

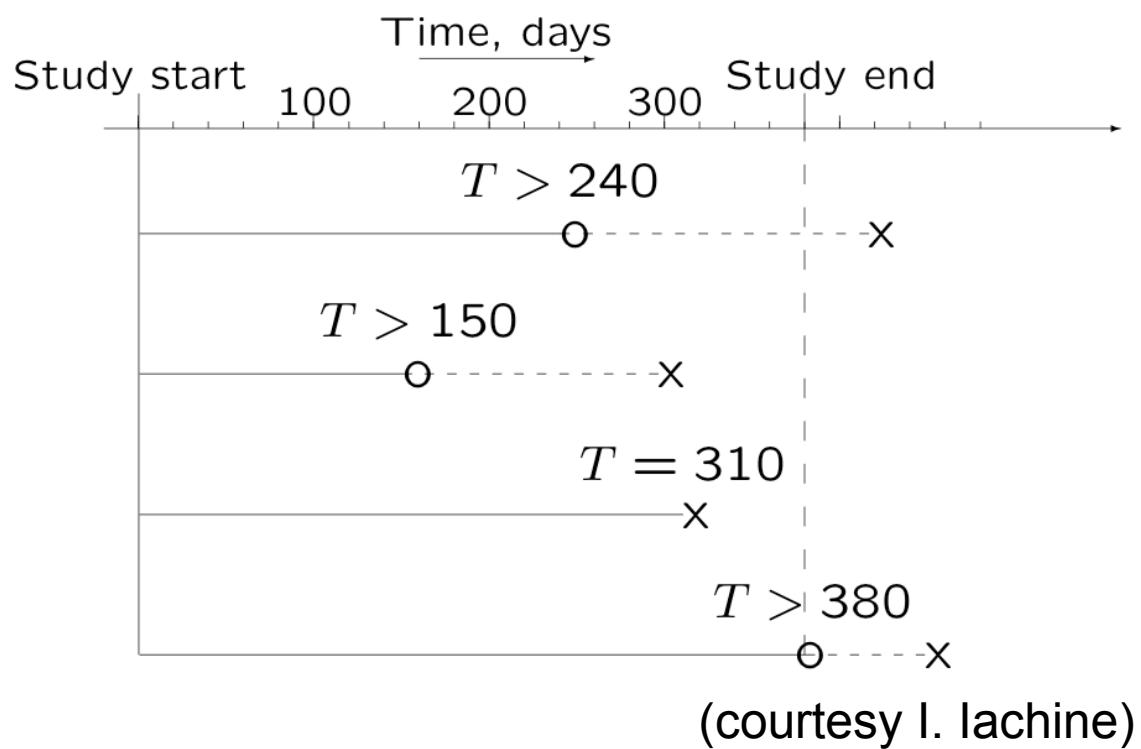
# **Gene expression predictors of breast cancer outcome**

Vincent Detours

IRIBHM, Université Libre de Bruxelles (U.L.B.)  
[vdetours@ulb.ac.be](mailto:vdetours@ulb.ac.be)

# Survival analysis

# Censored time-to-event data



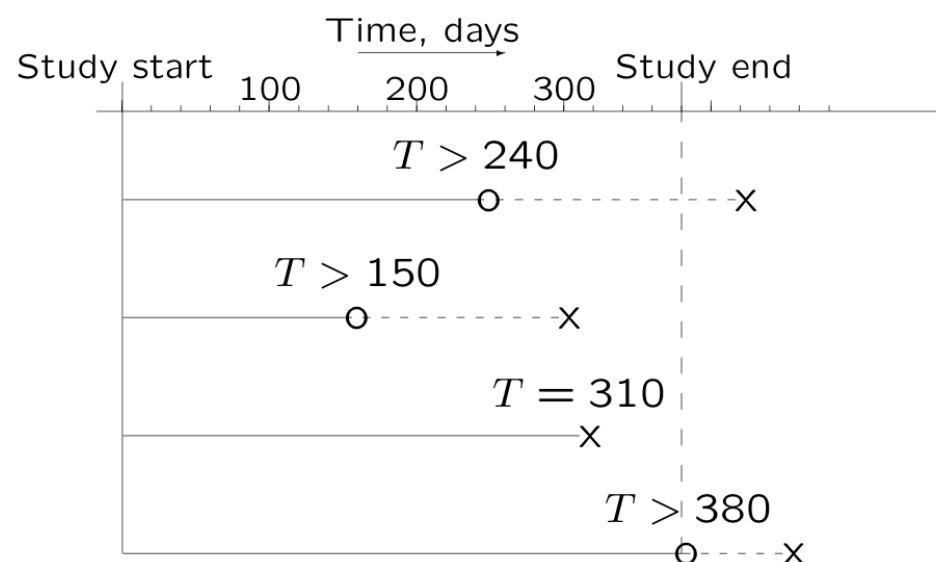
$\times$  = 'event' (e.g.,  
death from cancer,  
relapse, etc.)

$\circ$  = drop out of  
study

**Assumption:** censoring and events are independent

# Kaplan-Meier curves display censored survival profiles

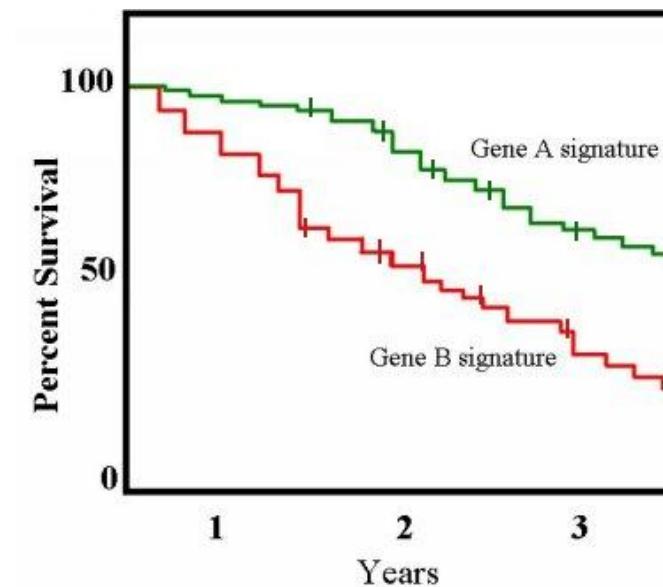
Censored time-to-event data



$\times$  = 'event'

$\circ$  = drop out of study

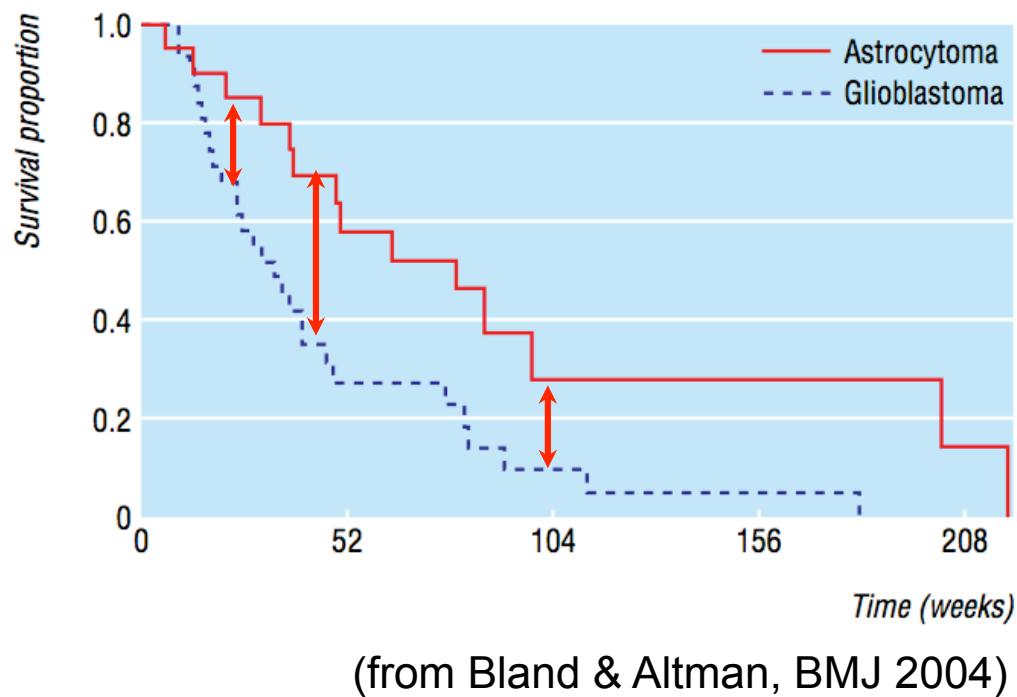
Kaplan-Meier curves



step down = 'event'

bar = drop out of study

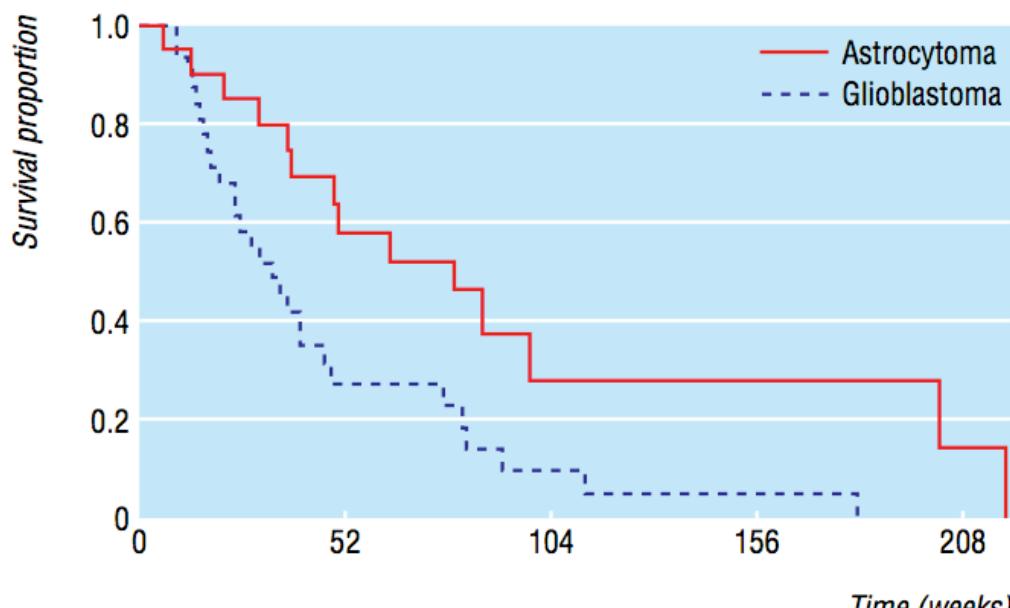
# The Kaplan-Meier estimator measure global, time-integrated, survival experience



Naive approach:  
measure survival  
difference at time  $t$

But we are interested in  
the global survival  
experience, i.e. some  
time-integrated measure

# The survival function measures global, time-integrated, survival experience



(from Bland & Altman, BMJ 2004)

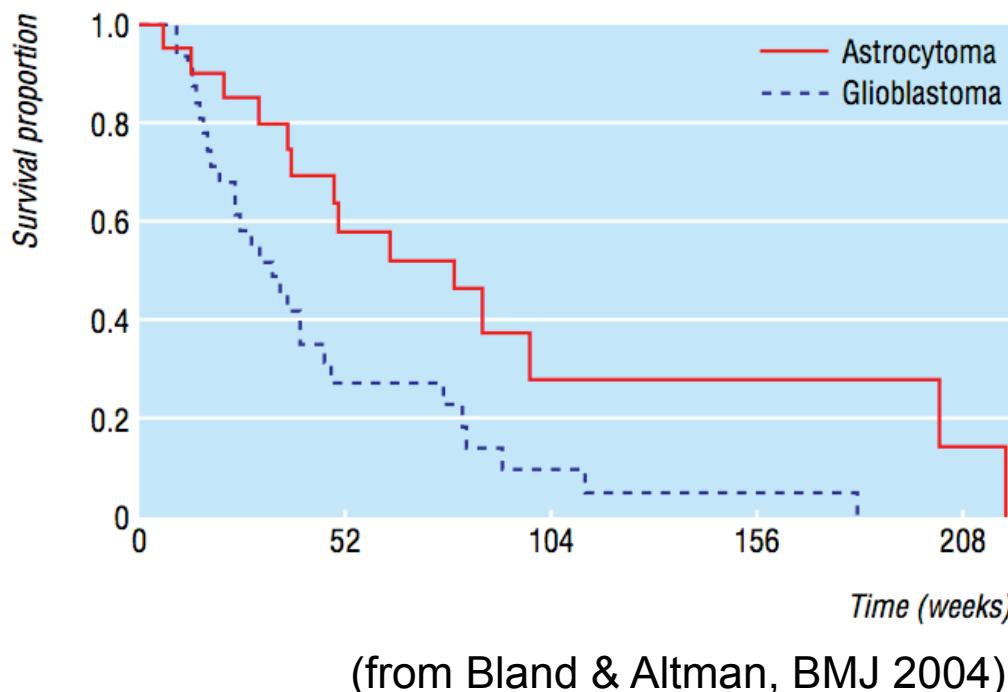
Survival function:  
 $S(t_i) = \text{Proba}(\text{survival to } t_i)$

Given

$p_i = \Pr(\text{survival in } [t_i, t_{i+1}] \mid \text{alive at } t_i)$   
= 1, if no death in  $[t_i, t_{i+1}]$ ,  
else  $(n_i - 1)/n_i$ ,  
with  $n_i = \text{number alive at } t_i$ ,

$S(t_i)$  can be rewritten as:  
 $S(t_i) = \prod p_i$

# The logrank test



Do the two groups have  
same survival functions?

$$H_0: S_A = S_G$$

$$H_A: S_A \neq S_G$$

- It is nonparametric
- It assumes independent censoring

# Cox regression model

- Assesses the association of covariates (e.g. disease state, treatment, biomarker, etc.) with survival
- Assumes proportional hazard, i.e. the ratio of survival probabilities between groups is constant with time

# Cox regression model

Let the *hazard function*,  $h(t)$ , be the percentage of participants experiencing an event (e.g. relapse, death, etc.) per time unit:

$$h(t) = \frac{d \ln(S(t))}{dt}$$

# Cox regression model

Under the *proportional hazard* hypothesis,  $h(t)$  can be written as

$$h(t|x_1, \dots, x_n) = h_0(t) \cdot e^{\beta_1 x_1 + \dots + \beta_n x_n}$$

where

- $x_i$  are the covariates (e.g. age, treatment status, etc.)
- $\beta_i$  are corresponding regression coefficients
- $h_0$  is the baseline hazard, i.e. hazard when all covariates are null

The goal is to estimate  $\beta_i$ , the contributions of covariates to  $h$

# Cox regression model

$$h(t|x_1, \dots, x_n) = h_0(t) \cdot e^{\beta_1 x_1 + \dots + \beta_n x_n}$$

The trick:

- proper coding of the covariates such that there is a baseline covariate state  $\{x_1, \dots, x_n\} = \{0, \dots, 0\}$  obviating the need to actually know  $h_0$ : one simply refers to  $h(t)$  relatively to this baseline state
- in the example above, we encode glioma with 0, astrocytoma with 1.  $h_0$  cancels out when computing the hazard ratio,  $HR$ :

$$HR = h(t|x=1)/h(t|x=0) = e^\beta$$

# Cox regression model

Regression coefficients are interpreted as

- $\beta < 0$ :  $HR < 1$ , glioma have worst prognostic
- $\beta = 0$ :  $HR = 1$ , equal prognostic
- $\beta > 0$ :  $HR > 1$ , glioma have better prognostic
- In the continuous case,  $\beta_i$  is the effect of a unit increase of  $x_i$ , other covariates being constant
- Beware,  $HR$  values depend on the unit of the covariates!

