



# **IDENTIFICAZIONE DI GENI ESTROGENO RESPONSIVI CON METODI COMPUTAZIONALI**

Francesca Cordero

Fulvio Lazzarato

Raffaele A. Calogero

# Promoters analysis

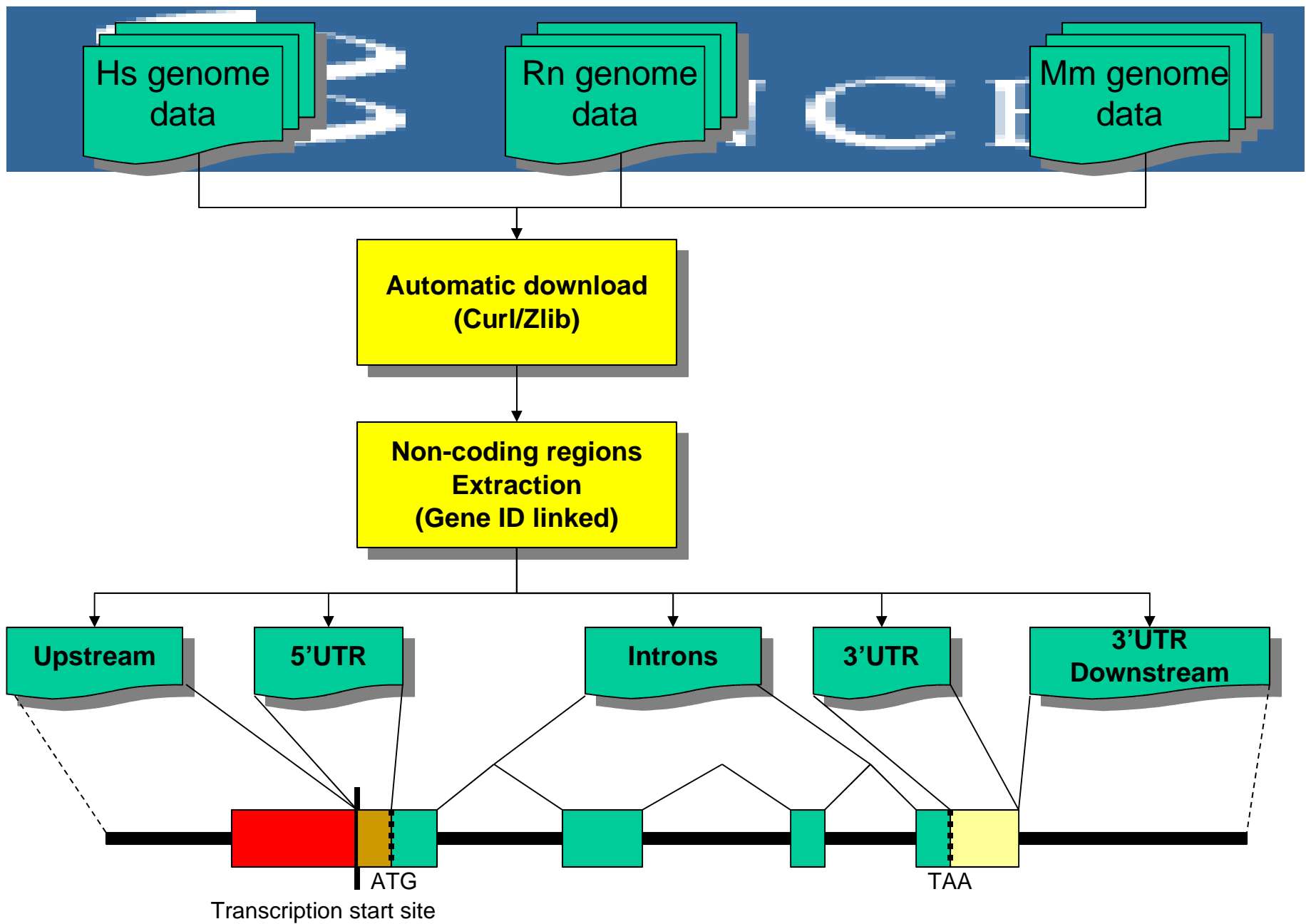
- Identification of common transcriptional elements, within co-regulated genes, might help to unmask hidden regulative mechanisms.
- Identification of **transcriptional regulative elements (TRE) clusters** linked to co-regulated genes might help to identify, in the genome, uncharacterized genes regulated by the same mechanism.

# Promoters analysis: limitations

- Computational tools are quite inefficient in the identification of eukaryotic promoters.

## Our approach:

- Relay on NCBI genomic annotations.
- Repeating the analysis as annotations are upgraded.



