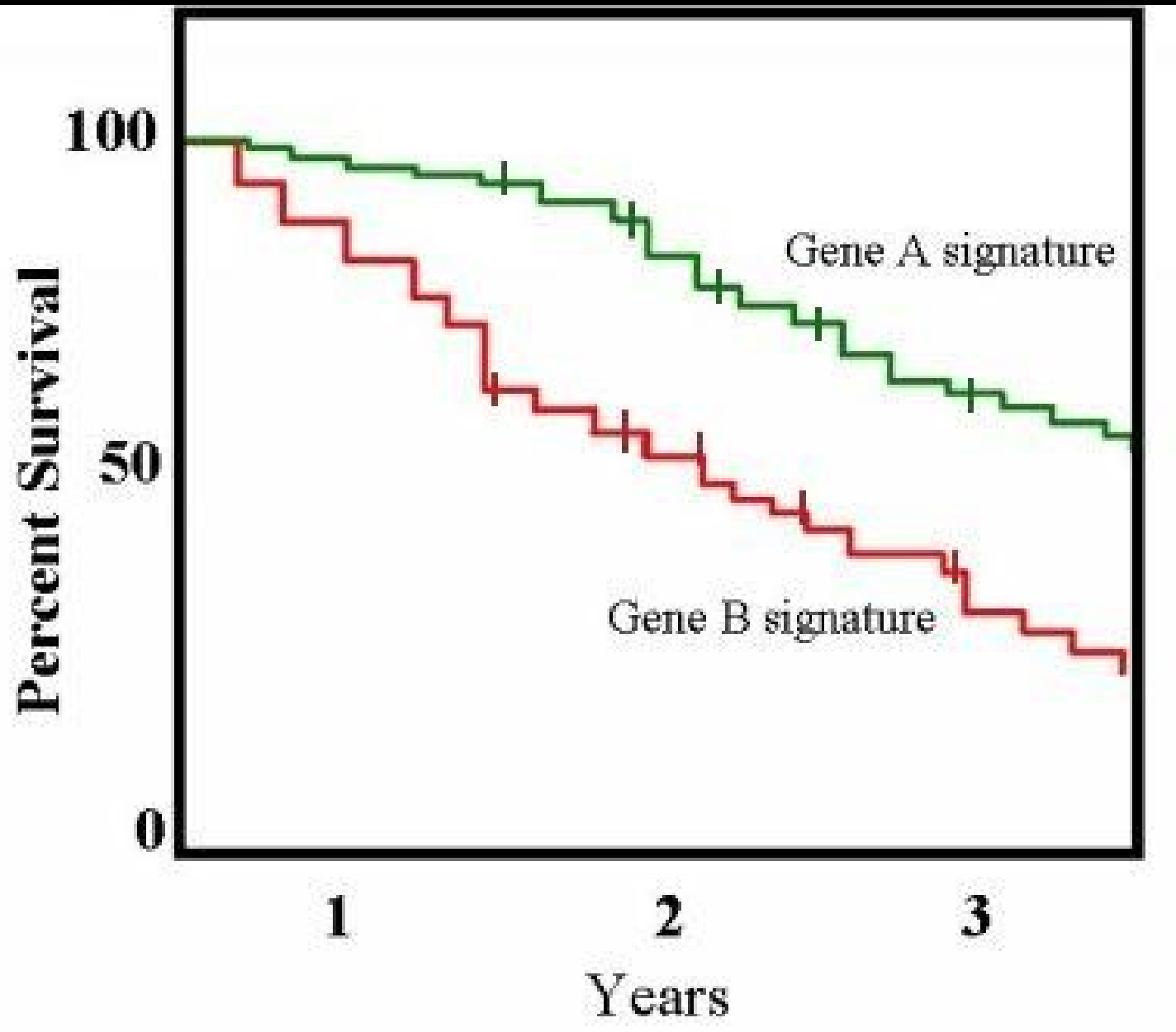


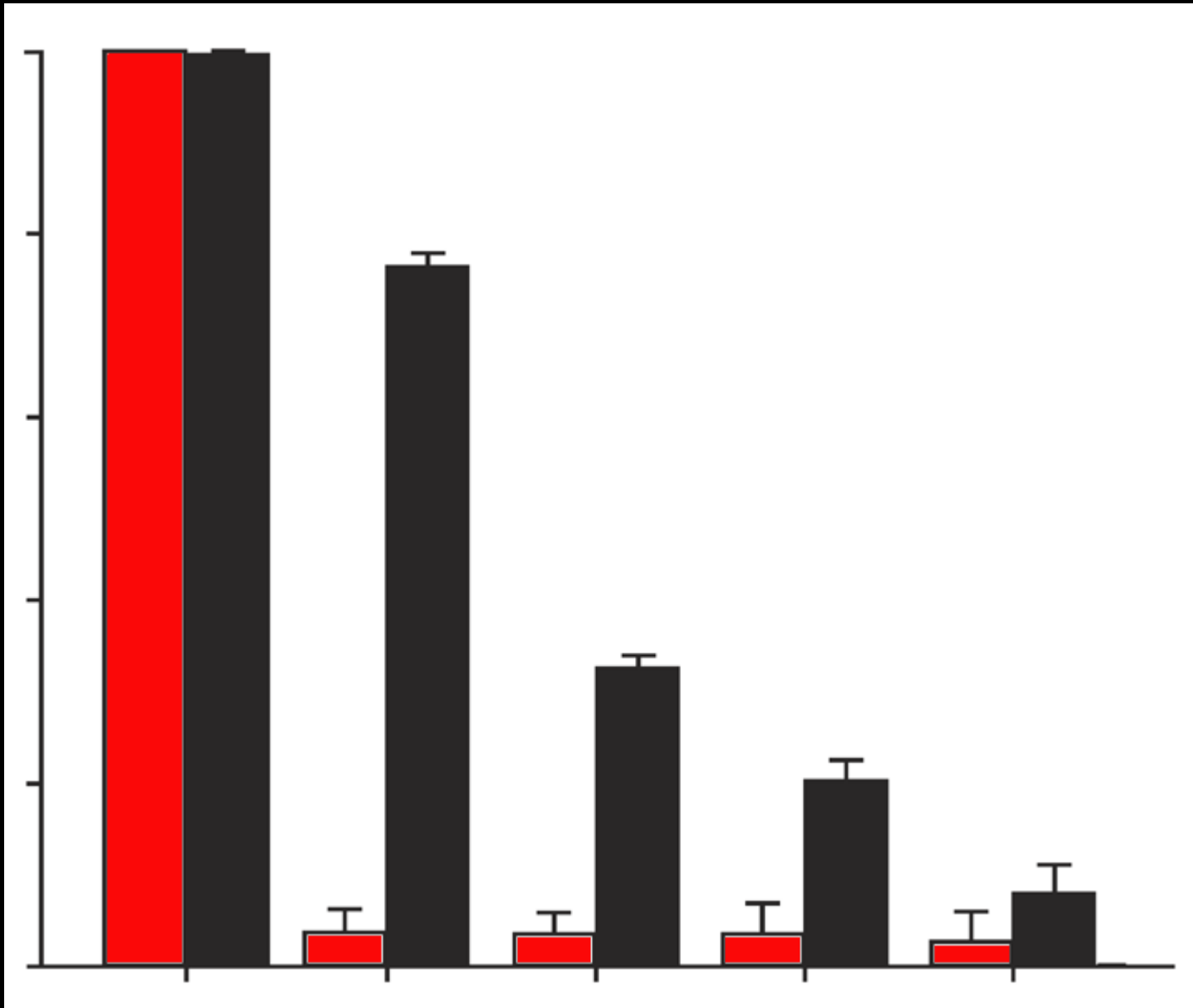
CAD Laboratory

February 7th, 2009

gene expression

gene expression signatures





Oncogenic pathway signatures in human cancers as a guide to targeted therapies

Andrea H. Bild^{1,2}, Guang Yao^{1,2}, Jeffrey T. Chang^{1,2}, Quanli Wang¹, Anil Potti^{1,4}, Dawn Chasse^{1,2}, Mary-Beth Joshi³, David Harpole³, Johnathan M. Lancaster⁷, Andrew Berchuck⁵, John A. Olson Jr^{1,3}, Jeffrey R. Marks³, Holly K. Dressman^{1,2}, Mike West⁶ & Joseph R. Nevins^{1,2}

