

Target Discovery in OncoPrint: Integrating -omics data to prioritize targets



Dan Rhodes

May 28, 2008

Agenda

- What makes a good target?
- Available data
 - Gene expression
 - DNA copy number
 - Somatic mutations
 - Functional screens
- Three target discovery analyses that integrate data types

What makes a good target?

- Druggable
- Mutated in cancer
 - Driver vs. passenger
- Over-expressed in cancer
 - Necessary role in tumorigenesis
- Functional dependence / vulnerability

Descriptive & Functional Genomics

- **Mutation Types**
 - Amplifications / Deletions
 - aCGH, gene expression
 - Point mutations
 - Sequencing
 - Translocations
 - Cytogenetics, sequencing, gene expression
- **Functional dependence / vulnerability**
 - siRNA / shRNA screens

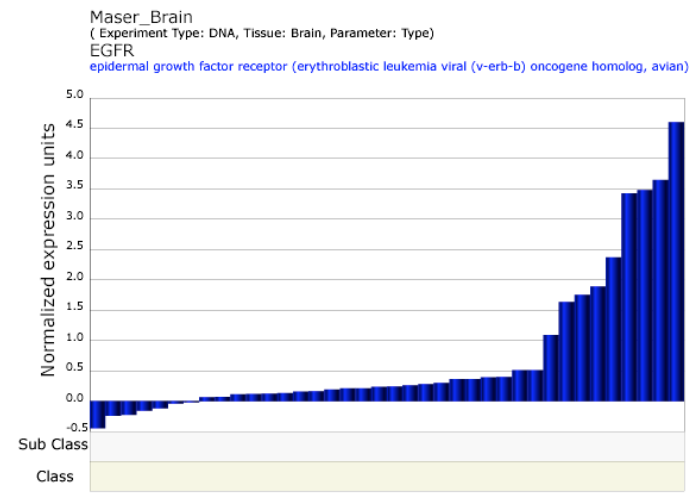
High-level DNA amplifications

■ Array CGH

- Minimal common region analysis
- COPA applied to DNA copy number
- Co-amplified with genomic neighbors

■ Examples

- EGFR in glioblastoma
- ERBB2 in breast cancer



Somatic Mutations

- COSMIC Cancer Gene Census
 - Causal cancer genes
 - Literature-review
 - Translocations
 - Dominant mutations
 - Recessive mutations

Nature Reviews Cancer 4, 177-183 (2004); doi:10.1038/nrc1299

A CENSUS OF HUMAN CANCER GENES

P. Andrew Futreal, Lachlan Coin, Mhairi Marshall, Thomas Down, Timothy Hubbard, Richard Wooster, Nazneen Rahman & Michael R. Stratton [about the authors](#)

Abstract

A central aim of cancer research has been to identify the mutated genes that are causally implicated in oncogenesis ('cancer genes'). After two decades of searching, how many have been identified and how do they compare to the complete set of genes in the human genome? We have conducted a census of cancer genes that indicates that many more genes contribute to human cancer. The census identifies the types of sequence alteration, and the number of mutations have been identified and their frequency by cancer genes.

Summary

- We have conducted a census of mutated and causally implicated genes).
- So far, 291 cancer genes have been identified in the human genome.
- 90% of cancer genes show somatic mutations and 10% show germline mutations and 10% show translocations.
- The most common mutation of cancer genes is a chromosomal translocation.

Cancer Gene Census

Sorted By	Number
Amplification	7
Chromosome	367
Frameshift mutation	67
Germline mutation	68
Large deletion	28
Missense mutation	87
Nonsense mutation	63
Other mutation	10
Somatic mutation	330
Splicing mutation	42
Symbol	367
Translocation	275

Somatic Mutations

- Sanger Kinase Screen
 - Coding sequence
 - 518 kinases
 - 25 breast cancers

1: [Nat Genet.](#) 2005 Jun;37(6):590-2. Epub 2005 May 22.

A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer.

[Stephens P](#), [Edkins S](#), [Davies H](#), [Greenman C](#), [Cox C](#), [Hunter C](#), [Bignell G](#), [Teague J](#), [Smith R](#), [Stevens C](#), [O'Meara S](#), [Parker A](#), [Tarpey P](#), [Avis T](#), [Barthorpe A](#), [Brackenbury L](#), [Buck G](#), [Butler A](#), [Clements J](#), [Cole J](#), [Dicks E](#), [Edwards K](#), [Forbes S](#), [Gorton M](#), [Gray K](#), [Halliday K](#), [Harrison R](#), [Hills K](#), [Hinton J](#), [Jones D](#), [Kosmidou V](#), [Laman R](#), [Lugg R](#), [Menzies A](#), [Perry J](#), [Petty R](#), [Raine K](#), [Shepherd R](#), [Small A](#), [Solomon H](#), [Stephens Y](#), [Tofts C](#), [Varian J](#), [Webb A](#), [West S](#), [Widaa S](#), [Yates A](#), [Brasseur F](#), [Cooper CS](#), [Flanagan AM](#), [Green A](#), [Knowles M](#), [Leung SY](#), [Looijenga LH](#), [Malkowicz B](#), [Pierotti MA](#), [Teh B](#), [Yuen ST](#), [Nicholson AG](#), [Lakhani S](#), [Easton DF](#), [Weber BL](#), [Stratton MR](#), [Futreal PA](#), [Wooster R](#).

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.