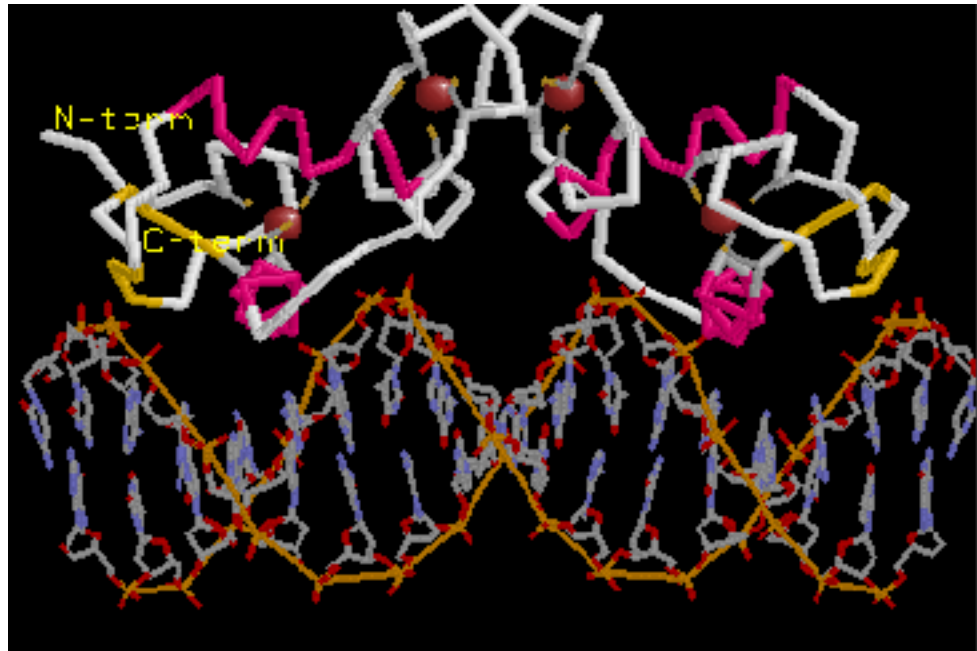
<https://www6.unito.it/RRE/EN/>

# Promoters analysis

- Transcriptional signatures might be linked to co-regulated genes.
- Transcriptional signatures might be characterized by the presence of TRE clusters.
  - The identification of TRE clusters (e.g. CISTER, CREME) relies on availability of a set of potential TREs mapped on promoters.

# TRE clusters in ERE containing genes

- The first step to search for TRE clusters is the availability of a set of genes sharing some common regulative behaviour(s).
- Our study:
  - The set we are interested is a set of genes directly controlled by the estrogens in breast cancer.



Il complesso estrogeno-ER, in forma dimerica, riconosce la sequenza ERE:

5' nnA**G**GGTCANNNT**G**ACCTnn 3'

3' nnTCCA**G**TNNNACT**G**GAnn 5'

L'interazione tra ER-DNA è principalmente guidata dai residui evidenziati in rosso.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
A	7	4	4	4	12	5	4	3	1	2	1	1	2
C	2	3	4	1	3	0	7	4	2	12	1	9	9
G	2	2	5	4	5	4	4	0	12	0	1	3	2

Count matrix

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
A	0.29	0.17	0.17	0.17	0.50	0.21	0.17	0.13	0.04	0.08	0.88	0.04	0.08
C	0.08	0.3	0.17	0.46	0.13	0.42	0.29	0.17	0.08	0.04	0.04	0.79	0.79
G	0.50	0.50	0.21	0.17	0.21	0.17	0.17	0.42	0.04	0.83	0.04	0.13	0.08

Frequency matrix

	P0	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
A	22	58	58	58	00	26	58	-1.00	-2.58	-1.58	1.81	-2.58	-1.58
C	58	-1.00	0.58	0.87	-1.00	0.74	0.22	0.58	1.58	2.58	5.85	6.60	6.61
G	00	00	26	58	26	58	58	74	81	0.74	82	02	582
T	00	26	87	26	58	-0.26	58	22	74	58	58	58	58

Log-Odds matrix

$$\log_2 \left( \frac{f_i}{e_i} \right)$$

BI

-1.22

0.22

8

A

-

A

R

T

M

P

C

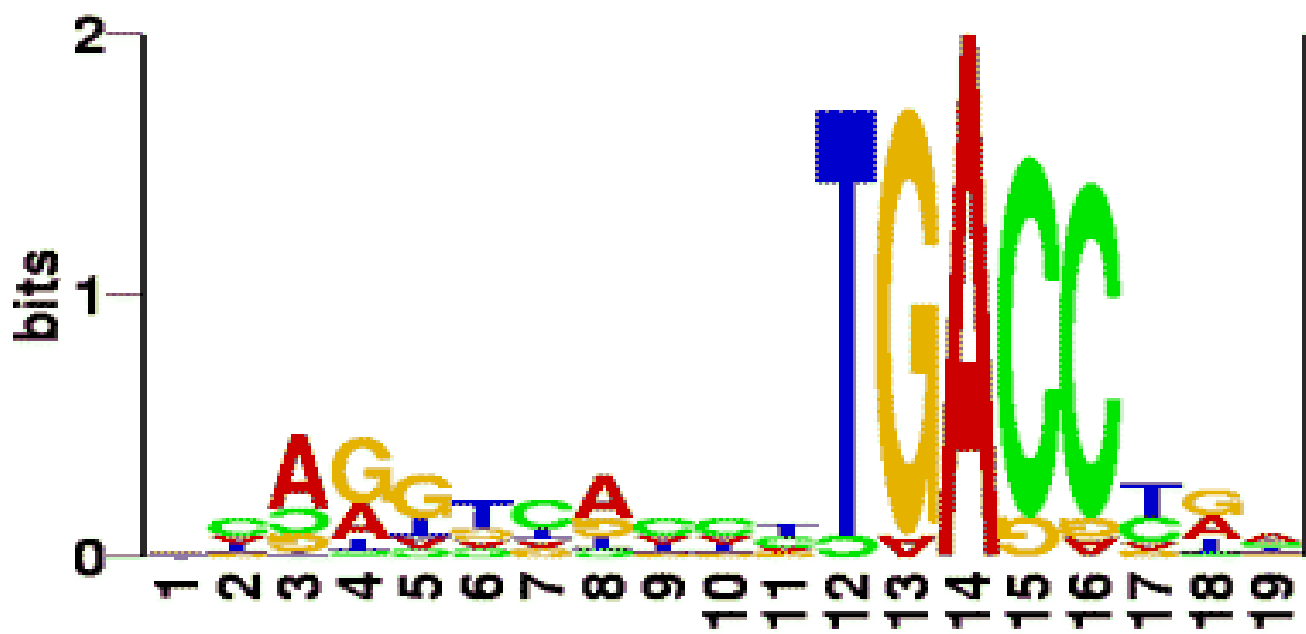
A

C

G

T

A



nnnGGTCANNNTGACCnnn  
 nnnCCAGTNNNACTGGnnn

-1.1

1.1

a

c

c

c

E' possibile migliorare la matrice  
ERE di TRANSFAC ?