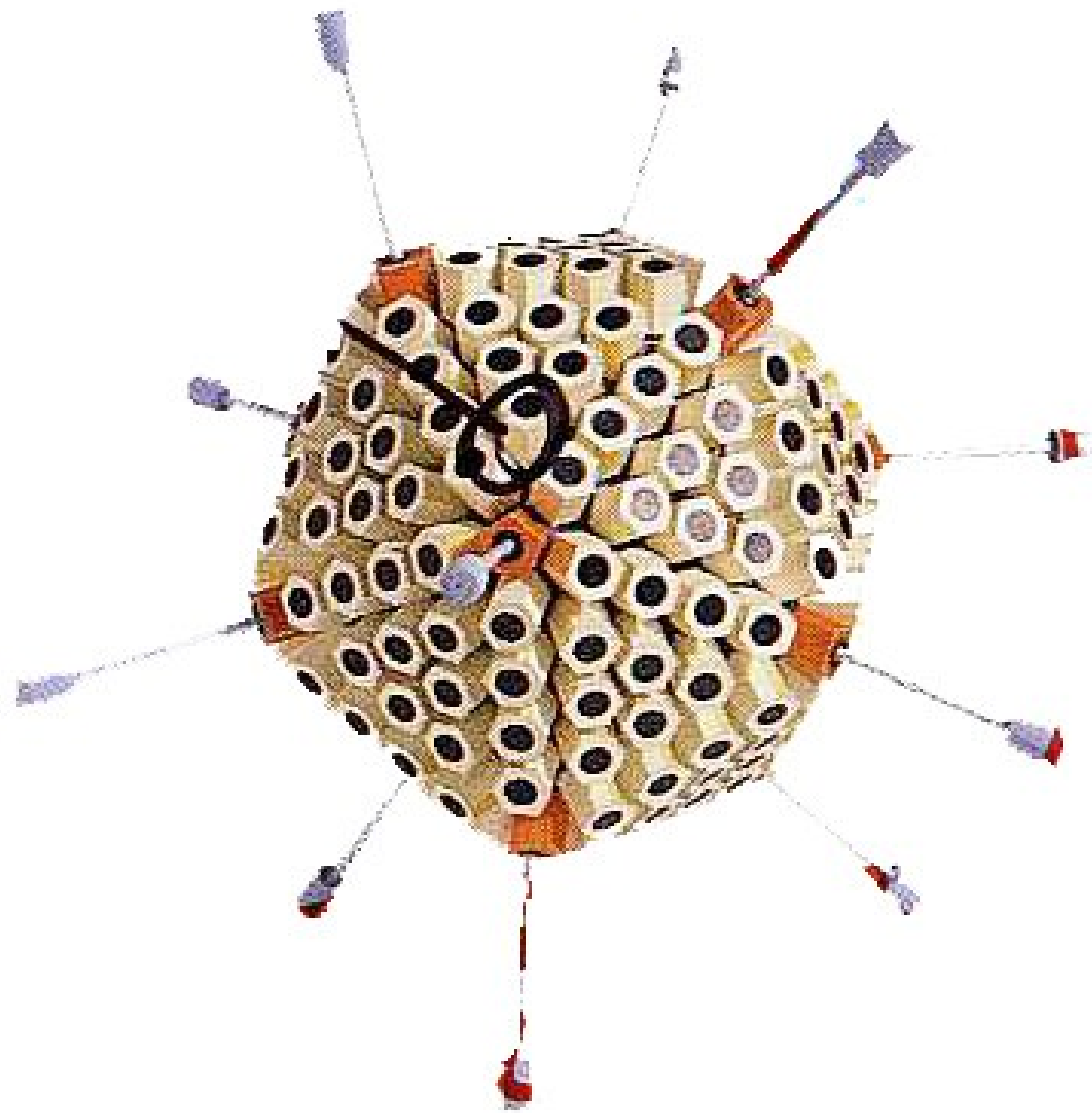
MYC

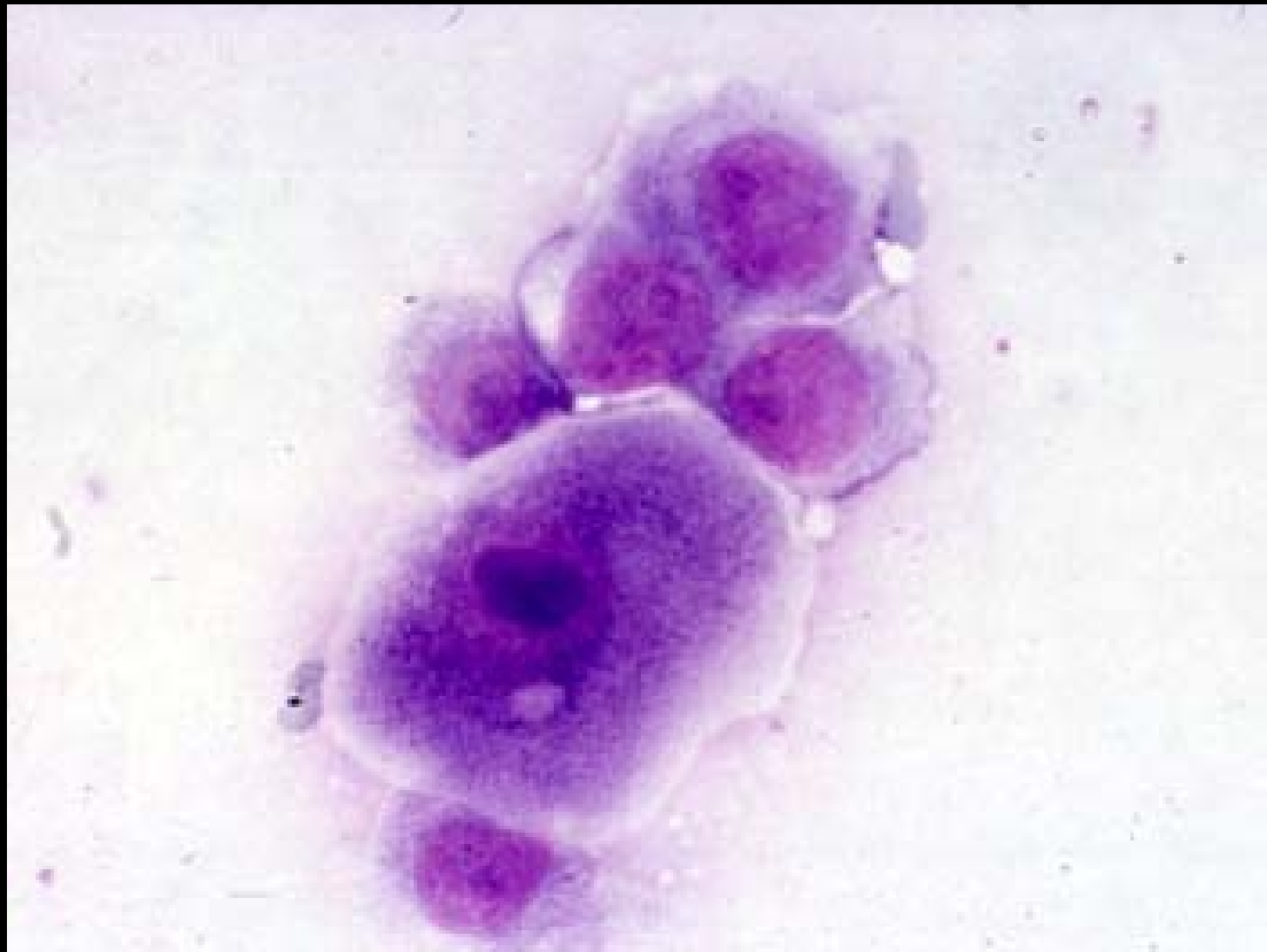
