Cancer genetics/genomics

I. Germline: predisposing mutations - best appreciated
are inactivating mutations in tumor suppressor genes
A. Rb (retinoblastoma)
B. BRCA1
C. p53 (Li-Fraumeni)

II. Somatic mutation and progression same genes and many others
A. mutation rate changes
B. activating mutations, oncogenes, Ras
C. gene dose changes, gene amplification N-myc
D. classic translocations, rhabdomyosarcomas
E. methylation changes "epigenetic", p53 (lecture 4)
F. metastasis:
1. Twist (Weinberg et al);
2. ras/mycconditional (Varmus and colleagues 2009)

III. Large scale tumor genomics: contributions and limitations to date
A. inherent biases of design
implications of cellularity - differential sensitivity of assays
B. "exome" focus: strengths and weaknesses.
C. Case studies. Glioma (GBM) global and IDH1

Expression pattern of a predisposition gene is sometimes a good clue about function (think dystrophin in muscle) -- but there are many exceptions to this. Consider here Lesch Nyhan syndrome (HGPRT gene) and Retinoblastoma (Rb).

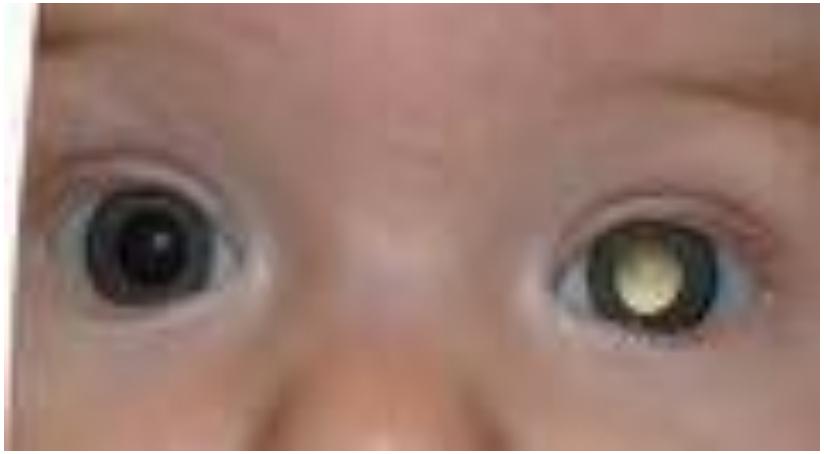
Both genes are widely expressed, their disease phenotypes are highly specific for a particular cell type and developmental time.

Rb is classic case of general, essential molecular function in cell cycle progression.

a. expressed in virtually all cell types -
yet highly specific mutant phenotype. Why?

b.reminder of basics: sporadic retinoblastoma vs inherited
monocular vs binocular

c. window of developmental vulnerability;
surgical therapy in the window is effective
long-term frequency consequences as affected individuals survive and pass allele on to offsrping who also survive



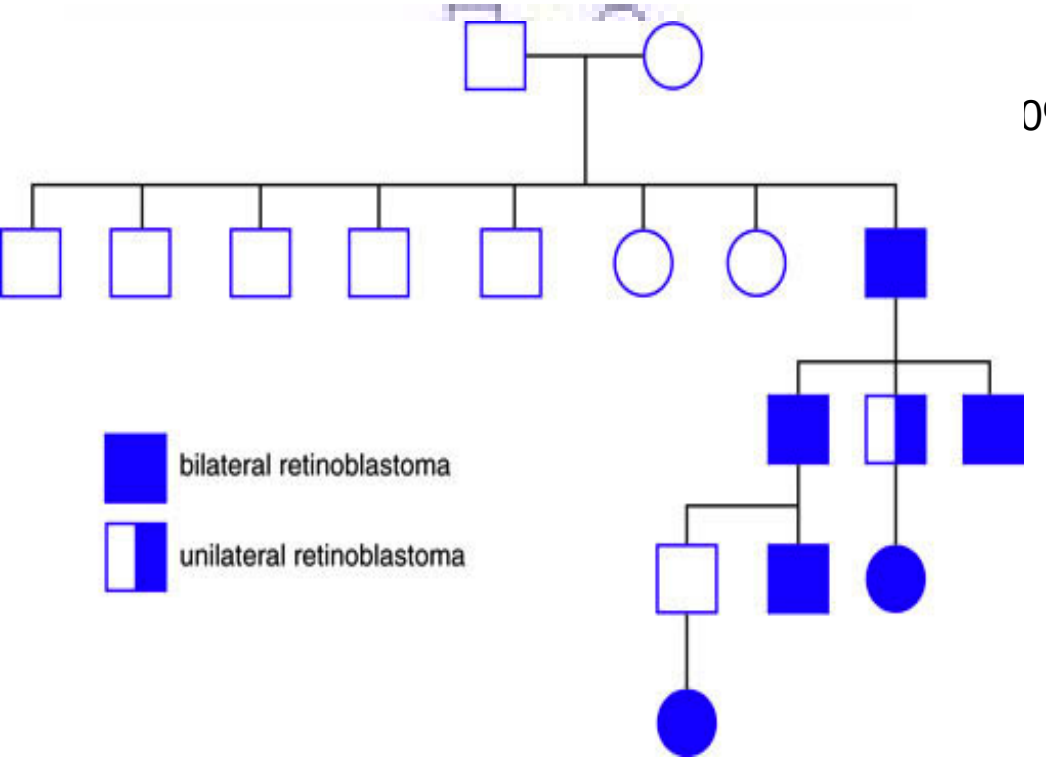
Untreated - fatal by 2 yrs post diagnosis
Treatment early is quite successful 93% 5yr

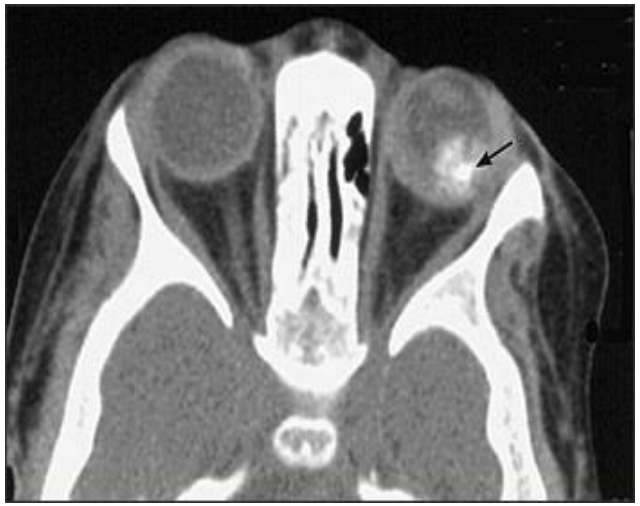
60% is unifocal, unilateral and sporadic
onset 1yr-18mos

40% is inherited, multifocal, often bilateral
onset usually earlier ~6mos

Window of vulnerability until ~3yrs of age
when cells become postmitotic
threat is largely eliminated

0% of patients cured of germ-line
Retinblastoma will develop a
second unrelated tumor by age 50





- Braca 1 and Braca2
  - Major early insight by Mary Claire King: ask for familial AND early onset criteria to select families to find genes. Do not conflate familial with hereditary
  - Surprise: Unlike p53 (TP53) BRACAs are not prominent in sporadic breast CA. In molecular terms, why?
  - Introduce here Exome sequencing as a tool of genetic/genomic discovery and as a clinical test at a locus of major interest
  - Use of genetic screening once its possible:
    - Benefits of knowing status – adjust watchfulness, prophylaxis
      - prophylaxis (surgical) much more important for ovarian
        - » Why? Inability to diagnose ovarian CA early orto treat effectively
        - » Because of ovarian risk, DNA test on patients presenting WITH breast CA is beneficial
    - Liabilities of knowing status

      Individual and in family

      Uncertainties for unusual alleles

Consider expression analysis across many diverse conditions to discover Relationships hypothesized - here the link of disregulated proliferation to Normal growth factor stimulation of cell cycle entry and progression