# Profiling of Estrogen Up- and Down-Regulated Gene Expression in Human Breast Cancer Cells: Insights into Gene Networks and Pathways Underlying Estrogenic Control of Proliferation and Cell Phenotype

JONNA FRASOR, JEANNE M. DANES, BARRY KOMM, KEN C. N. CHANG, C. RICHARD LYTTLE, AND BENITA S. KATZENELLENBOGEN

*Department of Molecular and Integrative Physiology, University of Illinois (J.F., J.M.D., B.S.K.), Urbana, Illinois 61801; and Women's Health Research Institute, Wyeth Research (B.K., K.C.N.C., C.R.L.), Collegeville, Pennsylvania 19426*

Endocrinology. 2003 Oct;144(10):4562-74. Epub 2003 Jul 10.

146  
Frasor et al.

17

297  
Cicatiello et al.

## A genomic view of estrogen actions in human breast cancer cells by expression profiling of the hormone-responsive transcriptome

Luigi Cicatiello, Claudio Scafoglio<sup>1</sup>, Lucia Altucci<sup>1</sup>, Massimo Cancemi<sup>1</sup>, Guido Natoli<sup>1</sup>, Angelo Facchiano<sup>2</sup>, Giovanni Iazzetti<sup>3</sup>, Raffaele Calogero<sup>4</sup>, Nicoletta Biglia<sup>6</sup>, Michele De Bortoli<sup>5,7</sup>, Christian Sfiligoi<sup>7</sup>, Piero Sismondi<sup>6,7</sup>, Francesco Bresciani<sup>1</sup> and Alessandro Weisz<sup>1</sup>

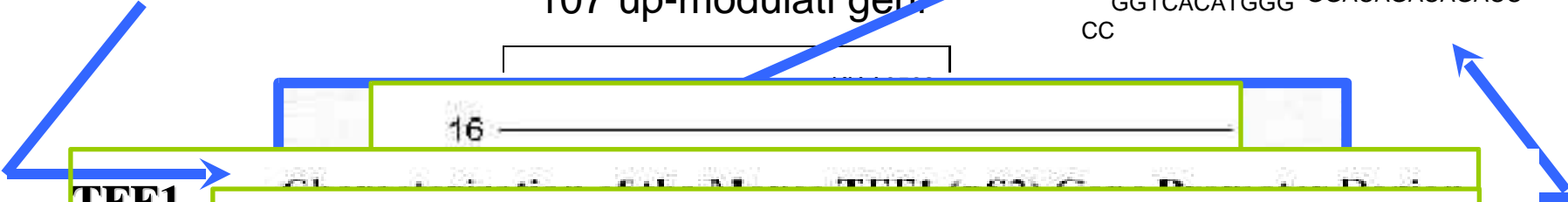
J Mol Endocrinol. 2004 Jun;32(3):719-75.

# 82 ERE in 45 geni

```

GGTACAATGA
GC GGCATTCTGA
CT GCTCAGGCTGA
GC GATGAGTCTGA
CC GGTCAGGGCCA
CC
GCTCAGAATAACC
AGTCACTGTGG
CC AGTGATTTTGAC
C AGTCACACTGA
CG GGTCACATGGG
GTGCATTCTGACC
GGACACCCAGACC
GGACATACTG
ATC
GGTCACGGTGGC
C GCCAAGATGA
CC ACCAGCCTGA
CC
GGACAGACAGACC
CC
    
```

107 up-modulati geni



TFF1

A  
C  
G  
T  
ERBB2

## Sequence Requirements for Estrogen Receptor Binding to Estrogen Response Elements\*

(Received for publication, June 5, 1998)

Mark D. Driscoll<sup>1</sup>, G. Sathya, Mesut Muyan, Carolyn M. Klinge<sup>2</sup>, Russell Hill,  
and Robert A. Bambara<sup>1</sup>

J Biol Chem. 1998 Nov 6;273(45):29321-30.

Valentina Perissi<sup>1,2</sup>, Nadia Menini<sup>2</sup>, Erika Cottone<sup>1,2</sup>, Daniela Capello<sup>2</sup>, Marco Sacco<sup>2</sup>,  
Fabrizio Montaldo<sup>2</sup> and Michele De Bertoli<sup>1,2</sup>