**Extremely  
brief introduction to  
cancer progression**

# Cancers develop slowly

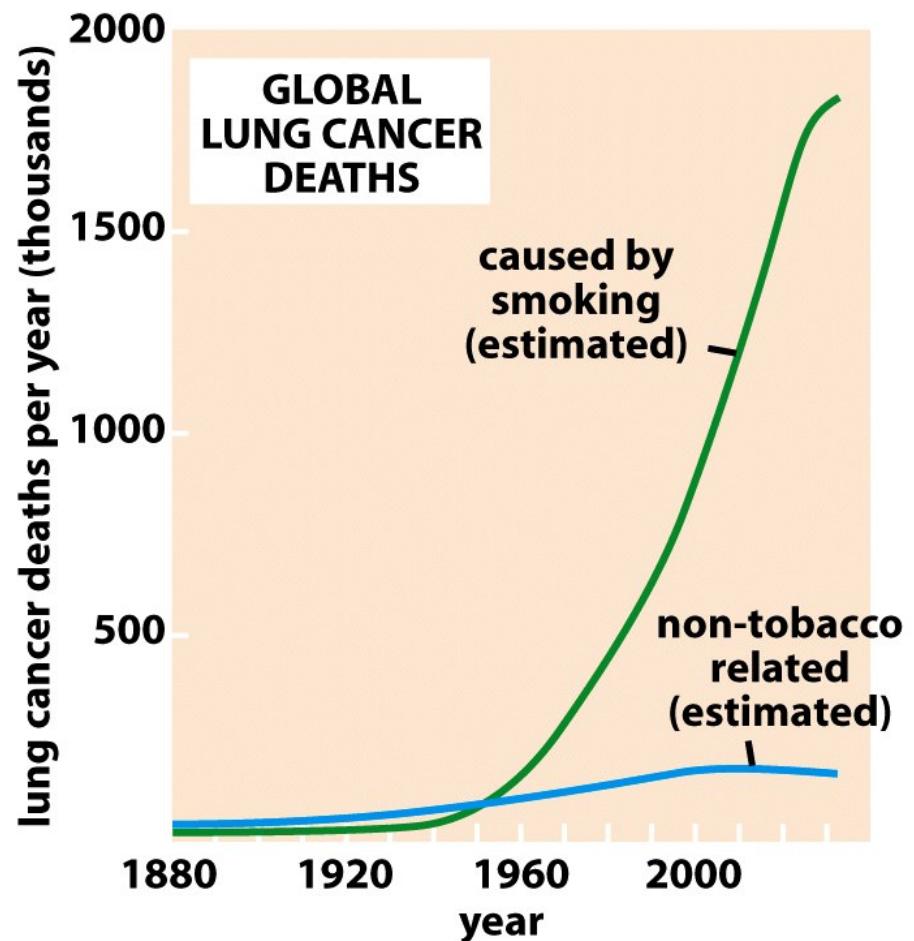
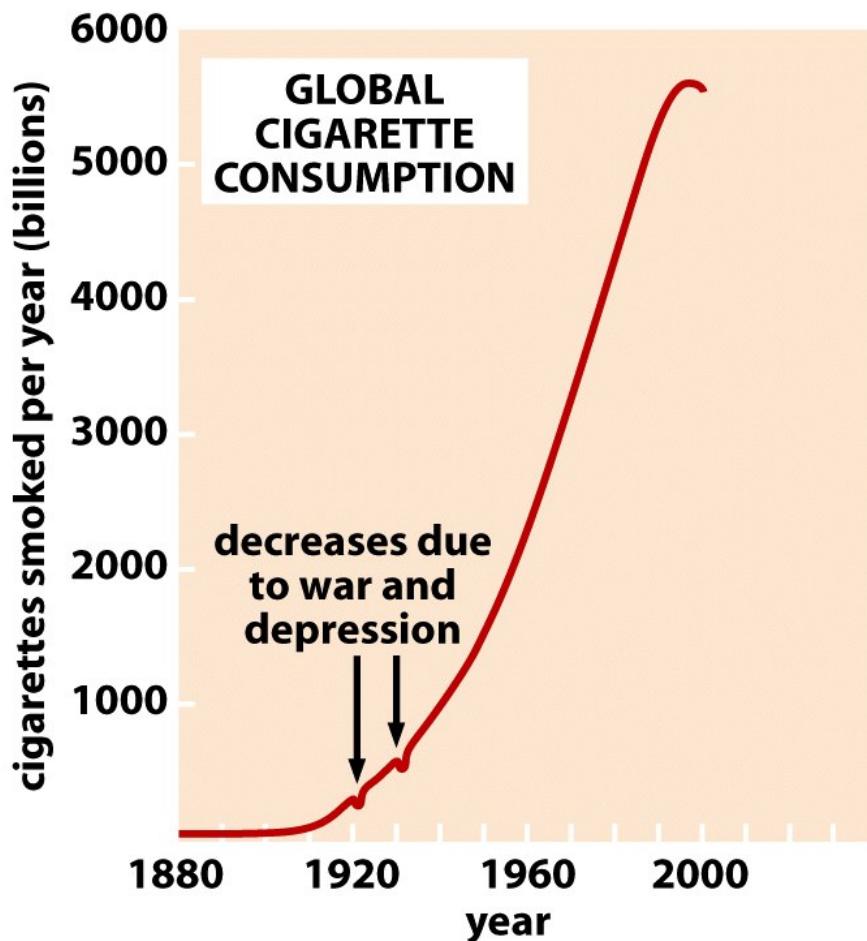
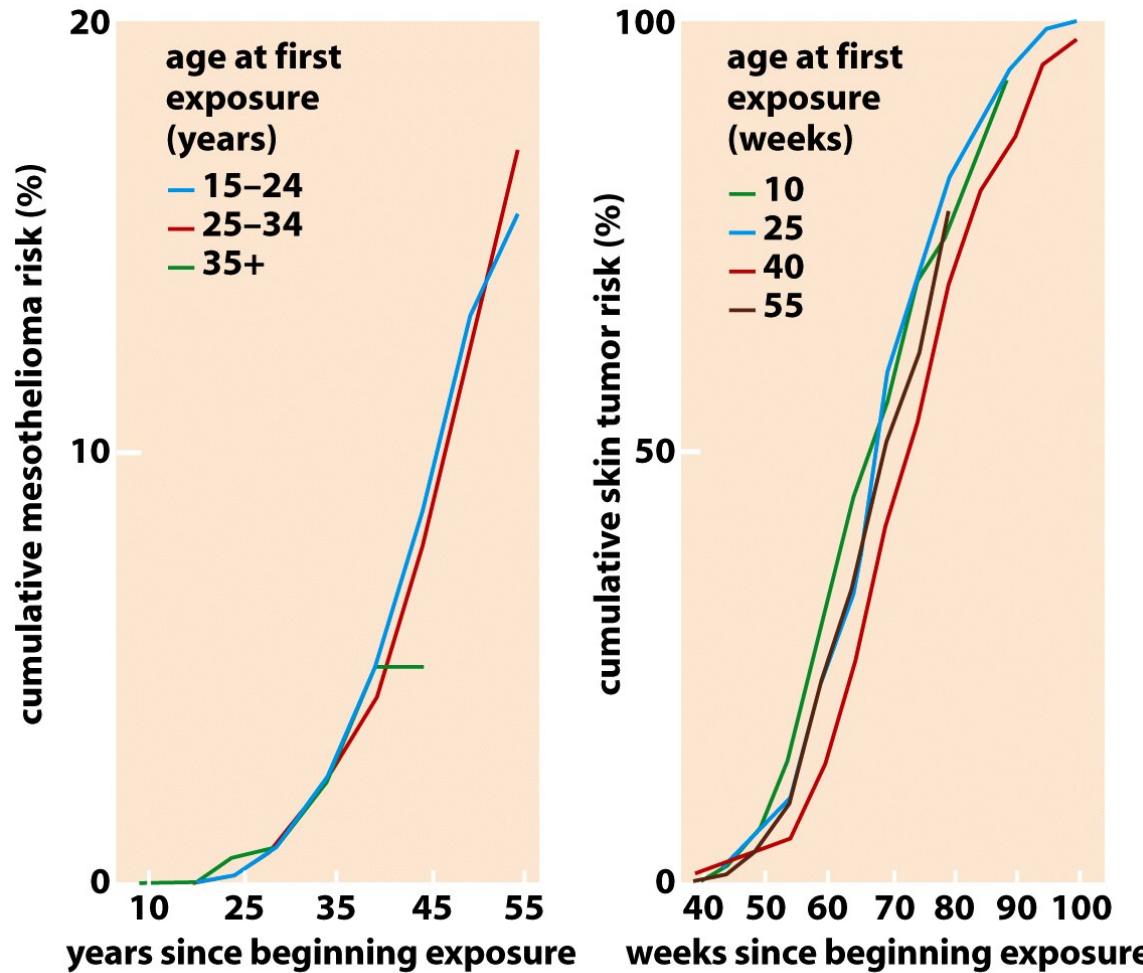


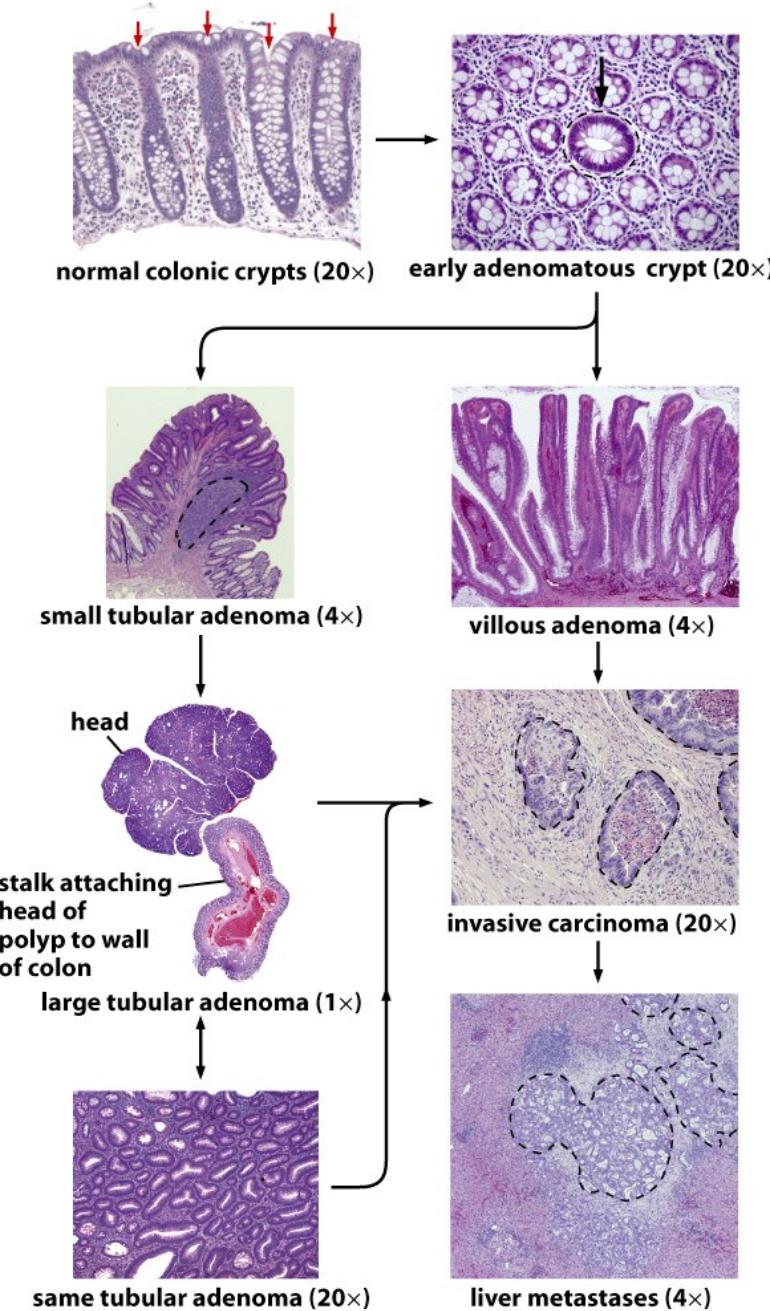
Figure 11-2 The Biology of Cancer (© Garland Science 2007)

# Cancers develop slowly

mesothelioma

skin cancer (mice)

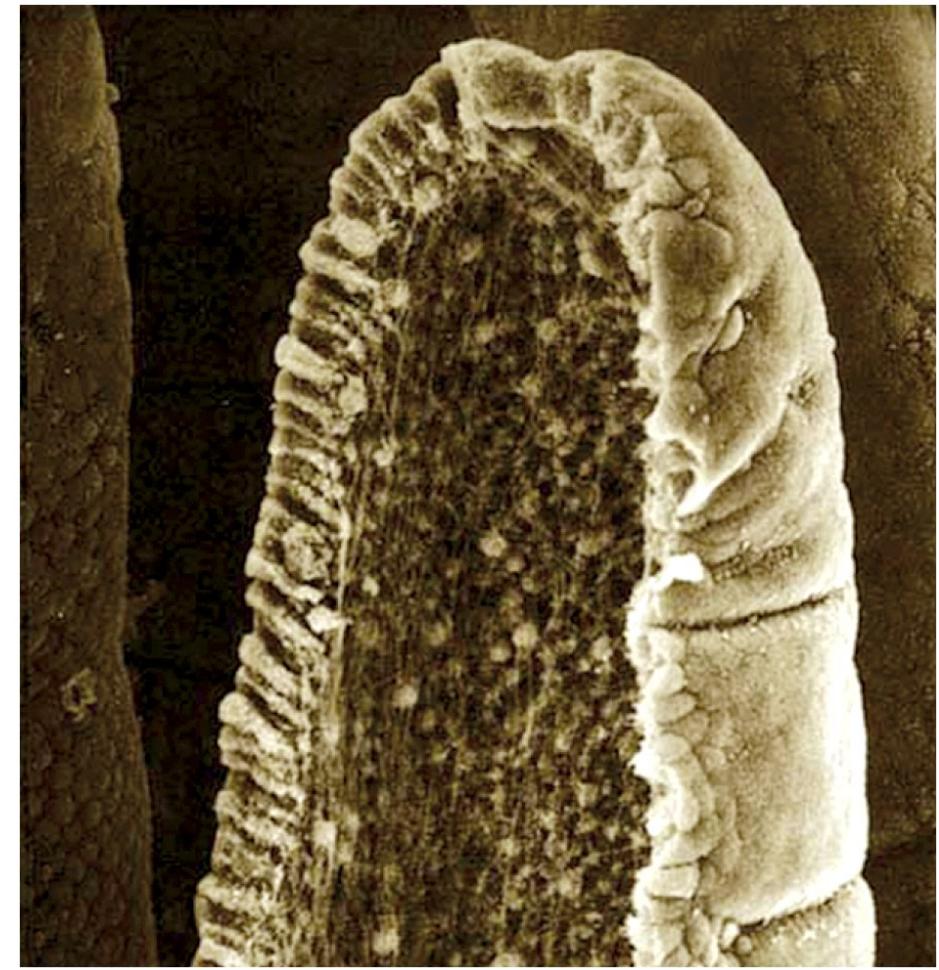




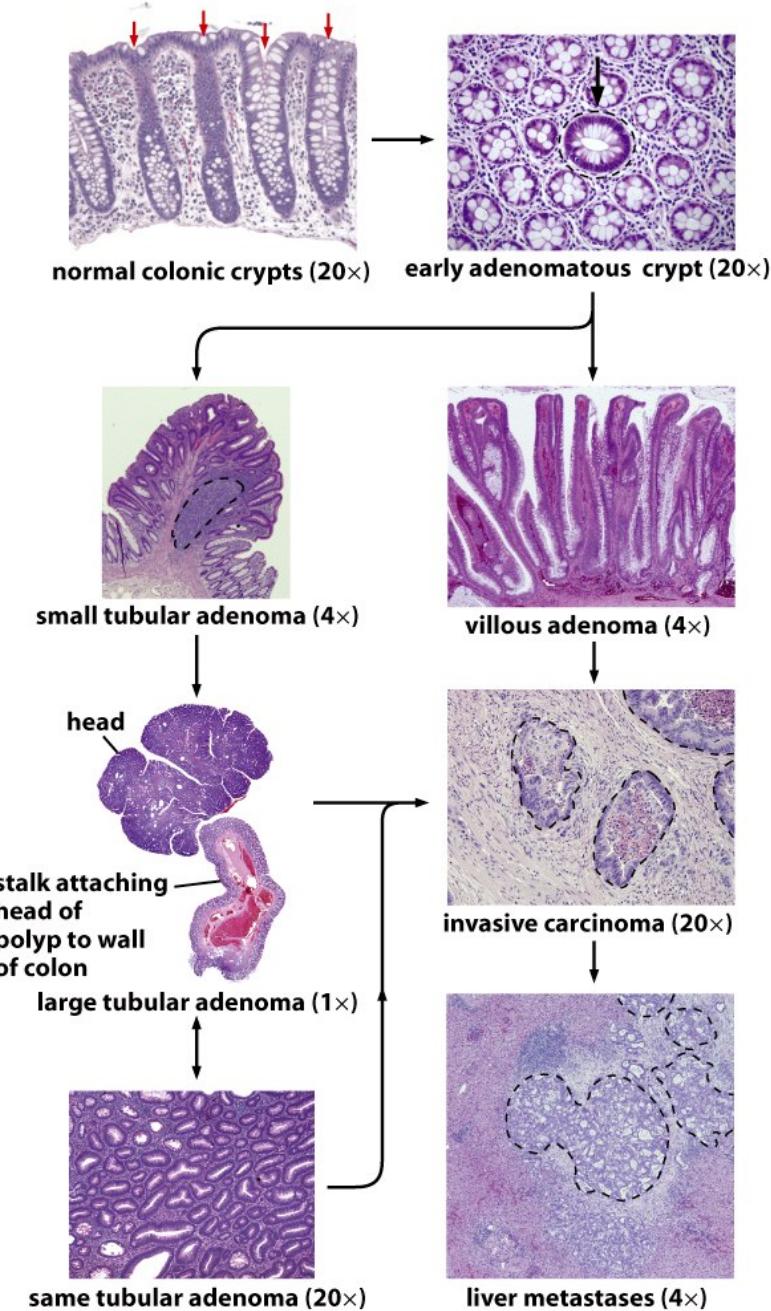
Cancer progression is a multi-step process