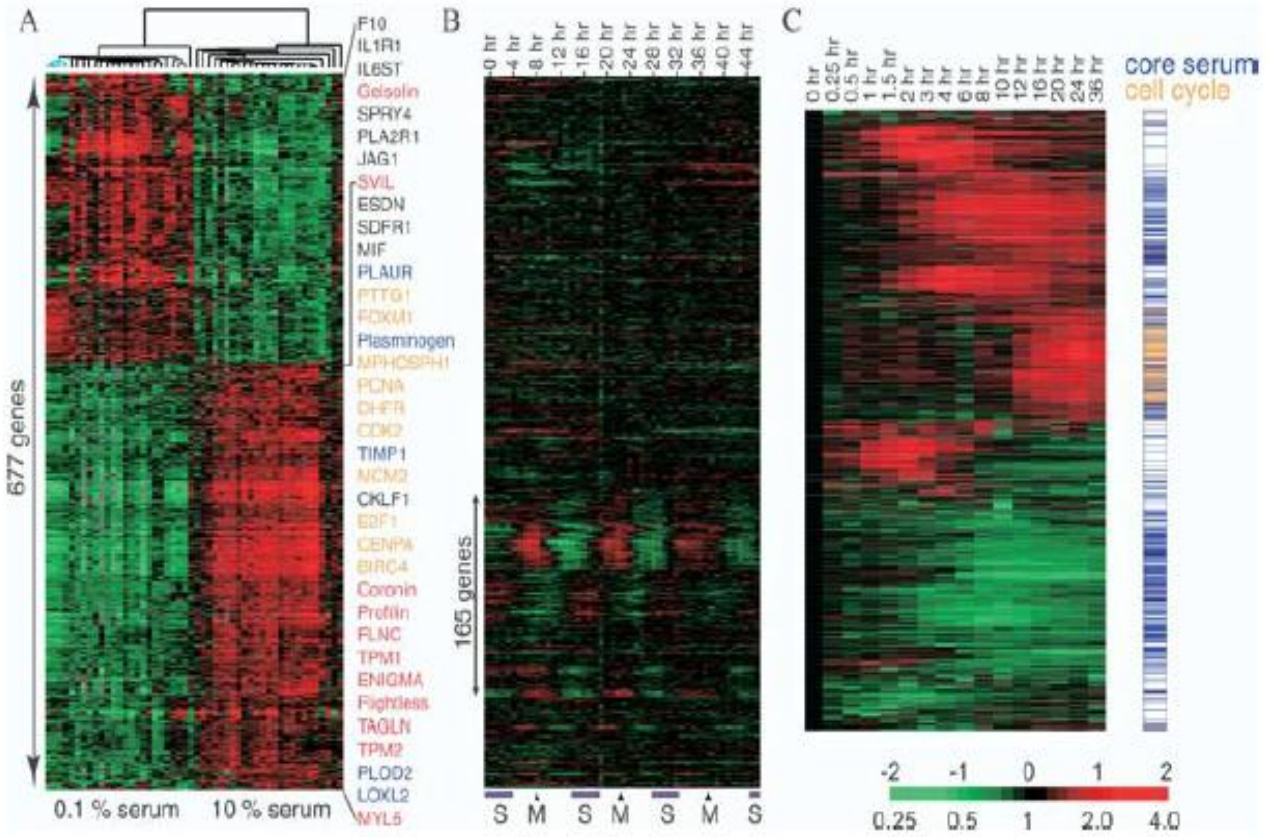


Figure 1. Identification and Annotation of a Common Serum Response in Fibroblasts

(A) The fibroblast common serum response. Genes with expression changes that demonstrate coordinate induction or repression by serum in fibroblasts from ten anatomic sites are shown. Each row represents a gene; each column represents a sample. The level of expression of each gene in each sample, relative to the mean level of expression of that gene across all the samples, is represented using a red–green color scale as shown in the key; gray indicates missing data. Representative genes with probable function in cell cycle progression (orange), matrix remodeling (blue), cytoskeletal rearrangement (red), and cell–cell signaling (black) are highlighted by colored text on the right. Three fetal lung fibroblast samples, cultured in low serum, which showed the most divergent expression patterns among these samples (in part due to altered regulation of lipid biosynthetic genes [Chang et al. 2002]), are indicated by blue branches.

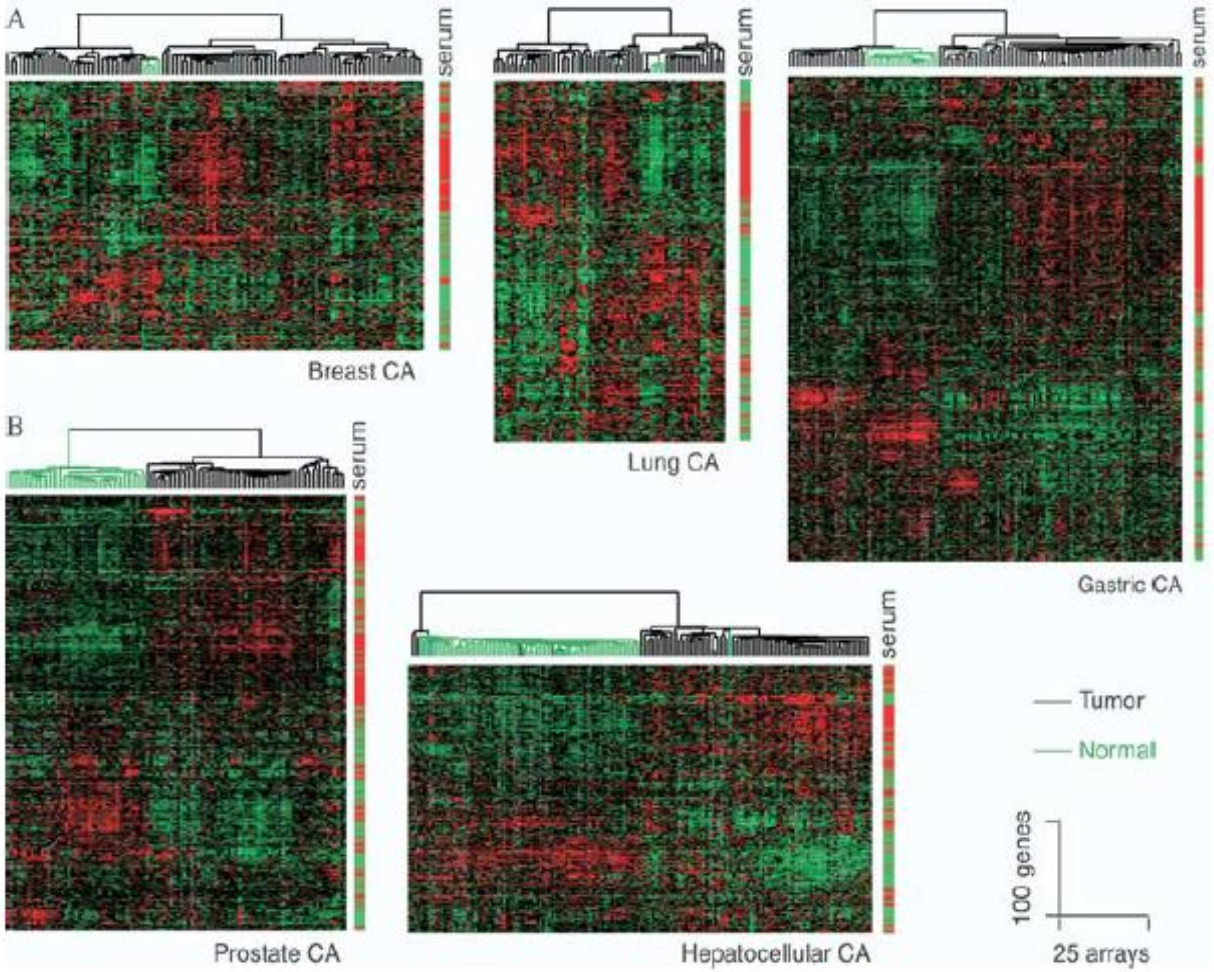


Figure 2. Survey of Fibroblast CSR Gene Expression in Human Cancers

Expression patterns of available CSR genes in over 500 tumors and corresponding normal tissues were extracted, filtered as described in Materials and Methods, and organized by hierarchical clustering. The response of each gene in the fibroblast serum response is shown on the right bar (red shows activated; green shows repressed by serum). The strong clustering of the genes induced or repressed, respectively, in fibroblasts in response to serum exposure, based solely on their expression patterns in the tumor samples, highlights their coordinate regulation in tumors. The dendrograms at the top of each data display represent the similarities among the samples in their expression of the fibroblast CSR genes; tumors are indicated by black branches, normal tissue by green branches.

DOI: 10.1371/journal.pbio.0020007.g002

2-way hierarchical clustering – note general patterns – Yet!

there is also high Variation in individual genes and samples – this is commonly seen

Consider implications for expectations in new samples

Consider principle of gene groups being reliable and individual genes not being so reliable.