Oncogene. 2000 Jan 13;19(2):280-8

379 1679 1350 1090 637 577 318 58 205

ERE consensus location with respect to 1st transcribed nt

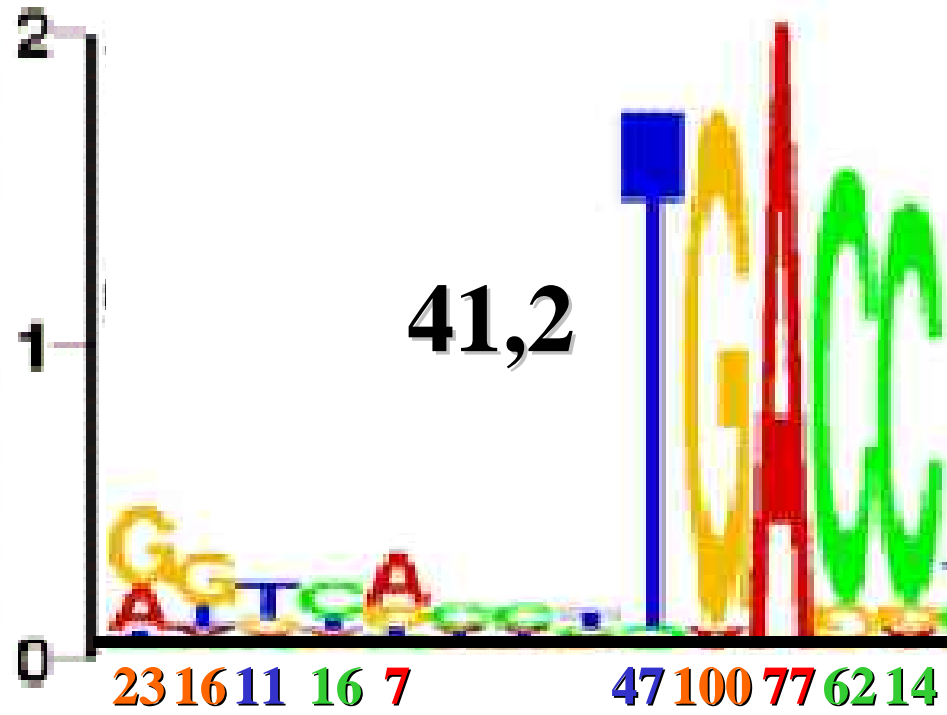
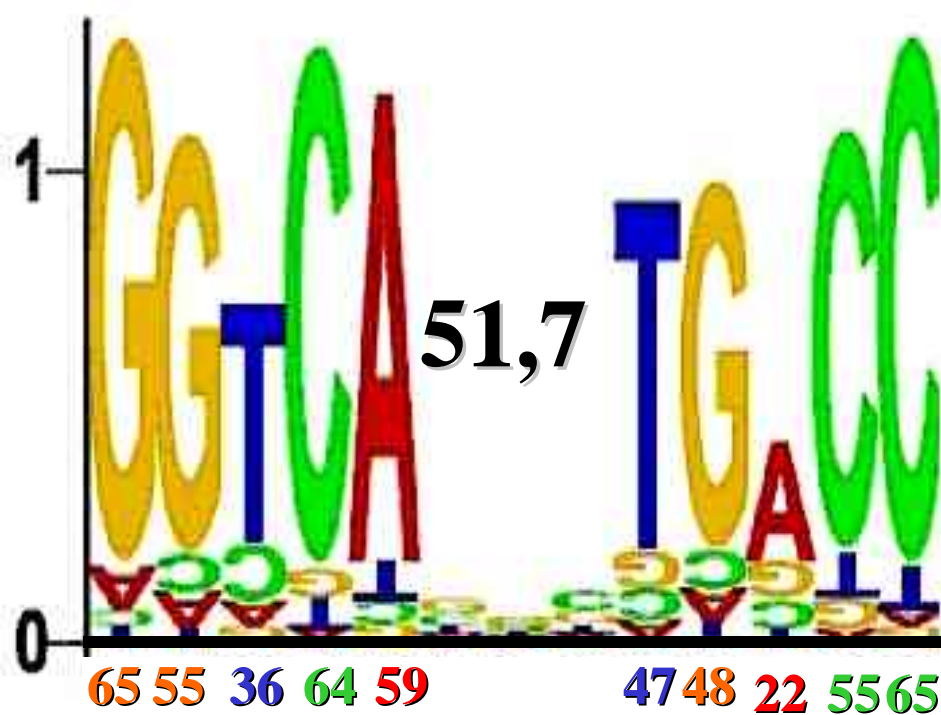
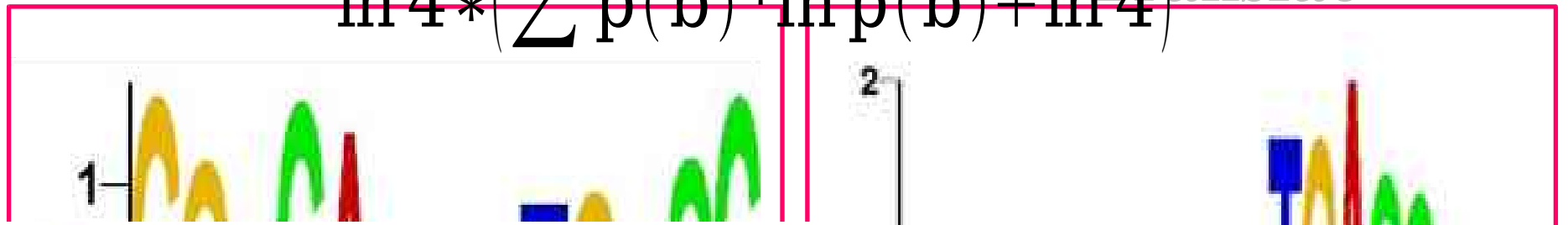
# CONTENUTO INFORMATIVO (Ci)