We examined the coding sequence of 518 protein kinases, approximately 1.3 Mb of DNA per sample, in 25 breast cancers. In many tumors, we detected no somatic mutations. But a few had numerous somatic mutations with distinctive patterns indicative of either a mutator phenotype or a past exposure.

Somatic Mutations

- Full genome sequencing
 - 11 breast & 11 colon tumors
 - 13,000 genes
 - CAN genes

1: [Science](#). 2006 Oct 13;314(5797):268-74. Epub 2006 Sep 7.

Comment in:

[Science](#). 2007 Feb 9;315(5813):762-4; author reply 764-5.

[Science](#). 2007 Feb 9;315(5813):762; author reply 764-5.

[Science](#). 2007 Feb 9;315(5813):762; author reply 764-5.

[Science](#). 2007 Sep 14;317(5844):1500.

[Science](#). 2007 Sep 14;317(5844):1500.

[Science](#). 2007 Sep 14;317(5844):1500; author reply 1500.

The consensus coding sequences of human breast and colorectal cancers.

[Sjöblom T](#), [Jones S](#), [Wood LD](#), [Parsons DW](#), [Lin J](#), [Barber TD](#), [Mandelker D](#), [Leary RJ](#), [Ptak J](#), [Silliman N](#), [Szabo S](#), [Buckhaults P](#), [Farrell C](#), [Meeh P](#), [Markowitz SD](#), [Willis J](#), [Dawson D](#), [Willson JK](#), [Gazdar AF](#), [Hartigan J](#), [Wu L](#), [Liu C](#), [Parmigiani G](#), [Park BH](#), [Bachman KE](#), [Papadopoulos N](#), [Vogelstein B](#), [Kinzler KW](#), [Velculescu VE](#).

Ludwig Center and Howard Hughes Medical Institute, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD 21231, USA.

The elucidation of the human genome sequence has made it possible to identify genetic alterations in cancers in unprecedented detail. To begin a systematic analysis of such alterations, we determined the sequence of well-annotated human protein-coding genes in two common tumor types. Analysis of 13,023 genes in 11 breast and 11 colorectal cancers revealed that individual tumors accumulate an average of approximately 90 mutant genes but that only a subset of these contribute to the neoplastic process. Using stringent criteria to delineate this subset, we identified 189 genes (average of 11 per tumor) that were mutated at significant frequency. The vast majority of these genes were not known to be genetically altered in tumors and are predicted to affect a wide range of cellular functions, including transcription, adhesion, and invasion. These data define the genetic landscape of two human cancer types, provide new targets for diagnostic and therapeutic intervention, and open fertile avenues for basic research in tumor biology.

Functional dependence

- **shRNA screens**
 - mRNAs necessary for viability of cell lines
 - Large pools of shRNA transfected into cell lines
 - Dropout screen via barcode microarray

Cancer Proliferation Gene Discovery Through Functional Genomics

Michael R. Schlabach,^{1*} Ji Luo,^{1*} Nicole L. Solimini,^{1*} Guang Hu,^{1*} Qikai Xu,¹ Mamie Z. Li,¹ Zhenming Zhao,¹ Agata Smogorzewska,^{1,2} Mathew E. Sowa,³ Xiaolu L. Ang,³ Thomas F. Westbrook,¹ Anthony C. Liang,¹ Kenneth Chang,⁴ Jennifer A. Hackett,¹ J. Wade Harper,³ Gregory J. Hannon,⁴ Stephen J. Elledge^{1†}

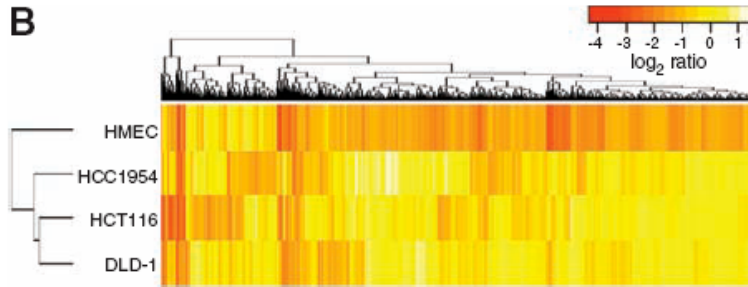
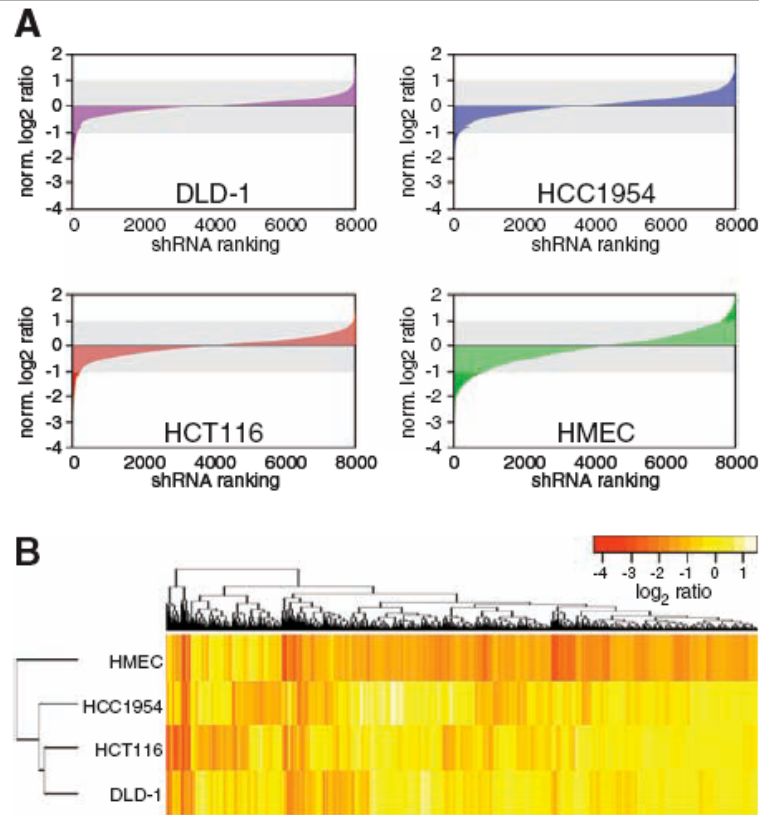
Retroviral short hairpin RNA (shRNA)-mediated genetic screens in mammalian cells are powerful tools for discovering loss-of-function phenotypes. We describe a highly parallel multiplex methodology for screening large pools of shRNAs using half-hairpin barcodes for microarray deconvolution. We carried out dropout screens for shRNAs that affect cell proliferation and viability in cancer cells and normal cells. We identified many shRNAs to be antiproliferative that target core cellular processes, such as the cell cycle and protein translation, in all cells examined. Moreover, we identified genes that are selectively required for proliferation and survival in different cell lines. Our platform enables rapid and cost-effective genome-wide screens to identify cancer proliferation and survival genes for target discovery. Such efforts are complementary to the Cancer Genome Atlas and provide an alternative functional view of cancer cells.