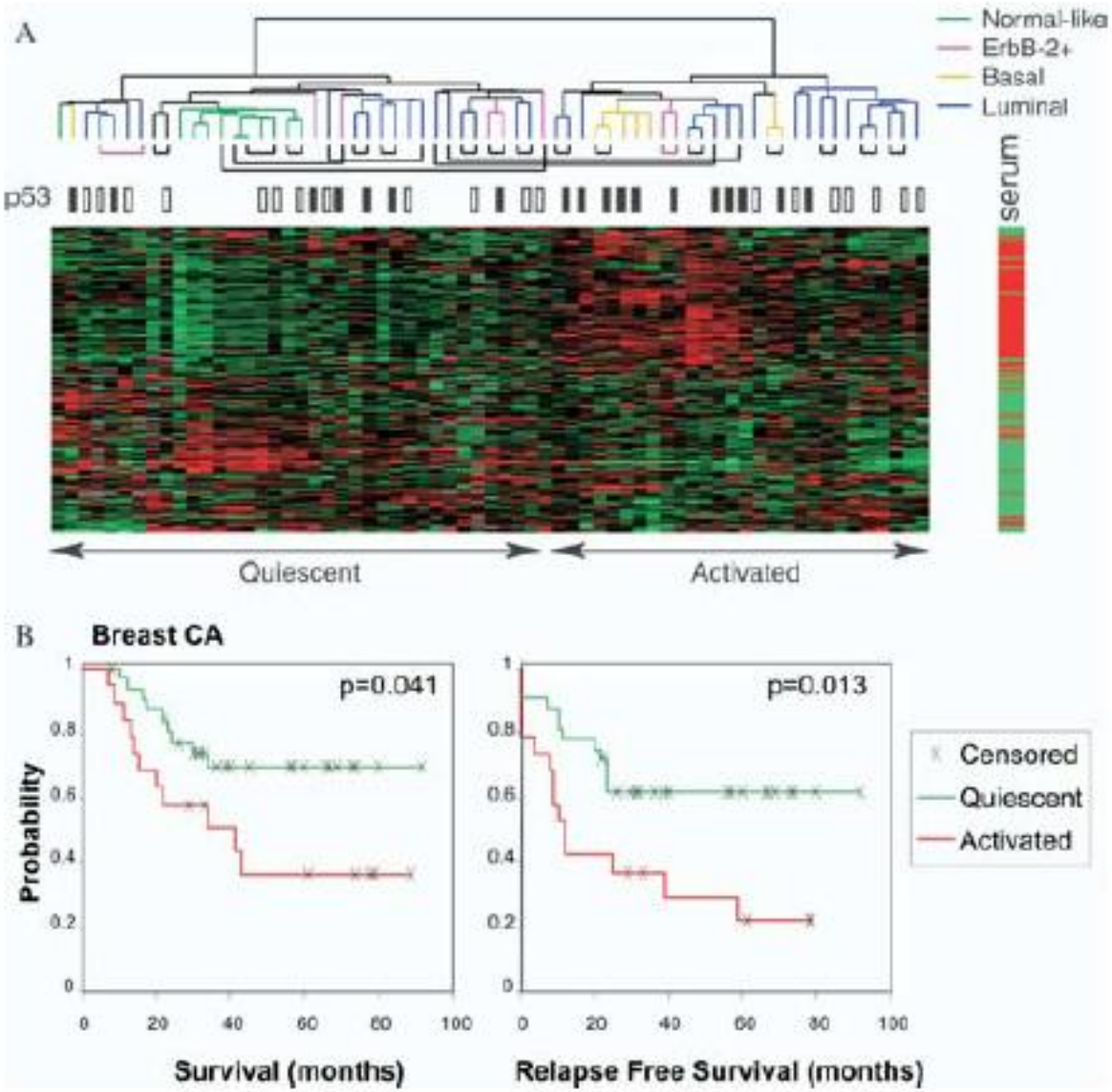


Figure 3. Context, Stability, and Prognostic Value of Fibroblast CSR in Breast Cancer

(A) Expression patterns of CSR genes in a group of breast carcinomas and normal breast tissue previously described in Perou et al. (2000). Genes and samples were organized by hierarchical clustering. The

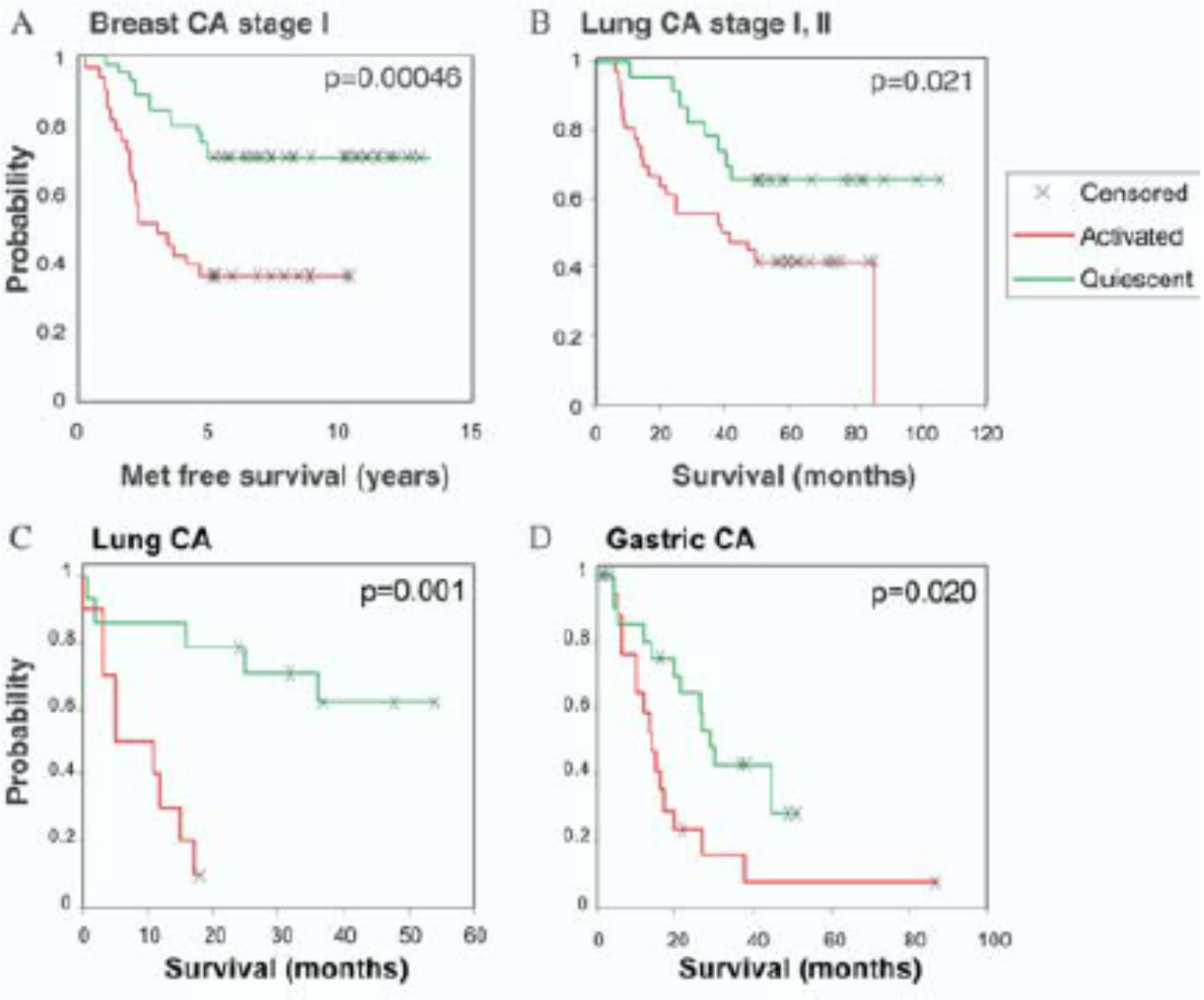
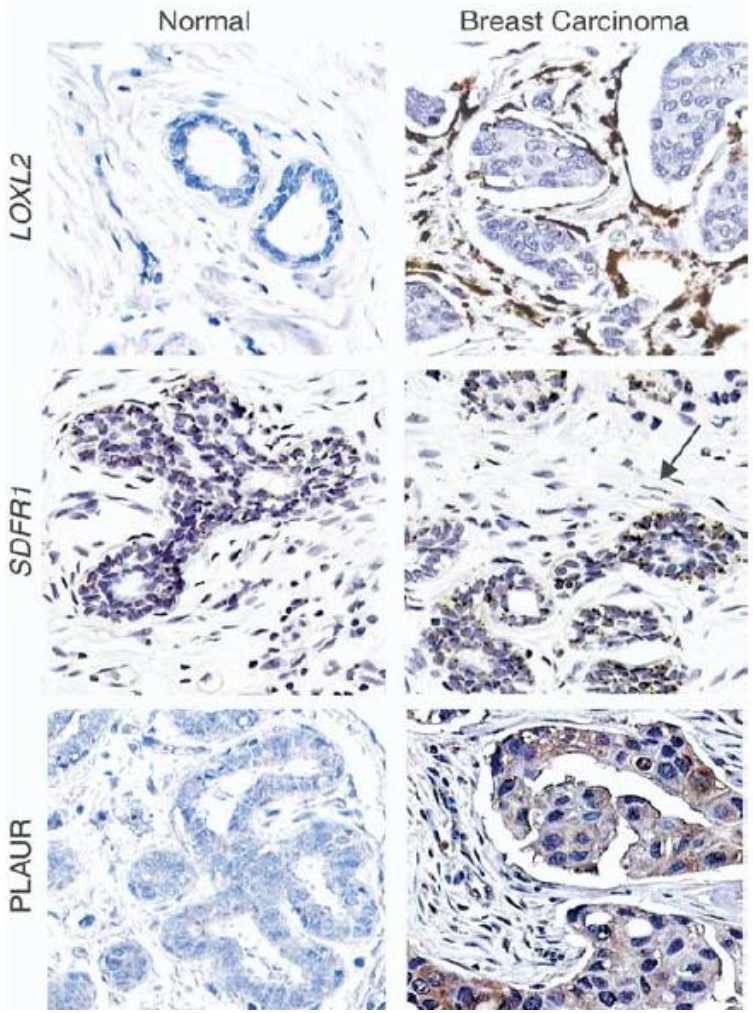


Figure 4. Prognostic Value of Fibroblast CSR in Epithelial Tumors

Kaplan–Meier survival curves of tumors stratified into two classes using the fibroblast CSR are shown for stage I and IIA breast cancer (van 't Veer et al. 2002) (A), stage I and II lung adenocarcinoma (Bhattacharjee et al. 2001) (B), lung adenocarcinoma of all stages (Garber et al. 2001) (C), and stage III gastric carcinoma (Leung et al. 2002) (D).

DOI: 10.1371/journal.pbio.0020007.g004

Key point is confirmation at protein level and analysis
Of cellular distribution within samples --- the issue of
Extracting the right biology conclusion from assays
Of millions of cells (i.e. most genomic assays).