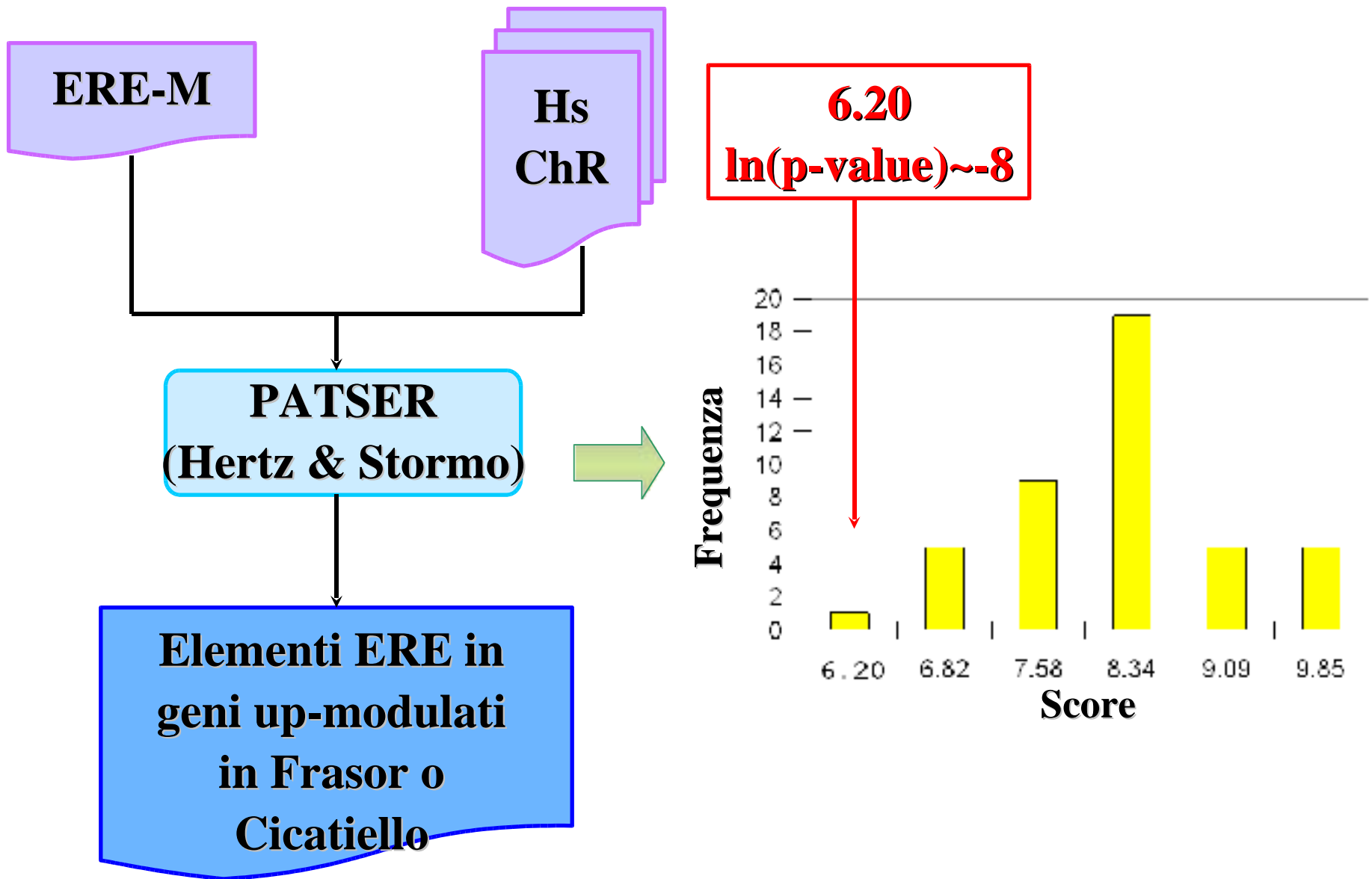
$C_i = \frac{100}{\text{ERE-M}}$

Matrice di  $b = A, T, C, G$

$$\ln 4 * \left( \sum p(b) * \ln p(b) + \ln 4 \right)$$

Transfac





- 🕒 6.20 è lo score minimo necessario per identificare ERE nelle sequenze -2000/+500 dei 45 geni utilizzati per generare la matrice ERE-M.

**ERE-M**

**Hs  
Chrs**

**Mm  
Chrs**

**Rn  
Chrs**

# rVista for Comparative Sequence-Based Discovery of Functional Transcription Factor Binding Sites

Gabriela G. Loots,<sup>1,4</sup> Ivan Ovcharenko,<sup>1</sup> Lior Pachter,<sup>2</sup> Inna Dubchak,<sup>1,3,4</sup>  
and Edward M. Rubin<sup>1</sup> Genome Res. 2002 May;12(5):832-9.

**(HERTZ & STORINO)**

**Geni conteneenti ERE**

# Breast cancer classification and prognosis based on gene expression profiles from a population-based study

Christos Sotiriou<sup>\*†</sup>, Soek-Ying Neo<sup>‡</sup>, Lisa M. McShane<sup>§</sup>, Edward L. Korn<sup>§</sup>, Philip M. Long<sup>‡</sup>, Amir Jazaeri<sup>\*</sup>, Philippe Martiat<sup>†</sup>, Steve B. Fox<sup>¶</sup>, Adrian L. Harris<sup>¶</sup>, and Edilson T. Liu<sup>\*\*¶</sup>

Proc Natl Acad Sci U S A. 2003 Sep 2;100(18):10393-8. Epub 2003 Aug 13.

# Gene expression profiling predicts clinical outcome of breast cancer

Laura J. van 't Veer<sup>\*†</sup>, Hongyue Dai<sup>‡‡</sup>, Marc J. van de Vijver<sup>\*†</sup>, Yudong D. He<sup>‡</sup>, Augustinus A. M. Hart<sup>\*</sup>, Mao Mao<sup>‡</sup>, Hans L. Peterse<sup>\*</sup>, Karin van der Kooy<sup>\*</sup>, Matthew J. Marton<sup>‡</sup>, Anke T. Witteveen<sup>\*</sup>, George J. Schreiber<sup>‡</sup>, Ron M. Kerkhoven<sup>\*</sup>, Chris Roberts<sup>‡</sup>, Peter S. Linsley<sup>‡</sup>, René Bernards<sup>\*</sup> & Stephen H. Friend<sup>‡</sup>

Nature. 2002 Jan 31;415(6871):530-6.

**Sotiriou  
(3914)**

**van't Veer  
(12652)**

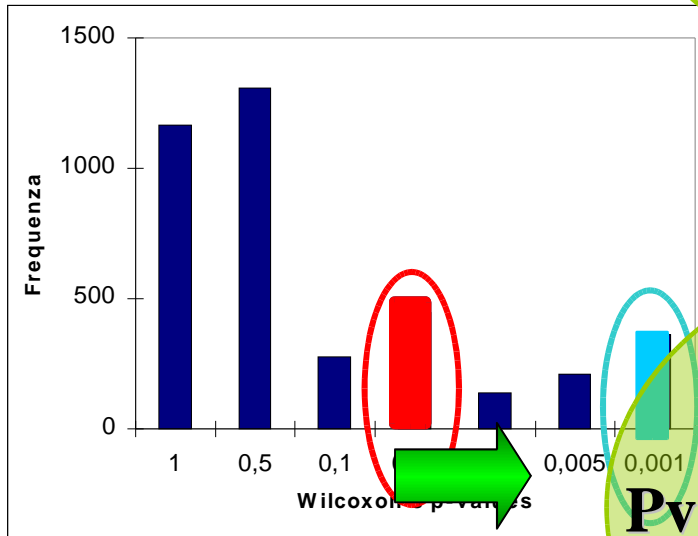
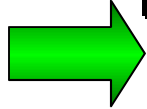
1996

