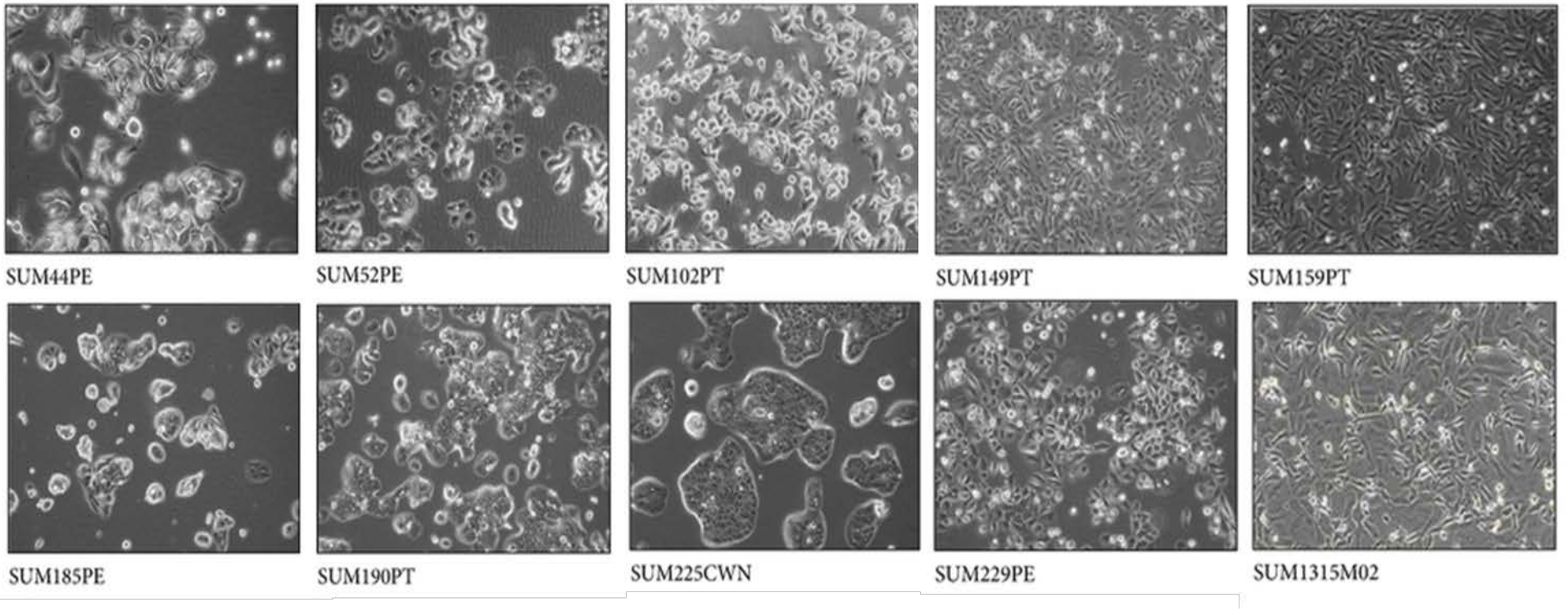
SUM Cell Line Knowledge Base (SLKBase): Using functional genomics to elucidate oncogene signatures and to navigate critical pathways for growth and survival of breast cancer cells

Stephen P. Ethier, Ph.D.

In the 1990s, the Ethier lab developed a panel of 10 breast cancer cell lines known as the SUM cell lines, which represent all major breast cancer subtypes



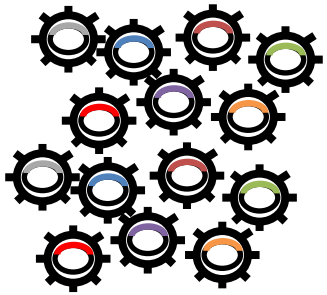
Descriptive Genomic Data sets associated with the SUM lines

- **Comparative genomic hybridization: copy number gain/loss**
- **Expression profiling: relative gene expression compared to normal cells/tissues**
- **Exome sequencing: SNP/point mutations**

Putting the "function" in functional genomics

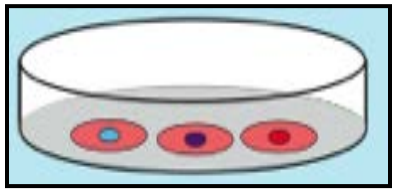
- **Genome scale shRNA screen data for all ten SUM lines**
  - **This identifies the genes that are essential for the proliferation and/or survival of each cell line**
- **Validation of screen hits using individual shRNA vectors for key genes identified in the screens**
- **Validation data obtained using drugs that target key genes identified in the screens**

RNAi-Based Loss of Function Screens



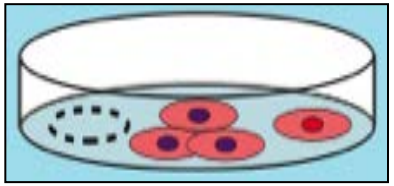
Pooled shRNA expressing lentivirus library

- **Total of 82,500 shRNA constructs**
- **Targets 15,377 genes**
- **On average 4-5 shRNAs/gene**
- **Each construct has a unique barcode sequence**