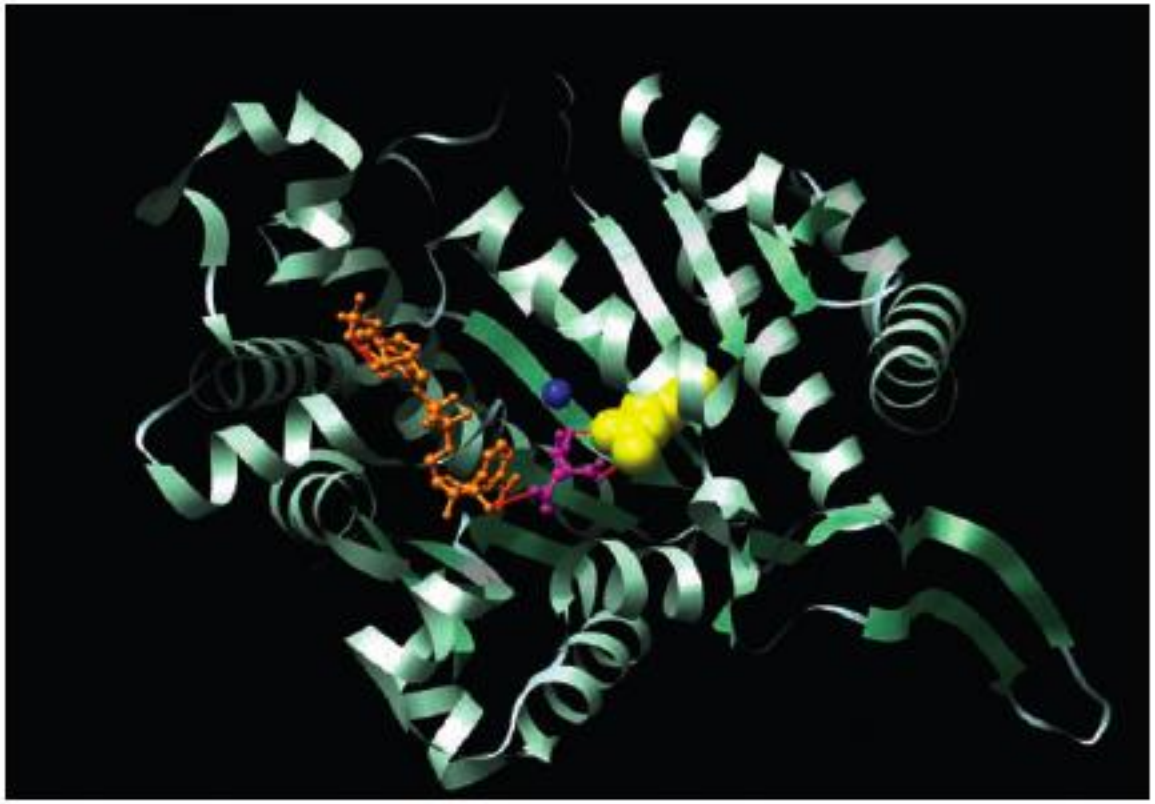
GBM = Glioblastoma multiform

- Became major test case for all our genomics assault – numbers large - treatment regimes differentially effective, but generally poor. Work from “TCGA” and major genome and cancer labs

- Need leads for drugs
- Need guidance among treatments –
  - Generally do not even have option of trying sequentially several possibilities

One early product of large scale tumor genomics
IDH1 mutation is protective - Parsons, Jones,
Et al 2008 Science VOL 321 26 SEPTEMBER 2008

Fig. 1. Structure of the active site of IDH1. The crystal structure of the human cytosolic NADP(+)–dependent IDH is shown in ribbon format (PDBID: 1T0L) (*44*). The active cleft of IDH1 consists of a NADP-binding site and the isocitrate-metal ion-binding site. The alpha-carboxylate oxygen and the hydroxyl group of isocitrate chelate the Ca^{2+} ion. NADP is colored in orange, isocitrate in purple and Ca^{2+} in blue. The Arg^{132} residue, displayed in yellow, forms hydrophilic interactions, shown in red, with the alpha-carboxylate of isocitrate. Displayed image was created with UCSF Chimera software version 1.2422 (*50*).



Isocitrate dehydrogenase 1 active site mutation in 12% of GBMs
Known genes (TP53(TS), EGFR (OG), PTEN(TS) already appreciated
Former mainly in secondary class; other two in primary tumors)

IDH1 found in exome sequencing
Multiple finds points to possible
Causal role rather than ride-along
Or random accumulated damage

Sequencing analysis	
Number of amplicons attempted	219,229 (100%)
Number of amplicons passing quality control*	208,311 (95%)
Fraction of bases in passing amplicons with PHRED > 20	98.3%
Number of genes analyzed	20,661
Number of transcripts analyzed	23,219
Number of exons analyzed	175,471
Total number of nucleotides successfully sequenced	689,071,123
Number of somatic mutations identified (n = 22 samples)	2,325
Number of somatic mutations (excluding Br27P)	993
Missense	622
Nonsense	43
Insertion	3
Deletion	46
Duplication	7
Splice site or UTR	27
Synonymous	245
Average number of sequence alterations per sample	47.3
Copy number analysis	
Total number of SNP loci assessed for copy number changes	1,069,688
Number of copy number alterations identified (n = 22 samples)	281
Amplifications	147
Homozygous deletions	134
Average number of amplifications per sample	6.7
Average number of homozygous deletions per sample	6.1

*Passing amplicons were defined as having PHRED20 scores or better over 90% of the target sequence in 75% of samples analyzed [see (*12*) for additional information].

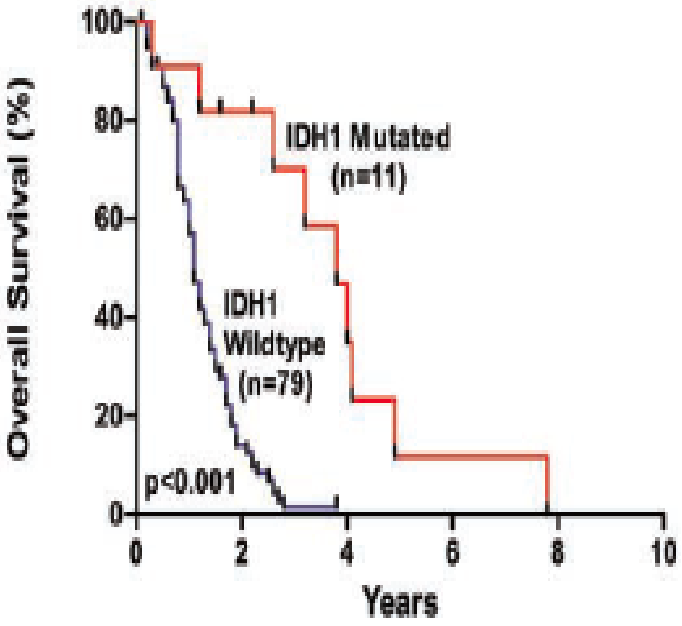


Fig. 2. Overall survival according to *IDH1* mutation status. The hazard ratio for death among patients with wild-type *IDH1* (n = 79), as compared to those with mutant *IDH1* (n = 11), was 3.7 (95 percent confidence interval, 2.1 to 6.5; P < 0.001). The median survival was 3.8 years for patients with mutated *IDH1*, as compared to 1.1 years for patients with wild-type *IDH1*.

CNV on GBM

Table 2. Most frequently altered GBM *CAN*-genes. All *CAN*-genes are listed in table S7.

Gene	Point mutations*		Amplifications†		Homozygous deletions†		Fraction of tumors with any alteration (%)	Passenger probability‡
	No. of tumors	Fraction of tumors (%)	No. of tumors	Fraction of tumors (%)	No. of tumors	Fraction of tumors (%)		
CDKN2A	0/22	0	0/22	0	11/22	50	50	<0.01
TP53	37/105	35	0/22	0	1/22	5	40	<0.01
EGFR	15/105	14	5/22	23	0/22	0	37	<0.01
PTEN	27/105	26	0/22	0	1/22	5	30	<0.01
NF1	16/105	15	0/22	0	0/22	0	15	0.04
CDK4	0/22	0	3/22	14	0/22	0	14	<0.01
RB1	8/105	8	0/22	0	1/22	5	12	0.02
IDH1	12/105	11	0/22	0	0/22	0	11	<0.01
PIK3CA	10/105	10	0/22	0	0/22	0	10	0.10
PIK3R1	8/105	8	0/22	0	0/22	0	8	0.10