**Pvalue  $\leq 0.05$   
Wilcoxon's Test**

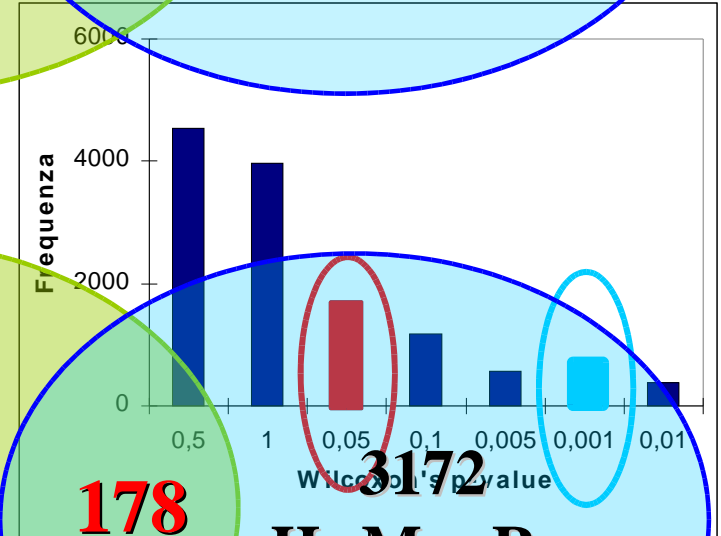
**662**

**3172**

**Hs Mm Rn**  
Contenenti ERE



**639**  
**Pvalue  $\leq 0.001$**



**178**

**3172**

**Hs Mm Rn**  
Contenenti ERE

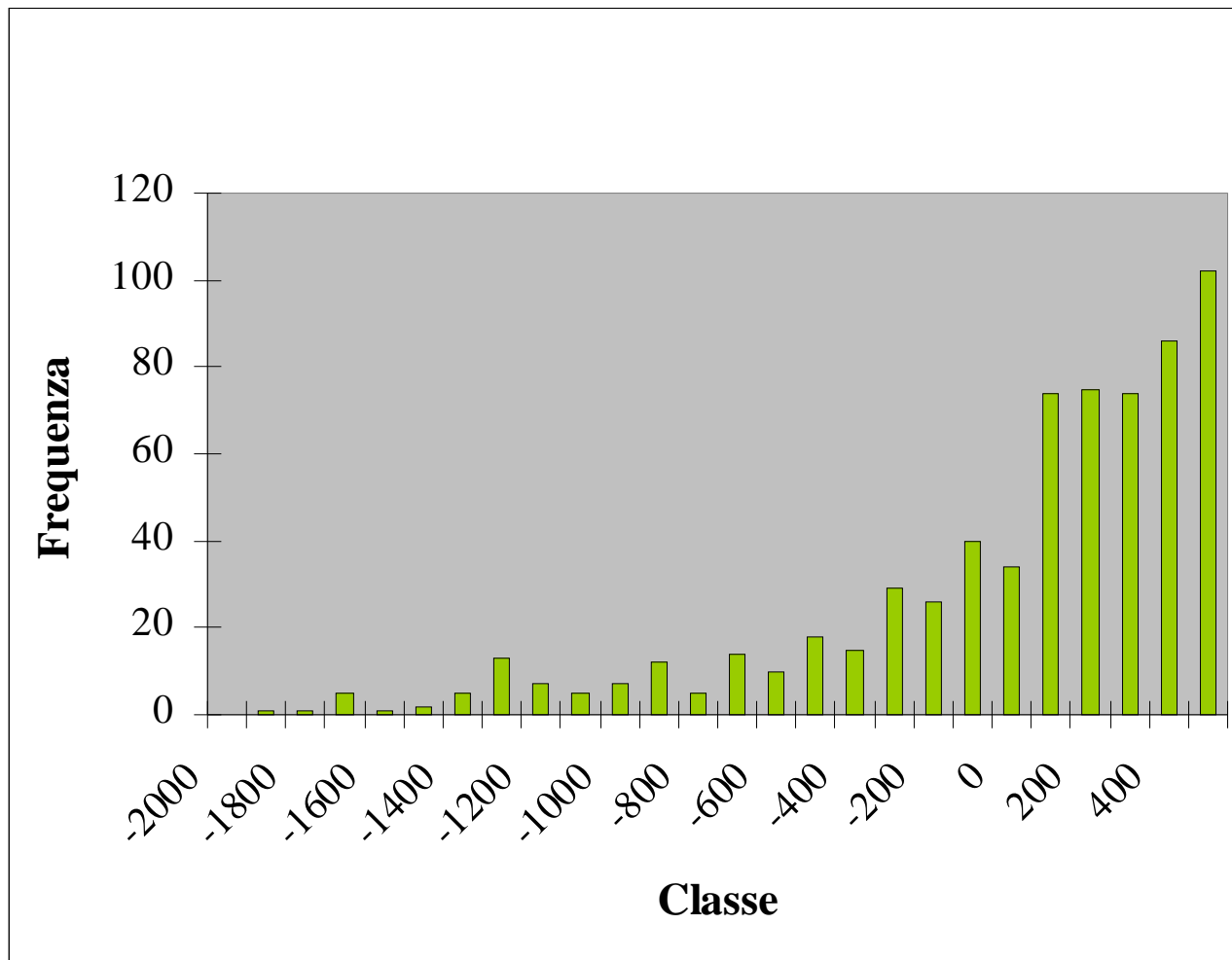
Identificazione di geni differenzialmente espressi tra tessuti tumorali ER+/ER-

# GENE ONTOLOGY

GO category	p-value in 3172	p-value in 662	p-value in 178
regulation of cell proliferation	.00161	Non significativo	.00161
apoptosis	.00272	.0017	.00272
programmed cell death	.00313	.00182	.00286
positive regulation of programmed cell death	.00401	.000161	.00313
positive regulation of apoptosis	.00401	.000161	.00361
induction of programmed cell death	.00401	.000161	.00401
induction of apoptosis	.00401	.000161	.00401
regulation of programmed cell death	.00401	.000161	.00401
cell growth and/or maintenance	.0048	.00355	.0048
regulation of apoptosis	.001	.000445	Non significativo
cell death	Non significativo	.00264	Non significativo
death	Non significativo	.00281	Non significativo

