Statistical Practice in High-Throughput siRNA Screens Identifying Genes Mediating Sensitivity to Chemotherapeutic Drugs

Fei Ye, PhD
Joshua Bauer, PhD

Division of Cancer Biostatistics, Cancer Biostatistics Center
Department of Biochemistry
Vanderbilt-Ingram Cancer Center

Cancer Biostatistics Workshop
November 20th, 2009

Background

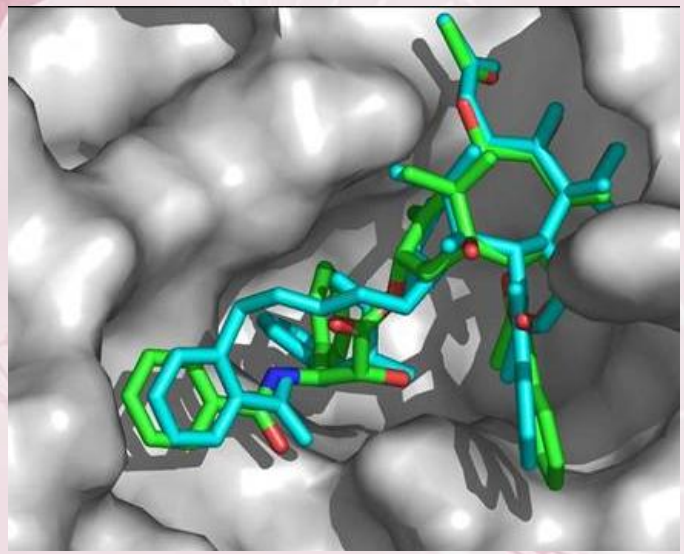
- Current treatment of locally advanced breast cancer consists of taxane-based neoadjuvant (presurgery) treatment (BRE9936 & BRE0368)
- Improves clinical outcome & overall survival
  - A small percentage of patients respond (5-40%)
- Clinically defined responders (ie, pCR) have good overall prognosis
- Important to find markers to target that would increase response of breast cancer patients to taxane-based therapy
- Drug combination with taxanes to improve treatment response and patient outcome

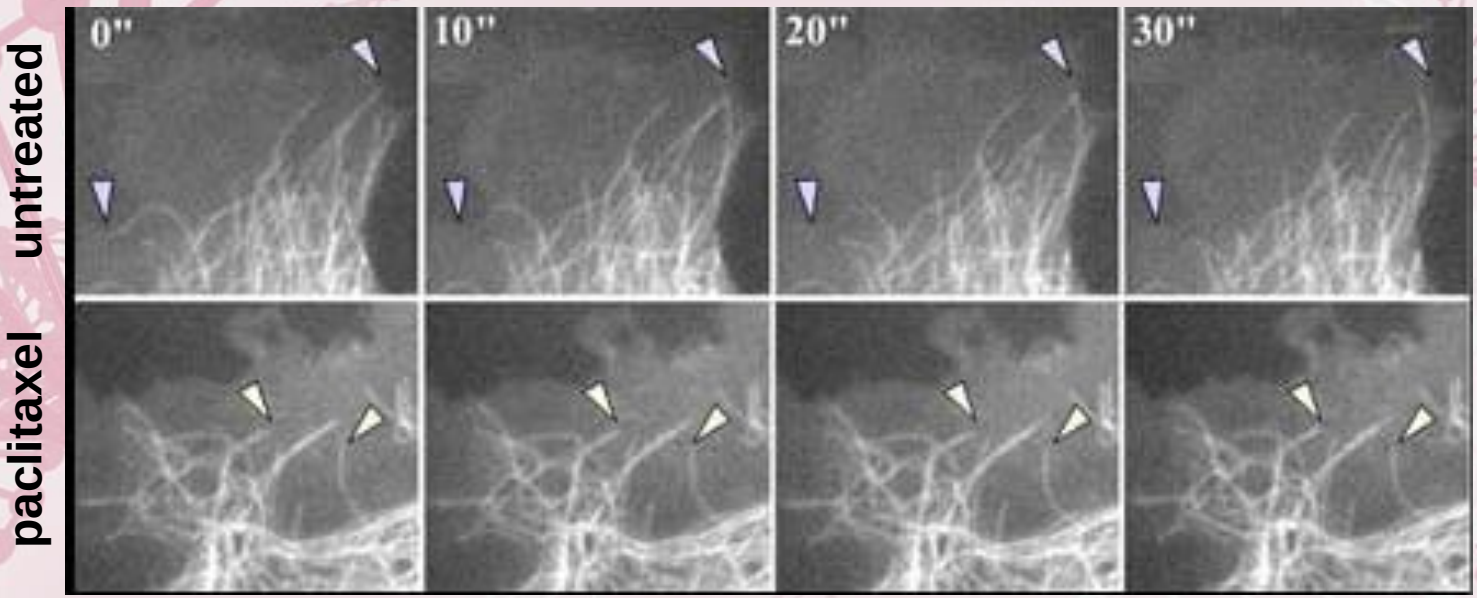
Taxane Mechanism of Action

- binds to microtubules and stabilizes their structure by shifting the dynamic equilibrium (increases polymerization and stabilizes depolymerization)