Target Cells
MOI = 0.1
Representation = 1000

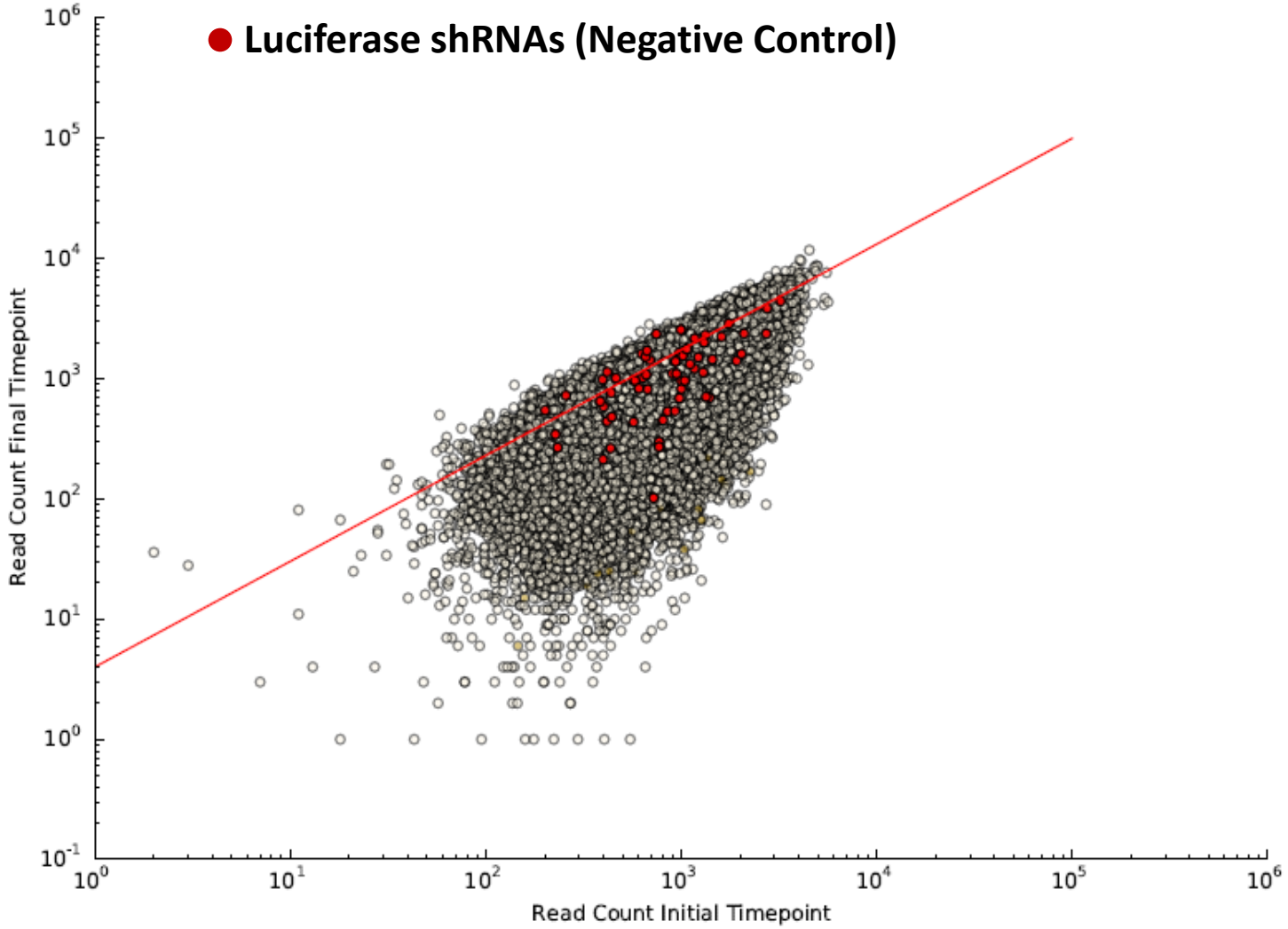
5-7 population doublings



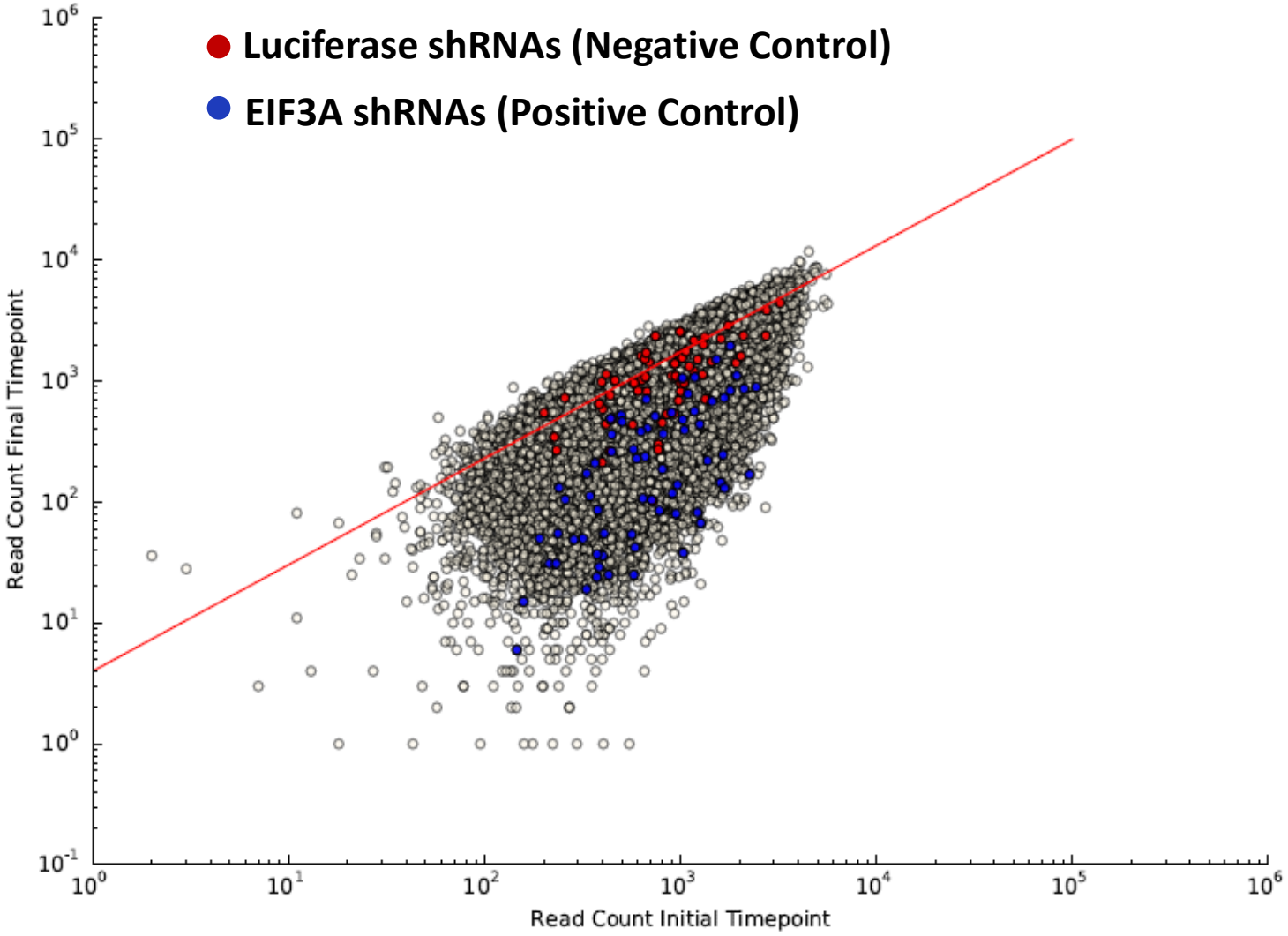
Sequence Barcodes

FASTQ Files

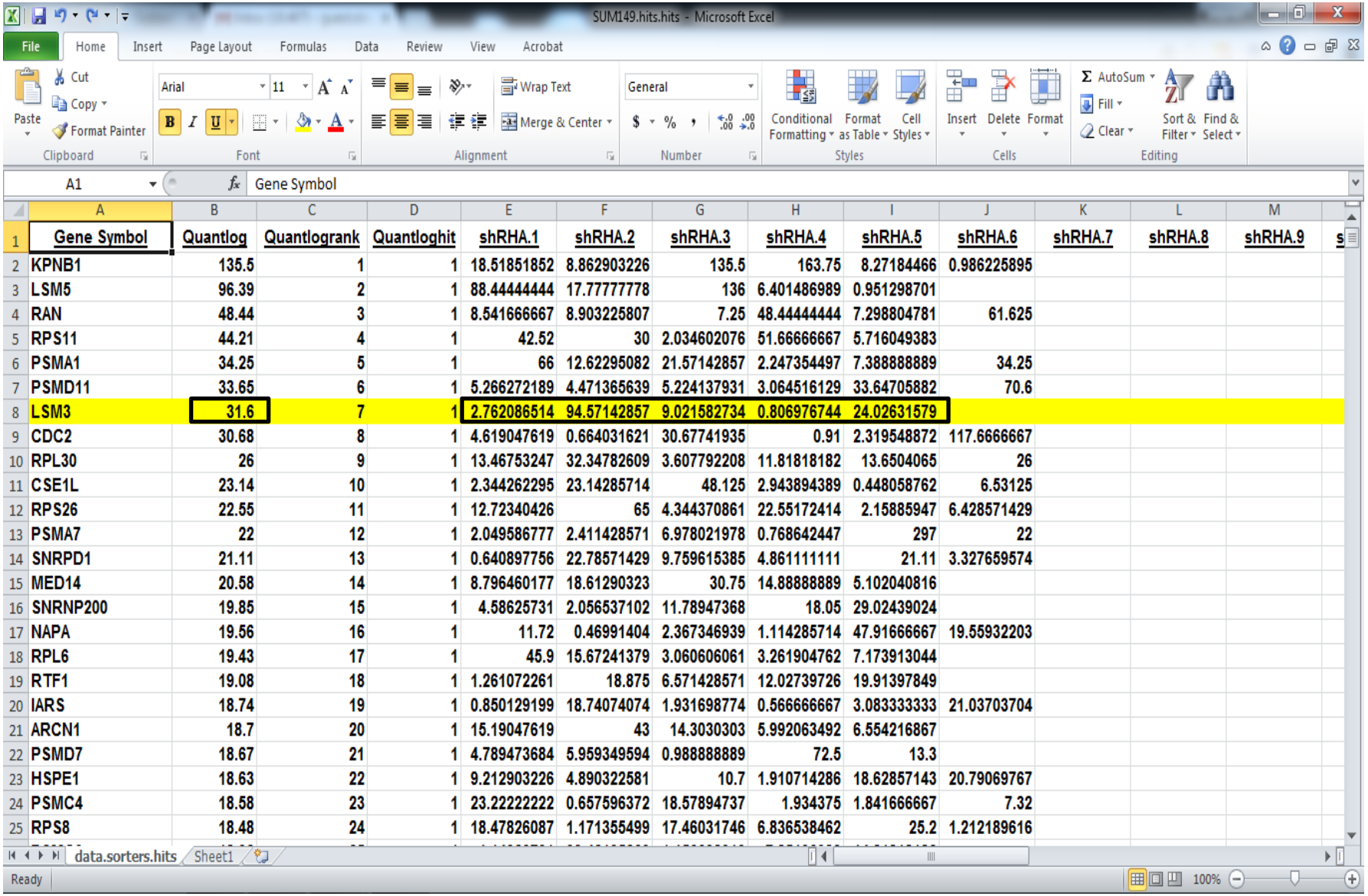
Functional Genomic Screen in a Breast Cancer Cell Line