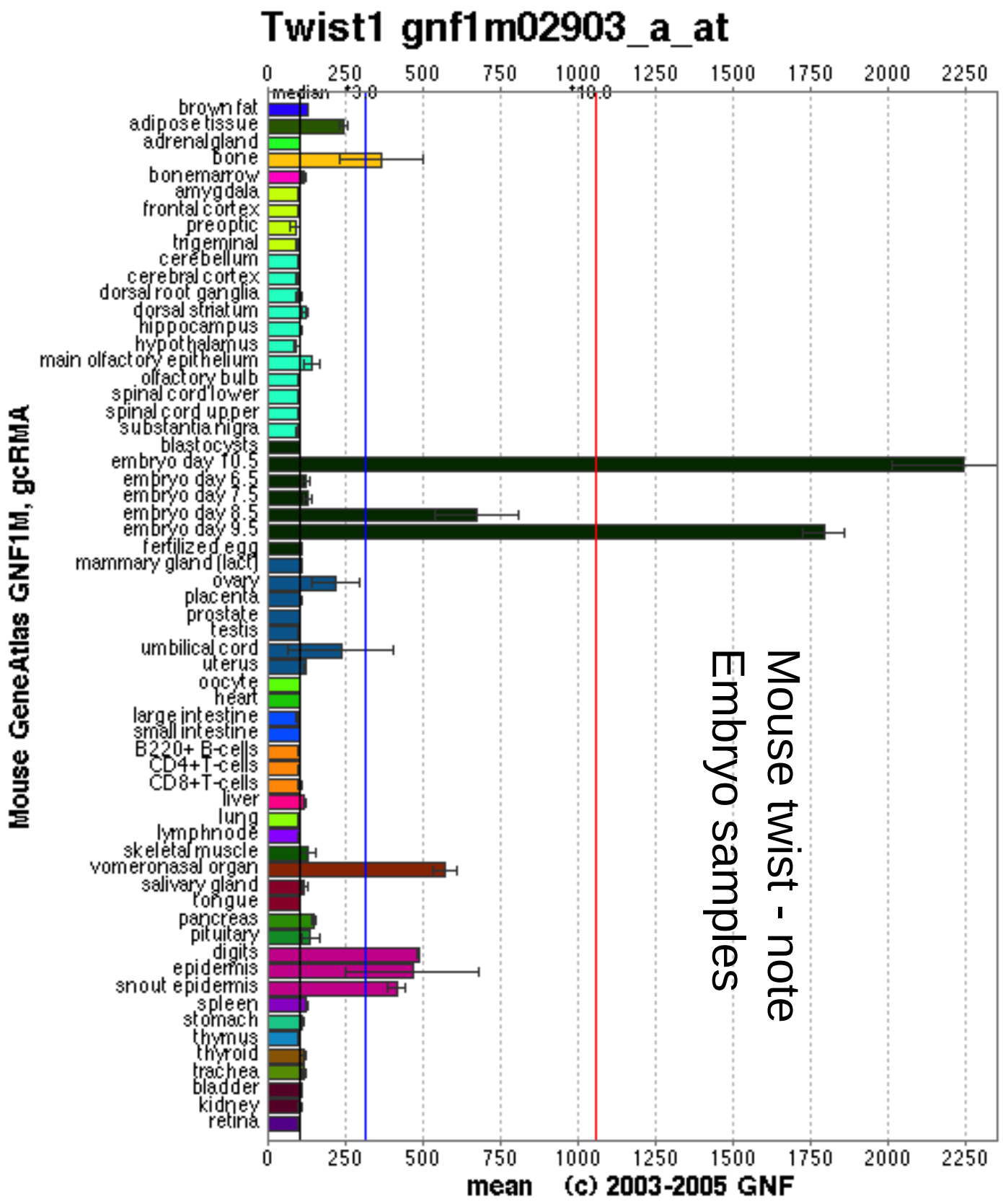
*Fraction of tumors with point mutations indicates the fraction of mutated GBMs out of the 105 samples in the Discovery and Prevalence Screens. CDKN2A and CDK4 were not analyzed for point mutations in the Prevalence Screen because no sequence alterations were detected in these genes in the Discovery Screen. †Fraction of tumors with amplifications and deletions indicates the number of tumors with these types of alterations in the 22 Discovery Screen samples. ‡Passenger probability indicates the probability obtained using the average of the lower and upper bound background mutation rates (*12*).

Note the way genes in a few signalling pathways show up Together. Can work that backwards – if one – look at others

Note this is conceptually similar to co-expression groups being more robust than single outliers

Table 3. Mutations of the TP53, PI3K, and RB1 pathways in GBM samples. Mut, mutated; Amp, amplified; Del, deleted; Alt, altered.

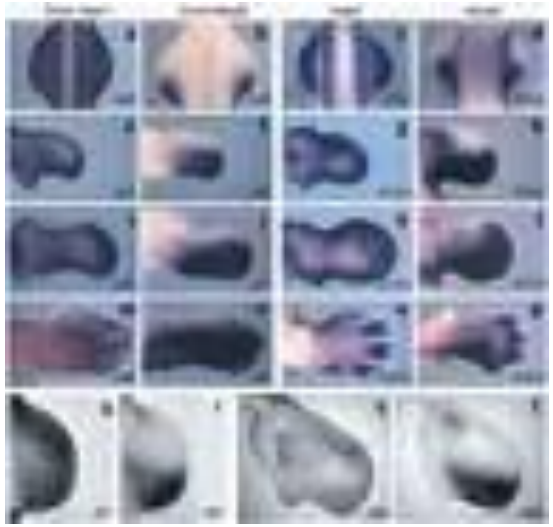
	TP53 pathway				PI3K pathway					RB1 pathway			
Tumor sample	TP53	MDM2	MDM4	All genes	PTEN	PIK3CA	PIK3R1	IRS1	All genes	RB1	CDK4	CDKN2A	All genes
Br02X	Del			Alt				Mut	Alt			Del	Alt
Br03X	Mut			Alt	Mut				Alt				
Br04X	Mut			Alt	Mut				Alt	Mut			Alt
Br05X			Amp	Alt		Mut			Alt			Del	Alt
Br06X												Del	Alt
Br07X	Mut			Alt	Mut				Alt	Del			Alt
Br08X												Del	Alt
Br09P	Mut			Alt							Amp		Alt
Br10P	Mut			Alt									
Br11P	Mut			Alt									
Br12P	Mut			Alt			Mut		Alt				
Br13X	Mut			Alt								Del	Alt
Br14X							Mut		Alt			Del	Alt
Br15X										Mut		Del	Alt
Br16X		Amp		Alt							Amp		Alt
Br17X					Mut				Alt			Del	Alt
Br20P													
Br23X	Mut			Alt	Del				Alt				
Br25X					Mut				Alt			Del	Alt
Br26X						Mut			Alt			Del	Alt
Br27P	Mut			Alt							Amp		Alt
Br29P	Mut			Alt									
Fraction	0.55	0.05	0.05	0.64	0.27	0.09	0.09	0.05	0.50	0.14	0.14	0.45	0.68


Twist1 gnf1m02903_a_at
0 250 500 750 1000 1250 1500 1750 2000 2250
median *3.0 *10.0
brown fat
adipose tissue
adrenal gland
bone
bone marrow
amygdala
frontal cortex
preoptic
trigeminal
cerebellum
cerebral cortex
dorsal root ganglia
dorsal striatum
hippocampus
hypothalamus
main olfactory epithelium
olfactory bulb
spinal cord lower
spinal cord upper
substantia nigra
blastocysts
embryo day 10.5
embryo day 6.5
embryo day 7.5
embryo day 8.5
embryo day 9.5
fertilized egg
mammary gland (lact)
ovary
placenta
prostate
testis
umbilical cord
uterus
oocyte
heart
large intestine
small intestine
B220+ B-cells
CD4+T-cells
CD8+T-cells
liver
lung
lymph node
skeletal muscle
vomeronasal organ
salivary gland
tongue
pancreas
pituitary
digits
epidermis
snout epidermis
spleen
stomach
thymus
thyroid
trachea
bladder
kidney
retina
Mouse GeneAtlas GNF1M, gcRMA
Mouse twist - note
Embryo samples
mean
(c) 2003-2005 GNF

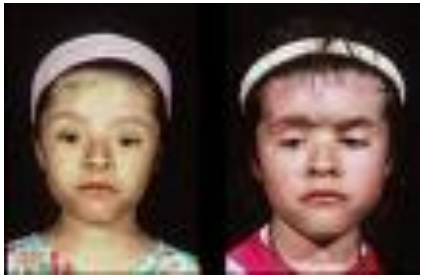
•Twist-null heterozygous mouse phenotype

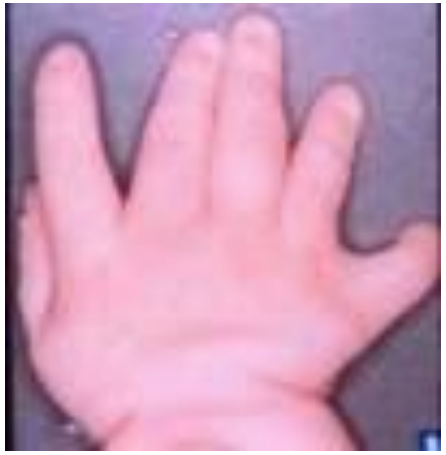
human Saethre-Chotzen syndrome (also has FGF families)