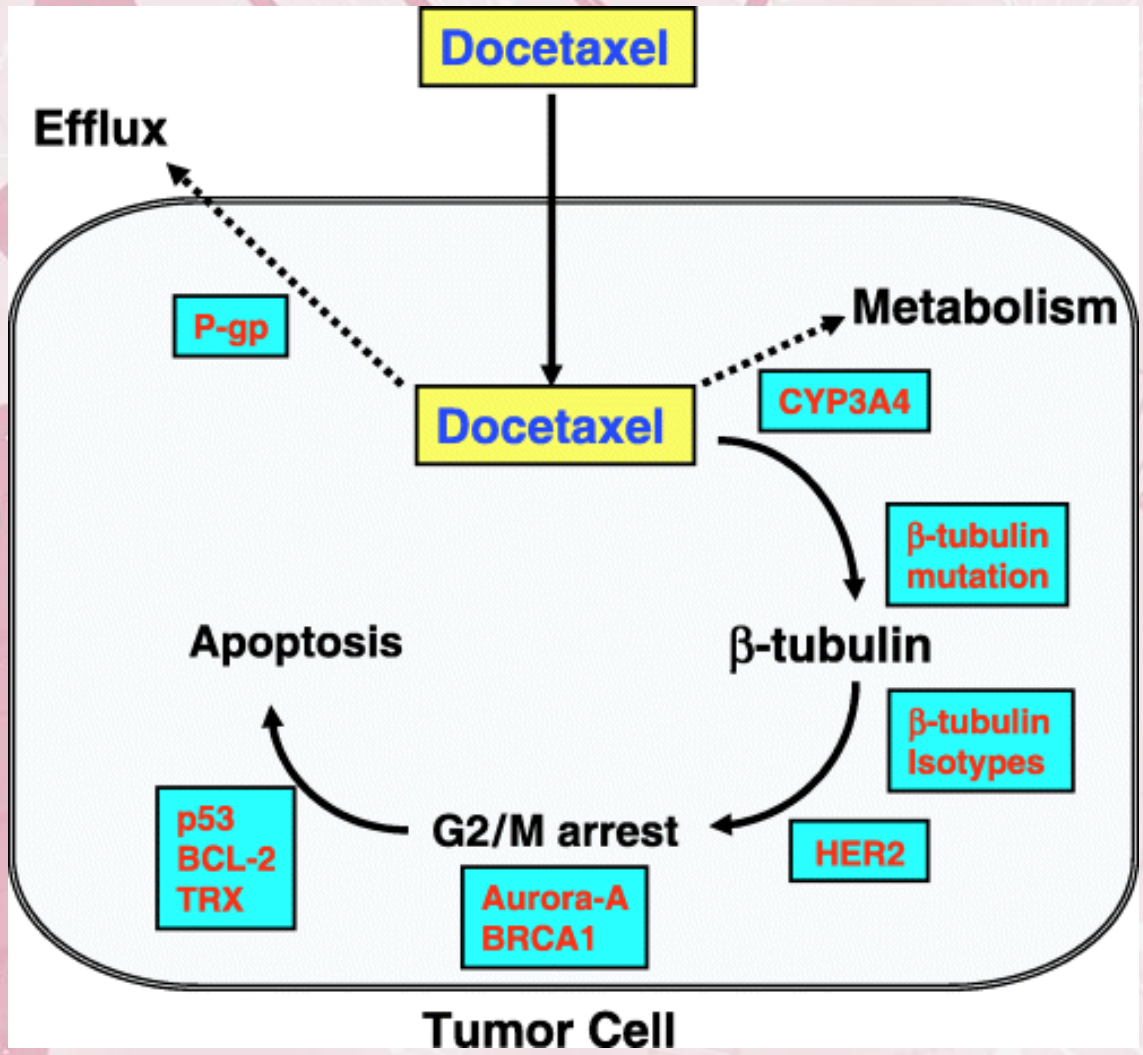
- enhancing microtubule assembly, resulting in inhibition of cellular replication

- inhibits cell processes that are dependent on microtubule turnover, such as mitosis, migration, endocytosis and secretion