Ras

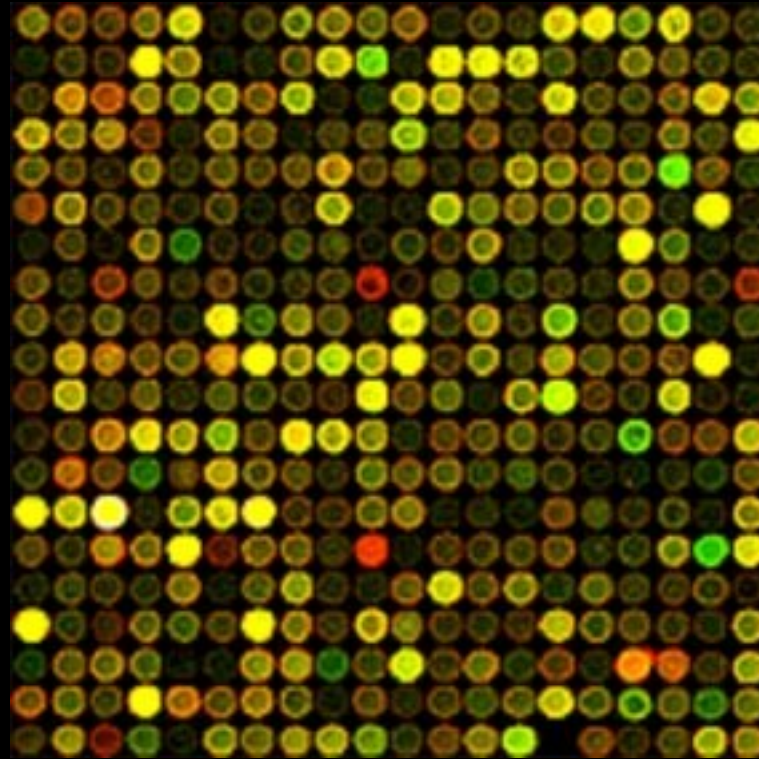