Example: progression of colon cancer

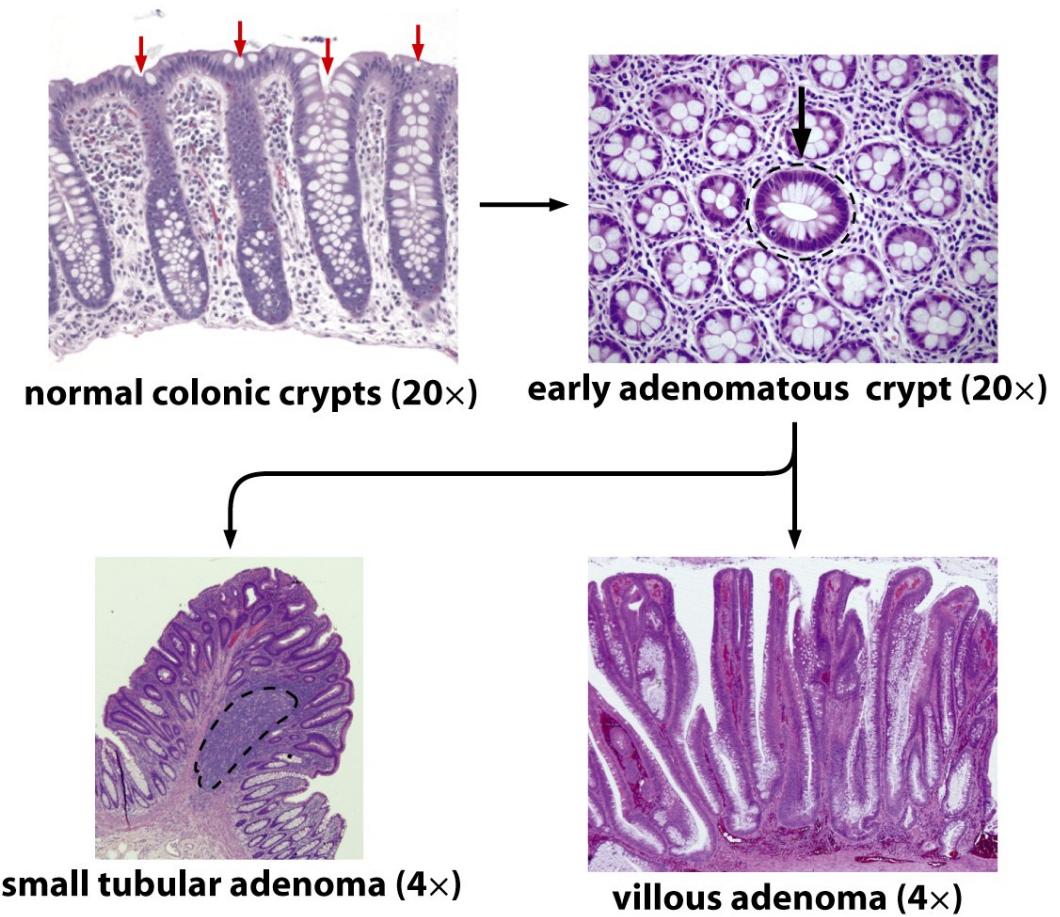
# Cancer progression is a multi-step process

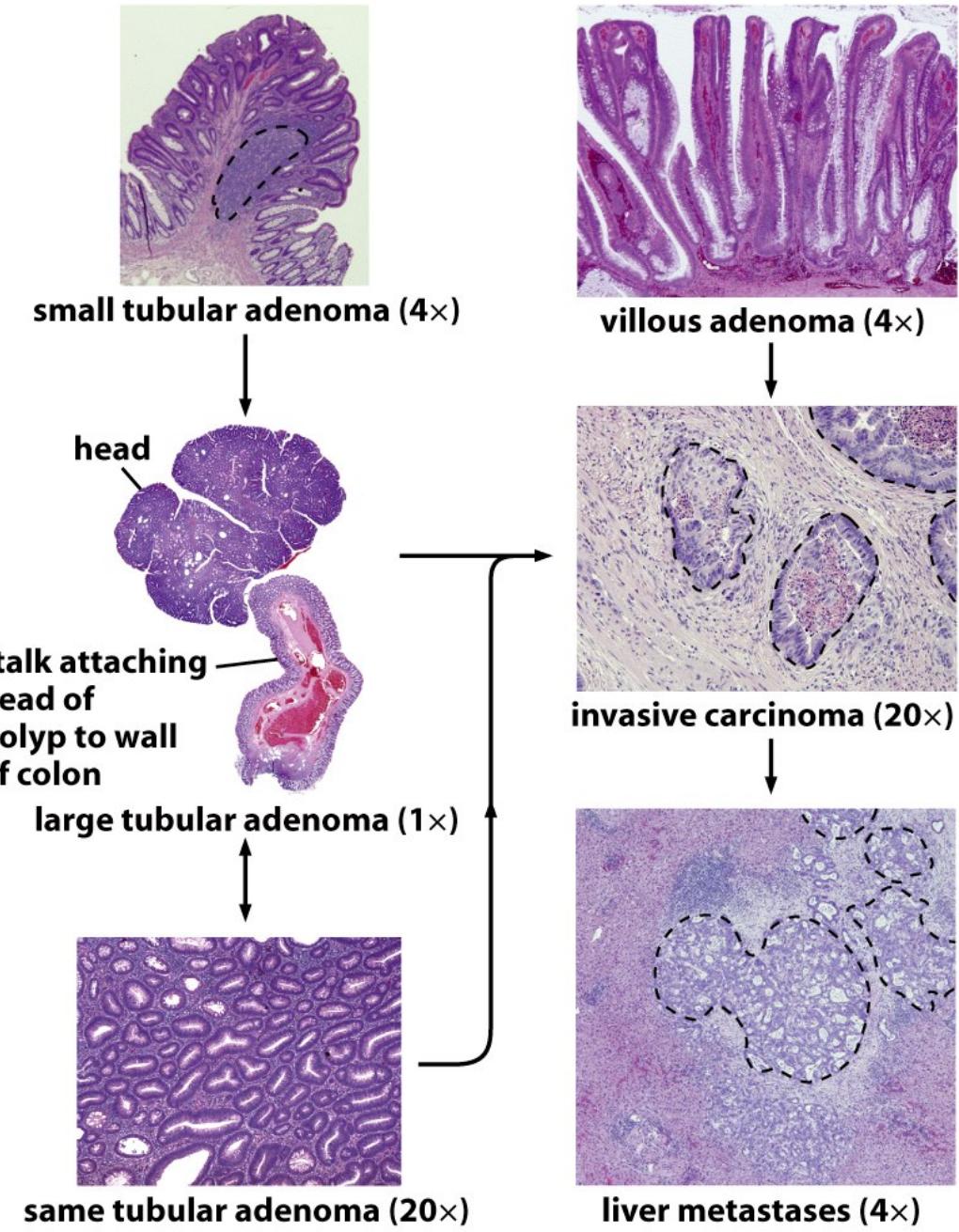
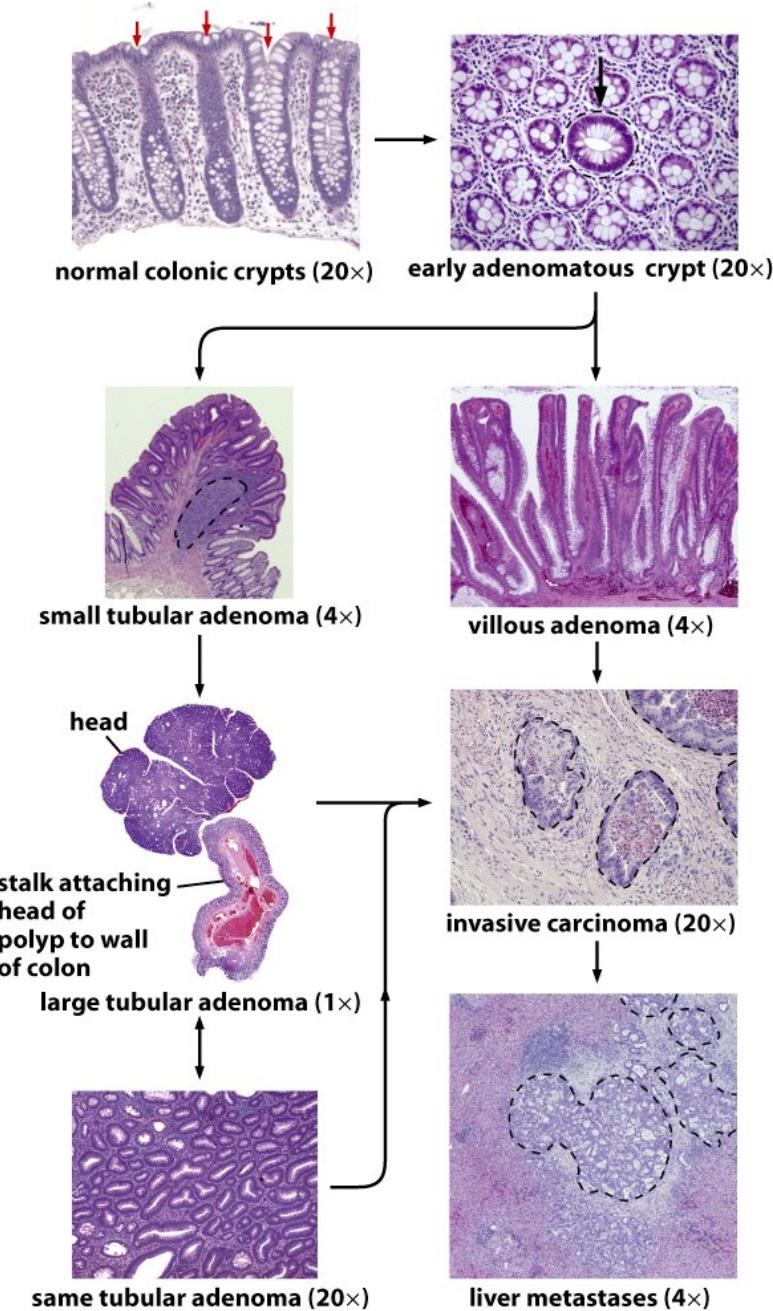


The normal colon is lined with cripts

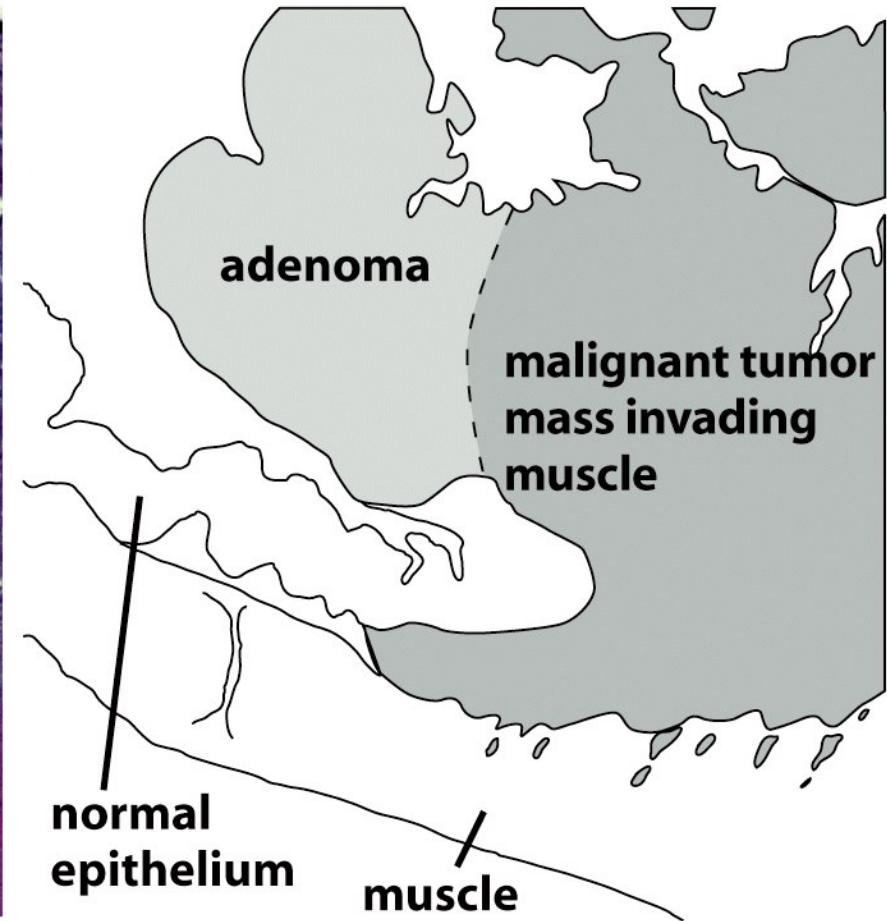
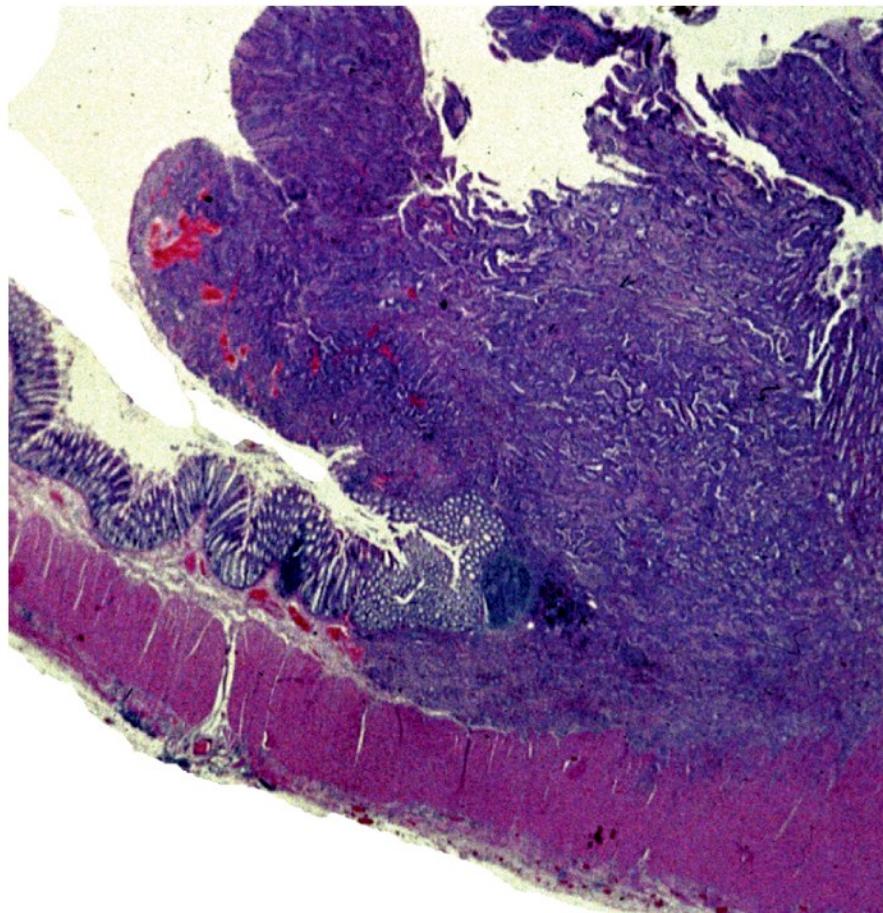