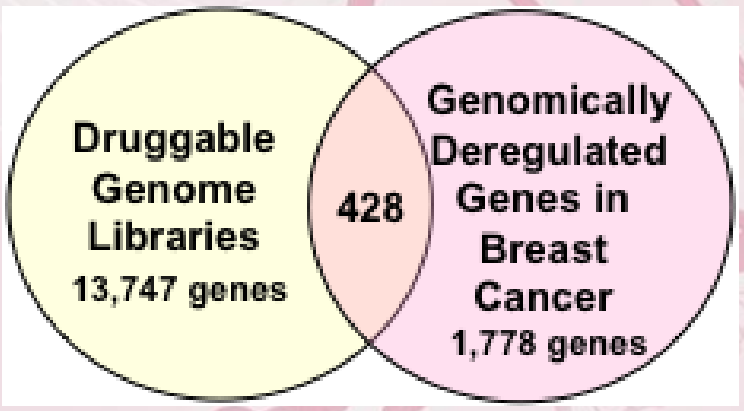


Known mechanisms of taxane resistance



Cancer Sci. 2006 Sep;97(9):813-20.

shRNA Screen to Identify Druggable Gene Targets of Taxol Sensitivity in Breast Cancer

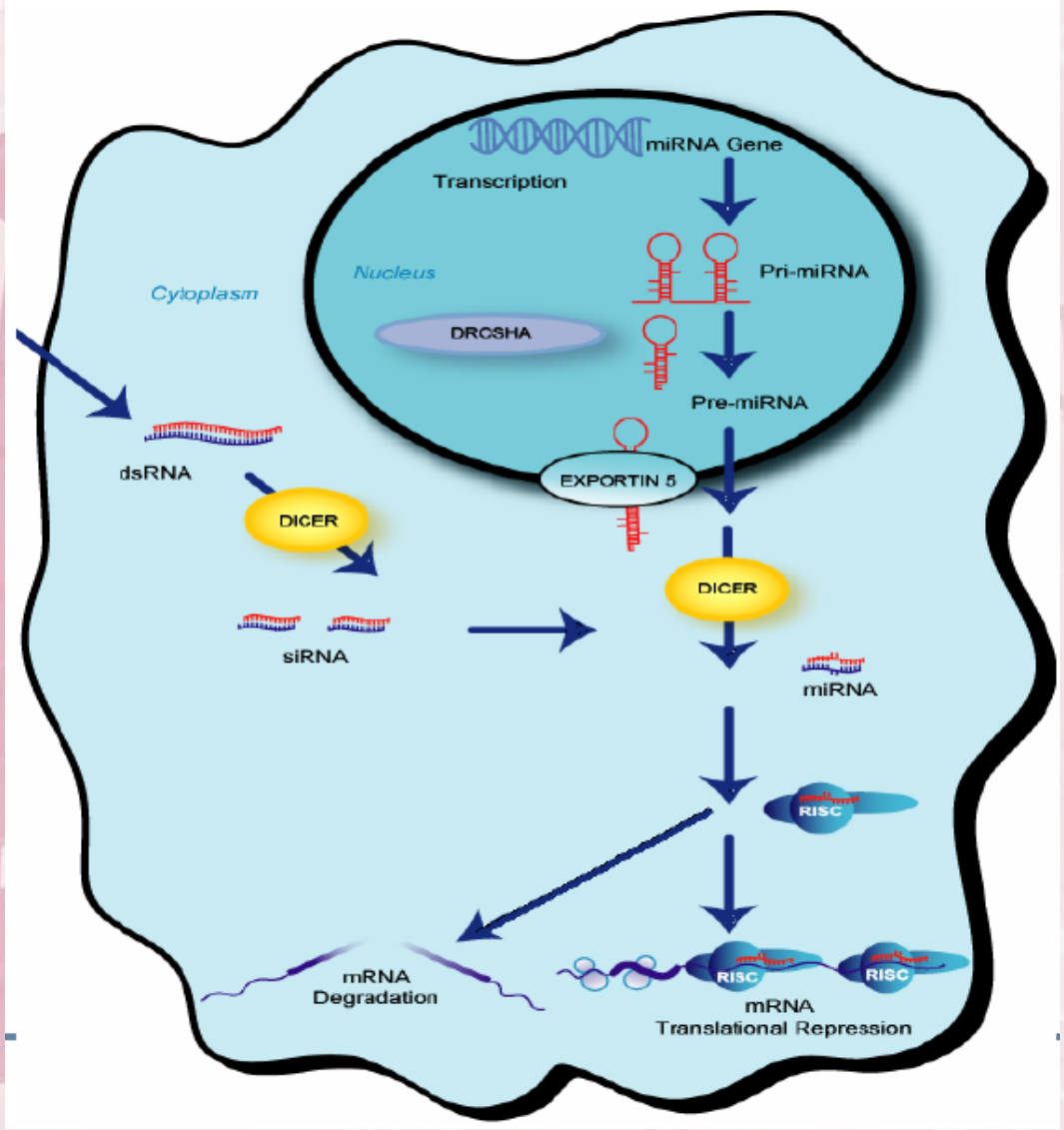