Functional Genomic Screen in a Breast Cancer Cell Line



Data Analysis Pipeline Output File

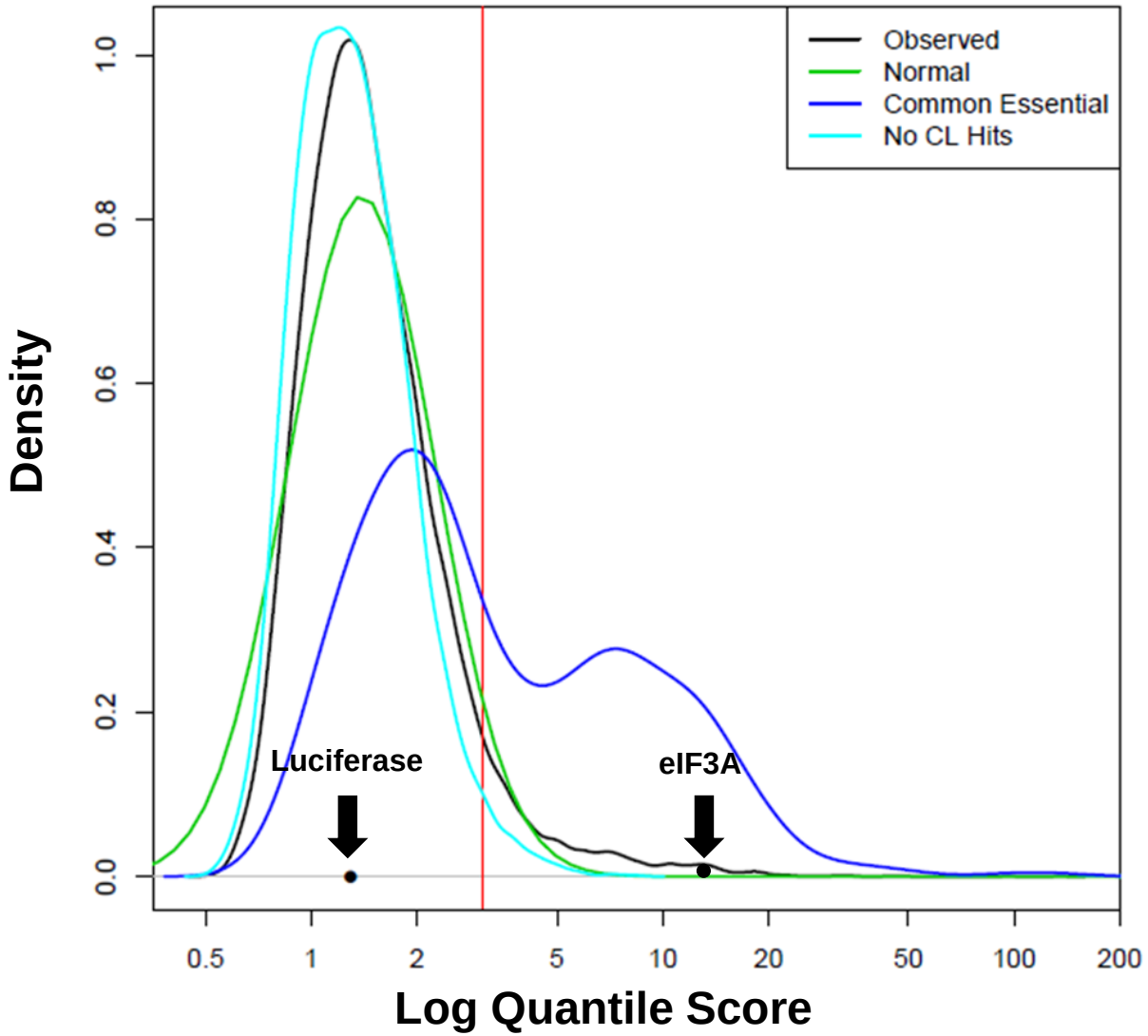


SUM149.hits.hits - Microsoft Excel

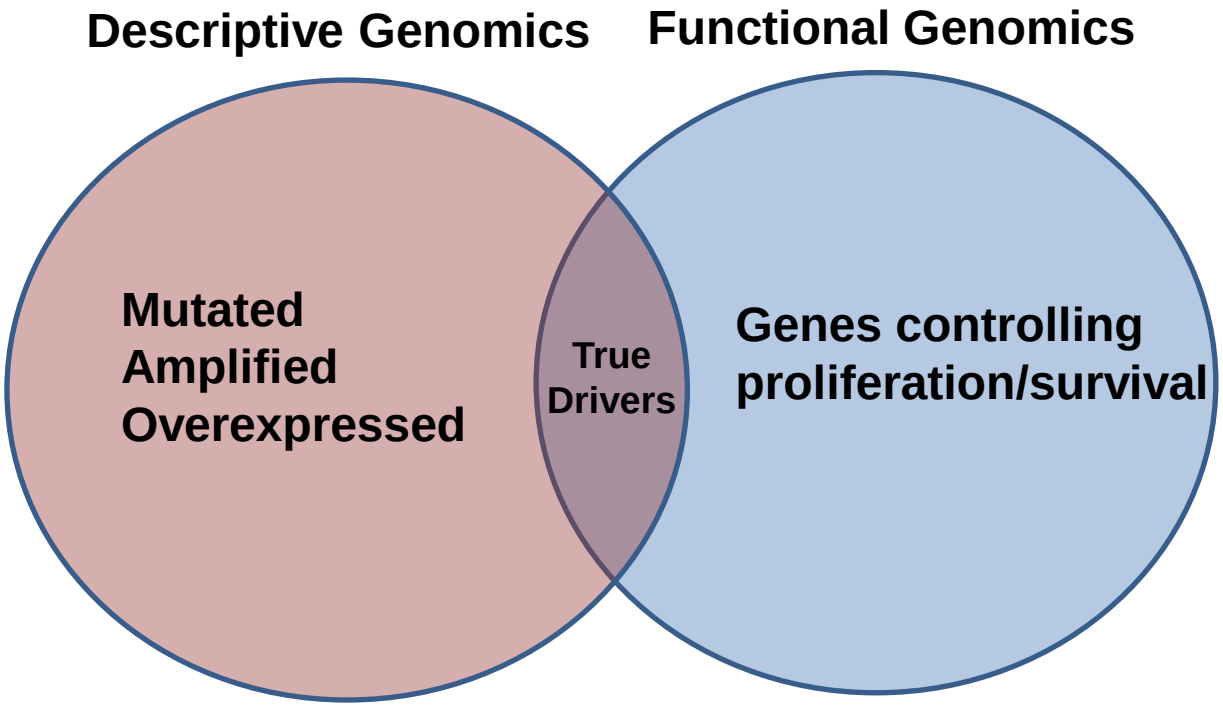
Gene Symbol	Quantlog	Quantlogrank	Quantloghit	shRHA.1	shRHA.2	shRHA.3	shRHA.4	shRHA.5	shRHA.6	shRHA.7	shRHA.8	shRHA.9
KPNB1	135.5	1	1	18.51851852	8.862903226	135.5	163.75	8.27184466	0.986225895			
LSM5	96.39	2	1	88.44444444	17.77777778	136	6.401486989	0.951298701				
RAN	48.44	3	1	8.541666667	8.903225807	7.25	48.44444444	7.298804781	61.625			
RPS11	44.21	4	1	42.52	30	2.034602076	51.66666667	5.716049383				
PSMA1	34.25	5	1	66	12.62295082	21.57142857	2.247354497	7.388888889	34.25			
PSMD11	33.65	6	1	5.266272189	4.471365639	5.224137931	3.064516129	33.64705882	70.6			
LSM3	31.6	7	1	2.762086514	94.57142857	9.021582734	0.806976744	24.02631579				
CDC2	30.68	8	1	4.619047619	0.664031621	30.67741935	0.91	2.319548872	117.6666667			
RPL30	26	9	1	13.46753247	32.34782609	3.607792208	11.81818182	13.6504065	26			
CSE1L	23.14	10	1	2.344262295	23.14285714	48.125	2.943894389	0.448058762	6.53125			
RPS26	22.55	11	1	12.72340426	65	4.344370861	22.55172414	2.15885947	6.428571429			
PSMA7	22	12	1	2.049586777	2.411428571	6.978021978	0.768642447	297	22			
SNRPD1	21.11	13	1	0.640897756	22.78571429	9.759615385	4.861111111	21.11	3.327659574			
MED14	20.58	14	1	8.796460177	18.61290323	30.75	14.88888889	5.102040816				
SNRNP200	19.85	15	1	4.58625731	2.056537102	11.78947368	18.05	29.02439024				
NAPA	19.56	16	1	11.72	0.46991404	2.367346939	1.114285714	47.91666667	19.55932203			
RPL6	19.43	17	1	45.9	15.67241379	3.060606061	3.261904762	7.173913044				
RTF1	19.08	18	1	1.261072261	18.875	6.571428571	12.02739726	19.91397849				
IARS	18.74	19	1	0.850129199	18.74074074	1.931698774	0.566666667	3.083333333	21.03703704			
ARCN1	18.7	20	1	15.19047619	43	14.3030303	5.992063492	6.554216867				
PSMD7	18.67	21	1	4.789473684	5.959349594	0.988888889	72.5	13.3				
HSPE1	18.63	22	1	9.212903226	4.890322581	10.7	1.910714286	18.62857143	20.79069767			
PSMC4	18.58	23	1	23.22222222	0.657596372	18.57894737	1.934375	1.841666667	7.32			
RPS8	18.48	24	1	18.47826087	1.171355499	17.46031746	6.836538462	25.2	1.212189616			

Identifying Hits in Functional Genomic Screen Data

Turquoise line is a density plot of the data for 4,389 genes that are not essential in any tested cell line



Integrating Functional and Descriptive Genomics: Oncogene Signatures



True drivers are BOTH genetically altered and functional.

Complete set of true drivers form an Oncogene Signature for that tumor

Signatures can illuminate the hidden biology

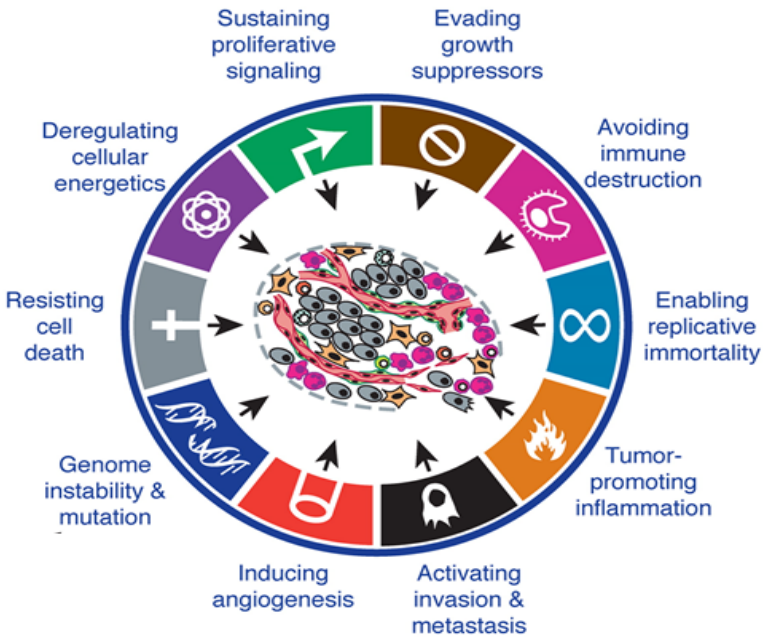
Signatures can identify therapeutic targets

Oncogene Signatures: The full complement of activated oncogenes that drive the phenotypes of human cancers