E2F3

b-cat

Src





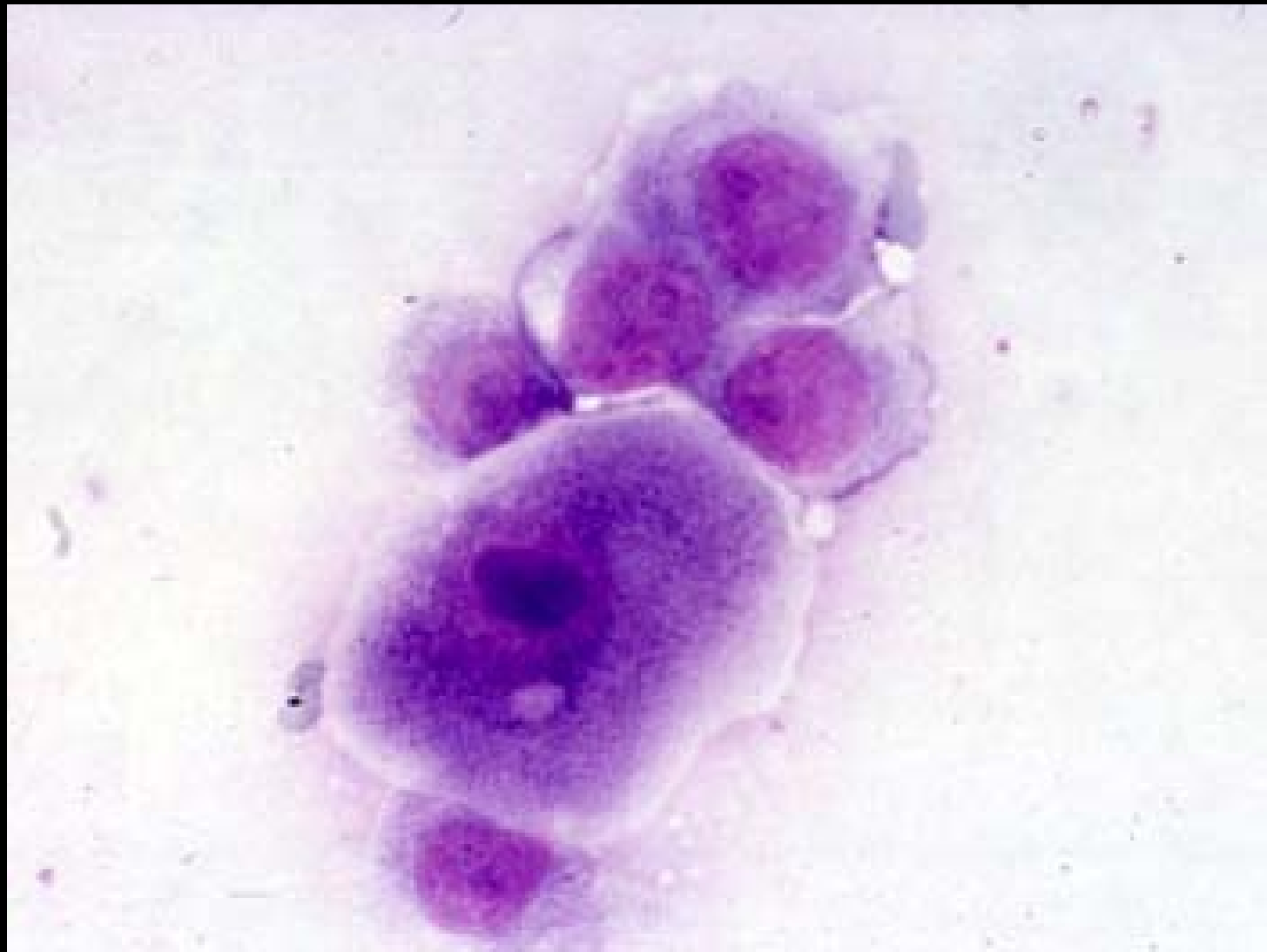