Profiling Essential Genes in Human Mammary Cells by Multiplex RNAi Screening

Jose M. Silva,¹ Krista Marran,¹ Joel S. Parker,³ Javier Silva,¹ Michael Golding,¹ Michael R. Schlabach,² Stephen J. Elledge,² Gregory J. Hannon,^{1*} Kenneth Chang¹

By virtue of their accumulated genetic alterations, tumor cells may acquire vulnerabilities that create opportunities for therapeutic intervention. We have devised a massively parallel strategy for screening short hairpin RNA (shRNA) collections for stable loss-of-function phenotypes. We assayed from 6000 to 20,000 shRNAs simultaneously to identify genes important for the proliferation and survival of five cell lines derived from human mammary tissue. Lethal shRNAs common to these cell lines targeted many known cell-cycle regulatory networks. Cell line-specific sensitivities to suppression of protein complexes and biological pathways also emerged, and these could be validated by RNA interference (RNAi) and pharmacologically. These studies establish a practical platform for genome-scale screening of complex phenotypes in mammalian cells and demonstrate that RNAi can be used to expose genotype-specific sensitivities.

shRNA Screen #1



C

Cell line	shRNAs	Genes
DLD-1	114	88
HCT116	202	115
HCC1954	177	159
HMEC	819	695

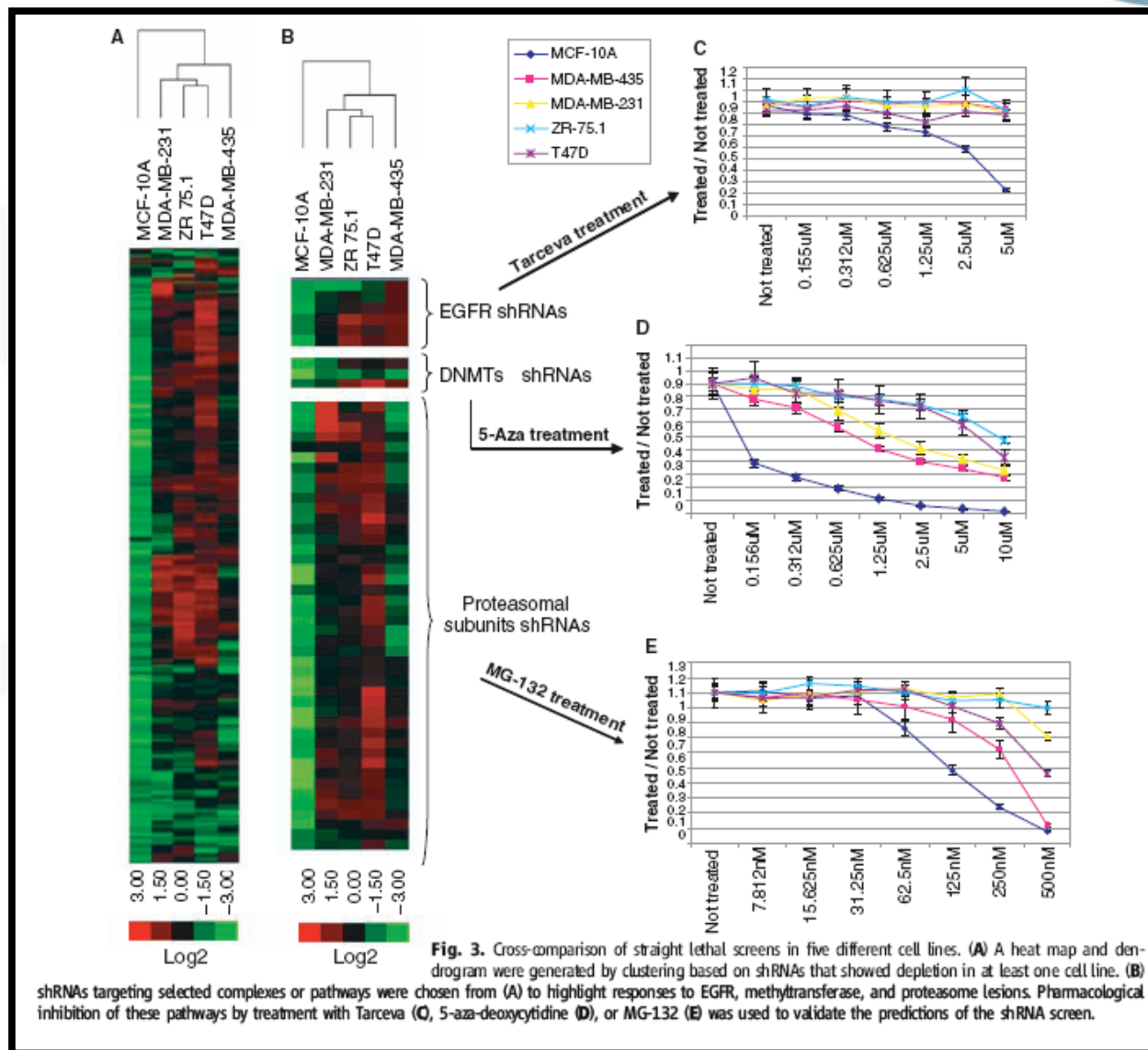
D

shRNAs	DLD-1	HCT116	HCC1954	HMEC	
DLD-1		44	33	88	19 common genes among all 4 cell lines
HCT116	61		53	95	
HCC1954	36	57		61	
HMEC	68	104	78		

23 common shRNAs among all 4 cell lines