Oncogene #1

Oncogene #2

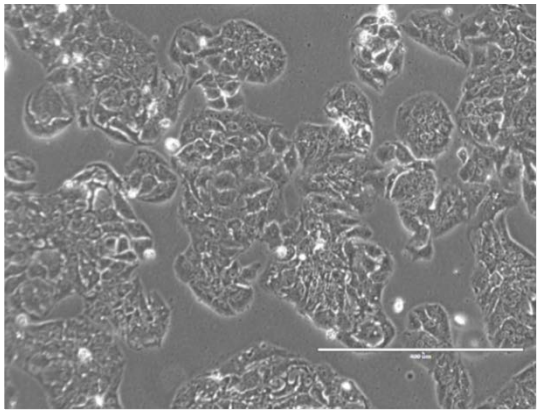
Oncogene #X



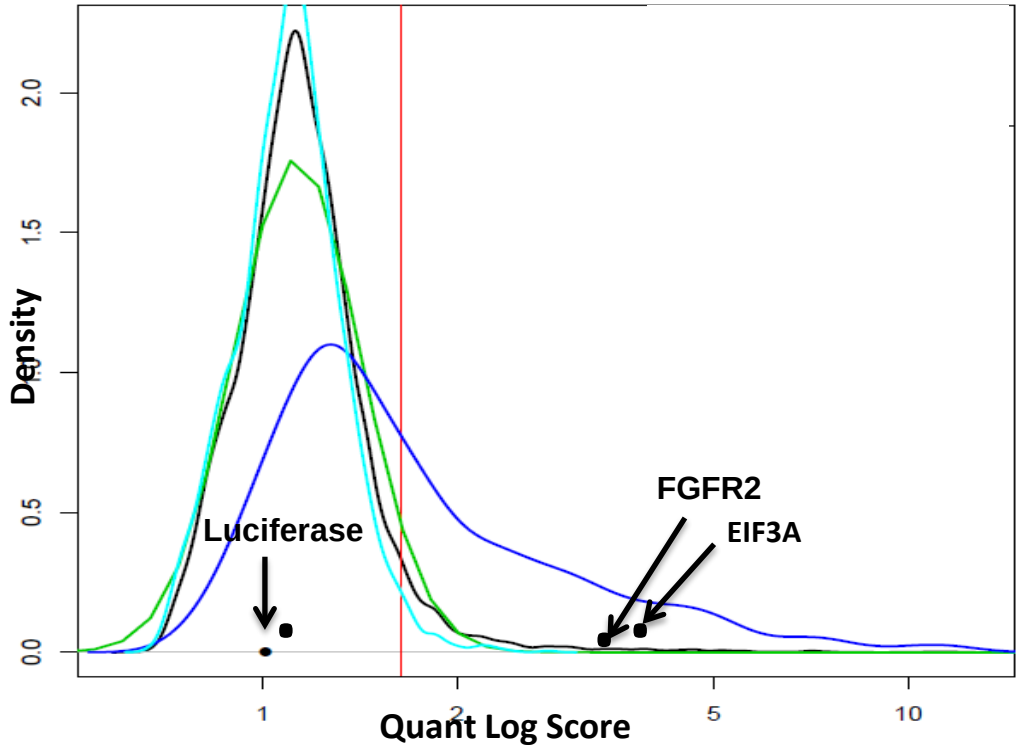
Oncogene Signatures can predict novel drug strategies

- Oncogene addiction is the basis for the efficacy of targeted drugs
- Drugs that effectively target driving oncogenes have potent effects at low doses with a high therapeutic index
- Targeting driving oncogenes can result in cell death, not just growth arrest
- Targeting multiple oncogenes in oncogene signatures can overcome the drug resistance that is inevitable with single drug-single target approaches

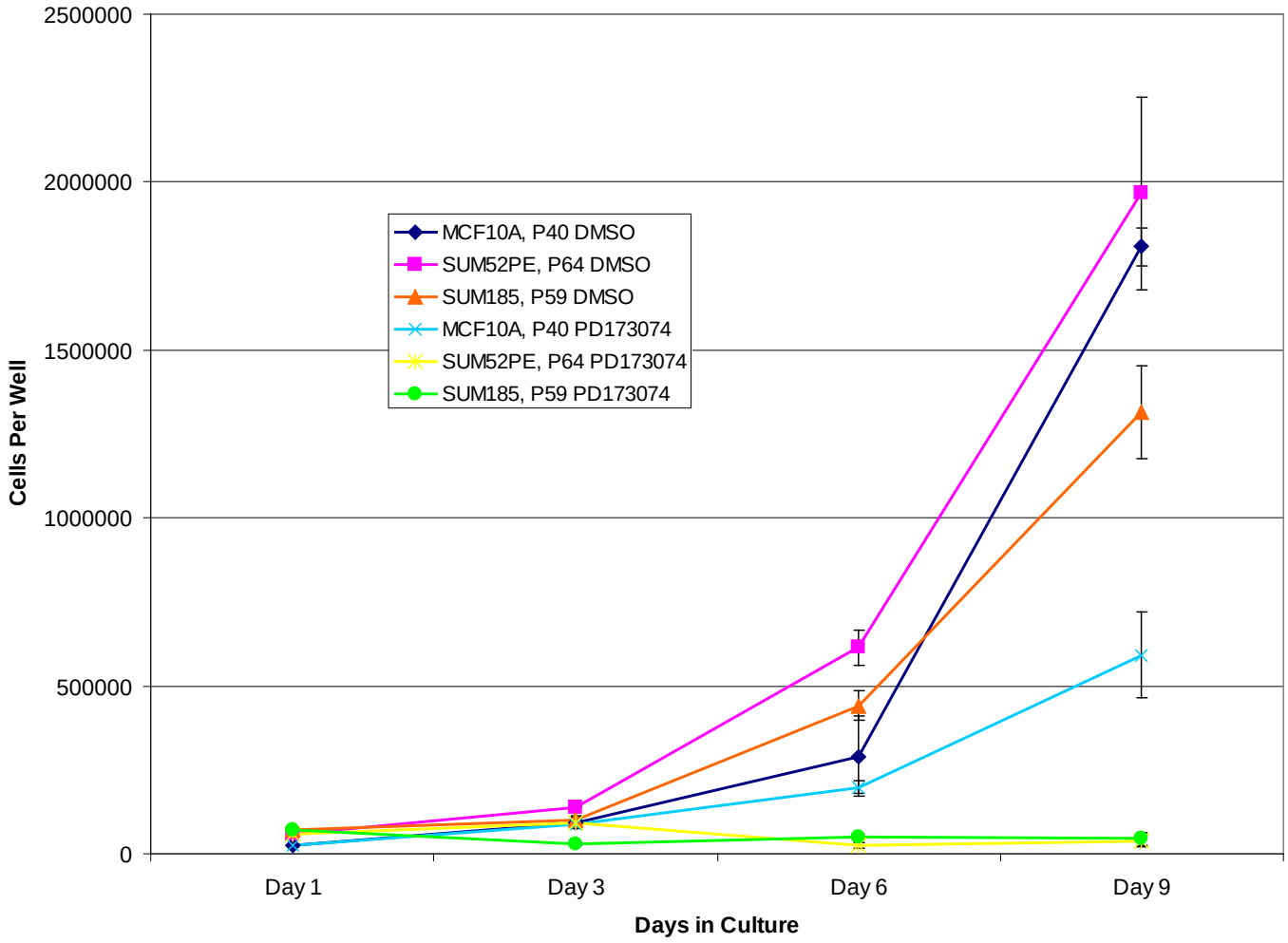
Assembling an Oncogene Signature: SUM-52



- Isolated from a pleural effusion metastasis
- Luminal B
- ER/PR negative
- High level FGFR2 amplification



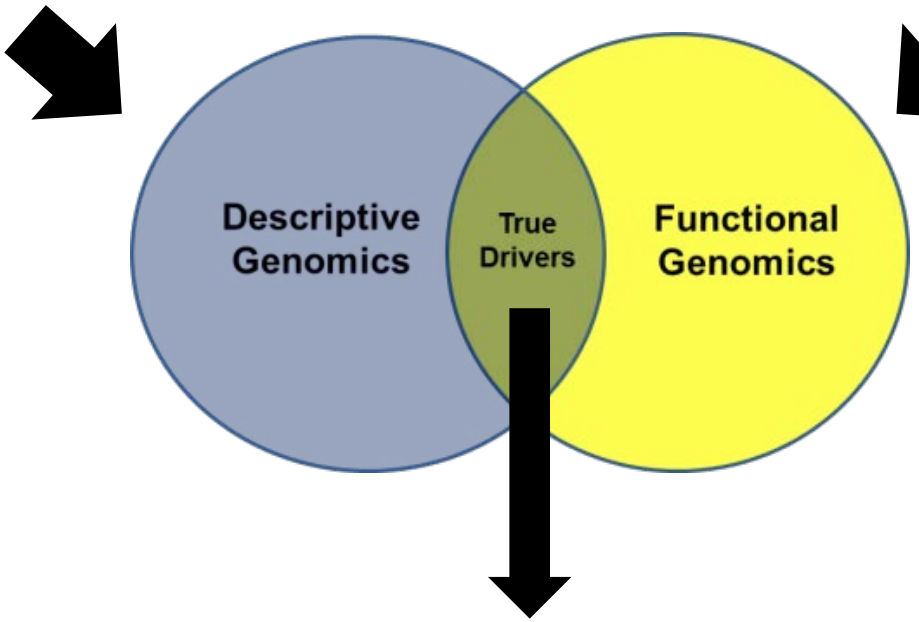
SUM-52 cells are highly responsive to FGFR targeted drugs



Generating Functional Oncogene Signature for SUM-52 cells

SUM-52 = 956 copy number amplified genes

SUM-52 = 863 hits in screen



Cancer Knowledge Databases
COSMIC/GISTIC

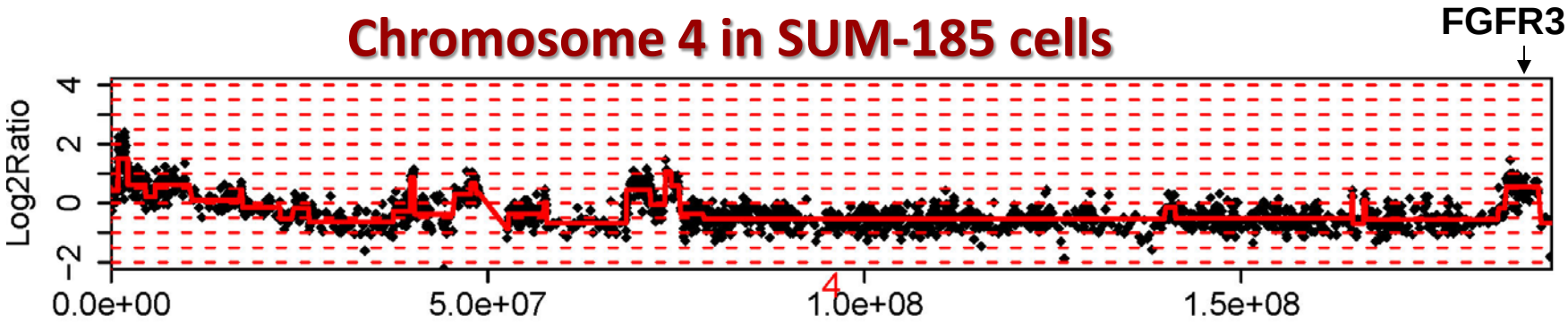
5 amplified driver oncogenes

Functional Oncogene Signatures

	SUM-52	SUM-185	SUM-190	SUM-225
Amplified	FGFR2	FGFR3	EPHA5	ERBB2
	CDK6	BCL2L1		RAD21
	KAT6A			HSF1
	KCNB1			
	DDX5			
Point Mutated	CDKN2A	PIK3CA	PIK3CA	BIRC6