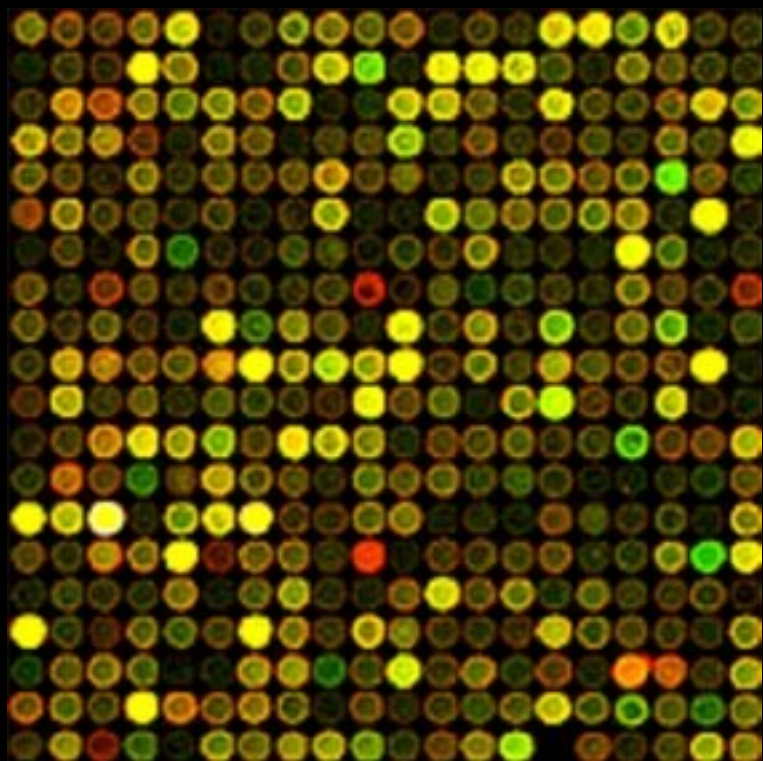
Which genes are downstream?