Fig. 2. Pool-based dropout screen for genes required for cancer cell viability. **(A)** Overview of shRNA pool behavior in the screen. For each cell line, shRNAs were ranked on the basis of their mean normalized \log_2 Cy3/Cy5 ratios. The shaded rectangle indicates the \log_2 ratio range within which an shRNA's abundance was considered unchanged. **(B)** Clustering of the four cell lines with the antiproliferative shRNAs identified in the screen. The color scale represents mean normalized \log_2 Cy3/Cy5 ratios of the probes. **(C)** Antiproliferative shRNAs and genes that scored in the screen for each cell line are shown. **(D)** Summary of the common shRNAs (blue) and genes (red) identified in the screen. Overlapping antiproliferative shRNAs/genes between pairwise combinations of cell lines are displayed (DLD-1 and HMEC have more overlapping genes than shRNAs because, in some cases, different sets of shRNAs targeting the same gene scored in each line).

shRNA Screen #2



Silva et al,
Science, 2008

Analysis #1

- Are there well known causal cancer genes that show marked over-expression / DNA copy number change in additional cancer subsets?
 - Activating mutation (excluding translocations) OR High-level amplification
 - Identify additional patient populations with driver mutation

Dominant Mutations

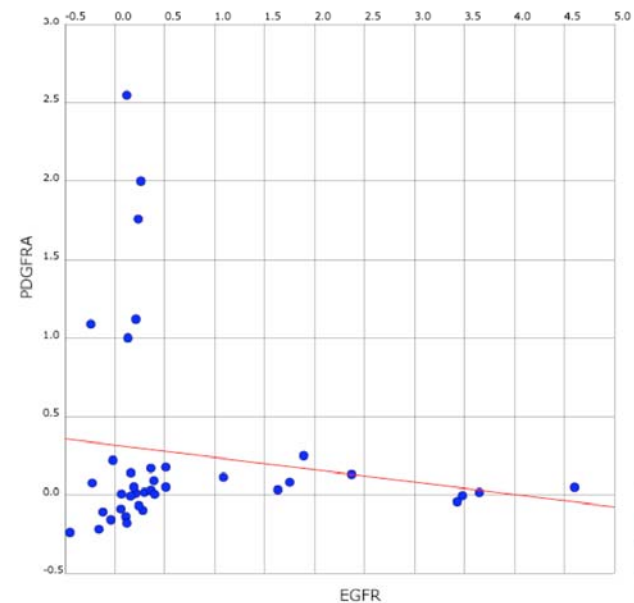
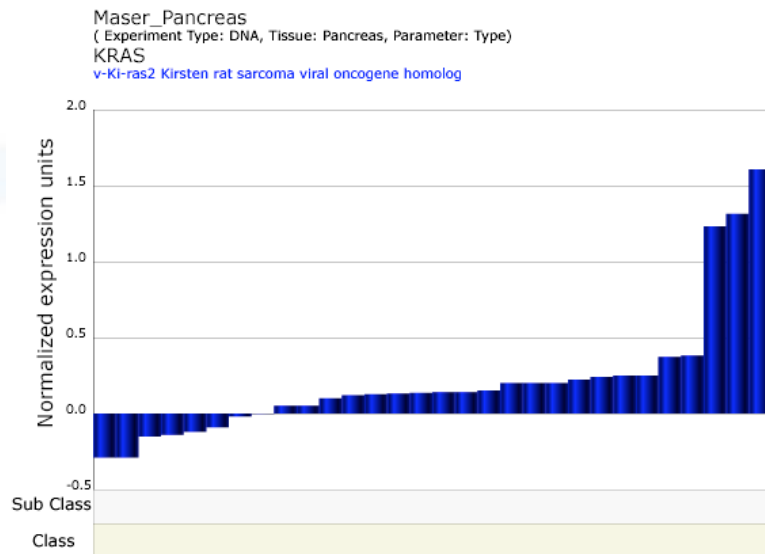
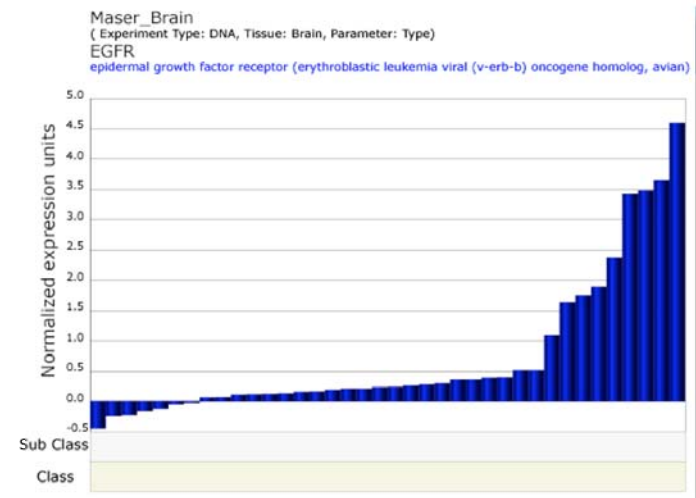
ABL1	FLT3	MYCL1
AKT2	GATA1	MYCN
ARHGAP26	GNAS	NOTCH1
BCL6	GOPC	NPM1
BRAF	HRAS	NRAS
CDK4	JAK2	PDGFRA
CEBPA	KIT	PIK3CA
CTNNB1	KRAS	PRKAR1A
EGFR	MET	PTPN11
ERBB2	MLL	REL
FGFR2	MPL	RET
FGFR3	MYC	SMO
		TSHR