Neve RN et al Cancer Cell

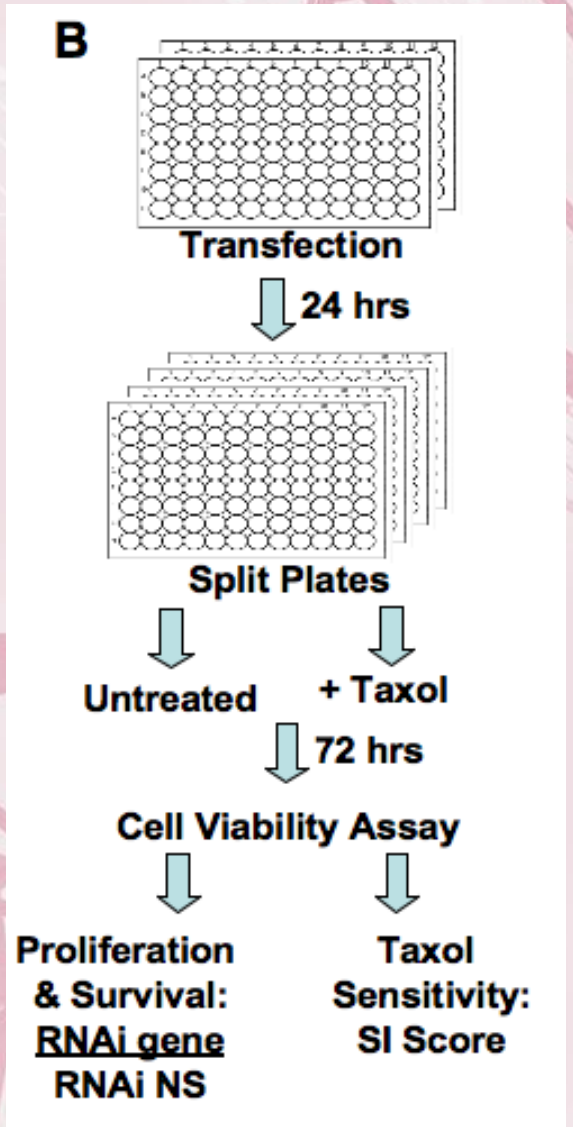
428 druggable deregulated in breast cancer genes

1078 pGIPZ shRNAmir

What is RNA interference (RNAi)?



shRNA Screen to Identify Druggable Gene Targets of Taxol Sensitivity in Breast Cancer