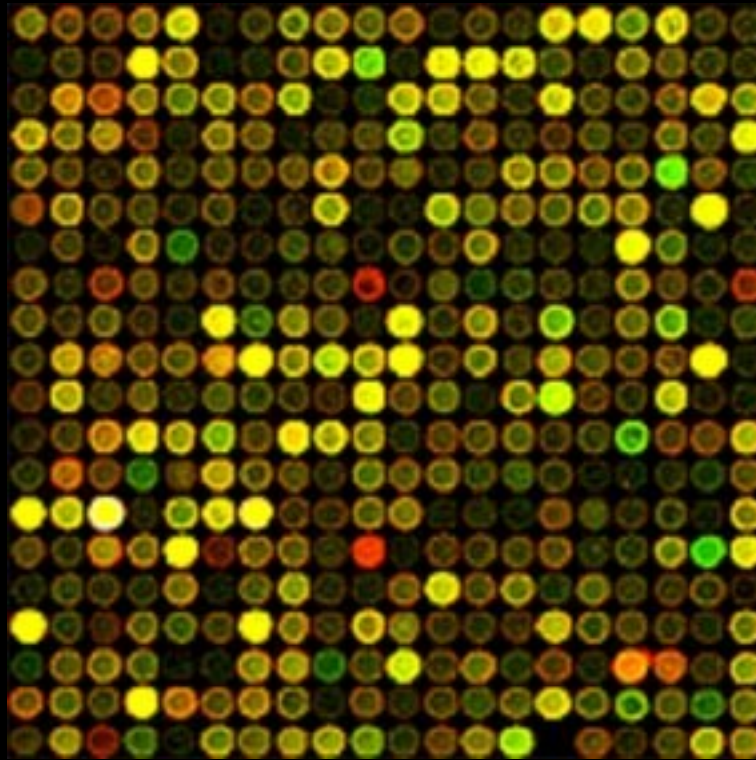
IGF-I gene signature in breast cancer

**Chad J. Creighton, Angelo Casa, ZaWaunyka Lazard, Shixia Huang, Anna
Tsimelzon, Susan G. Hilsenbeck, C. Kent Osborne, and Adrian V. Lee**

IGF-1







Which genes are affected?

homework

Bild dataset

determine gene signatures

IGF-1 dataset

determine gene signatures

compare

TTEST

`TTEST(array1, array2, tails, type)`

- `array1` is the first data set.
- `array2` is the second data set.
- `tails` specifies the # of distribution tails
(Use "2")
- `type` is the kind of t-Test to perform
(Use "2").

MATCH

MATCH(lookup_value,lookup_array,match_type)

- **lookup_value** **what value are you looking for?**
- **Lookup_array** **range of cells**
- **match_type** **should be 0 for our purposes.**

COUNT

COUNT (range)

- **range** **cells to count**

- **Only numbers in a range are counted. Empty cells, logical values, text, or error values in the array or reference are ignored.**

Data > Filter > AutoFilter

	A	B
1	Salesperson	Region
2	Buchanan	South
3	Davolio	West
4	Suyama	
5	Suyama	
6	Dodsworth	
7	Davolio	
8	Davolio	
9	Suyama	
10	Davolio	
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21	Davolio	South
22	Davolio	East
23		
24		
25		
26		
27		
28		
29		
30		

- arrows appear to the right of the column labels
- filtered items appear in blue.
- complex criteria: rows that contain values within a specific range (e.g. $p < 0.01$)

Table 1. Number of transcripts in each gene expression signature.