aCGH Data

Prostate Cancer
Breast Cancer
Glioblastoma
Pancreatic Cancer
Ewing's Sarcoma
T-Cell Leukemia
Small Cell Lung Cancer

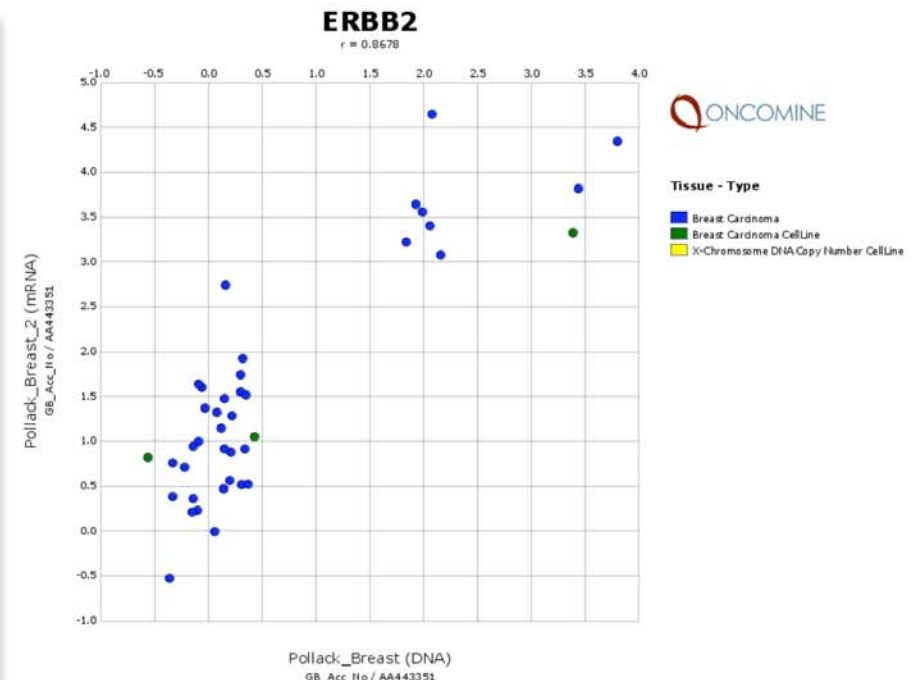
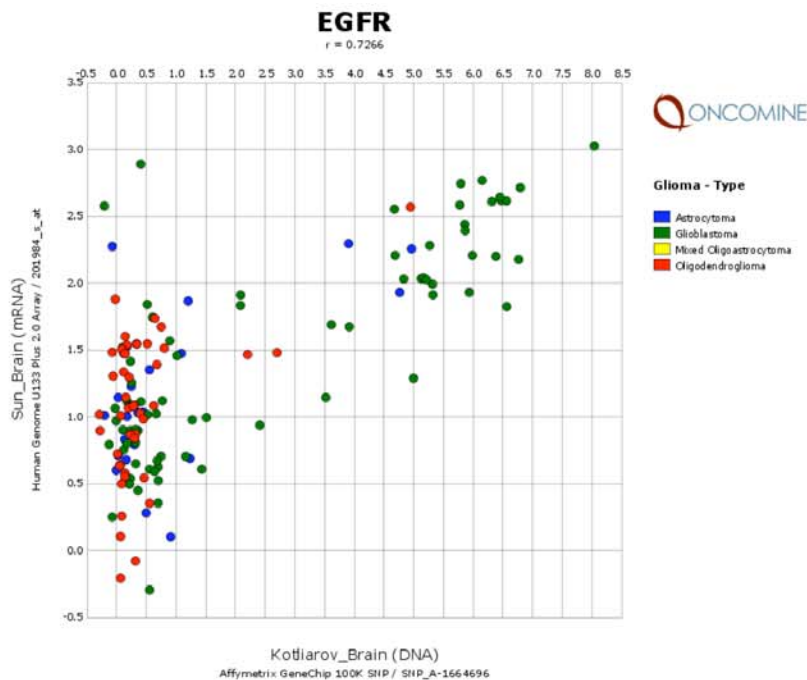
Analysis #1 Results