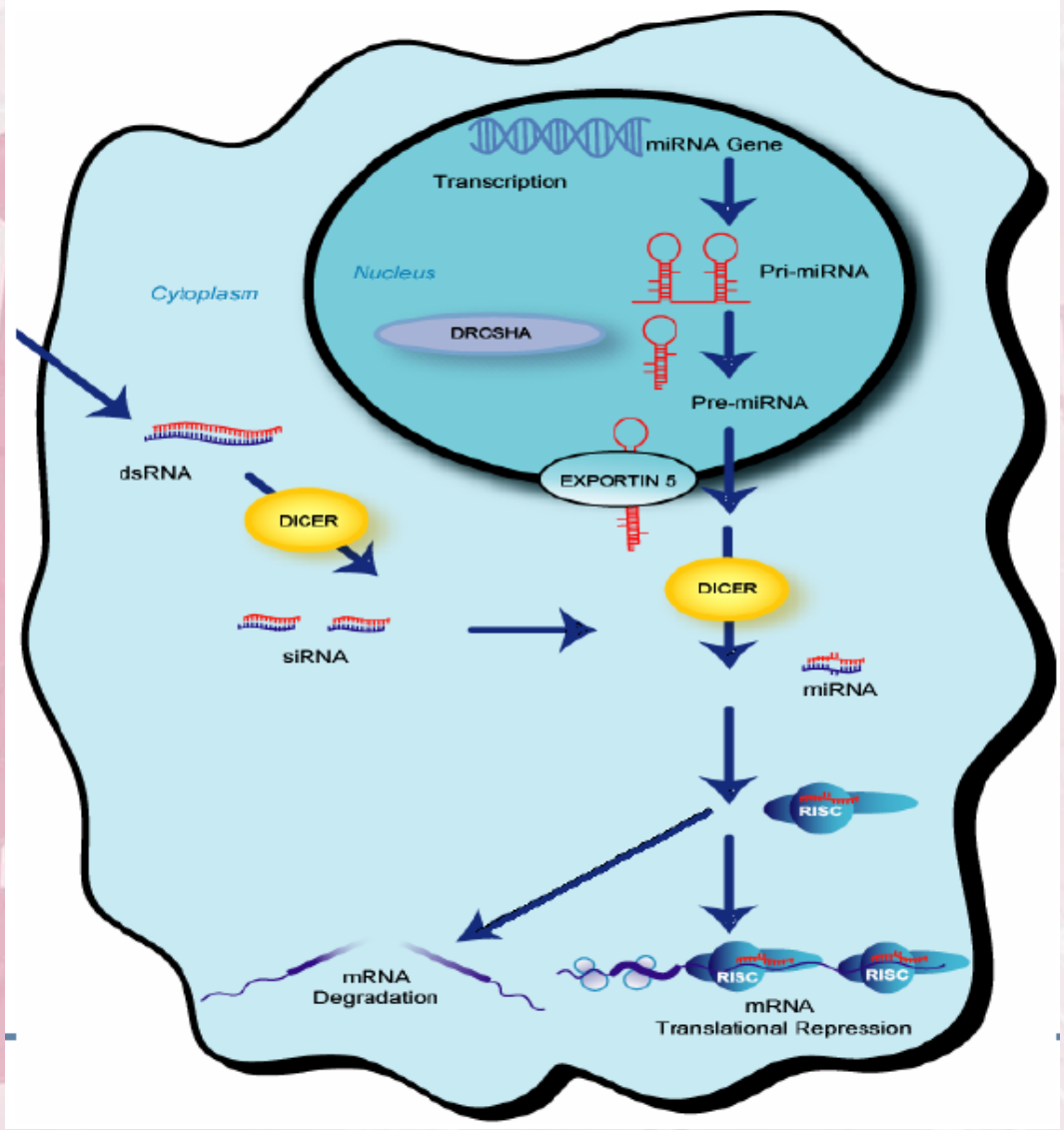
> 3 clones, mean SI > 0.075					< 3 clones, mean SI > 0.150			
Gene	shRNAs	Mean SI	95% CI	Genomic dereg*	Gene	shRNAs	Mean SI	Genomic dereg*
YWHAZ	6	0.193	0.154-0.242	amp/OE	**PCK1**	1	0.461	amp/OE
RPS6KB1	4	0.186	0.135-0.242	amp/OE	**SREBF2**	1	0.391	amp/OE
COG2	5	0.186	0.110-0.265	amp/OE	IL1B	1	0.388	del/UE
PTK2	5	0.184	0.121-0.287	amp/OE	COG6	1	0.325	del/UE
PPM1D	4	0.179	0.120-0.241	amp/OE	**SRC**	1	0.276	amp/OE
PTS	4	0.168	0.042-0.314	del/UE	**FNTA**	1	0.263	amp/OE
SKP1	6	0.166	0.075-0.259	amp/OE	**BCL2L1**	1	0.247	amp/OE
MARK1	5	0.157	0.006-0.345	amp/OE	TDG	1	0.231	del/UE
NFYB	4	0.148	0.085-0.186	amp/OE	**COMMD1**	3	0.220	amp/OE
RBBP4	4	0.139	0.080-0.198	amp/OE	KIF13B	3	0.214	del/UE
IL10	4	0.136	0.081-0.205	amp/OE	**COG8**	1	0.214	amp/OE
EGFR	4	0.136	0.087-0.378	amp/OE	**FRAP1**	1	0.212	NA
FZD4	4	0.134	0.101-0.167	del/UE	RPS6KA2	1	0.205	del/UE
SP1	5	0.130	0.080-0.138	amp/OE	**PCTK1**	1	0.199	amp/OE
STX16	4	0.107	0.053-0.175	amp/OE	EBP	2	0.188	del/UE
ATM	4	0.102	0.039-0.166	del/UE	**ERBB2**	1	0.180	amp/OE
PTPN7	4	0.095	0.019-0.170	amp/OE	**IQGAP1**	3	0.176	amp/OE
SENP1	5	0.095	0.022-0.168	del/UE	GTF2B	3	0.168	del/UE
UBE2N	7	0.093	0.040-0.142	del/UE	DECR2	1	0.165	del/UE
CENPF	4	0.085	0.030-0.140	amp/OE	TOP3A	1	0.164	del/UE
TOPORS	4	0.083	0.056-0.111	del/UE	**CTSC**	2	0.160	amp/OE
FBXW7	9	0.078	0.039-0.123	del/UE	PHB	3	0.159	del/UE
IGF1	4	0.077	0.016-0.137	amp/OE	**NDUFS6**	1	0.159	amp/OE
					COG1	1	0.159	amp/OE
CASP3 (control)	5	-0.042	-0.061--0.021	NA	**PRPF4B**	3	0.156	amp/OE
					FADD	3	0.156	amp/OE
					ERK1	1	0.154	amp/OE
					TGFB1	1	0.153	amp/OE
					TUBG1	3	0.152	amp/OE
					IKBKB	2	0.151	amp/OE