E' un vocabolario dinamico, che permette l'associazione dei geni differenzialmente espressi a specifiche classi funzionali

- Contenenti elementi ERE localizzati intorno al primo nucleotide trascritto



# Set Geni Estrogeno Responsivi 662:

- Geni con ERE in Hs Mm Rn

[www.bioinformatica.unito.it/bioinformatics/rre/rre.html](http://www.bioinformatica.unito.it/bioinformatics/rre/rre.html)

- **DIU** **RRE: a tool for the extraction of non-coding regions surrounding annotated genes from**

- **Arr** genomic datasets.

ulare

- **Cor**

leotide

traso <sup>1</sup>Lazzarato F, <sup>1</sup>Franceschini G, <sup>2</sup>Botta M, <sup>3</sup>Cordero F and <sup>3\*</sup>Calogero RA

- Presenza di ERE raggruppati in cluster <sup>Bioinformatics. 2848-2850, 20:2004</sup>

**[www.bioinformatica.unito.it/bioinformatics/rre/rre.html](http://www.bioinformatica.unito.it/bioinformatics/rre/rre.html)**

**RRE: a tool for the extraction of non-coding regions surrounding annotated genes from genomic datasets.**

<sup>1</sup>Lazarato F, <sup>1</sup>Franceschini G, <sup>2</sup>Botta M, <sup>3</sup>Cordero F and <sup>3\*</sup>Calogero RA

Bioinformatics. 2004 Apr 29

# RRE database

- RRE allows the extraction of non-coding regions surrounding a coding sequence [i.e. gene upstream region, 5-untranslated region (5-UTR), introns, 3-UTR, downstream region] from annotated genomic datasets available at NCBI.
- RRE parser and web-based interface are accessible at <http://www.bioinformatica.unito.it/bioinformaticas/rre/rre.html>

# RRE database

- Available genomes data:
  - *Arabidopsis thaliana*
  - *Caenorhabditis elegans*
  - *Drosophila Melanogaster*
  - *Homo Sapiens*
  - *Mus Musculus*
  - *Plasmodium falciparum*
  - *Rattus Norvegicus*
  - *Schizosaccharomyces pombe*

# RRE database

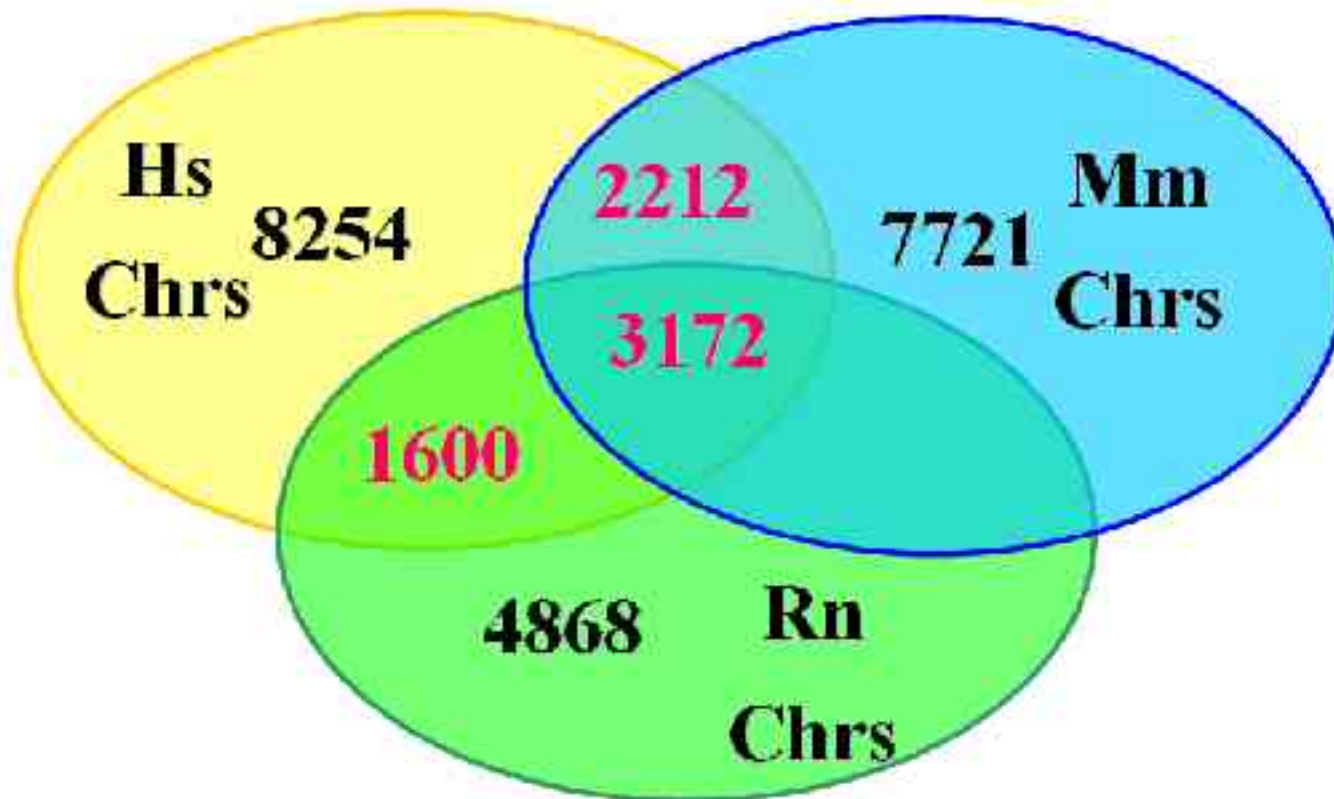
- Available retrieval options:
  - Upstream, -2000/+500, 5'UTR, Introns, Exons, 3'UTR, Downstream.
- Available queries:
  - By gene symbols list
  - By gene ID (Locus Link) list
  - By orthologs gene ID list
  - By genomic contig(s) list
  - By chromosome

# RRE database

- Other features:
  - Searching for a specific feature (e.g. 5'UTR) genes located within a certain distance from a list of genes.
  - This query is interesting for users guessing the presence of some regulative common mechanism associate to a specific chromosomal locus.