Study Name	Gene	Amplification Rank
Maser_Brain	EGFR	1
Kotliarov_Brain	EGFR	1
Pollack_Breast	ERBB2	2
Maser_Brain	PDGFRA	4
Maser_Pancreas	KRAS	7
Maser_Brain	KIT	12
Pollack_Breast	MYC	48
Maser_Brain	CDK4	58
Lapointe_Prostate_2	CDK4	76
Pollack_Breast	GNAS	81



Analysis #1 Followup

- DNA / mRNA correlation
 - High-level DNA amplifications of oncogenes should show evidence for over-expression in the same cases

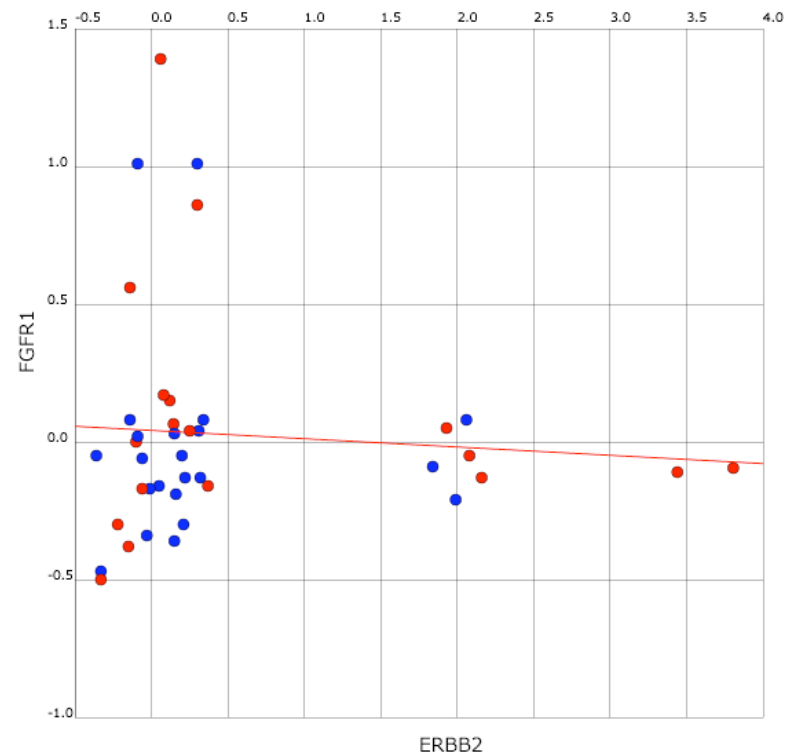


Analysis #1b

- Are there causal cancer genes (including translocations) that show high-level amplification in cancer subsets?

Study Name	Gene	Amplification Rank
Pollack_Breast	ERBB2	2
Pollack_Breast	FGFR1	26
Pollack_Breast	CLTC	52
Pollack_Breast	MSH2	69
Pollack_Breast	GNAS	81
Pollack_Breast	PATZ1	82

- FGFR1 involved in translocations in NHL and MPD



Analysis #1c

- Are there causal cancer genes markedly over-expressed in cancer subsets?

Cancer Type	# Studies	Gene	Best COPA Rank
Breast	19	ERBB2	6
Prostate	8	ERG	2
Leukemia	7	PBX1	1
Myeloma	6	FGFR3	1
Leukemia	6	HOXA9	2
Leukemia	5	IGL@	31
Leukemia	5	MYH11	25
Leukemia	5	HOXA11	36
Breast	5	SSX2	30
Breast	5	HOXA9	42
Myeloma	5	WHSC1	1
Brain	5	TRD@	5
Brain	5	HOXA9	5
Breast	5	IGL@	16
Leukemia	4	MUC1	14
Brain	4	IGL@	4
Sarcoma	4	MYH11	1
Leukemia	4	EVI1	8
Prostate	4	ETV1	1
Leukemia	4	MAF	9
Breast	4	EGFR	12

Cancer Type	# Studies	Gene	Best COPA Rank
Lymphoma	3	KIT	41
Leukemia	3	RUNX1T1	1
Myeloma	3	CCND1	10
Myeloma	3	MAF	22
Bladder	3	FGFR1	78
Breast	3	TCL1A	18
Breast	3	NUMA1	6
Breast	3	NTRK3	75
Breast	3	MYH11	14
Breast	3	SMOX	8
Brain	3	TRA@	2
Brain	3	GPC3	25
Brain	3	CHN1	11
Brain	3	EGFR	9
Lung	2	CBL	16
Lung	2	WT1	36
Lung	2	BIRC3	84
Lung	2	MYH11	41
Brain	2	PAX3	46
Brain	2	HMGA2	67
Colon	2	SET	38
Colon	2	EGFR	67
Colon	2	PPBP	5
Liver	2	PML	48

Related Questions

- Large scale tumor sequencing efforts often identify rare mutations
 - Difficult to discern drivers from passengers
- Evidence for high-level amplification or over-expression in subsets provides additional evidence
 - e.g. EGFR mutation or high-level amplification

Analysis #2

- Can functional genomics data serve as a filter to identify amplified / over-expressed genes that are drivers of cancer?