# Cancer progression is a multi-step process

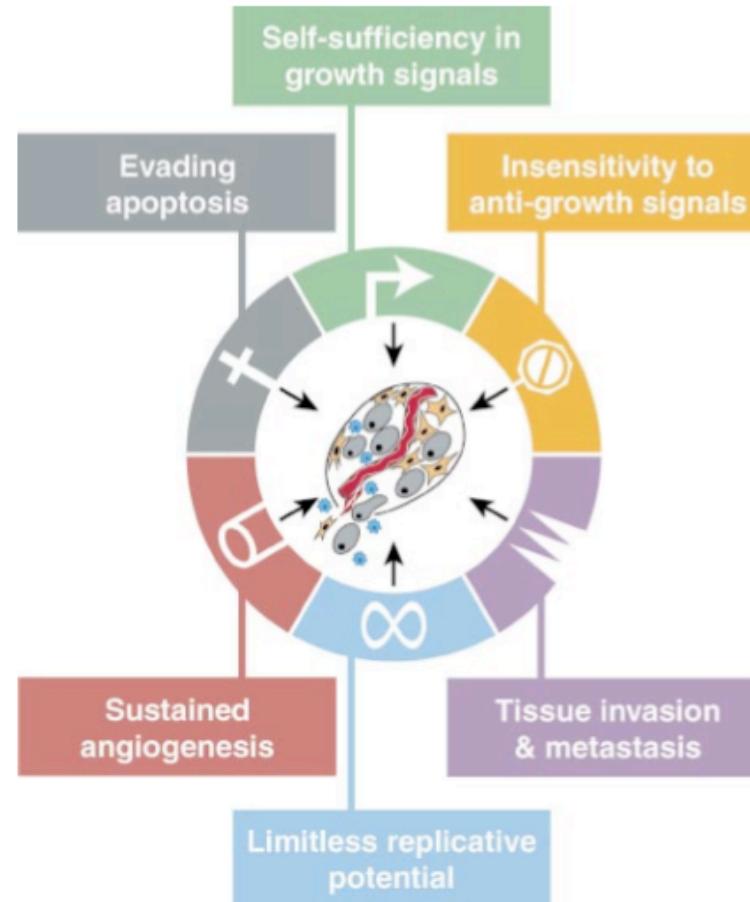




# Cancer progression is a multi-step process



# Cancer progression is a multi-step process



(Hanahan & Weinberg, Cell 2000)

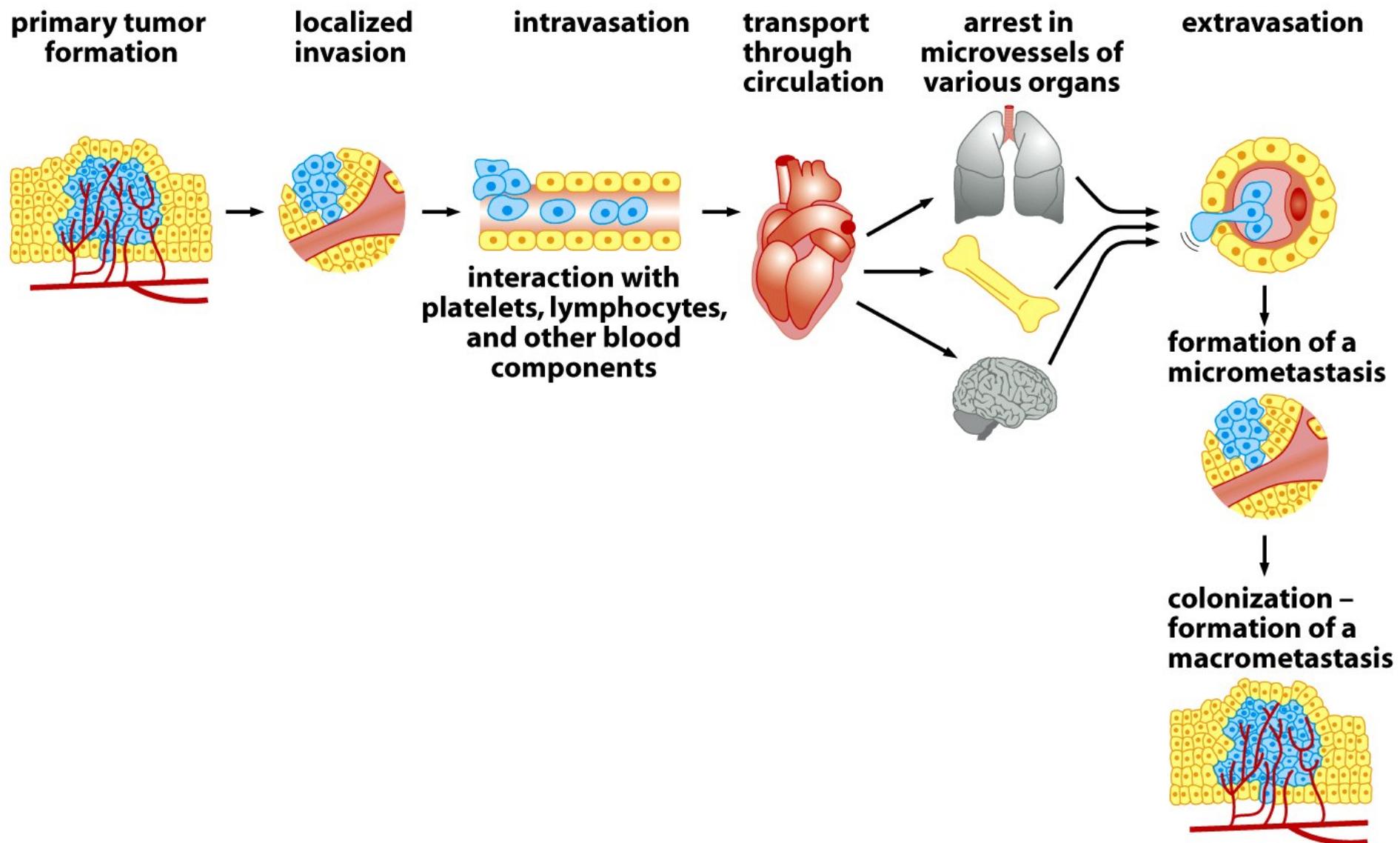
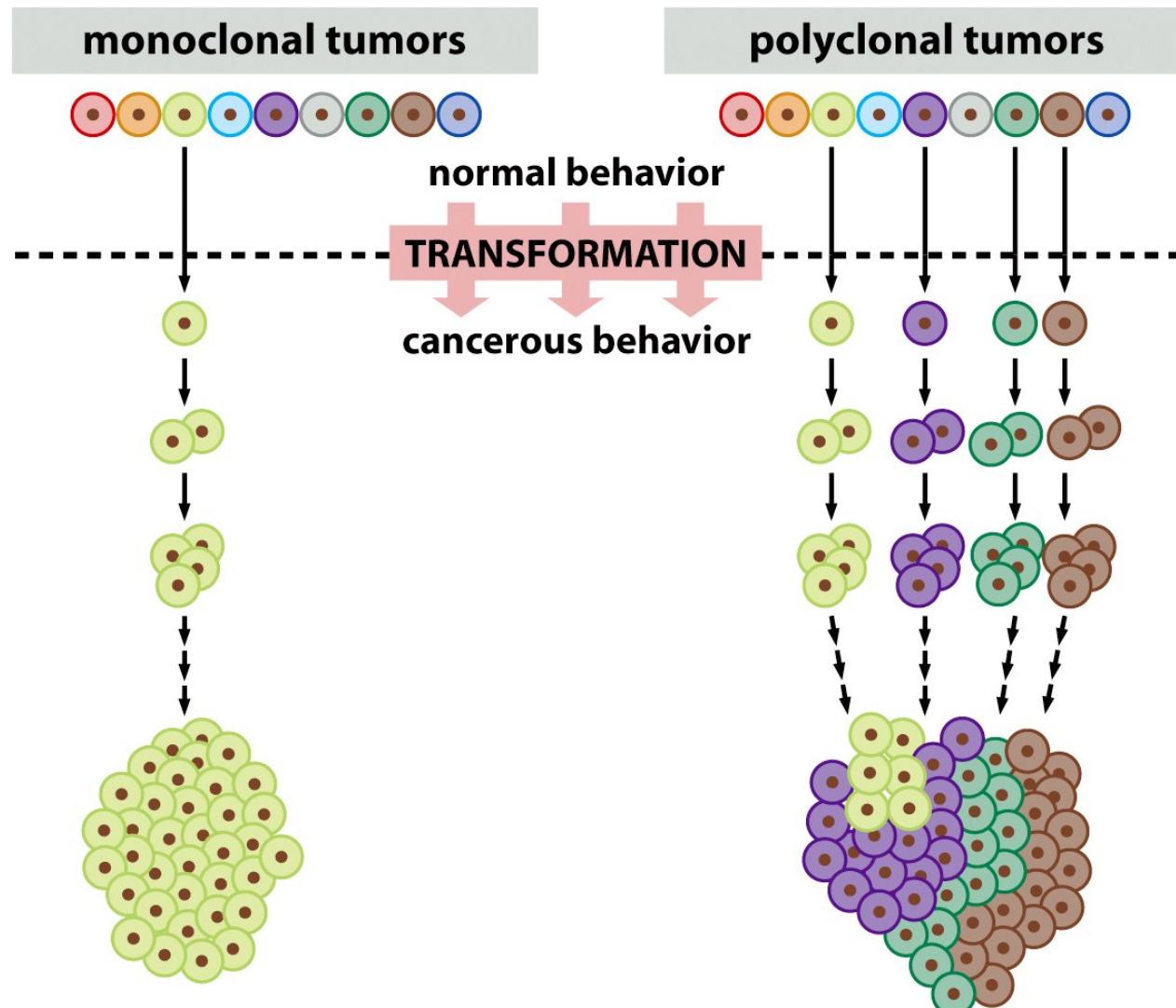
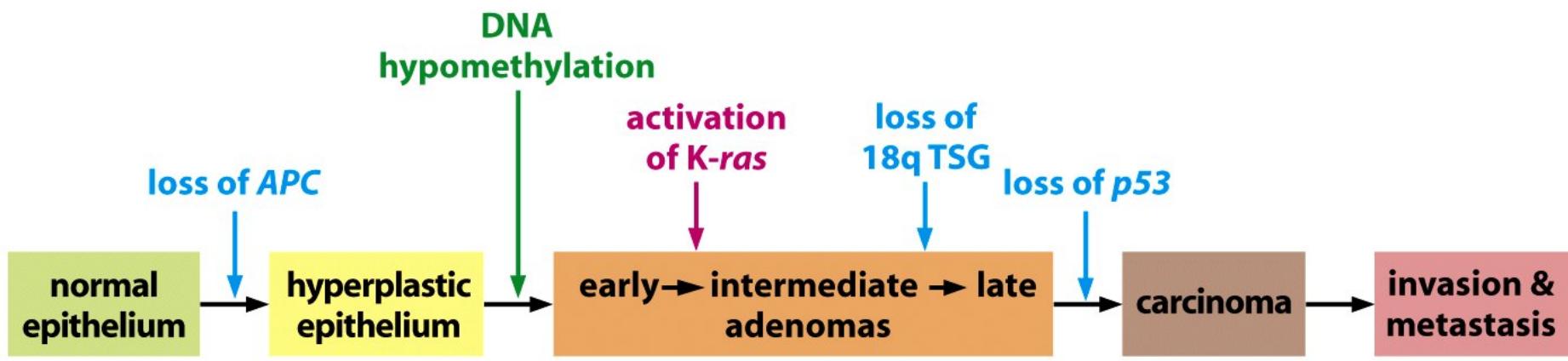


Figure 14-4 The Biology of Cancer (© Garland Science 2007)

# Tumor clonality

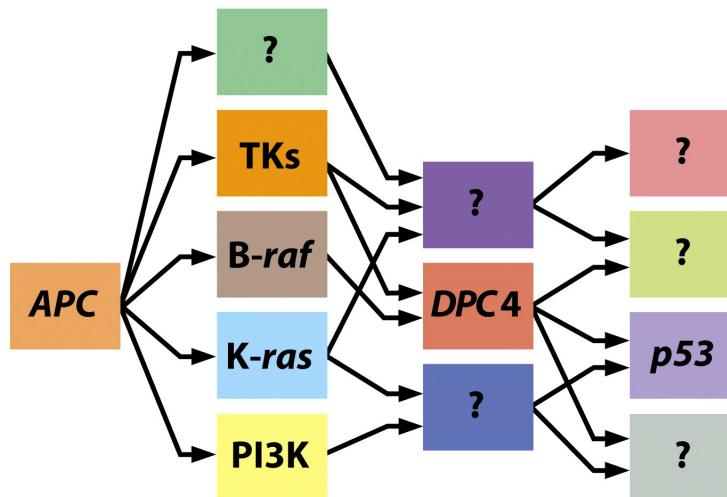
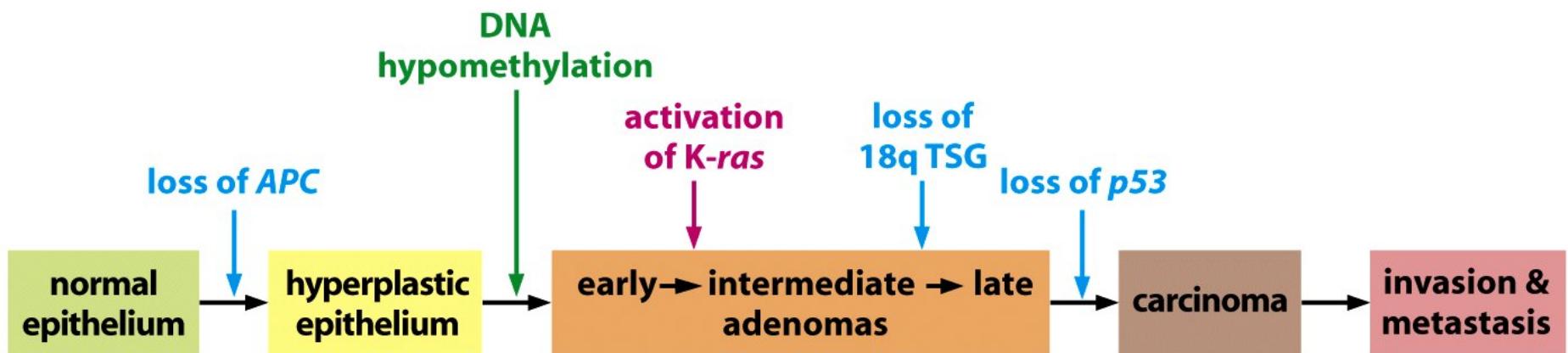


# Multiple genetic events underly cancer progression

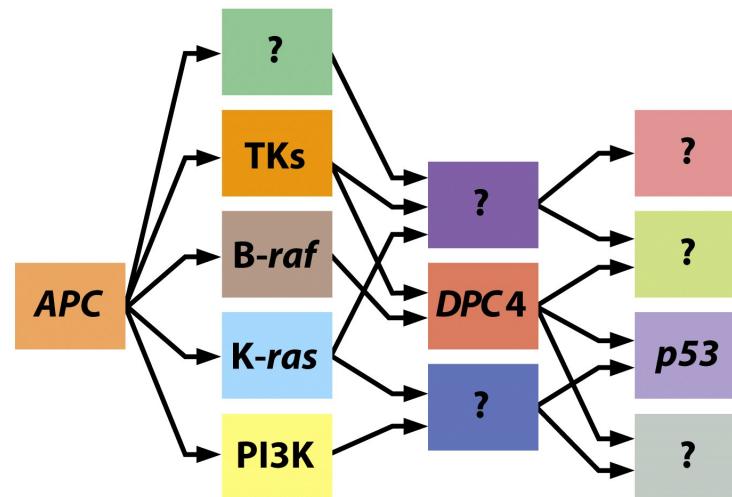
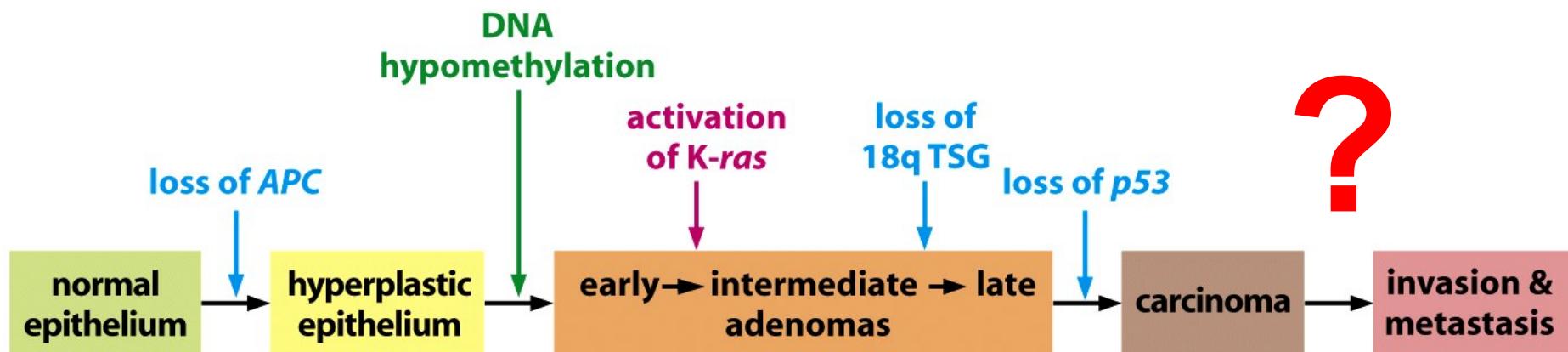