* *as determined by DNA copy number and gene expression analysis by Neve et al.*

amp/OE= amplification/overexpressed; del/UE= deletion/underexpressed

NA= not available

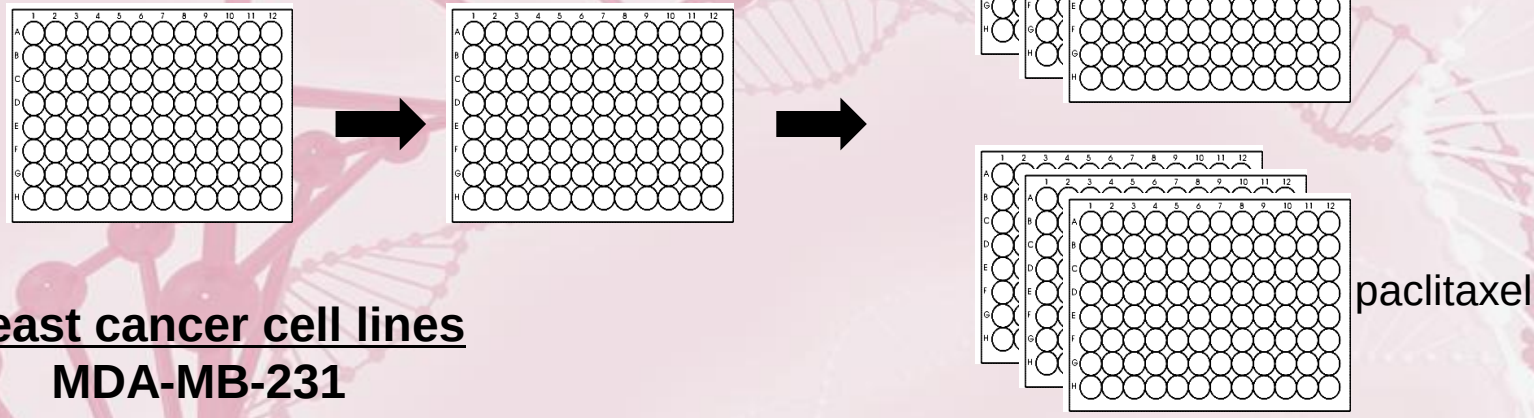
What is RNA interference (RNAi)?