- variable expressivity
- incomplete penetrance



mouse





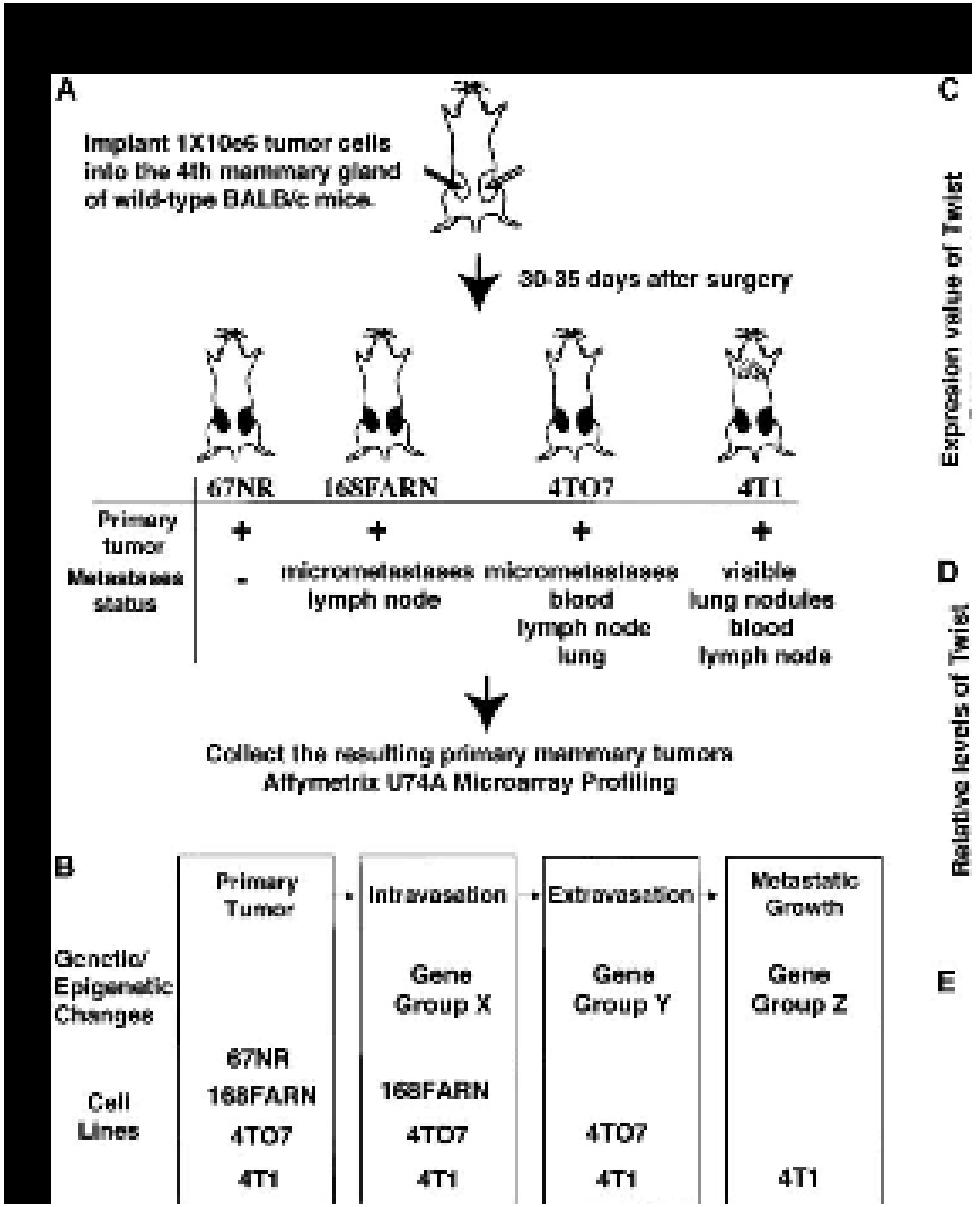
Possibly working both ways - a common confounding pattern in human genetics in general and in cancer genetics in particular

Twist haploinsufficiency in Saethre–Chotzen syndrome induces calvarial osteoblast apoptosis due to increased TNF{alpha} expression and caspase-2 activation

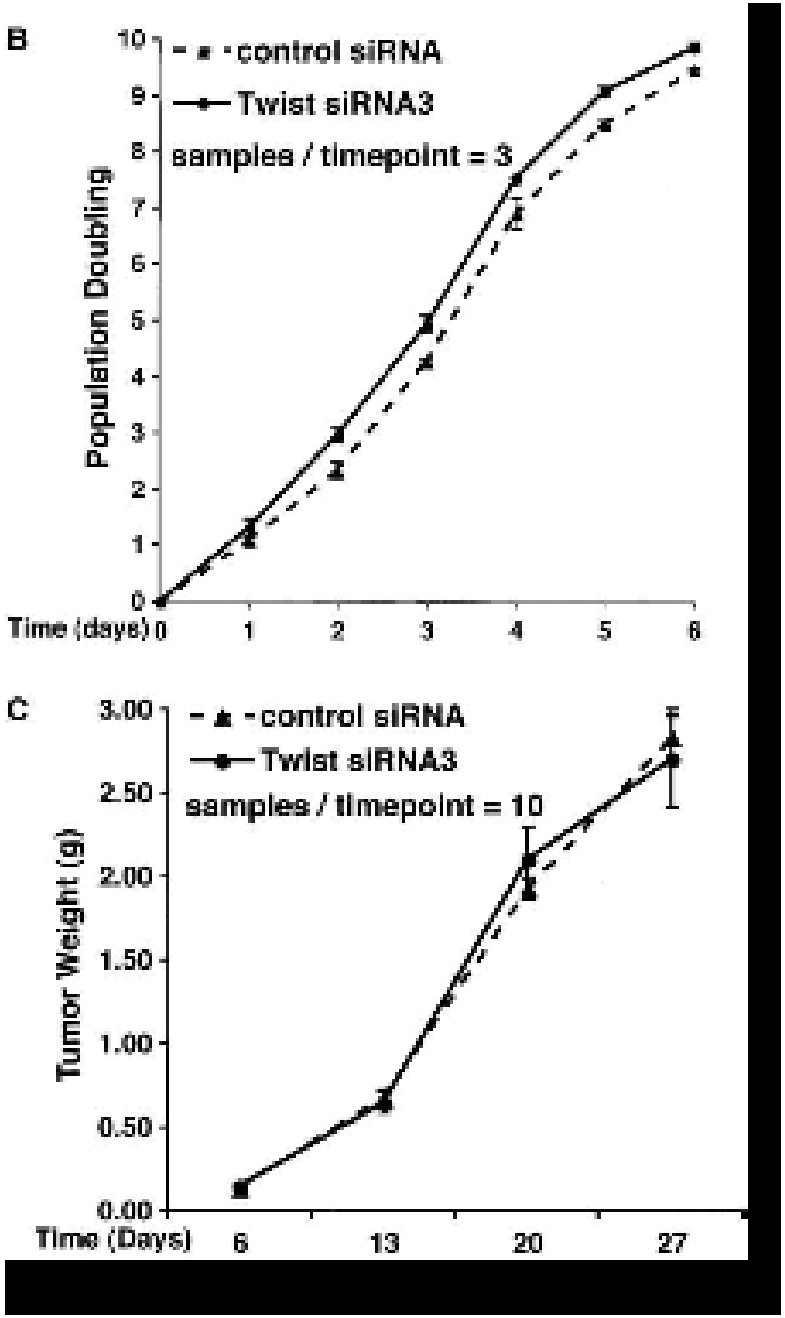
(Malika Yousfi, Francoise Lasmoles, Vincent El Ghouzzi1 and Pierre J. Marie. Human Molecular Genetics, 2002, Vol. 11, No. 4 359, 2002)

Down-Regulation of the Ubiquitin Ligase Cbl is Induced by Twist Haploinsufficiency in Saethre-Chotzen Syndrome leading to PI3K/Akt Signaling and Osteoblast Proliferation

(H. Guenou, K. Kaabeche, C. Dufour, H. Miraoui, and P. J. Marie. Am. J. Pathol., October 1, 2006; 169(4): 1303 - 1311)



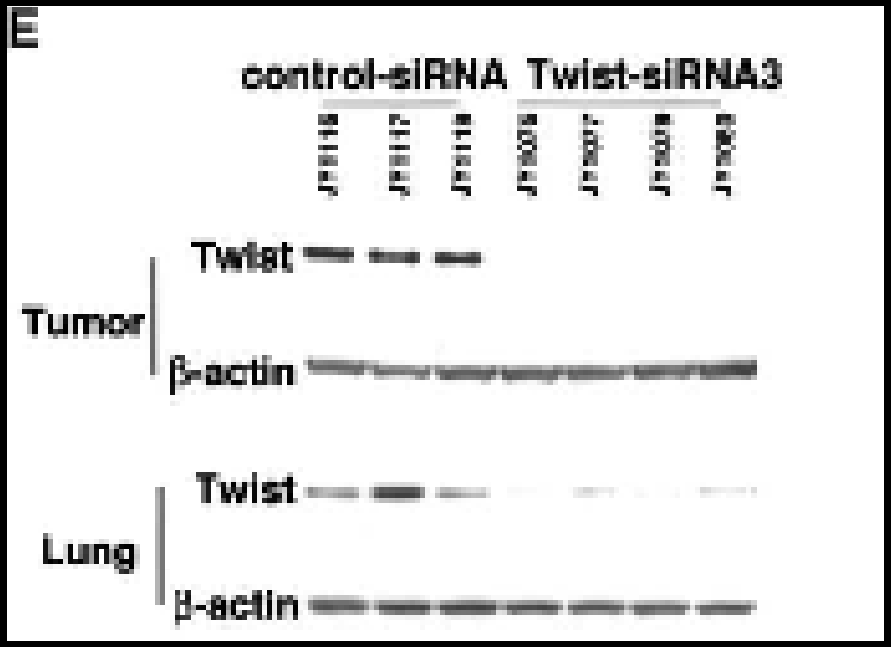
Yany et al 2004 (weinberg) Cell 188 [lect 3 list]