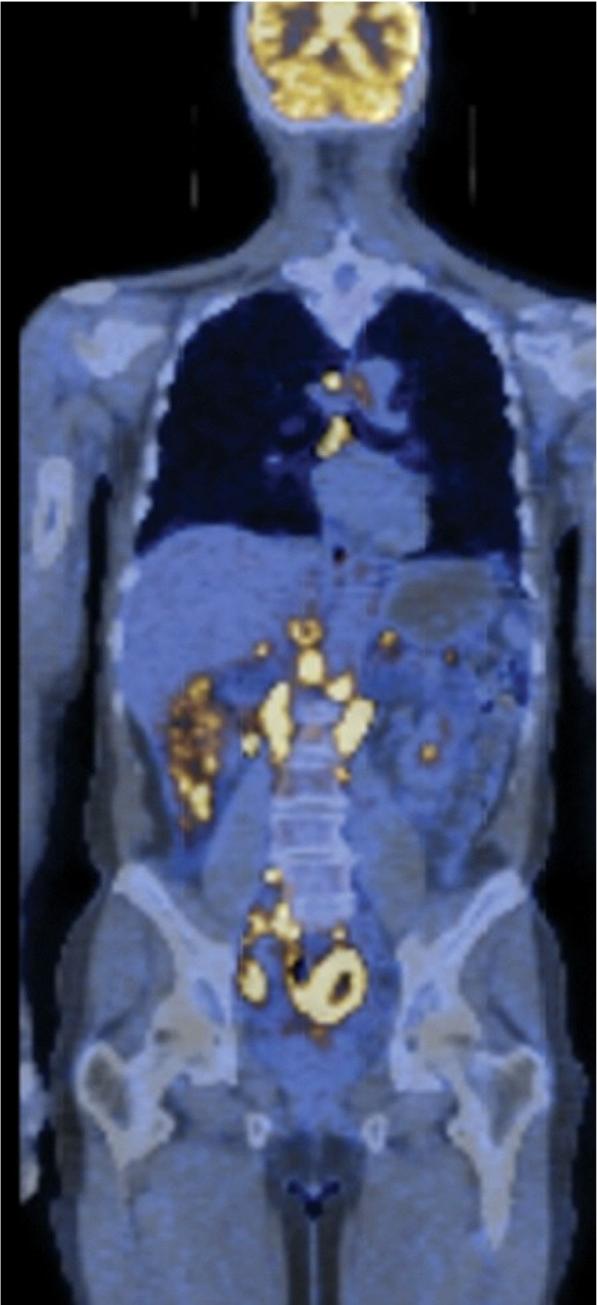
(grossly simplified scenario)

# Multiple genetic events underly cancer progression



# Multiple genetic events underly cancer progression





- 90% of cancer deaths are linked to distant metastases

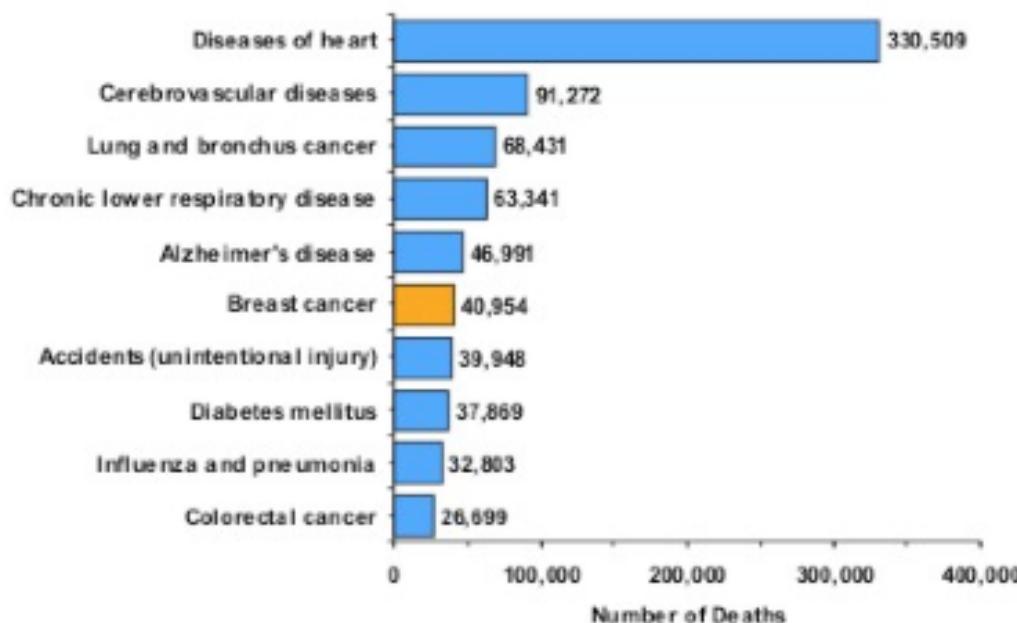
Figure 14-1 The Biology of Cancer (© Garland Science 2007)

Application

# Predicting breast cancer outcome with microarrays

# Breast cancer incidence and mortality

- Lifetime probability of developing BC is 1/8
- 190,000 women diagnosed with BC in 2004 in the USA
- Five year survival rate is 80-85%
- 41,000 BC-related deaths in the USA in 2004



CDC statistics, 2004

# Adjuvant therapy reduces the risk of disease relapse after surgery

- Adjuvant therapy is a systemic treatment given to patients soon after surgery
- The goal is to kill undetected cancer cells missed by surgery and/or may that have spread out of the primary tumor
- It increases survival rate by 20-30%
- Treatment decision depends on a number of factors such as tumor size, invasion of lymph nodes, estrogen receptor status, patient age, etc.

# Adjuvant therapy incurs significant side effects

Hormone therapy interferes with the estrogen receptor

- Menopausal symptoms
- Increased risk of endometrial cancer
- Typically prescribed for 5 years

Chemotherapy damages replicating cells

- Fatigue and impaired quality of life
- Myelosuppression
- Cardiac toxicity
- Ovarian failure with risk of irreversible infertility
- **Accurate prediction of disease outcome would spare adjuvant therapy to patients at low risk of relapse.**
- **Can existing clinical predictors be improved?**

# Microarray expression profiles predict survival in early breast cancer patients

## The New England Journal of Medicine

---

Copyright © 2002 by the Massachusetts Medical Society

VOLUME 347

DECEMBER 19, 2002

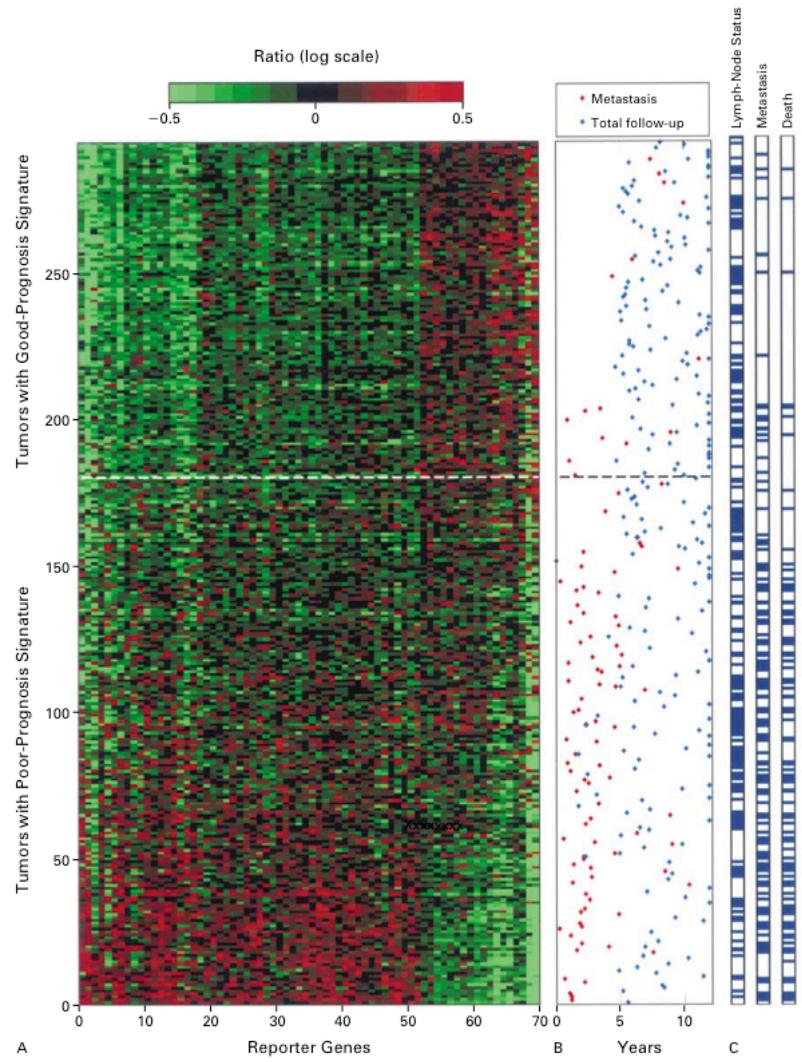
NUMBER 25



### A GENE-EXPRESSION SIGNATURE AS A PREDICTOR OF SURVIVAL IN BREAST CANCER

MARC J. VAN DE VIJVER, M.D., PH.D., YUDONG D. HE, PH.D., LAURA J. VAN 'T VEER, PH.D., HONGYUE DAI, PH.D.,  
AUGUSTINUS A.M. HART, M.Sc., DORIEN W. VOSKUIL, PH.D., GEORGE J. SCHREIBER, M.Sc., JOHANNES L. PETERSE, M.D.,  
CHRIS ROBERTS, PH.D., MATTHEW J. MARTON, PH.D., MARK PARRISH, DOUWE ATSMA, ANKE WITTEVEEN,  
ANNUSKA GLAS, PH.D., LEONIE DELAHAYE, TONY VAN DER VELDE, HARRY BARTELINK, M.D., PH.D.,  
SJOERD RODENHUIS, M.D., PH.D., EMIEL T. RUTGERS, M.D., PH.D., STEPHEN H. FRIEND, M.D., PH.D.,  
AND RENÉ BERNARDS, PH.D.

# Microarray expression profiles predict survival in early breast cancer patients

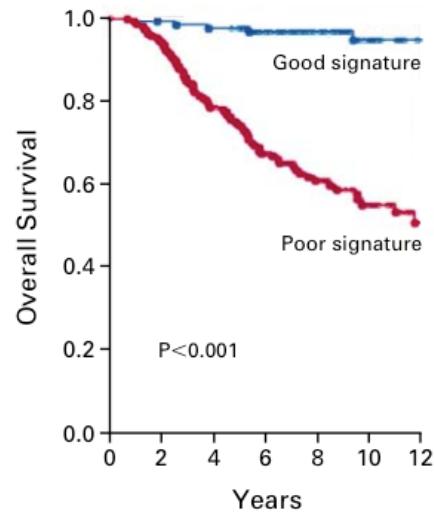


- Genome-wide expression profiles of 295 early breast cancers
  - Supervised search for genes associated with survival
- ⇒ 70 genes outcome predictor

(van de Vijver *et al.* NEJM, 2002)

# Microarray expression profiles predict survival in early breast cancer patients

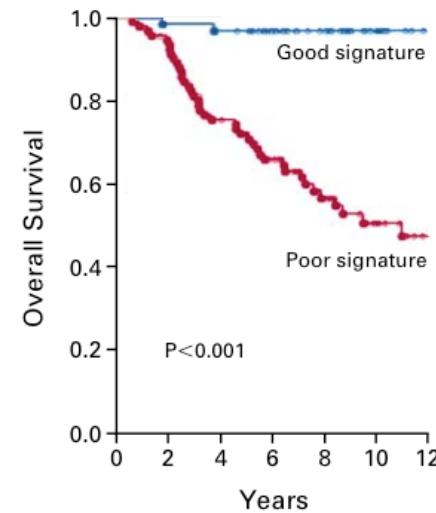
B All Patients



No. AT RISK

Low risk	115	114	112	91	65	43	23
High risk	180	167	134	100	62	40	19

D Lymph-Node-Negative Patients



No. AT RISK

Good signature	60	59	58	48	35	24	12
Poor signature	91	86	66	50	33	21	10

(van de Vijver *et al.* NEJM, 2002)

# The presence of metastasis predictors in the bulk of primary tumors challenges the ‘rare seeds’ theory of metastatic progression

## concepts

### A progression puzzle

René Bernards and Robert A. Weinberg

Most, if not all, human tumours develop through a succession of genetic and epigenetic changes that confer increasingly neoplastic (cancer-like)

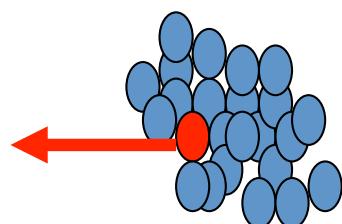
conceptual inconsistency: the genes that specify the final step in tumour progression — metastasis — would not seem to confer increased proliferative benefit at the primary site. That is, there is no reason to think that a metastatic phenotype enables cells to proliferate more effectively within the primary

### Metastasis genes

*The prevailing model of tumour progression carries with it a striking conceptual inconsistency.*

(Bernards & Weinberg, *Nature*, 2002)

### Rare “seed” model



○ =cell with *no* metastatic potential

● =cell with metastatic potential

# The presence of metastasis predictors in the bulk of primary tumors challenges the ‘rare seeds’ theory of metastatic progression

## concepts

### A progression puzzle

René Bernards and Robert A. Weinberg

Most, if not all, human tumours develop through a succession of genetic and epigenetic changes that confer increasingly neoplastic (cancer-like)

conceptual inconsistency: the genes that specify the final step in tumour progression — metastasis — would not seem to confer increased proliferative benefit at the primary site. That is, there is no reason to think that a metastatic phenotype enables cells to proliferate more effectively within the primary

### Metastasis genes

*The prevailing model of tumour progression carries with it a striking conceptual inconsistency.*

(Bernards & Weinberg, *Nature*, 2002)

Microarrays measure bulk expression

Thus, it seems that the metastatic potential of a primary tumor is present in the bulk of tumor cells, *not* rare ‘seeds’

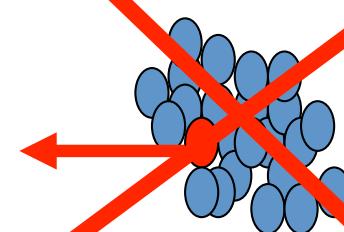


=cell with *no* metastatic potential

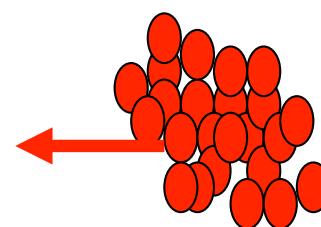


=cell with metastatic potential

Rare “seed” model



“Bulk” model



# The presence of metastasis predictors in the bulk of primary tumors challenges the ‘rare seeds’ theory of metastatic progression

## concepts

### A progression puzzle

René Bernards and Robert A. Weinberg

Most, if not all, human tumours develop through a succession of genetic and epigenetic changes that confer increasingly neoplastic (cancer-like)

conceptual inconsistency: the genes that specify the final step in tumour progression — metastasis — would not seem to confer increased proliferative benefit at the primary site. That is, there is no reason to think that a metastatic phenotype enables cells to proliferate more effectively within the primary