siRNA screen to validate gene targets that sensitize to taxol

transfect

divide



<u>Breast cancer cell lines</u>

MDA-MB-231

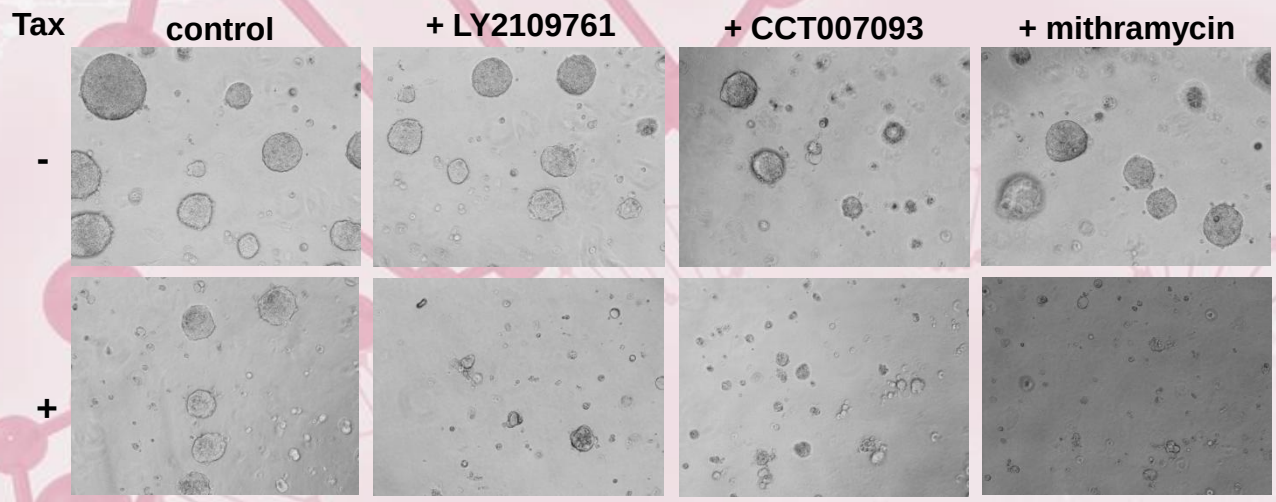
MDA-MB-468

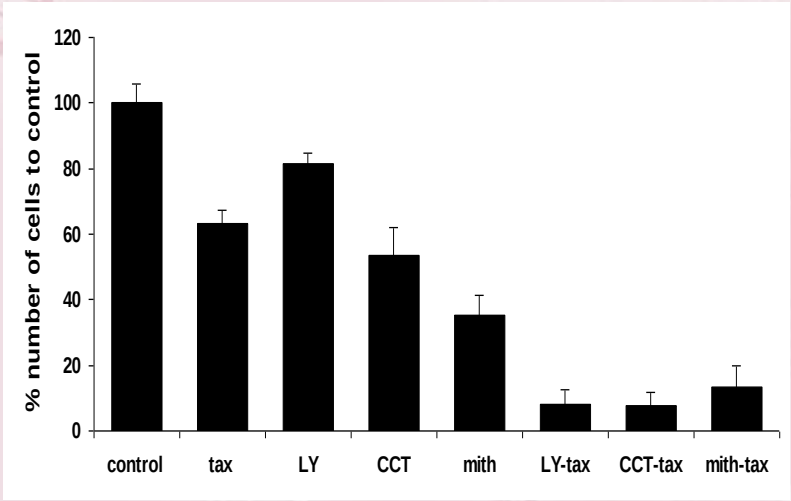
Top gene targets of paclitaxel sensitivity in breast cancer cells

	MDA-MB-231		MDA-MB-468			Previous combination with paclitaxel		
	% viab rel to		% viab rel to					
Gene	NS ctrl	Mean SI	NS ctrl	Mean SI	Drug/chemical inhibitor	Pre-clinical	Clinical	Ref
PPM1D	82.6	0.055	91.8	0.136	thioxanthen-9-one; CCT007093; anti-estrogens*	no	no	
CENPF	96.1	0.049	106.6	0.113	farnesyltransferase inhibitors*	yes	yes	
BCL2L1	88.4	0.041	82.7	0.093	ABT-737; AT-101	yes	yes	
FRAP1	78.2	0.037	82.1	0.078	rapamycin; RAD001	yes	yes	
IGF1	80.9	0.038			NVP-AEW541; 9-cis-retinoic acid*; raloxifene*	no	no	
MARK1	90.0	0.053				-	-	
EGFR			82.7	0.154	erlotinib; gefitinib; cetuximab	yes	yes	
ERK1			98.0	0.148	ERK/MEK inhibitors	yes	yes	
RPS6KB1			105.0	0.140	celecoxib*; staurosporine*	yes	no	
TGFB1			104.0	0.121	LY2109761; LY2157299; SD-208	no	no	
SP1			92.4	0.085	mithramycin; arsenic trioxide*	no	no	