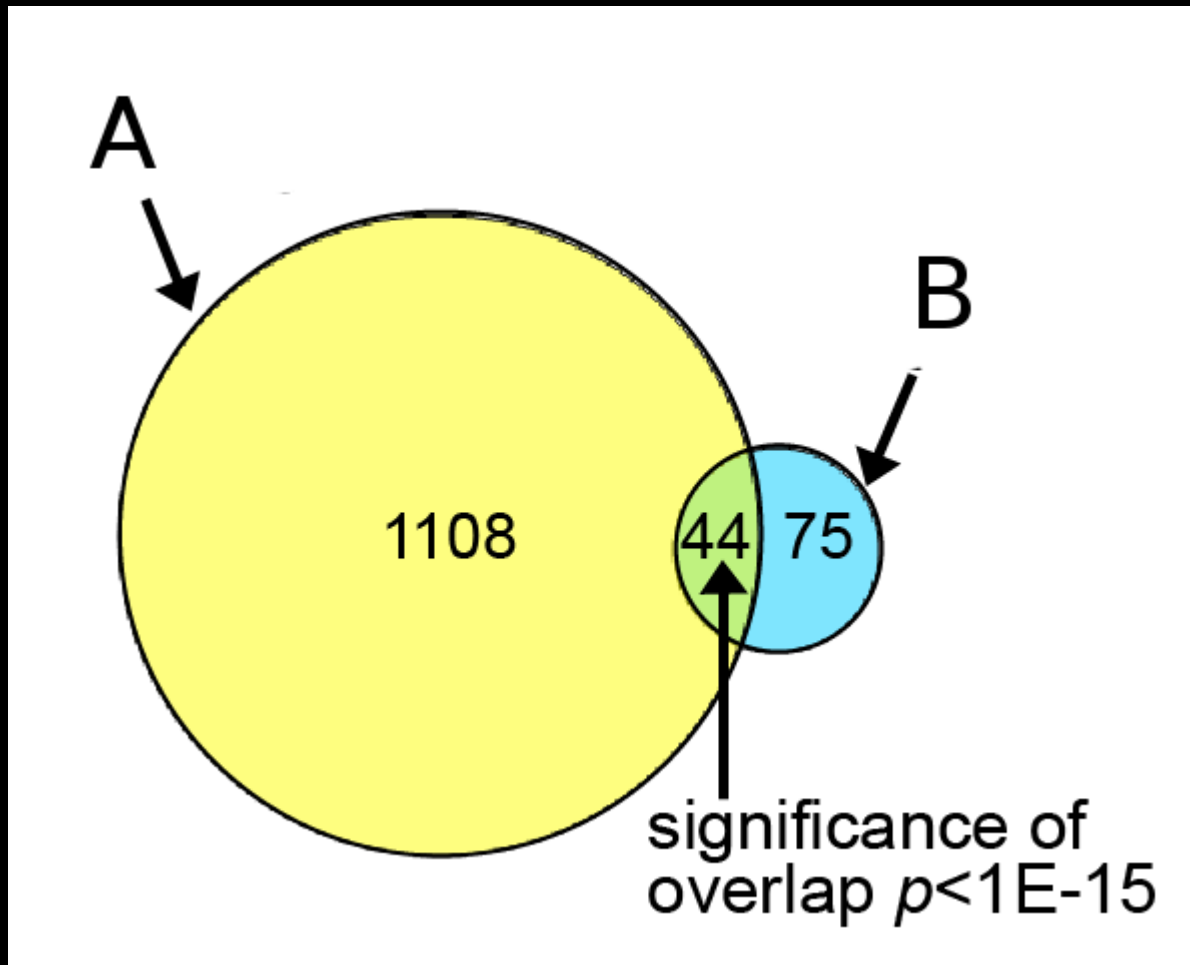
	up	down
IGF		
Myc		
Ras		
E2F3		
beta-catinin		
Src		

Table 2. Transcripts intersecting between gene sets.

	Number intersecting	P-value*
IGF up X Myc up		
IGF up X Ras up		
IGF up X E2F3 up		
IGF up X Bcat up		
IGF up X Src up		
IGF down X Myc down		
IGF down X Ras down		
IGF down X E2F3 down		
IGF down X Bcat down		
IGF down X Src down		

* by one-sided Fisher's exact test.

Compare two signatures



Sig A: **1152**

Sig B: **119**

Genes on both
platforms: **11079**

Genes shared by both
gene signatures: **44**

one-sided
Fisher's exact test

dhypcr

Sig A: 1162

Sig B: 119

Genes on both platforms: 11079

Genes shared by both gene signatures: 44

```
> m<-119           #number of Sig B genes
> n<-11079-119     #number of other genes
> k<-1162          #number of Sig A genes
> x<-0:k           #vector of successes
> 1-sum(dhyper(x,m,n,k)[1:44])
```

```
[1] 1.265654e-14
```

Table 1. Number of transcripts in each gene expression signature.

	up	down
IGF		
Myc		
Ras		
E2F3		
beta-catinin		
Src		

Table 2. Transcripts intersecting between gene sets.

	Number intersecting	P-value*
IGF up X Myc up		
IGF up X Ras up		
IGF up X E2F3 up		
IGF up X Bcat up		
IGF up X Src up		
IGF down X Myc down		
IGF down X Ras down		
IGF down X E2F3 down		
IGF down X Bcat down		
IGF down X Src down		

* by one-sided Fisher's exact test.

GSEA

Chris Miller

chrisamiller@gmail.com