### Metastasis genes

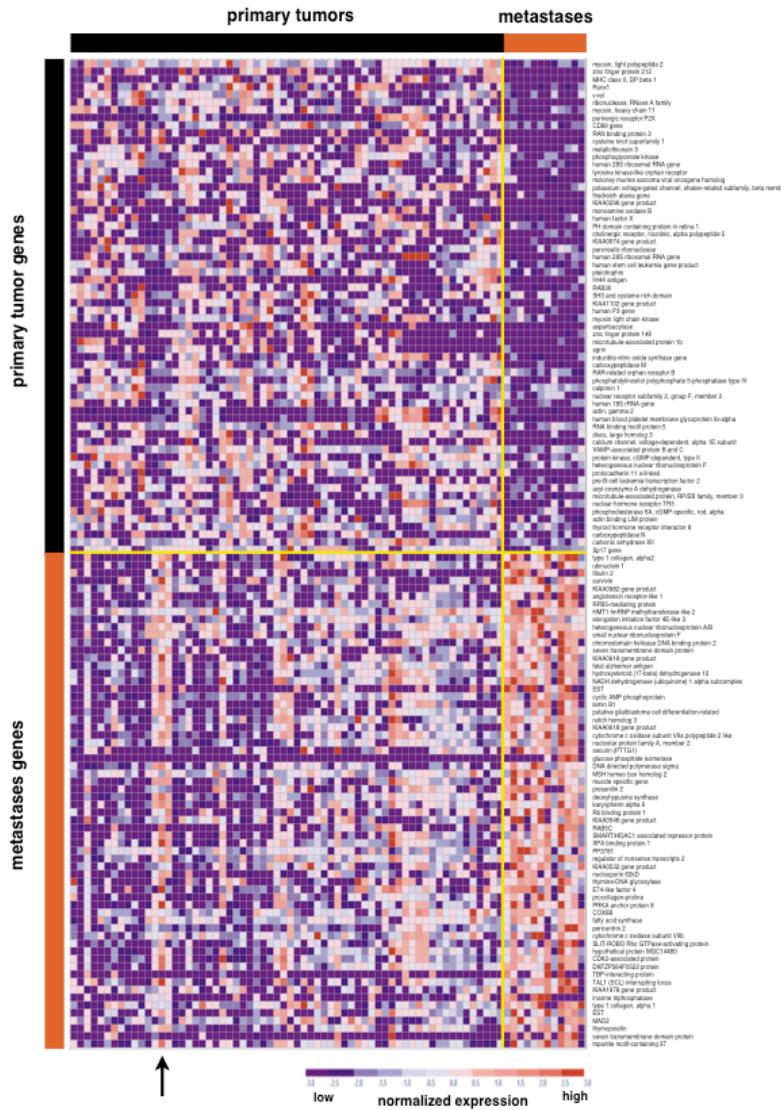
*The prevailing model of tumour progression carries with it a striking conceptual inconsistency.*

(Bernards & Weinberg, *Nature*, 2002)

**The progression puzzle:**  
How can metastatic cells outcompete others cells in the primary tumor in the absence of *local* Darwinian pressure to metastasize, i.e. to settle in a *distant* site they cannot see?

**Bernards-Weinberg stance:**  
*There is no metastasis genes as such.* Genes promoting metastasis also provide a growth advantage in the primary tumor. They are the classical oncogenes and tumor suppressors.

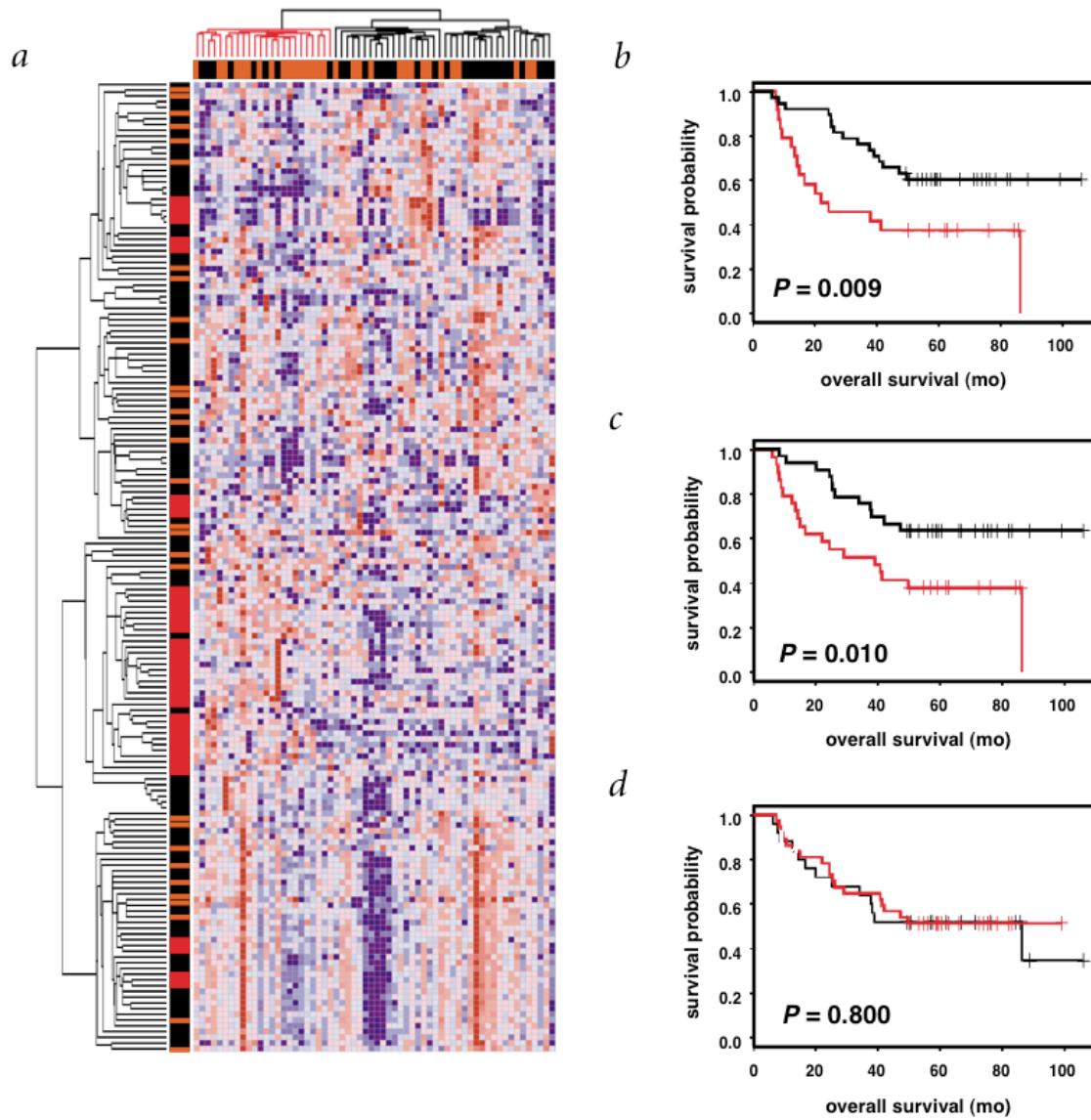
# Some metastasis markers are expressed in low grade, but bad prognosis, primary tumors



- 128 genes separate primary stage I/II from metastatic lung adenocarcinomas

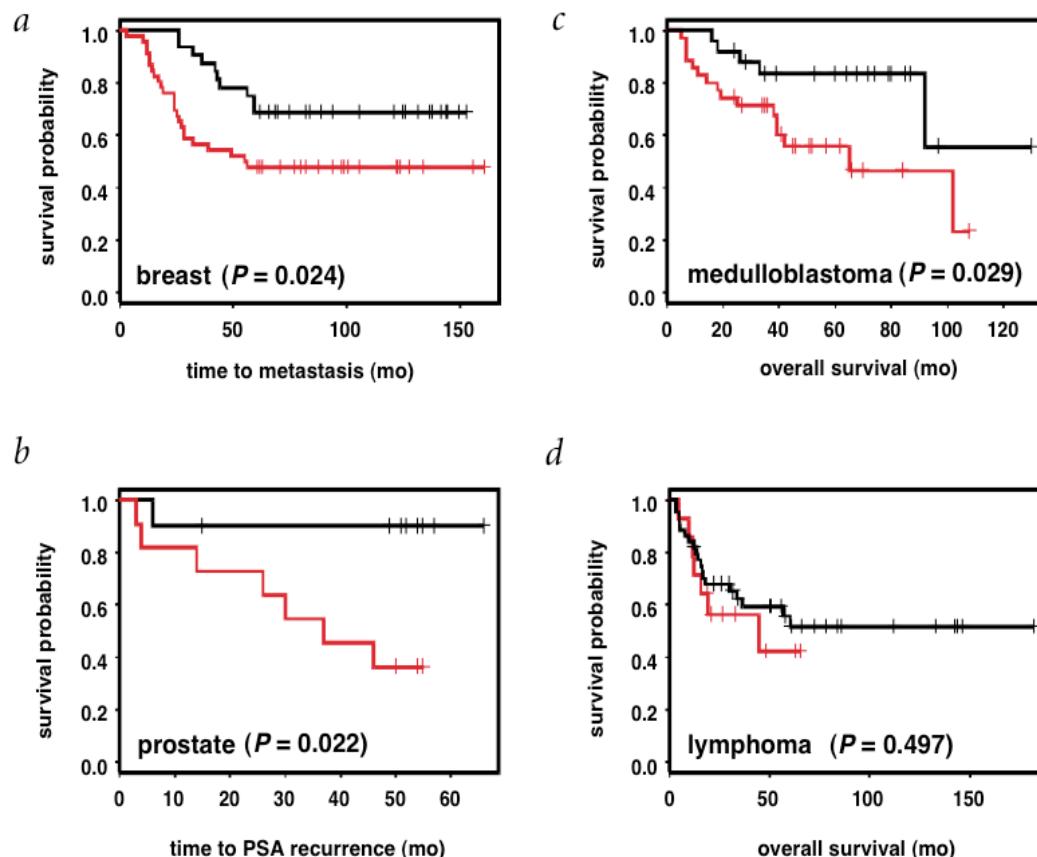
(Ramaswamy *et al.*, *Nature Genetics*, 2003)<sup>38</sup>

# Some metastasis markers are expressed in low grade, but bad prognosis, primary tumors



- The expression of the 128 genes metastasis signature in primary lung tumors is associated with outcome
- It may be reduced to a more tractable 17 genes predictor

# Some metastasis markers are expressed in low grade, but bad prognosis, primary tumors



- The metastasis signature seems to be universal
- It was extracted from independent expression data on the basis of a biological rationale, *not* a supervised approach
- Reuse of published data opens the door to unprecedented synthesis in cancer research

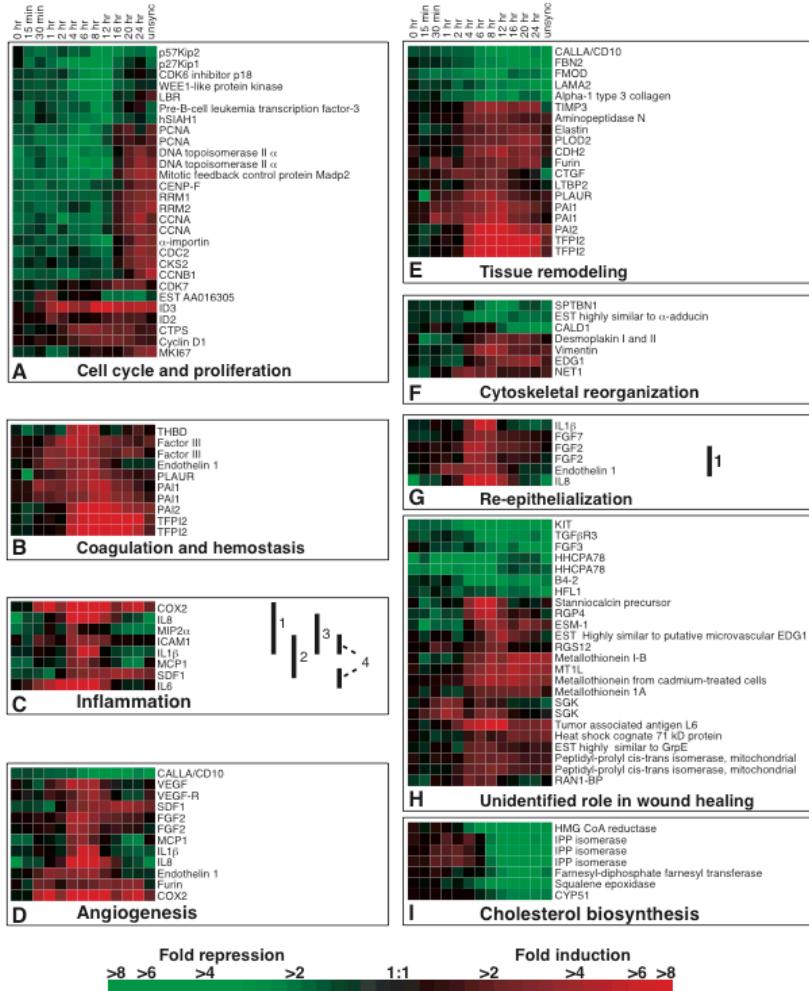
(Ramaswamy *et al.*, *Nature Genetics*, 2003)

# A vexing question

**What biological processes are  
the outcome predictors involved in?**

Beside sheer curiosity, this question has a bearing on devising treatments to control aggressive cancers

# The gene expression signature of a biological process may be obtained from *in vitro* experiments profiling

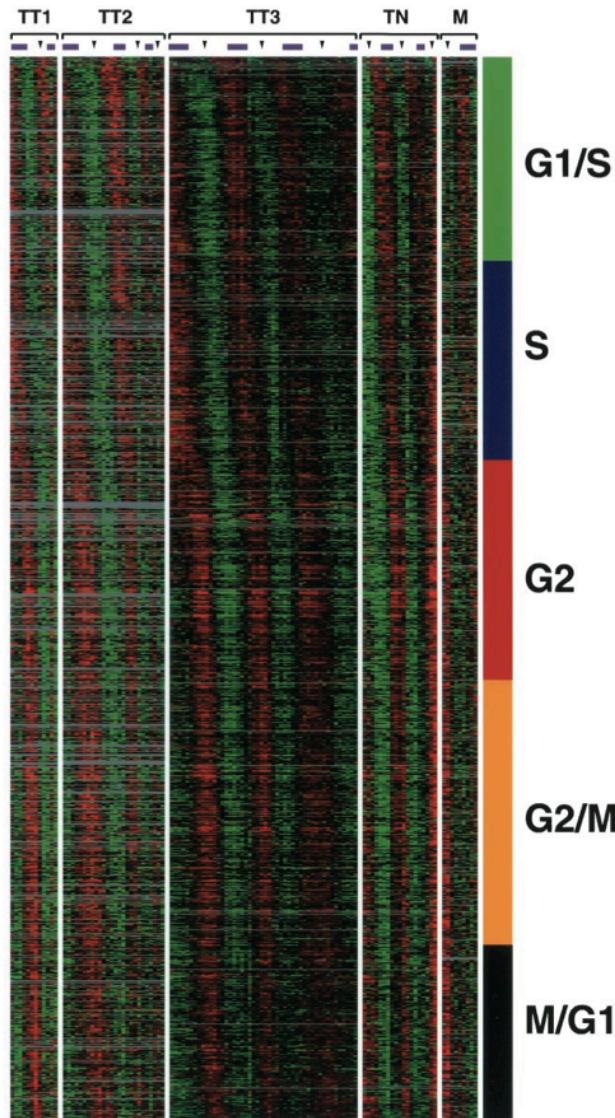


## Example #1: the serum response

- Serum response is the canonical *in vitro* model to study cell proliferation control
- Genome-wide expression was measured over time in serum-exposed skin fibroblasts, and compared to the expression profile of serum-deprived cells
- Surprisingly, proliferation genes expression is part of a much wider program reminiscent of the wound healing process

(Iyer et al., *Science*, 1999)

The gene expression signature of  
a biological process may be obtained from  
*in vitro* experiments profiling