- **Signatures significantly reduce the complexity of each cancer genome**
- **Signatures identify well-known driver oncogenes (e.g. FGFRs, HER2, CDK6)**
- **Signatures identify previously uncharacterized drivers (e.g. KAT6A & BCL2L1)**

Chromosome 4 in SUM-185 cells



SYMBOL	GENE NAME	FOLD CHANGE	P VALUE (FDR)
TACC3	transforming, acidic coiled-coil containing protein 3	9.9	0.00015
CTBP1	C-terminal binding protein 1	3.2	0.00030
FGFR3	fibroblast growth factor receptor 3	16.0	0.00077
MAEA	macrophage erythroblast attacher	3.3	0.00083
FGFRL1	fibroblast growth factor receptor-like 1	3.7	0.00087
WHSC2	Wolf-Hirschhorn syndrome candidate 2	2.0	0.00315
MGC21675	hypothetical protein MGC21675	2.8	0.00836
TMEM129	transmembrane protein 129	3.5	0.00938
LETM1	leucine zipper-EF-hand containing transmembrane protein 1	2.8	0.01186
KIAA1530	KIAA1530 protein	2.4	0.01256
IDUA	iduronidase, alpha-L-	2.5	0.01670
SLBP	stem-loop (histone) binding protein	3.7	0.01819
WHSC1	Wolf-Hirschhorn syndrome candidate 1	2.0	0.01959
CRIPAK	cysteine-rich PAK1 inhibitor	2.8	0.02187
SPON2	spondin 2, extracellular matrix protein	-6.8	0.05501
DGKQ	diacylglycerol kinase, theta 110kDa	1.1	0.18712
NAT8L	N-acetyltransferase 8-like	1.7	0.21110
POLN	polymerase (DNA directed) nu	-1.3	0.23409
RNF212	ring finger protein 212	-1.5	0.23886
SLC26A1	solute carrier family 26 (sulfate transporter), member 1	-1.1	0.32088
MXD4	MAX dimerization protein 4	1.1	0.42822
TMEM175	transmembrane protein 175	-1.0	0.43751

SUM-185 cells.
Screen hits and amplified

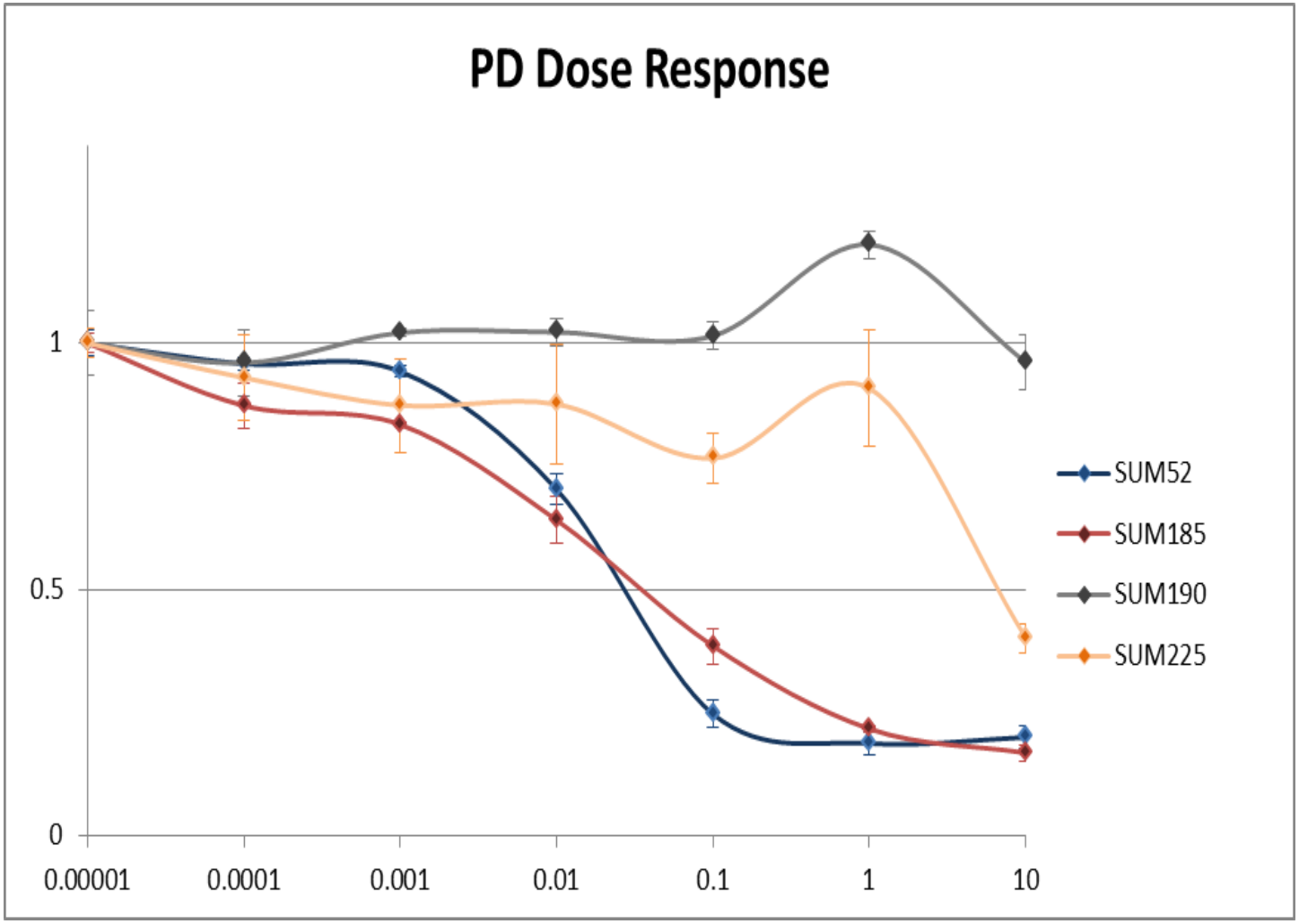
Gene symbol	Locus	RNAi Screen Rank[a]	Expression[b]	Gene Ranker Score[c]	GISTIC FA[d]	GISTIC Peak[e]
CTBP1	4p16.3c	938	1.63	2.5	No	No
FGFR3	4p16.3b	261	2.87	8.5	Yes	Yes
ANKRD17	4q13.3d	213	0.94	1	Yes	Yes
IDUA	4p16.3c	1245	2.02	2	No	No
GPT2	16q11.2i-q12.1a	911.5	2.51	1	No	No
HSPBP1	19q13.42b	997	0.48	0	Yes	Yes
NDUFA13	19p13.11a	216	0.48	2.25	No	No
ELL	19p13.11c	2842	0.95	3.25	Yes	No
ACSS1	20p11.21a	195	1.24	2	No	No
BCL2L1	20q11.21b	5	1.17	3.5	Yes	Yes
COX4I2	20q11.21b	1159	-0.08	1	Yes	Yes
POFUT1	20q11.21b	252	0.64	2.25	Yes	No
ID1	20q11.21b	194	-0.69	2.5	Yes	No
NKX2-2	20p11.22b	876	1.56	0	No	No
SSTR4	20p11.21c	980	0.13	0	No	No

Functional Oncogene Signatures

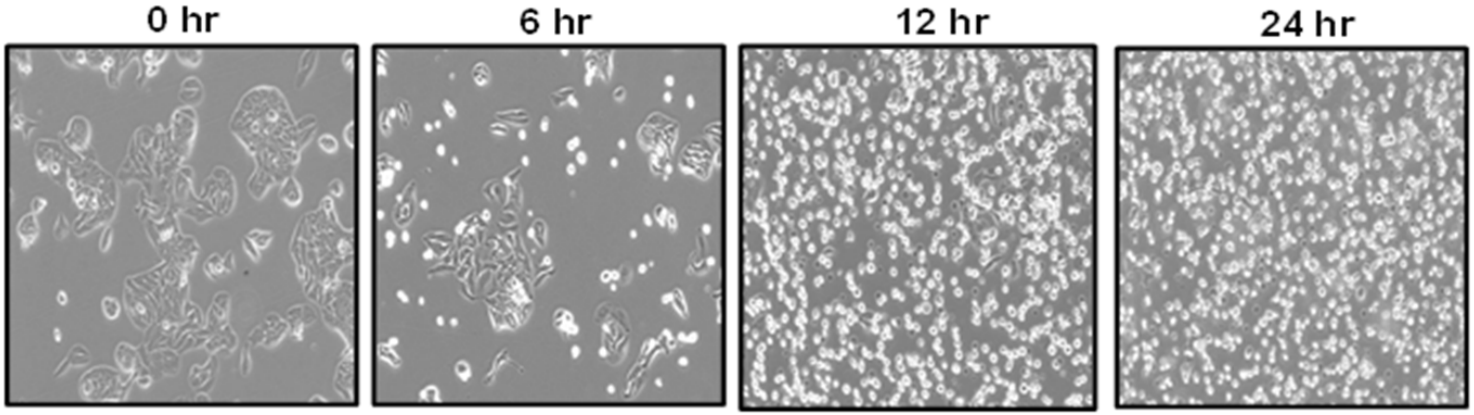
	SUM-52	SUM-185	SUM-190	SUM-225
Amplified	FGFR2	FGFR3	EPHA5	ERBB2
	CDK6	BCL2L1		RAD21
	KAT6A			HSF1
	KCNB1			
	DDX5			
Point Mutated	CDKN2A	PIK3CA	PIK3CA	BIRC6