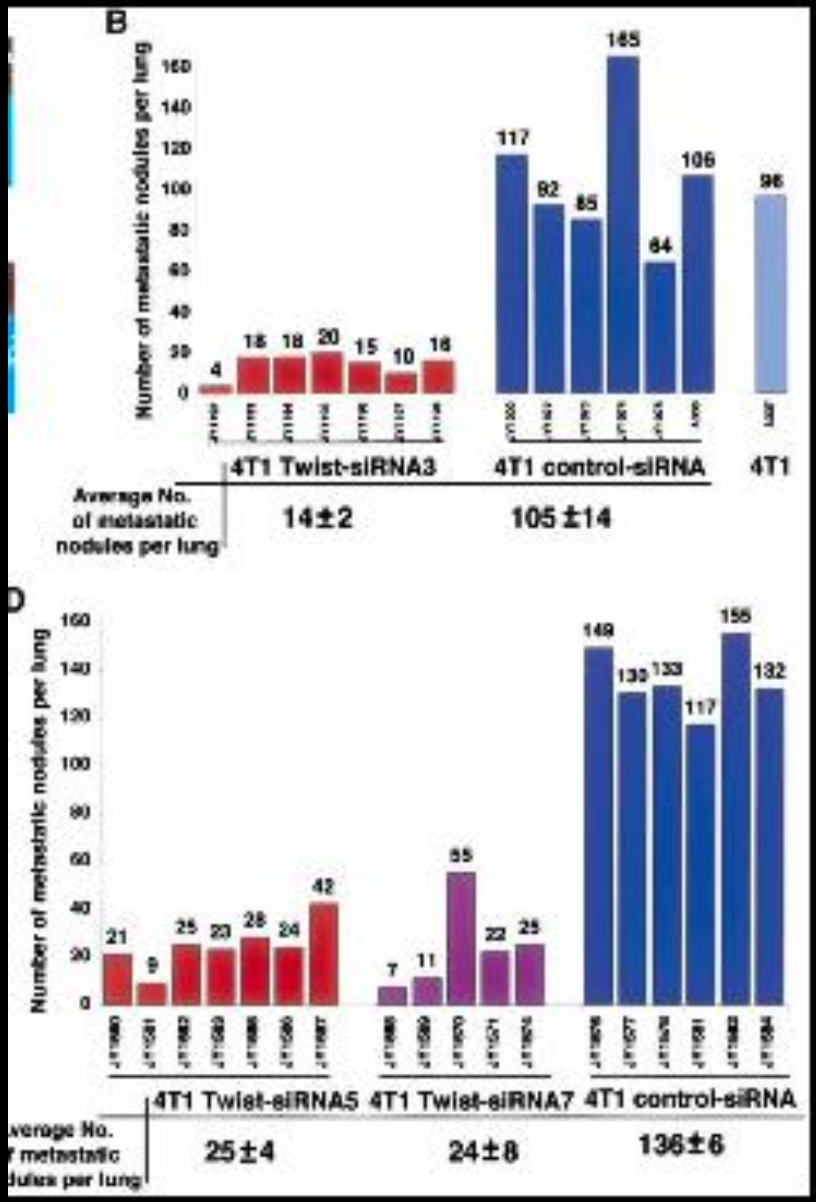


siRNA against
Twist does not affect tumor size

Nor does it affect cell proliferation rate

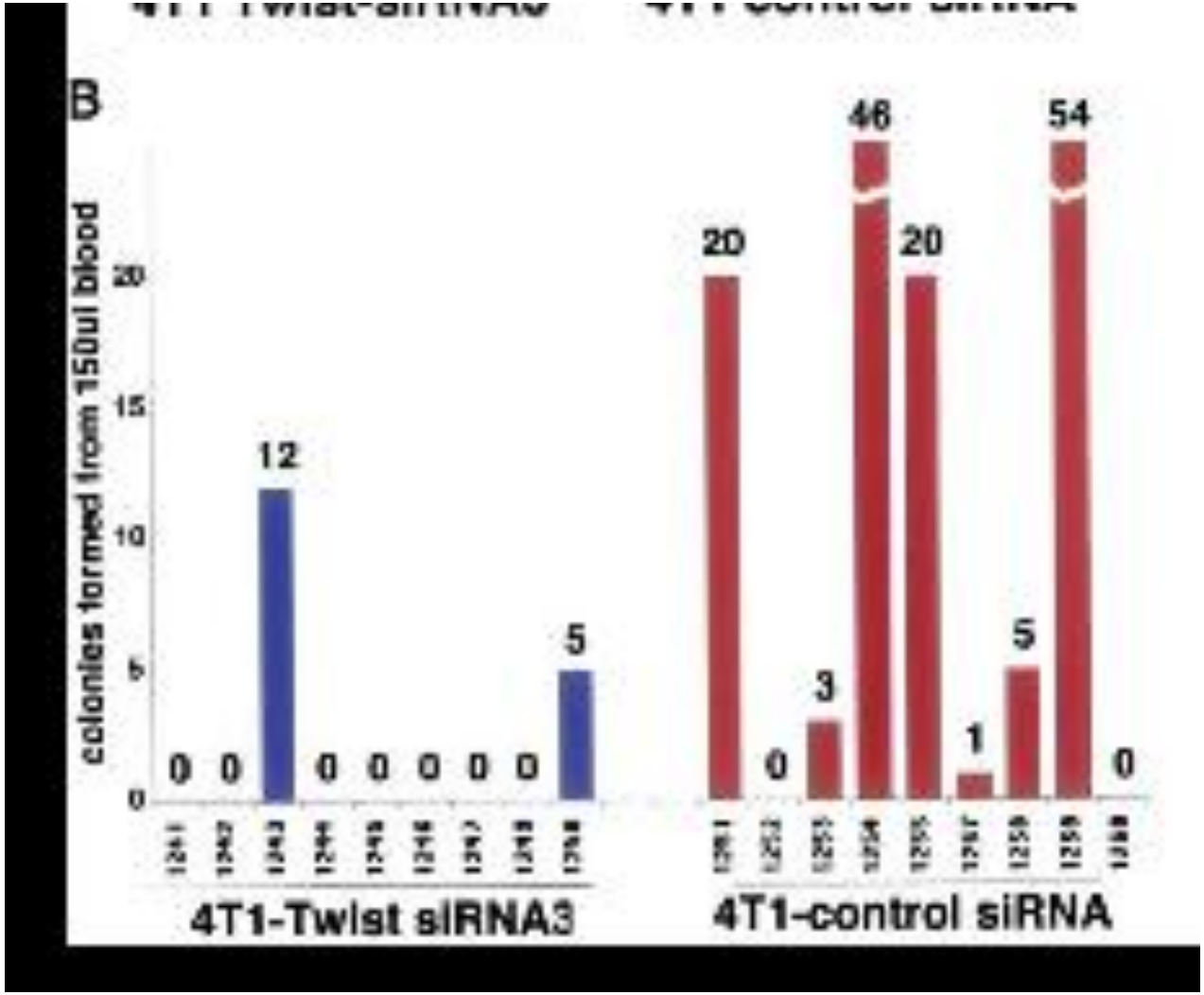
[Measured in breast tumors in mouse model]