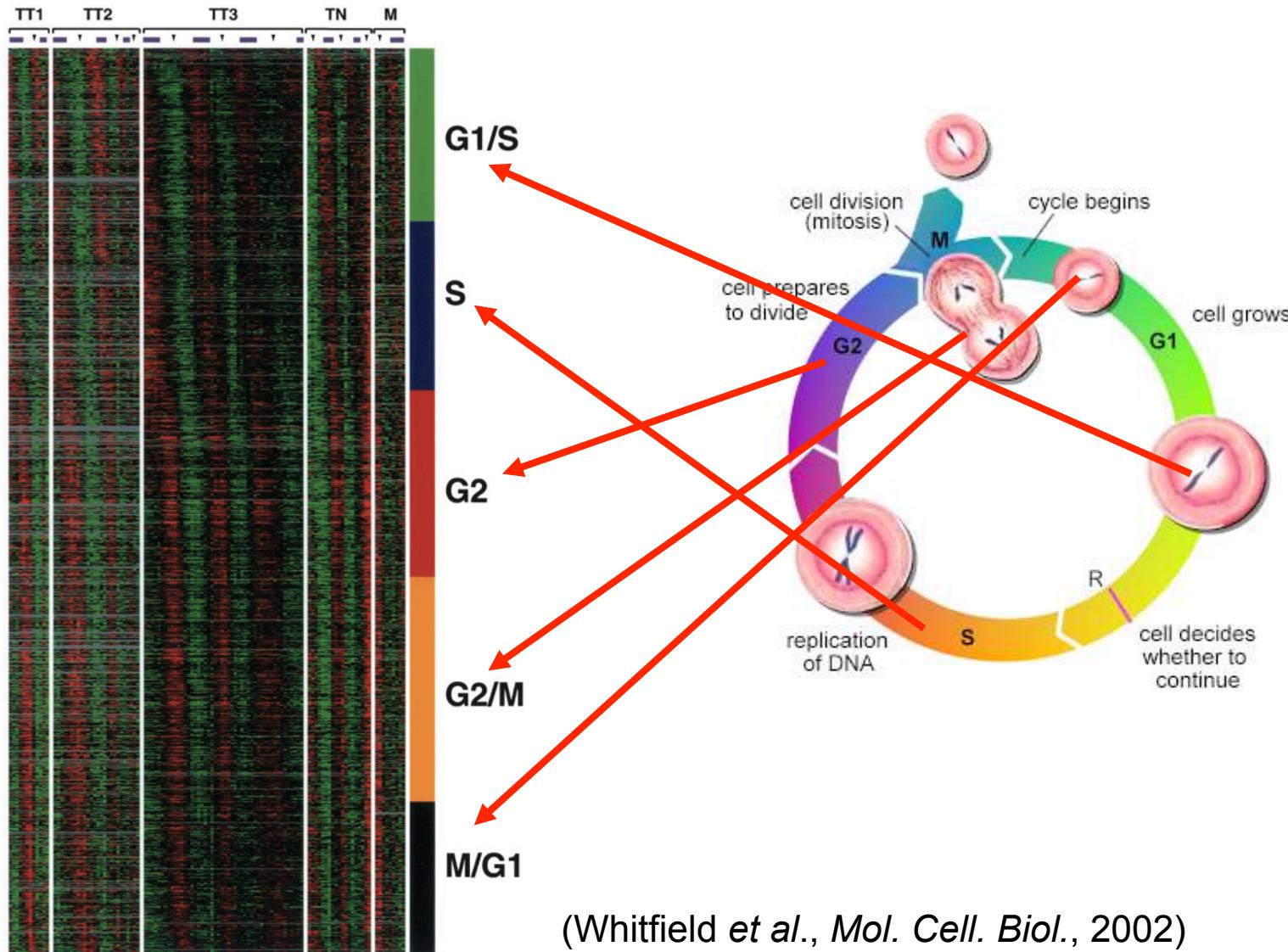


Example #2: the cell cycle

1. Measure genome-wide expression at evenly spaced time points in synchronized and nonsynchronized (i.e., control) cell lines
2. Detect genes showing periodic differential expression pattern with Fourier transform
3. Order periodic genes according to phase

(Whitfield *et al.*, *Mol. Cell. Biol.*, 2002)

# The gene expression signature of a biological process may be obtained from *in vitro* experiments profiling



# A ‘wound’ signature in cancers

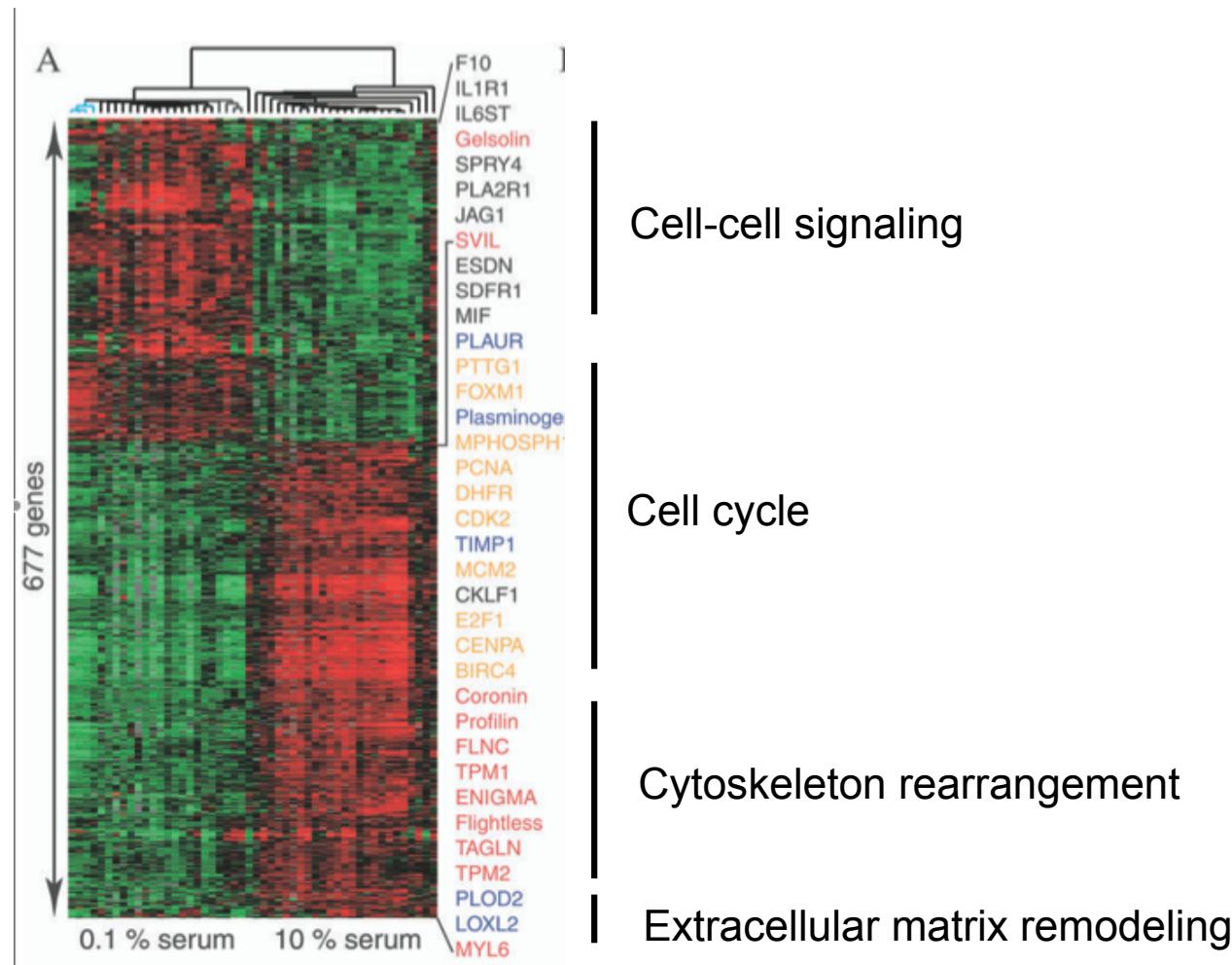
Malignant tumors resemble ‘wounds that do not heal’

- History of chronic inflammatory
- Fibroblasts and epithelial cells proliferation
- Local invasion (epithelial-mesenchymal transition)
- Angiogenesis

Genome-wide expression profiling may help capture comprehensively the many processes involved in wound healing

# A ‘wound’ signature in cancers

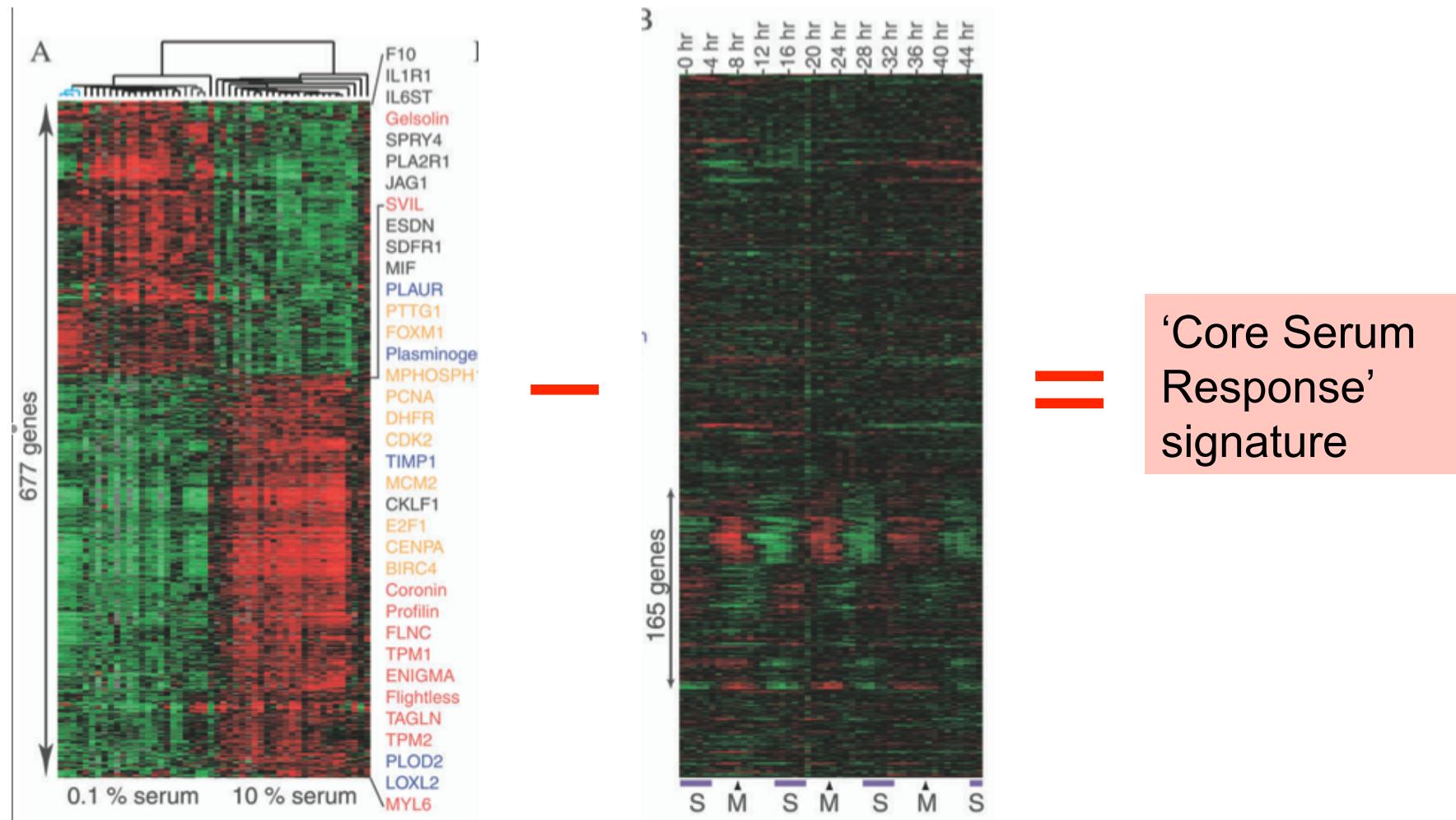
Genes differentially expressed  
in fibroblasts from 10 sites  
in 10% vs. 0.1% serum



(Chang et al., PLoS Biol., 2004)

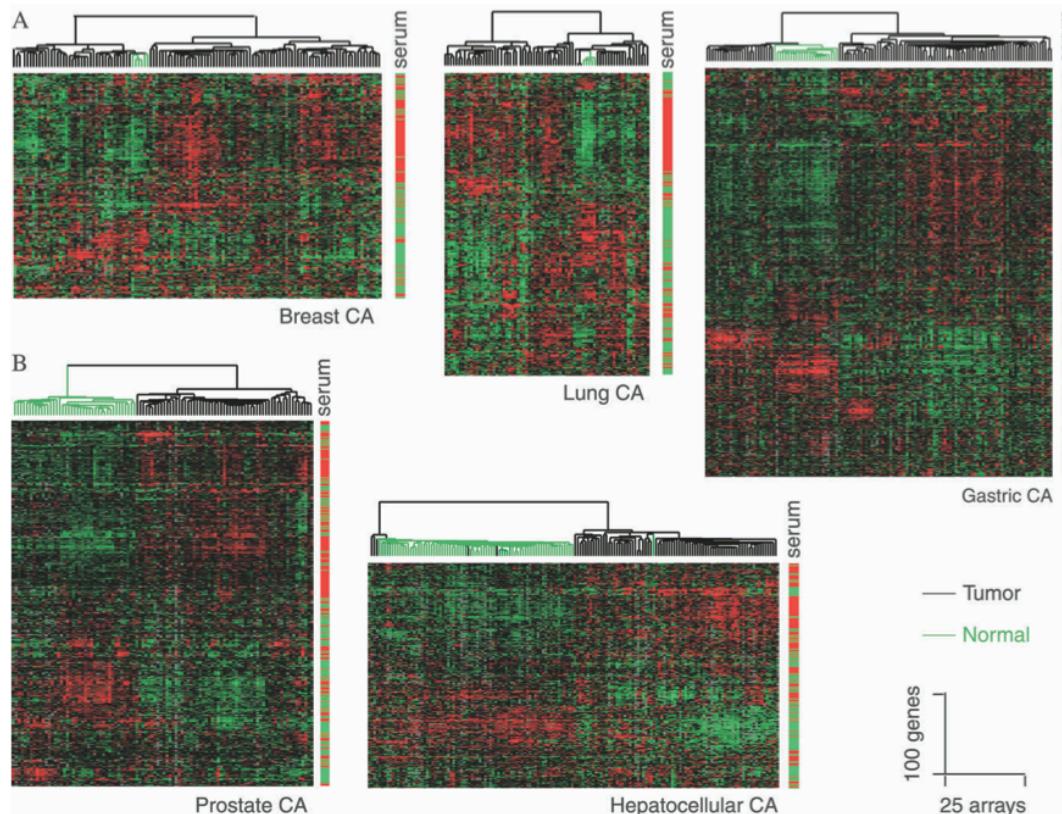
# A ‘wound’ signature in cancers

Genes differentially expressed  
in fibroblasts from 10 sites  
in 10% vs. 0.1% serum



(Chang et al., *PLoS Biol.*, 2004)

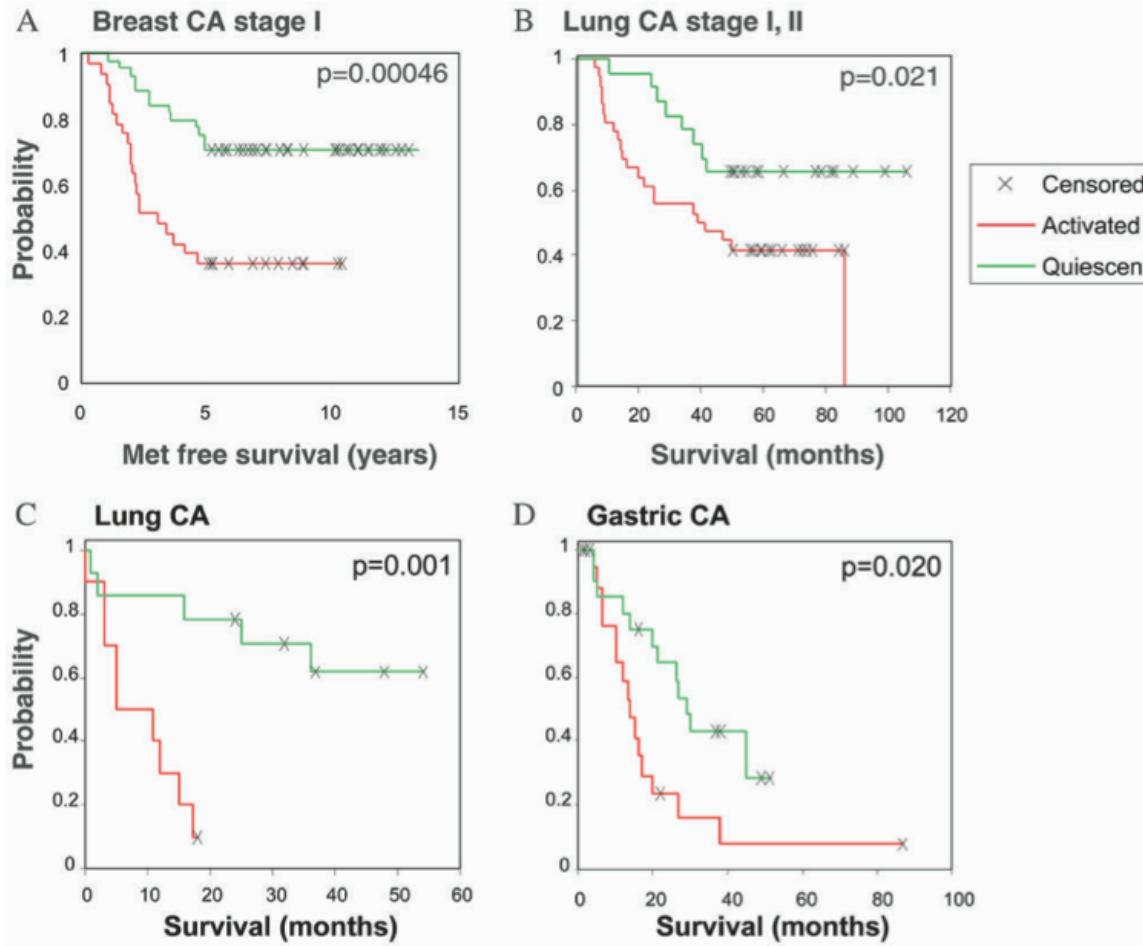
# A ‘wound’ signature in cancers



(Chang *et al.*, PLoS Biol., 2004)

