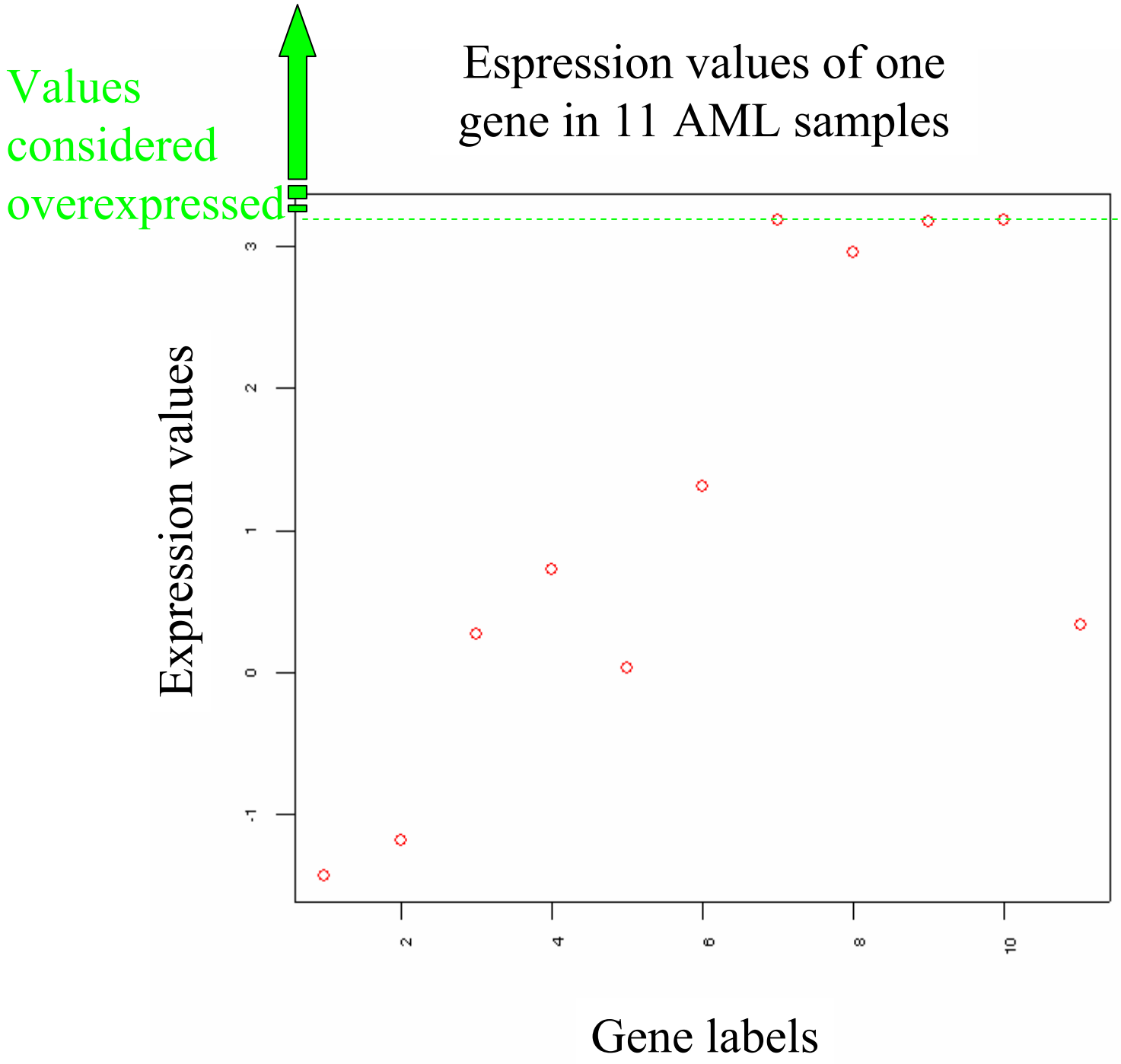
Human vasoactive intestinal peptide (VIP) mRNA, complete cds.
Human desmin gene, complete cds.
H.sapiens mRNA for GCAP-II/uroguanylin precursor
MYOSIN REGULATORY LIGHT CHAIN 2, SMOOTH MUSCLE ISOFORM

It has been suggested to promote the growth and proliferation of tumor cells (Fujarewicz & Wiench, 2003).