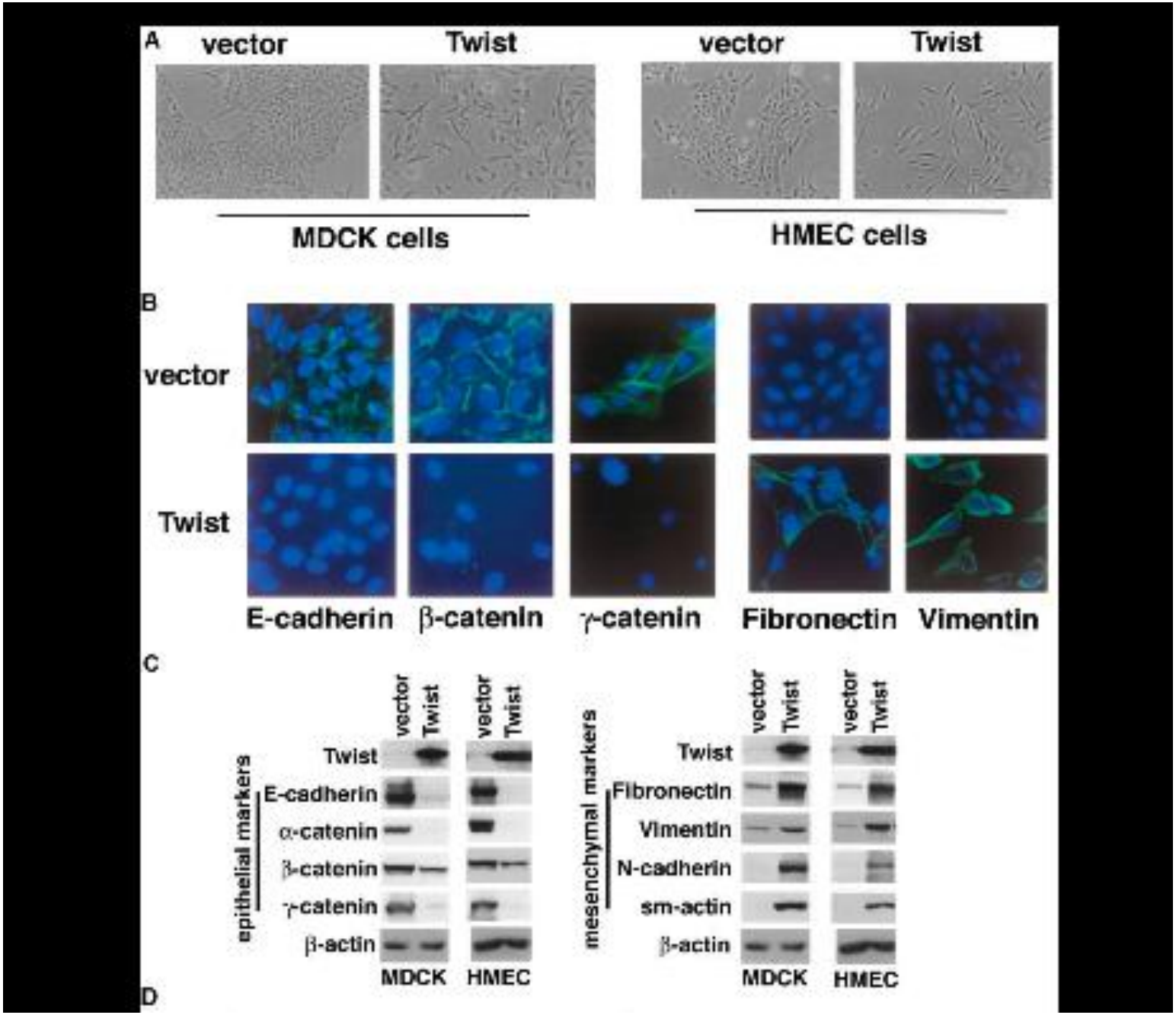


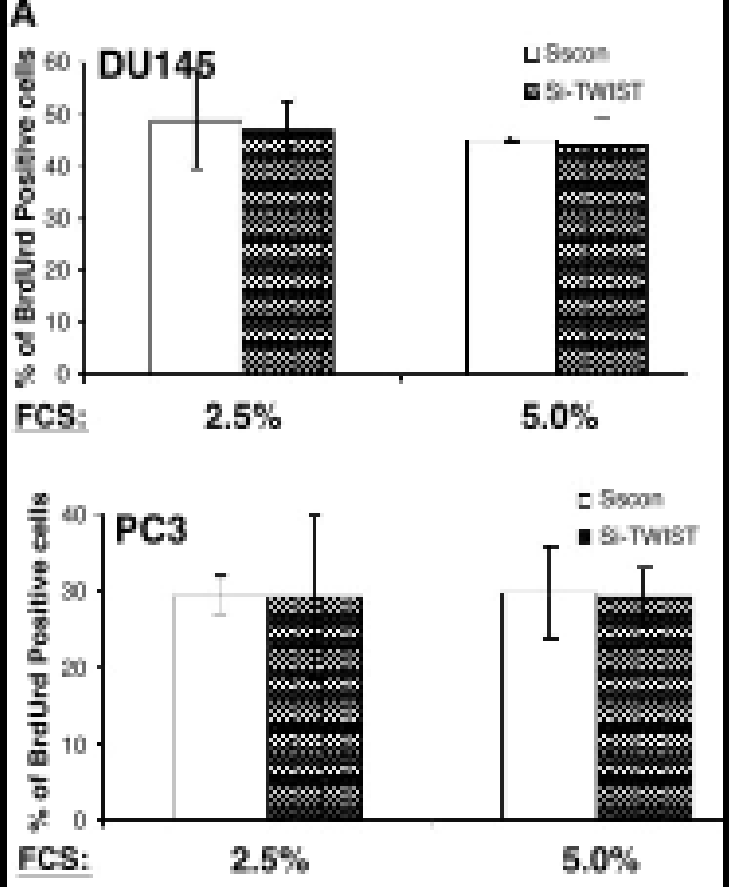
siRNA analysis
points to metastasis

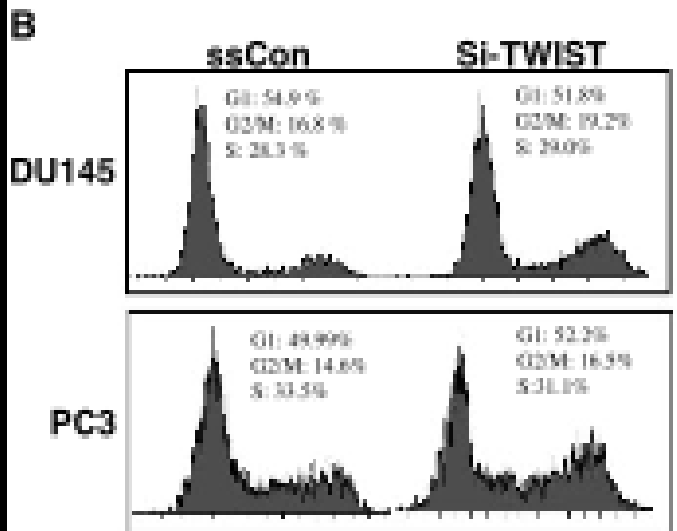
Consider what target genes
Might be important….and why…



Cells in circulation as a function of twist inhibition (siRNA)

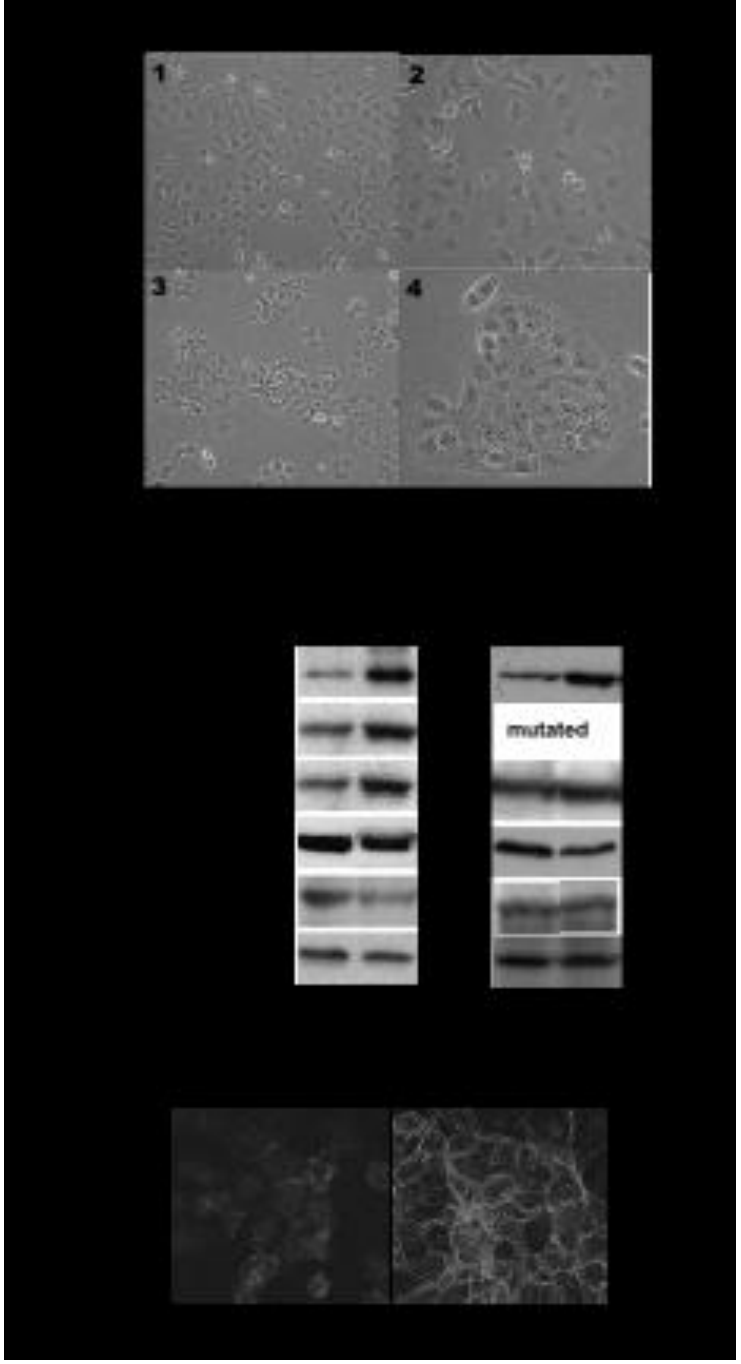

A
vector
Twist
vector
Twist
MDCK cells
HMEC cells
B
vector
Twist
E-cadherin
β-catenin
γ-catenin
Fibronectin
Vimentin
C
vector
Twist
vector
Twist
epithelial markers
Twist
E-cadherin
α-catenin
β-catenin
γ-catenin
β-actin
MDCK
HMEC
vector
Twist
vector
Twist
mesenchymal markers
Twist
Fibronectin
Vimentin
N-cadherin
sm-actin
β-actin
MDCK
HMEC
D





Twist in human prostate cancer: androgen independent effects

Proliferation measures not affected, as in mouse Breast CA model



Molecular and cell col morphol
Phenotypes are affected by twist

Bottom panel is e-cadherin immunostaining

Kwok et al., 2007