- The core serum response signature is expressed in many cancers, but not in normal tissues
- Its expression is robust across several cancer sites

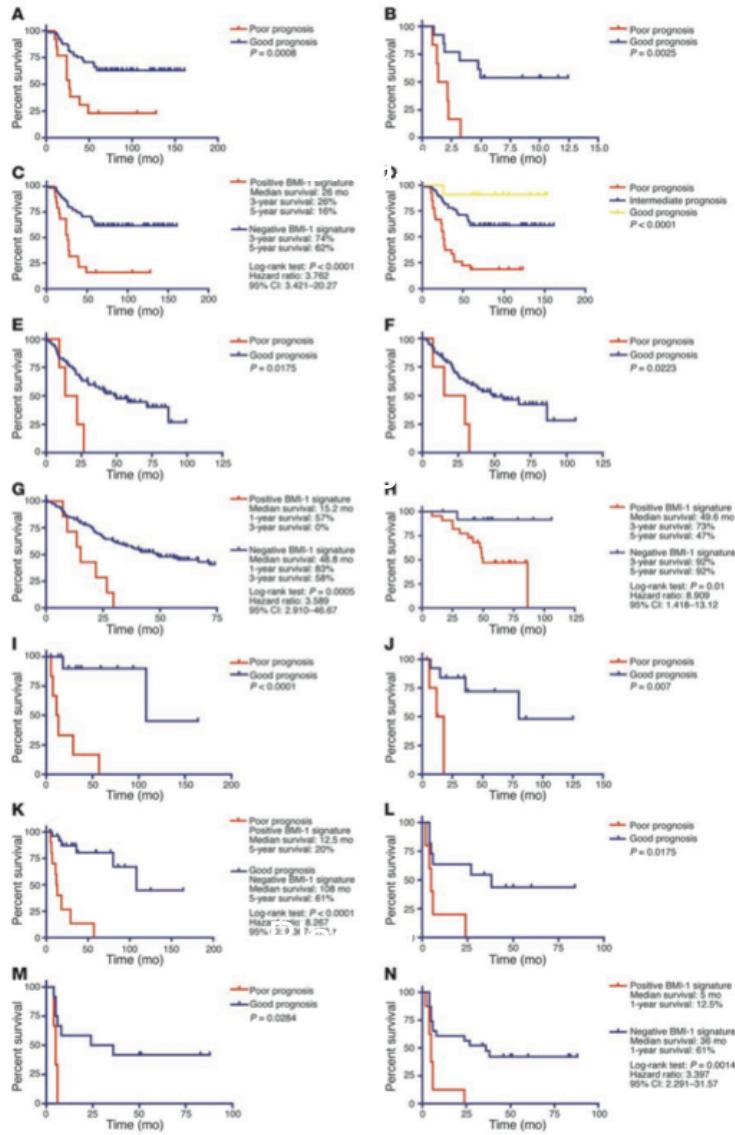
# A ‘wound’ signature predicts outcome



- Expression of the core serum response signature predicts outcome
- It suggests that aggressive cancers display a wound-like phenotype (we'll come back to this in a while...)
- A qualitative biological idea has been turned into quantitative molecular biology, *in vivo*

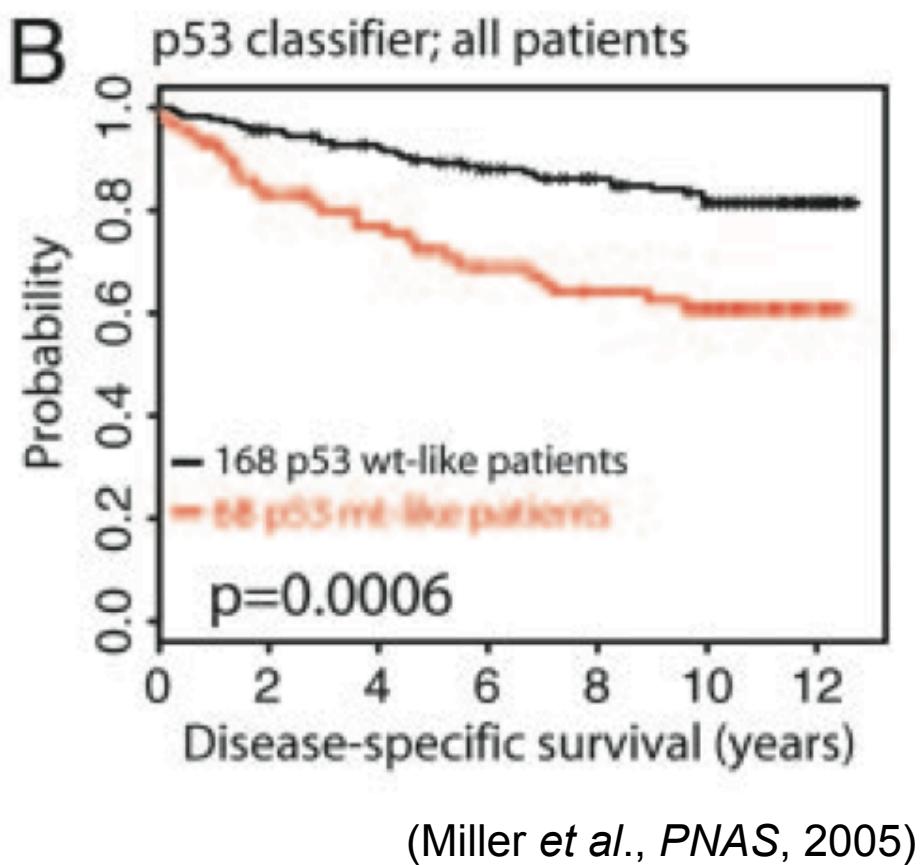
(Chang *et al.*, PLoS Biol., 2004)

# Poor outcome seems to be associated with a ‘stem cell-like’ phenotype



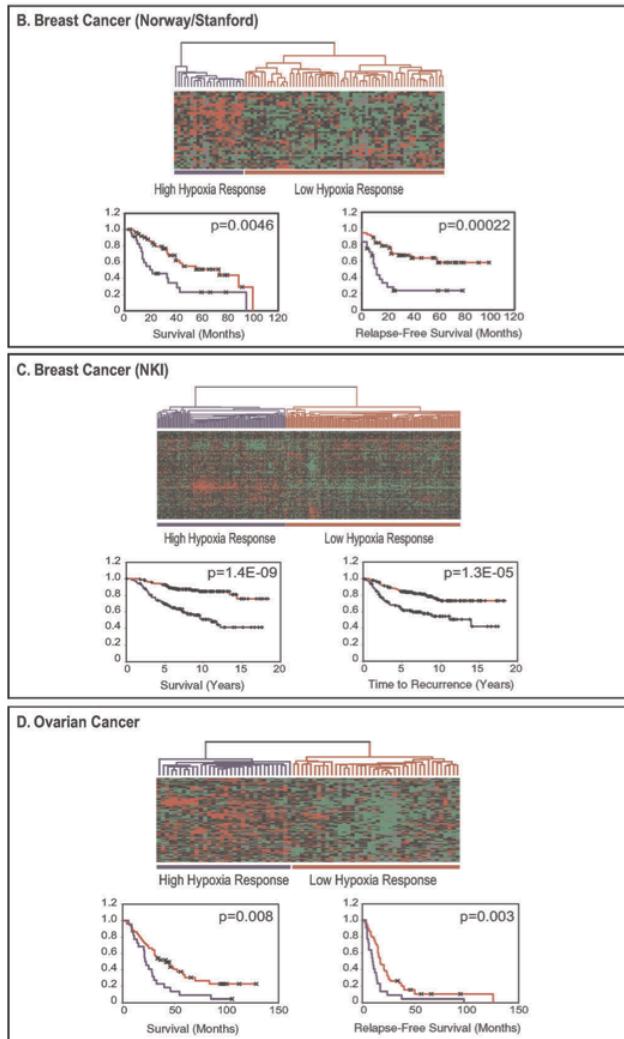
- 11 genes associated with stem cell marker BMI-1 in a mouse model of metastatic prostate cancer
- It seems to be a universal predictor of cancer outcome

# Poor outcome seems to be associated with a ‘p53 mutant’ phenotype



- p53 is a major regulator of proliferation and cell death
- 32 genes optimally classify primary tumors with p53 mutation from tumors with intact p53
- The p53 ‘mutant’ signature is expressed in several nonmutant tumors, these tumors tend to be aggressive

# Poor outcome seems to be associated with an hypoxia response



- 176 genes expressed in *in vitro*  $O_2$ -deprived epithelial cells from various anatomical sites
- Predict outcome of breast and ovarian cancers

(Chang *et al.*, PLoS Med., 2005)

>100 gene expression predictors of breast cancer outcome have been proposed so far

### Proliferation-related

	Name	Ref.	Biological/clinical interpretation	Survival analysis	N
Proliferation	Whitfield.2002	27	Genes periodically expressed in synchronized HeLa cells	None	1300
	Dai.2005	9	Genes associated with outcome in breast cancer and periodically expressed in synchronized HeLa cells	Breast	50
	super.PCNA	-	1% genes most correlated with PCNA in normal tissues	Breast	131
	NCI60.pd	36	Genes correlated with the doubling time of NCI60 cell lines (we recomputed this signature)	Breast	178

>100 gene expression predictors of breast cancer outcome have been proposed so far

### grade-related

Grade	Ma.2003	8	Genes associated with high grade and with preinvasive to invasive transition	None	38
	Ivshina.2006	10	Genes associated with grade	Breast	19
	Sotiriou.2006	11	—	Breast	97

>100 gene expression predictors of breast cancer outcome have been proposed so far

outcome-related (supervised approaches)

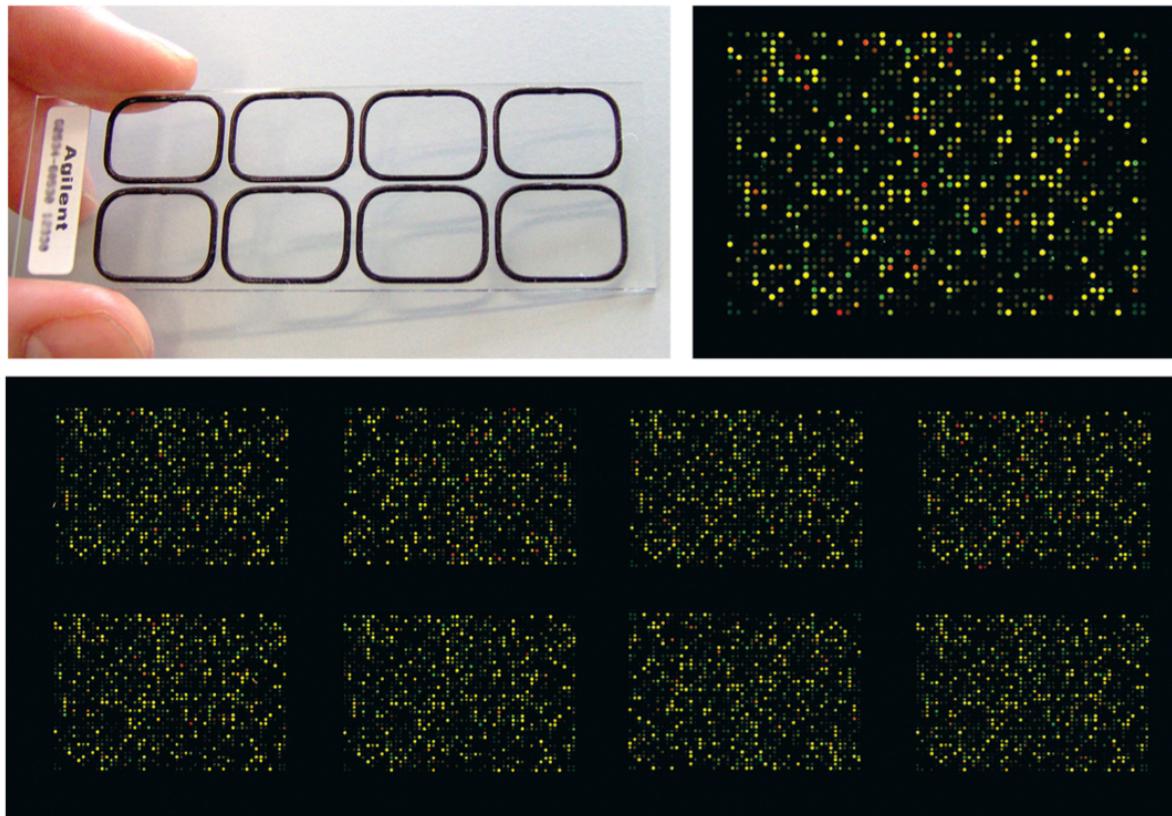
Outcome	van't Veer.2002	2	Genes associated with outcome	Breast	70
	Pawitan.2005	5	—	Breast	64
	Korkola.2007	7	—	Breast	21
	Paik.2004	4	Genes associated with outcome in node negative patients	Breast	21
	Wang.2005	6	—	Breast	76

>100 gene expression predictors of breast cancer outcome have been proposed so far

## Hypothesis-driven

hypothesis-driven	Ramaswamy.2003	12	Genes differentially expressed in metastases vs. primary tumors	Breast, prostate, glioma, lung	17
	Chang.2004	13	Gene associated with fibroblasts growth in 0.1% vs. 10% serum, cell cycle genes are removed. Reflect wound healing.	Breast, prostate, lung, glioma	512
	Rhodes.2004	26	Genes associated with dedifferentiation in lung, bladder, breast, ovarian and prostate cancers, and in glioma	None	69
	Glinsky.2005	14	Genes associated with stem cell-like phenotype	Breast, prostate, breast, lung, ovarian, bladder cancers, lymphoma, mesothelioma, medulloblastoma, glioma, and acute myeloid leukemia	11
	Miller.2005	15	Genes associated with p53 mutation	Breast	32
	Minn.2005	16	Genes associated with lung metastasis in mice	Breast	54
	Carter.2006	17	Genes associated with aneuploidy in the NCI60 cell lines	Breast, glioma, lung, lymphoma, mesothelioma, medulloblastoma	70
	Liu.2007	18	Genes associated with the CD44+CD24-/low phenotype in breast cancer cells	Breast	186
	Saal.2007	19	Genes associated with PTEN loss	Breast	246
	Welm.2007	20	Genes associated with macrophage stimulating protein expression, reflects chemotaxis and EMT	Breast	3

Several breast cancer outcome predictors, have been approved by the USA Food & Drug Administration and are marketed



(Glas *et al.*, *BMC Genomics*, 2006)

The 70 genes signature of the NKI is now a commercial product

It is expected to spare adjuvant therapy to 10% of the node-negative patients who will use it

# But do these signatures have different predictive abilities?

- Four signatures tested in van de Vijver data showed nearly identical predictive value (Fan *et al.*, NEJM, 2006)
- The same patients tended to be misclassified by all four signatures
- This suggested that although they have different biological rationale, the signature predictive ability rely on the *same* biological parameter
- **What biological parameter underlies outcome predictions?**
- **Is it possible to predict significantly better than Mammaprint's 70 genes signature?**