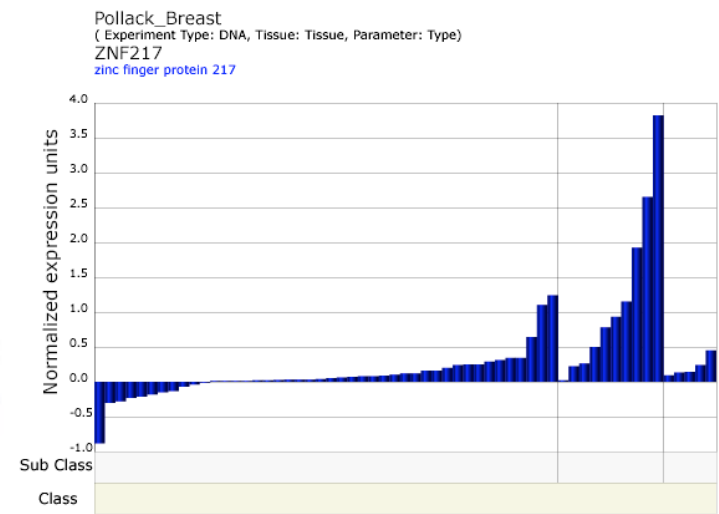
The screenshot shows a web interface with a search bar and a 'Concept Search' button. Below the search bar, it states 'Returned 8 concepts for shRNA Literature-defined Concepts'. The results are listed in a table with 8 rows, each containing a document icon, a text description of the concept, and a gear icon.

Concept	Count (n)
shRNAs that reduce viability of MCF-10A non-tumorigenic breast epithelial cells	626
shRNAs that reduce viability of normal human mammary epithelial cells (HMECs)	591
shRNAs that reduce viability of HCC-1954 breast cancer cells	145
shRNAs that reduce viability of HCT116 colon cancer cells	131
shRNAs that reduce viability of DLD-1 colon cancer cells	77
shRNAs that reduce viability of MDA-MB-231 breast adenocarcinoma cells	67
shRNAs that reduce viability of MDA-MB-435S ductal breast adenocarcinoma cells	64
shRNAs that reduce viability of T47D ductal breast adenocarcinoma cells	24

Amplified & Functional Driver

- Genes with significant DNA amplifications & evidence for functional relevance
 - Enrichment for known oncogenes

Study	# shRNA	Gene	Amplification Rank
Maser_Brain	3	EGFR	1
Kotliarov_Brain	3	EGFR	1
Maser_Pancreas	1	KRAS	7
Lapointe_Prostate_2	3	CASP3	10
Pollack_Breast	1	PSMB3	11
Maser_Brain	1	KIT	12
Maser_Pancreas	1	BPTF	12
Maser_Brain	1	KDR	14
Lapointe_Prostate_2	1	AKT3	14
Pollack_Breast	2	ZNF217	15
Maser_Brain	1	PPP1R15B	17
Lapointe_Prostate_2	1	TNFRSF1A	19
Maser_Brain	1	MDM4	23
Lapointe_Prostate_2	1	SMARCA2	23
Kotliarov_Brain	2	PRPS2	24
Maser_Brain	2	PIK3C2B	26
Pollack_Breast	1	FGFR1	26
Kotliarov_Brain	2	NR0B1	27
Kotliarov_Brain	1	COL9A1	27
Pollack_Breast	2	RAD51C	28
Maser_Pancreas	1	CDH18	32
Pollack_Breast	1	STK3	32
Pollack_Breast	2	PCK1	38
Pollack_Breast	2	MYC	39
Kotliarov_Brain	2	NBN	39



Colon Cancer

- Functional relevance & over-expressed in subsets
 - Reduced viability when genes are targeted by shRNA

Gene	COPA Score	Datasets
MDM2	70	2
PCNA	26	1
PKM2	96	1
SAE1	61	1
COPS4	87	1
MAPKAPK5	57	1
EIF3B	76	1
EIF3E	91	1
EWSR1	94	1
PEA15	76	1
PRPS2	36	1
PPP1R7	93	1
MYC	53	1
CUL4A	73	1

Analysis #3

- Can functional data serve as a filter to identify driver mutations from mutation screens?
 - Breast & Colon CAN genes
 - Sanger kinase mutation screen

CAN Genes

- Genes with somatic mutations & functional dependence

Breast Cancer

<u>Gene</u>	<u>Name</u>	<u># shRNA</u>
HDAC4	histone deacetylase 4	2
PRPF4B	PRP4 pre-mRNA processing factor 4 homolog B (yeast)	2
BRCA1	breast cancer 1, early onset	2
ANAPC2	anaphase promoting complex subunit 2	2
ABCB10	ATP-binding cassette, sub-family B (MDR/TAP), member 10	2
KIAA0999	KIAA0999 protein	2
COL11A1	collagen, type XI, alpha 1	1
MAP3K6	mitogen-activated protein kinase kinase kinase 6	1
GPNMB	glycoprotein (transmembrane) nmb	1
EGFL6	EGF-like-domain, multiple 6	1
PIGU	phosphatidylinositol glycan anchor biosynthesis, class U	1
KEAP1	kelch-like ECH-associated protein 1	1