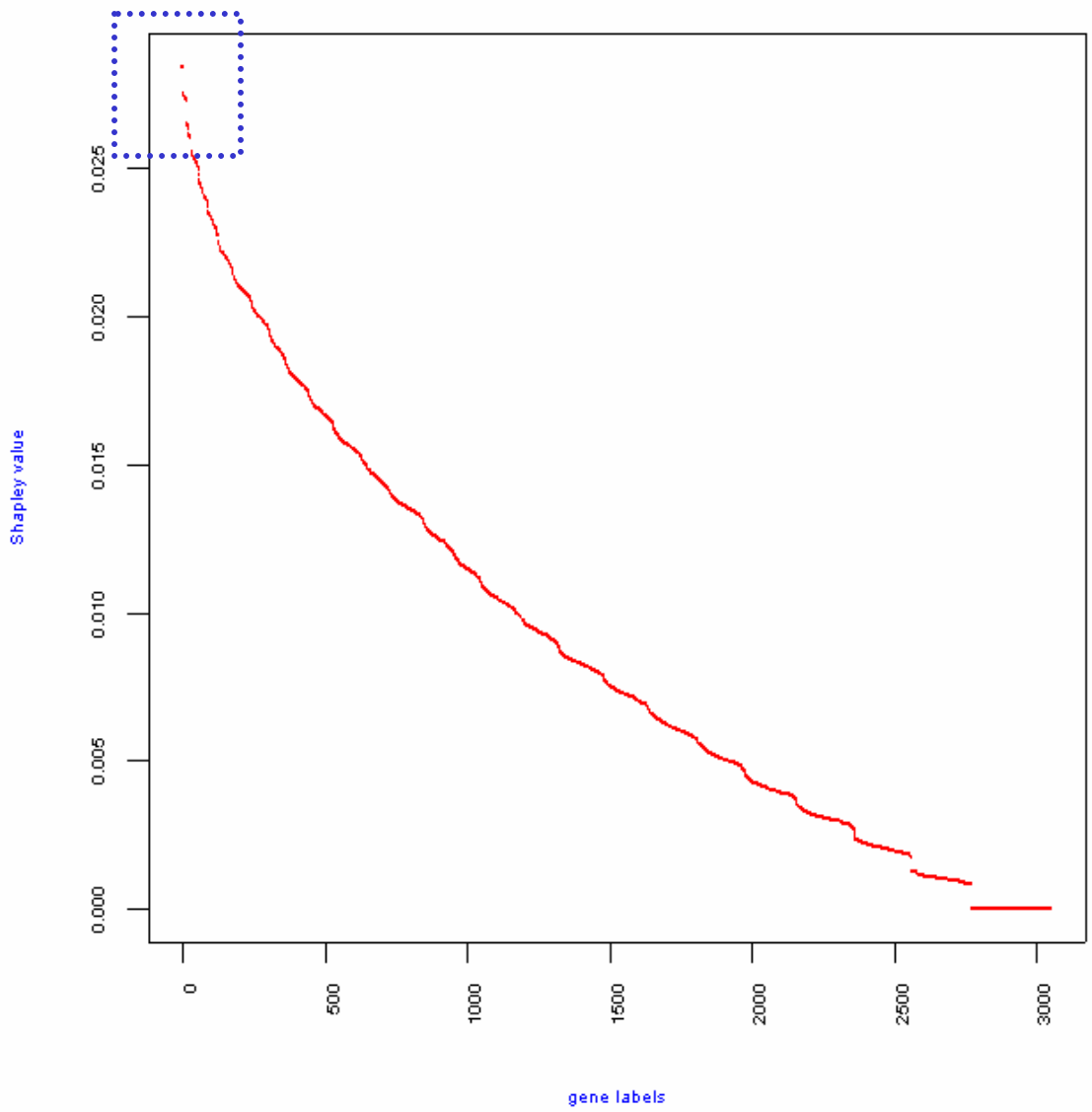
It might provide an indication propensity for metastasis of cells (Moler et al., 2000).