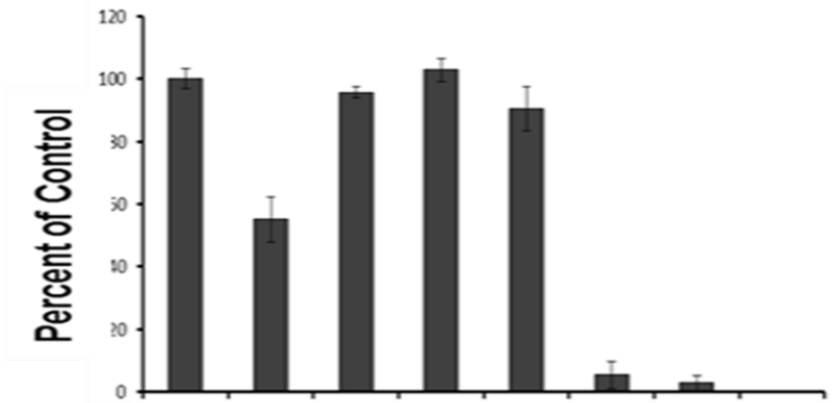
- **Signatures significantly reduce the complexity of each cancer genome**
- **Signatures identify well-known driver oncogenes (e.g. FGFRs, HER2, CDK6)**
- **Signatures identify previously uncharacterized drivers (e.g. KAT6A & BCL2L1)**

SUM-52 and SUM-185 cells are exquisitely sensitive to the FGFR inhibitor PD173074

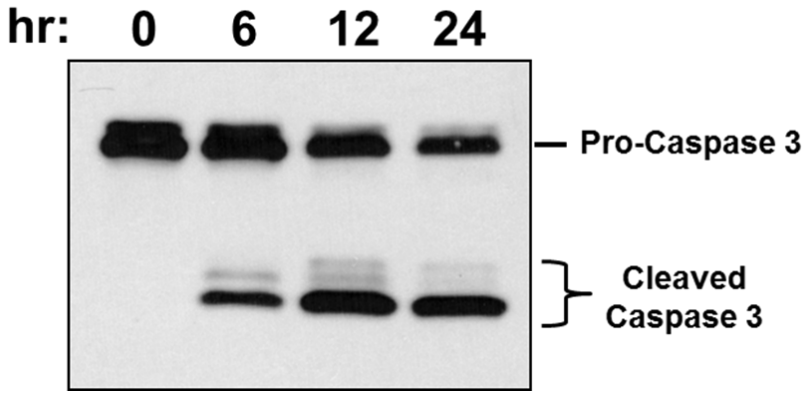


Three oncogenes in the SUM-185 oncogene signature are druggable; FGFR3, PIK3CA, BCL2L1





.05uM PD173074	-	-	+	-	+	-	+	+
.5uM A66	-	-	-	+	+	+	-	+
.1uM Navitoclax	-	+	-	-	-	+	+	+



Integrating Functional and Descriptive Genomics: Pathway identification

- Our shRNA screens generate approximately 1000 hits in every cell line
- It is difficult to analyze these hit lists simply by browsing through the gene list
- One can group sets of genes by Gene Ontology (GO) annotations, which is helpful
- A more powerful way to extract biological meaning from the screen hit data is to display the hits genes on KEGG pathways

The SUM Breast Cancer Cell Line Knowledge Base (SLKBase)

Using functional genomics to navigate critical pathways for growth and survival of breast cancer cells

SUM Cell Lines Gateway Page

The SUM Cell Lines Gateway is the heart and soul of the Knowledge Base. This is the space where we attempt to present all of the Omics data that we have generated with each of the SUM lines in a way that illuminates the biology of each cell line. Each Cell Line page contains information on how we developed the line and everything we know about the patient from which the cells came, and we also summarize what we know about the oncogenes that are likely driving their biology. But the key to all of this are the KEGG pathway pages associated with each cell line. In these KEGG pathways, we present the data from our shRNA genome scale screen in several KEGG pathways. This method is superior to providing long lists of hit genes by allowing easy visualization of the screen hit data in pathways of interest to breast cancer researchers. This is a powerful approach because it allows investigators to identify critical pathways, and not just critical genes, as elucidated by the screen. When multiple hit genes map to a pathway, this is clear evidence for an important aspect of the biology of the cells, and points to specific genes within the pathways that may be targetable. By viewing the same pathways in different cell lines, one can get a sense of the differing biology of the different ce;; lines, and all of this is directly related to the ultimate goal of this Knowledge Base, which is to develop novel strategies to reverse engineer breast cancer cells.

Within each KEGG Pathway page for each cell line, there is a link to the KEGG Pathway Engine and this is where cell line users can dig deeper into the data sets. Users can choose any KEGG pathway in any cell line, and see both how the screen hit data and the gene expression data map onto the KEGG pathways of interest. In addition, genes that have genomic alterations are marked in these pathways, and the data for each hit gene is presented in tabular form.

Here are the links that will take you to the Knowledge Base for each of the SUM breast cancer cell lines.

- SUM-44
- SUM-52
- SUM-102
- SUM-149
- SUM-159
- SUM-185
- SUM-190
- SUM-225
- SUM-229
- SUM-1315

Back to SLKBase Home

HERE IS WHAT YOU'LL FIND IN THE KNOWLEDGE BASE

CATEGORIES

Select Category

SEARCH

Search ... SEARCH

The SUM Breast Cancer Cell Line Knowledge Base (SLKBase)

Using functional genomics to navigate critical pathways for growth and survival of breast cancer cells

SUM-44 Home

Okay, let's delve more deeply into the SUM-44 cell line.

1. The development of the SUM-44PE cell line, with patient characteristics
2. Bibliography of published papers in which SUM-44 cells were used
3. Narrative summary of the SUM-44PE cells.
4. Oncogenes and candidate oncogenes in SUM-44PE cells
5. The KEGG canonical pathways enriched with data from our genome-scale shRNA screen
6. Comment on the SUM-44 cell line on the blog page

Back to the SUM cell line Gateway Page

Back to SUM-line Knowledge Base Home page

Edit

SLKBase!

About Steve Ethier

SLKBase functional genomics strategy for the SUM Breast cancer cell lines

SUM Cell Lines Gateway Page

SUM-line KEGG pathway engine

Blog posts on the SUM breast cancer cell lines

HERE IS WHAT YOU'LL FIND IN THE KNOWLEDGE BASE

Copy number data for candidate oncogenes in each cell line

Point mutation data for candidate oncogenes

Genome-wide expression data

Genome-scale shRNA interference screen data for all cell lines with hit data displayed on KEGG pathways

Narrative summary describing how each cell line was developed

A Bibliography of peer-reviewed papers containing data obtained with each cell line

The KEGG canonical pathways enriched with data from our genome-scale shRNA screen in SUM-44 cells

In the pages below, we present the results of our genome-scale RNAi drop-out screen that we performed in the SUM-44 cells. The genes that were "hits" in the screen are those genes that, when reduced in expression, had a significant negative effect on the growth or survival of these cells. By mapping the hit genes to the KEGG pathways, it is possible to directly observe the pathways and biological processes that play functional roles in the biology of these breast cancer cells.

Click on the KEGG pathways below to see the genes that hit in the shRNA screen that map to these pathways.

Cell Cycle Pathway

PI3'K/AKT Pathway

WNT Pathway

Hippo Pathway

Apoptosis

MAPK Signaling

ECM Receptor interaction

Cytokine-Cytokine Receptor interaction

For analysis of the screen hit data for any KEGG pathway, go to the KEGG Pathway Engine

Back to SUM-44 Home

SLKBase!

About Steve Ethier

SLKBase functional genomics strategy for the SUM Breast cancer cell lines

SUM Cell Lines Gateway Page

SUM-line KEGG pathway engine

Blog posts on the SUM breast cancer cell lines

HERE IS WHAT YOU'LL FIND IN THE KNOWLEDGE BASE

Copy number data for candidate oncogenes in each cell line

Point mutation data for candidate oncogenes

Genome-wide expression data

Genome-scale shRNA interference screen data for all cell lines with hit data displayed on KEGG pathways

Narrative summary describing how each cell line was developed

A Bibliography of peer-reviewed papers containing data obtained with each cell line