Colon Cancer

<u>Gene</u>	<u>Name</u>	<u># shRNA</u>
UHRF2	ubiquitin-like, containing PHD and RING finger domains, 2	1
FBXW7	F-box and WD repeat domain containing 7	1

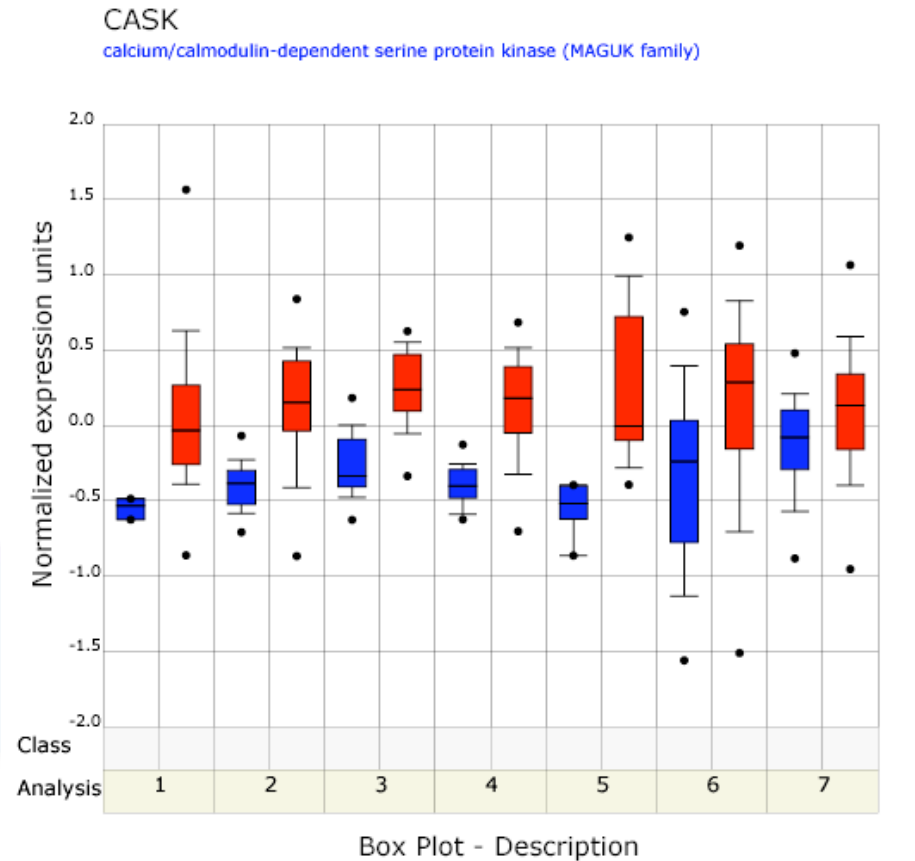
Kinase Screen

- Kinases with somatic mutations & functional dependence

Gene	Name	# shRNA
CASK	calcium/calmodulin-dependent serine protein kinase (MAGUK family)	4
JAK1	Janus kinase 1 (a protein tyrosine kinase)	4
PIM2	pim-2 oncogene	4
PRPF4B	PRP4 pre-mRNA processing factor 4 homolog B (yeast)	4
CDKL5	cyclin-dependent kinase-like 5	4
TRRAP	transformation/transcription domain-associated protein	4
PLK1	polo-like kinase 1 (Drosophila)	4
GAK	cyclin G associated kinase	3
ICK	intestinal cell (MAK-like) kinase	3
TTK	TTK protein kinase	3
DDR1	discoidin domain receptor family, member 1	3
PKN1	protein kinase N1	3
RAF1	v-raf-1 murine leukemia viral oncogene homolog 1	3
MAPK15	mitogen-activated protein kinase 15	3
DAPK1	death-associated protein kinase 1	3
ERBB4	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)	3
LMTK3	lemur tyrosine kinase 3	3
STK16	serine/threonine kinase 16	3
CAMKK2	calcium/calmodulin-dependent protein kinase kinase 2, beta	3
MAP3K3	mitogen-activated protein kinase kinase kinase 3	3
ULK1	unc-51-like kinase 1 (C. elegans)	3

Putting it all together

- Target characteristics
 - Druggable - kinases, etc.
 - Over-expressed in cancer
 - Somatic mutations in cancer
 - Functional dependence
- Ex: CASK
 - Kinase mutation screen
 - Multiple lines with functional dependence
 - Widely over-expressed in certain cancer types
 - Glioblastoma, cervical, head & neck, seminoma, bladder, liver



Summary

- Multiple levels of genomic data should be considered in target discovery
 - Gene expression
 - DNA copy number
 - Somatic mutations
 - Functional screens
- Integrated analysis approaches are necessary to fully utilize diverse genomic data
 - Filtering is just the first step

Summary #2

■ Improved Datasets

- aCGH - increased resolution, more samples
- Functional screens - increased shRNA library size, more cell lines
- Mutation sequencing - larger patient cohorts

■ Integrated Datasets

- Same samples profiled by multiple genomic technologies
 - Tumors - aCGH, gene expression, methylation, mutation sequencing, miRNA expression
 - Cell Lines - AND functional screens

Oncomine Access

- Register at www.oncomine.org
- Freely available to academic researchers
 - 360 Studies
 - 25,000+ arrays
- Data you'd like to see in Oncomine?
 - Check catalog
 - Profile search for first author
 - Tell us: support@compendiabiotech.com