# RRE database

(Genes containing ERE elements)



# RRE database

(Genes containing ERE elements)

- Available retrieval options:
    - Homo S. genes containing at least a predicted ERE in the -2000/+500 region.
- AND/OR
- P-value associated to ER-responsive breast cancer (Sotiriou/LJ van 't Veer)
  - P-value associated to early onset of metastasis (LJ van 't Veer)

# Breast cancer classification and prognosis based on gene expression profiles from a population-based study

Christos Sotiriou<sup>\*†</sup>, Soek-Ying Neo<sup>‡</sup>, Lisa M. McShane<sup>§</sup>, Edward L. Korn<sup>§</sup>, Philip M. Long<sup>‡</sup>, Amir Jazaeri<sup>\*</sup>, Philippe Martiat<sup>†</sup>, Steve B. Fox<sup>¶</sup>, Adrian L. Harris<sup>¶</sup>, and Edilson T. Liu<sup>\*\*¶</sup>

Proc Natl Acad Sci U S A. 2003 Sep 2;100(18):10393-8. Epub 2003 Aug 13.

# Gene expression profiling predicts clinical outcome of breast cancer

Laura J. van 't Veer<sup>\*†</sup>, Hongyue Dai<sup>‡‡</sup>, Marc J. van de Vijver<sup>\*†</sup>, Yudong D. He<sup>‡</sup>, Augustinus A. M. Hart<sup>\*</sup>, Mao Mao<sup>‡</sup>, Hans L. Peterse<sup>\*</sup>, Karin van der Kooy<sup>\*</sup>, Matthew J. Marton<sup>‡</sup>, Anke T. Witteveen<sup>\*</sup>, George J. Schreiber<sup>‡</sup>, Ron M. Kerkhoven<sup>\*</sup>, Chris Roberts<sup>‡</sup>, Peter S. Linsley<sup>‡</sup>, René Bernards<sup>\*</sup> & Stephen H. Friend<sup>‡</sup>

Nature. 2002 Jan 31;415(6871):530-6.

**Sotiriou  
(3914)**

**van't Veer  
(12652)**

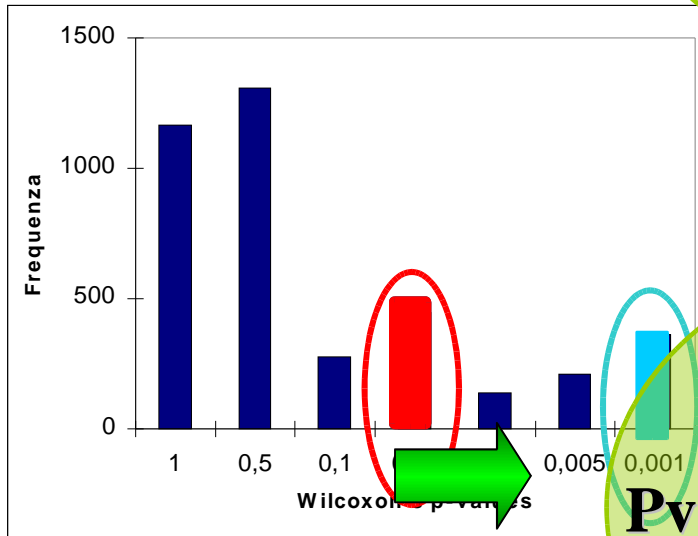
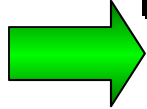
1996

**Pvalue  $\leq 0.05$**   
Wilcoxon's Test

**662**

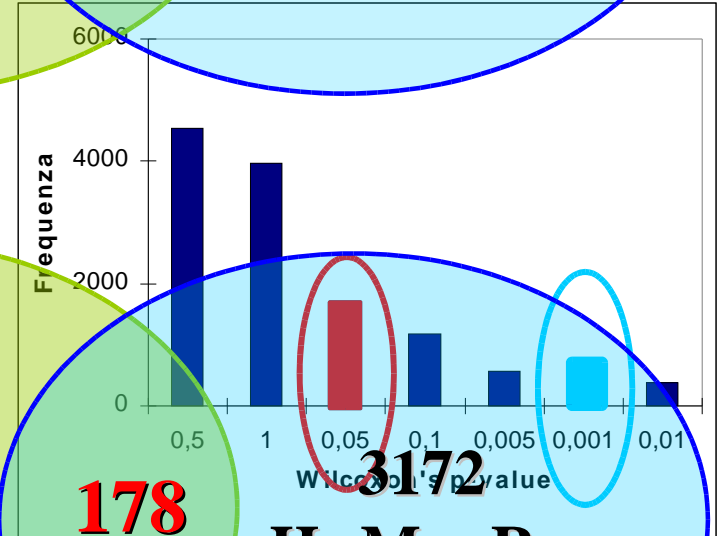
**3172**

**Hs Mm Rn**  
Contenenti ERE



**639**

**Pvalue  $\leq 0.001$**



**178**

**3172**

**Hs Mm Rn**  
Contenenti ERE

Identificazione di geni differenzialmente espressi tra tessuti tumorali ER+/ER-

# RRE database

(Genes containing ERE elements)

- Available retrieval options:
  - AND/OR
    - Genes found modulated by estrogen in microarray analysis in MCF7 cell line (Frasor et al. 2003/2004)
  - AND/OR
    - Genes found modulated by estrogen in microarray analysis in ZR75 cell line (Cicatiello et al. 2004).
  - AND/OR
    - Homo S. genes containing at least a predicted ERE also conserved in Mus musculus
  - AND/OR
    - Homo S. genes containing at least a predicted ERE also conserved in Rattus norvegicus