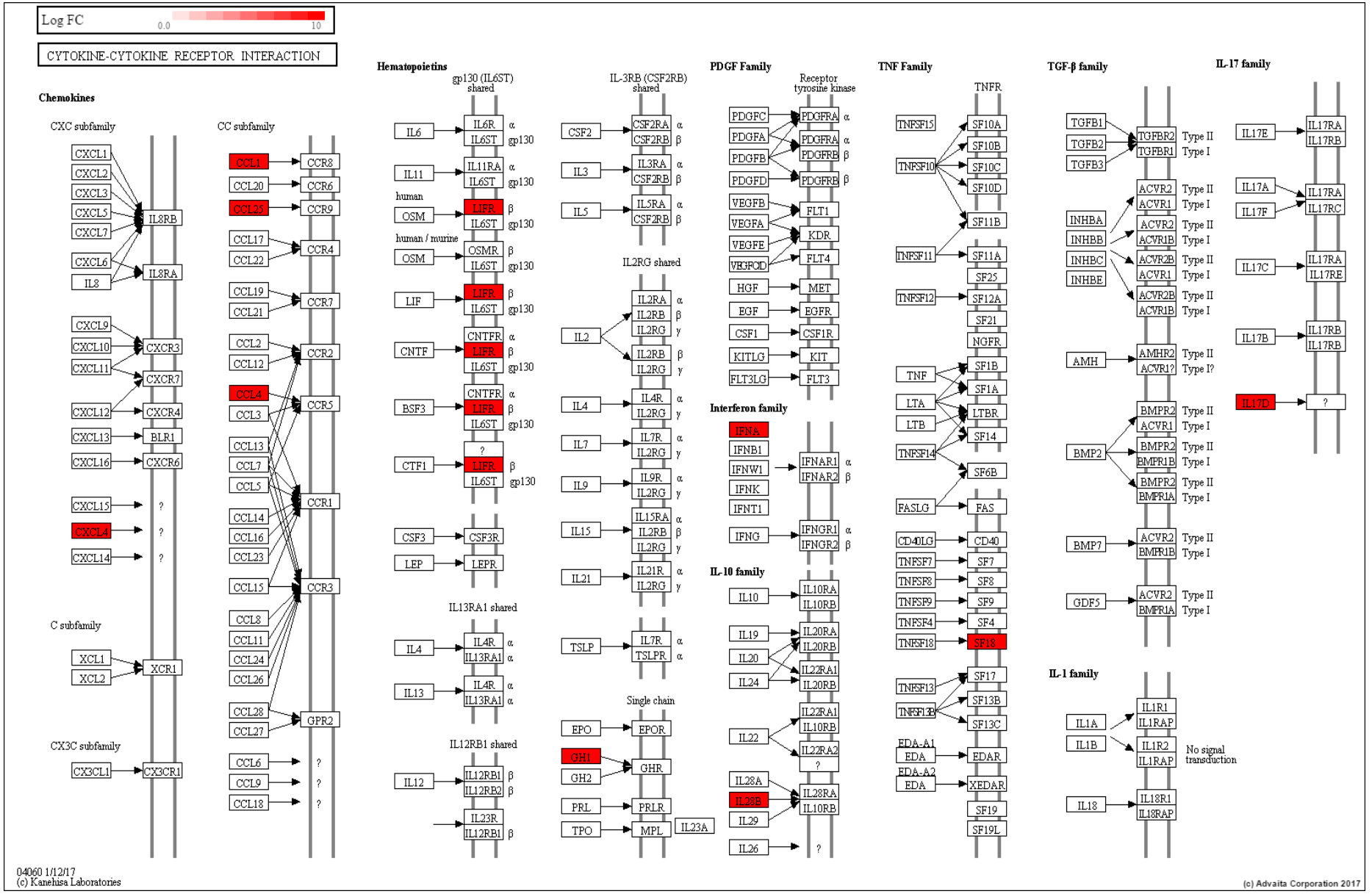
CATEGORIES

Select Category

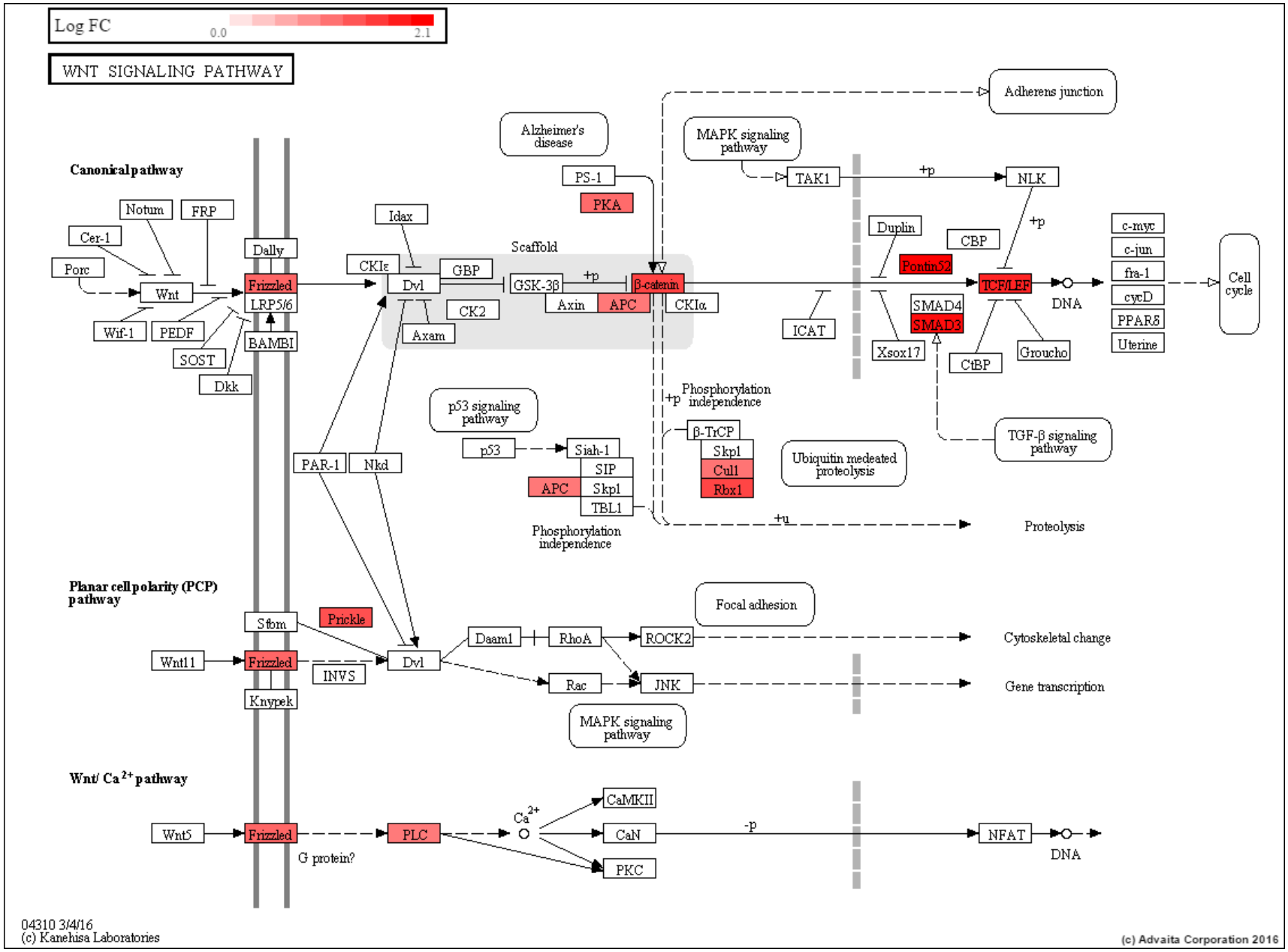
SEARCH

Search ... SEARCH

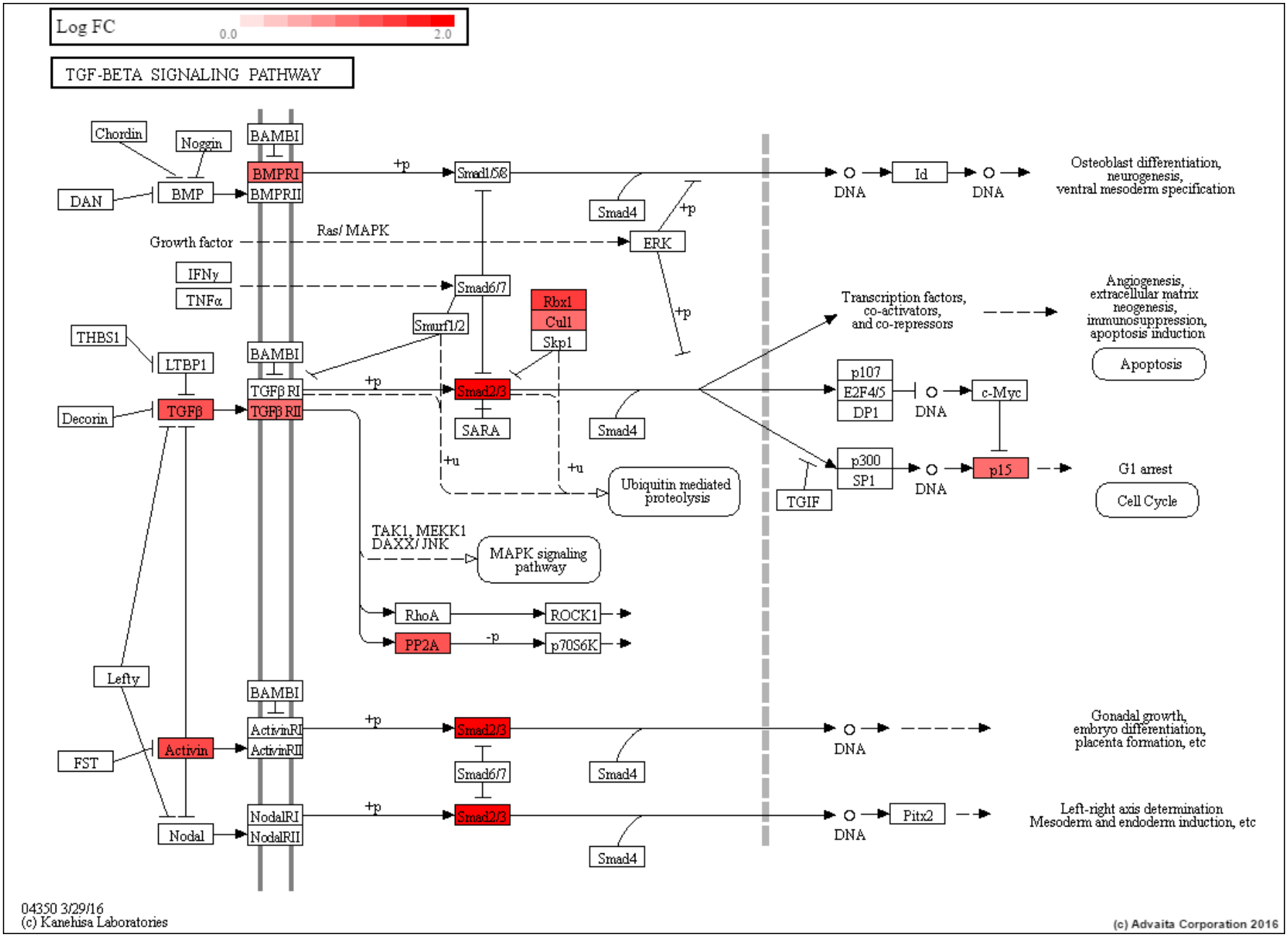
SUM-44 screen hits in the Cytokine-cytokine receptor interactions pathway