Application 2:
ALL versus AML

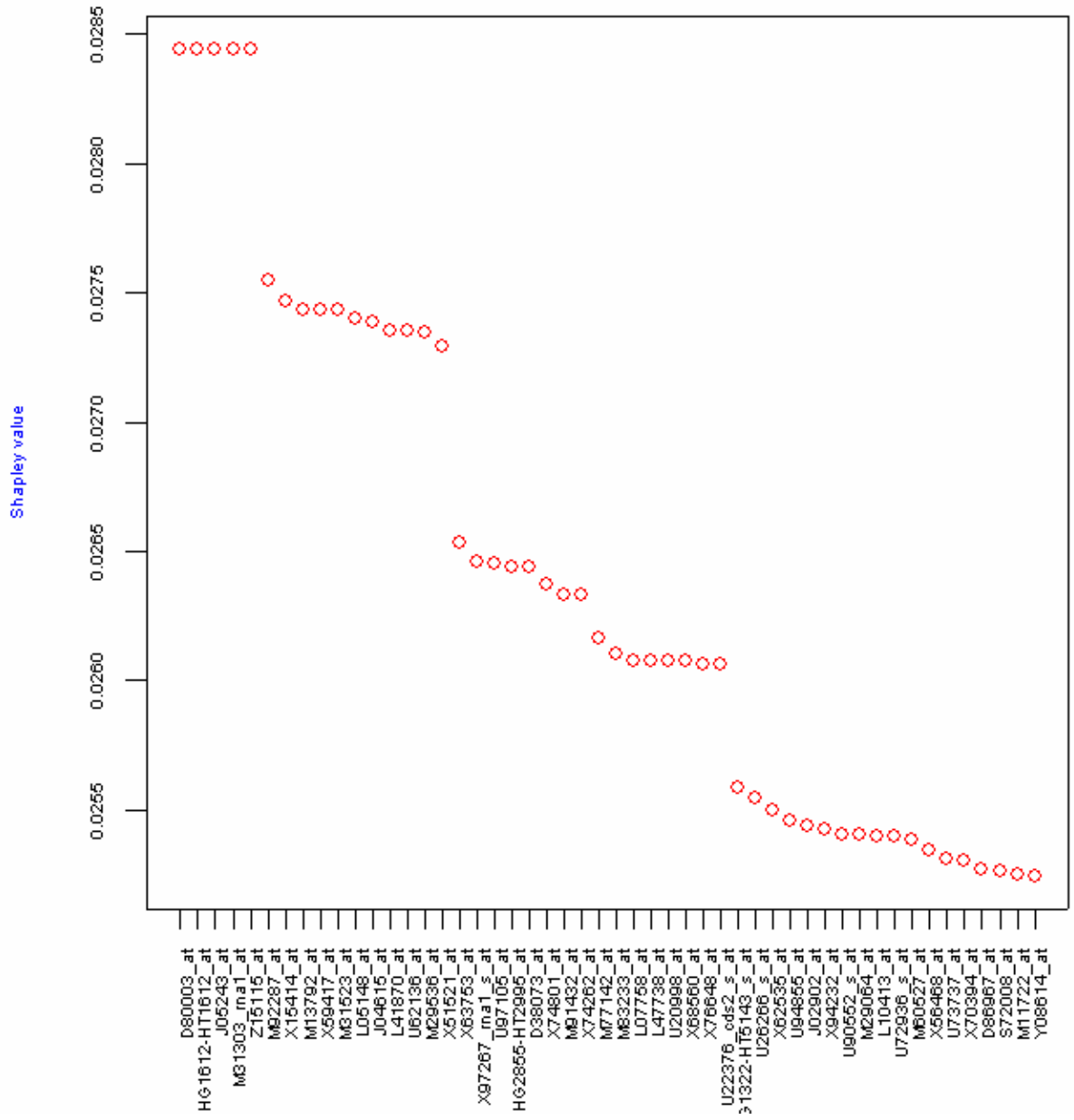
- Golub et al. (1999) were interested in identifying genes that are differentially expressed in patients with two type of leukemias, *Acute Lymphoblastic Leukemia* (ALL) and *Acute Myeloid Leukemia* (AML).
- They studied 6817 human genes.
- Genes were collected on 38 samples,
  - 27 ALL cases
  - 11 AML cases
- *We discretized the expression values of genes in ALL samples on the basis of expression values in AML samples*


Values considered overexpressed
Espression values of one gene in 11 AML samples
Expression values
3
2
1
0
-1
2
4
6
8
10
Gene labels

Shapley value of 3051 genes



Shapley value of 50 genes



Group of five genes with the same highest Shapley value.

KIAA0181 gene, partial cds
Macmarcks
Oncoprotein 18 (Op18) gene
SPTAN1 Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)
TOP2B Topoisomerase (DNA) II beta (180kD)

Encode a critical protein for the cell cycle progression related to leukemia (Glub et al., 1999)

Related references

- Kaufman, Kupiec and Ruppin: *Multi-Knockout Genetic Network Analysis: The Rad6 Example,* preprint
- Keinan, Kaufman, Sachs, Hilgetag and Ruppin: *Fair localization of function via multi-lesion analysis,* to appear on Neuroinformatics, 2004.

three comments

- Be **humble**: many approaches, a lot of knowledge around
- Be **determined**: if you have an idea, follow it to see really what it can give
- Be **critical**: a characteristic which is not as widespread as it should be

The END

Thanks for your attention