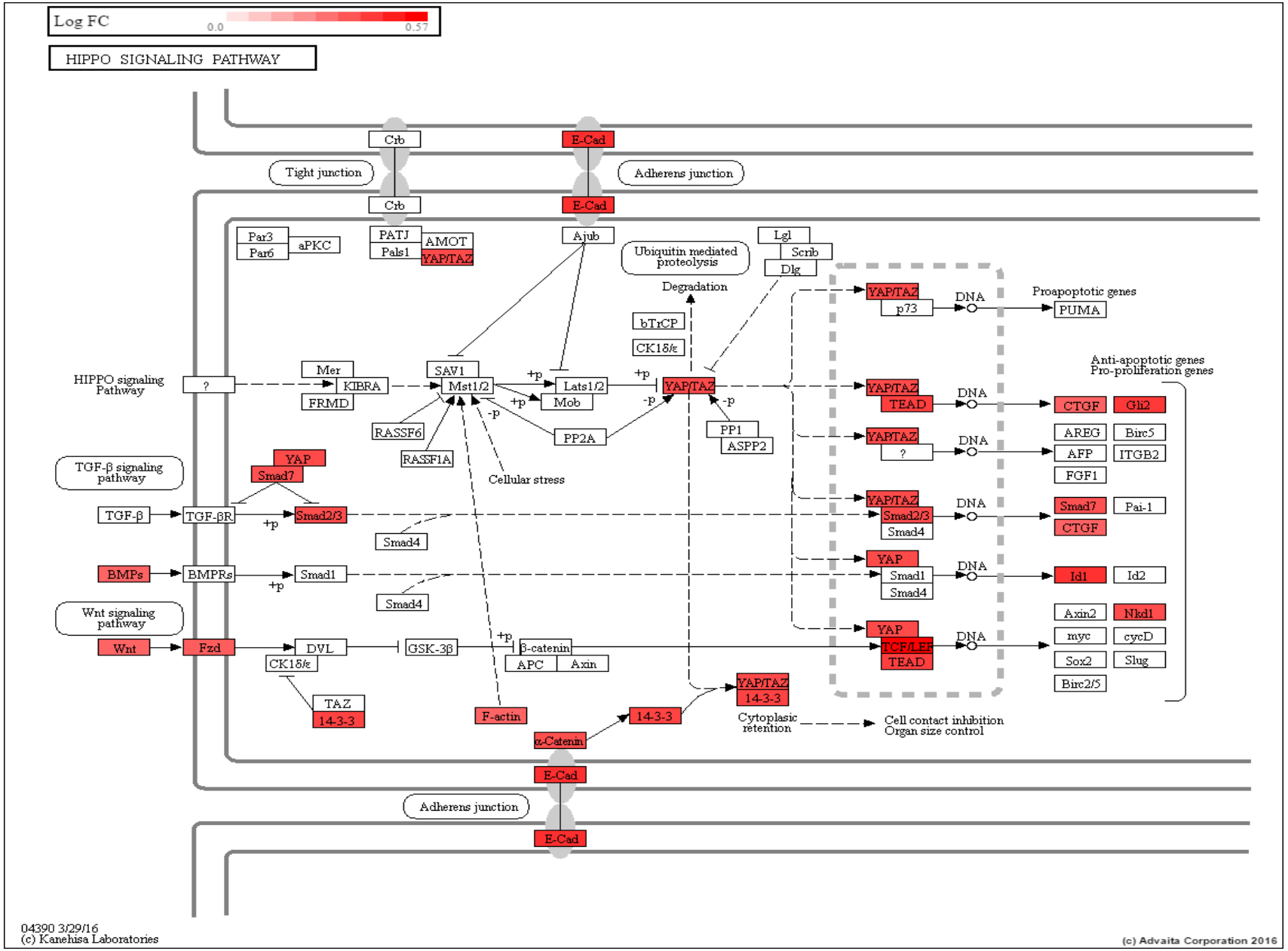
SUM-229 screen hits in the Wnt pathway



SUM-229 screen hits that map to the TGF-beta pathway



SUM-185 screen hits that map to the Hippo pathway



KEGG pathway files in this part of the KB are flat files that are not linked to the data, so….

The SUM Breast Cancer Cell Line Knowledge Base (SLKBase)

Using functional genomics to navigate critical pathways for growth and survival of breast cancer cells

KEGG pathways enriched with screen hit data for SUM-185

The KEGG canonical pathways enriched with data from our genome-scale shRNA screen.

- Cell Cycle
- PI3'Kinase/AKT pathway
- MAPKinase pathway
- Hippo signaling pathway
- TGF-beta pathway
- WNT signaling pathway
- Apoptosis pathway
- FoxO Signaling pathway
- Focal adhesion pathway

For analysis of the screen hit data for **any** KEGG pathway, go to the KEGG Pathway Engine

Back to SUM-185 Home

SLKBase!

About Steve Ethier

SLKBase functional genomics strategy for the SUM Breast cancer cell lines

SUM Cell Lines Gateway Page

SUM-line KEGG pathway engine

Blog posts on the SUM breast cancer cell lines

HERE IS WHAT YOU'LL FIND IN THE KNOWLEDGE BASE

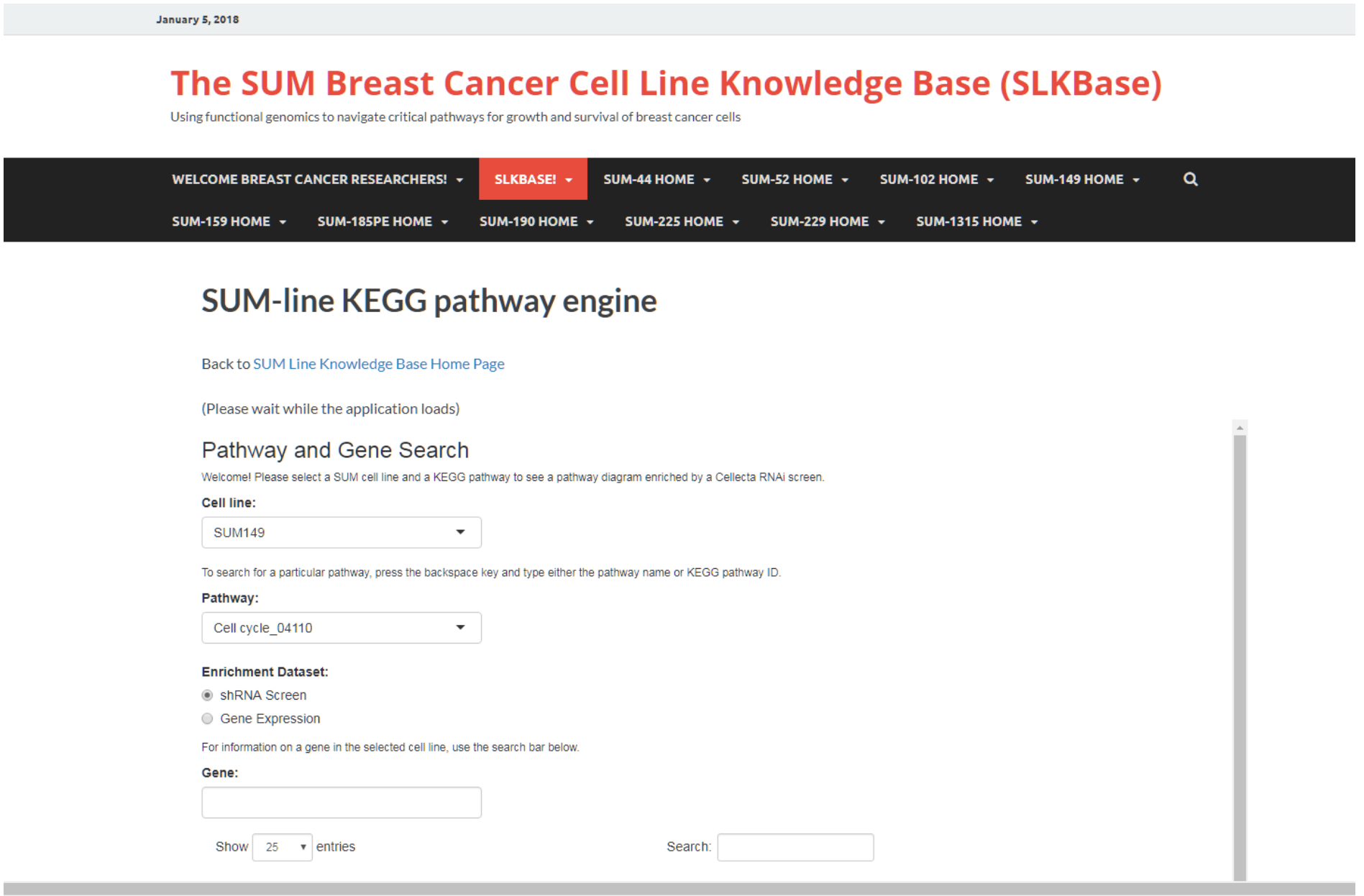
Copy number data for candidate oncogenes in each cell line

Point mutation data for candidate oncogenes

Genome-wide expression data

Genome-scale shRNA interference screen

…we generated a MySQL data base containing all of the omics data sets for all of the cell lines, and built a KEGG pathway engine



Summary and Conclusions

- Breast cancer cell lines remain a mainstay of breast cancer research
- Usefulness of cell lines is limited by the knowledge that most investigators have of the lines they work with
- The SUM Line Knowledge Base is designed to provide users access to everything that is known about these cells, and to provide a space for communication among other workers using the same cells
- This resource has the potential to transform the power of research carried out with breast cancer cell lines