* indirect inhibitors

Inhibitors of gene targets sensitize breast cancer cells to paclitaxel





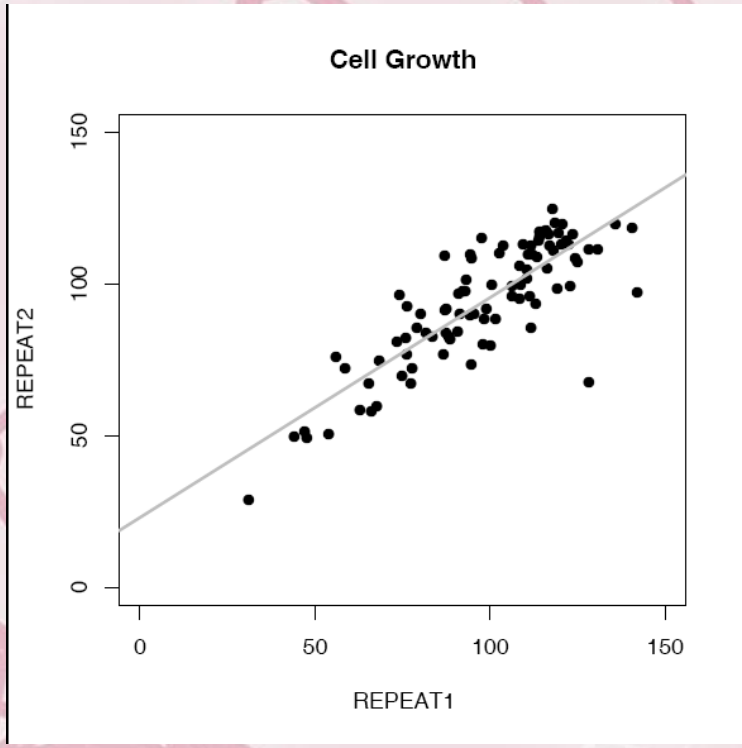
Statistical Considerations

Raw Data

MDA-MB-231		UT#1	UT#2	UT#3	Trt#1	Trt#2	Trt#3
CENPF	A01	5012	3786.092	3968.578	1228.001	1125.549	1051.839
SENP1	A02	1828.84	1923.381	2122.035	1494.562	791.093	765.842
CDKN1A	A03	4665.522	4783.95	4637.939	1692.481	1920.495	1719.794
PTK2	A04	3082.721	2769.689	2877.273	1236.828	1379.308	1258.688
FZD4	A05	3545	4021.945	4015.088	941.099	871.293	765.844
AKAP8	A06	3587.768	5559.518	3725.452	1584.797	1650.895	1594.591
TYRP1	A07	3805.166	3993.343	5342.36	1459.304	1634.373	1446.242
SH3D19	A08	3547.631	3682.168	3736.361	2306.86	1346.532	1356.79
STX7	A09	2820.625	5083.708	3011.929	957.291	961.347	1016.46
ATP2B4	A10	2946.101	2719.373	2685.778	985.01	1349.658	938.433
FRAP1	A11	1878.323	1859.208	1861.325	488.994	424.138	644.317
RAN	A12	3082.513	4587.782	3519.944	1185.351	1118.756	1025.09
CHUK	B01	2355.404	2599.342	2659.815	716.176	733.004	1964.698
FOS	B02	3014.853	3014.901	3458.796	1745.239	990.24	1120.715
RPS6KB1	B03	3636.493	2956.283	3193.836	1278.323	946.904	1167.472
STX7	B04	3133.092	3318.486	3405.048	956.05	732.37	1065.654
CCRL1	B05	3956.498	4058.943	4032.581	1054.981	1035.645	1070.517
MDM2	B06	2937.771	3181.747	3324.665	1143.142	2922.33	1146.291
BCL2L1	B07	3721.851	3891.899	4039.54	1059.517	1001.578	1012.5

Reproducibility?

1. Experimental variation



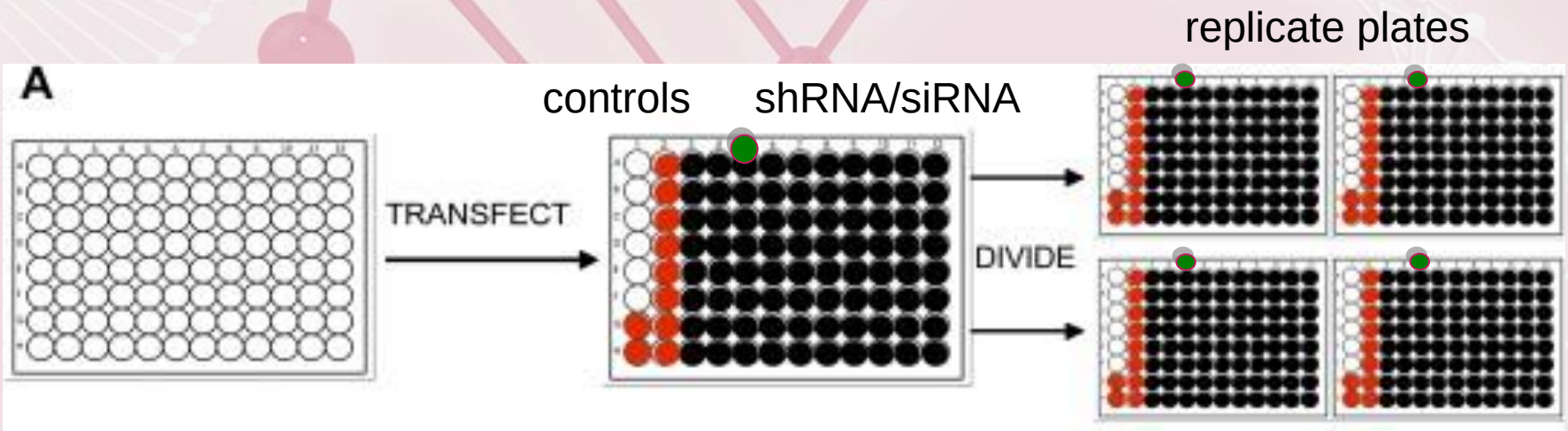
*Non-targeting siRNA control- does not target any mRNA in genome

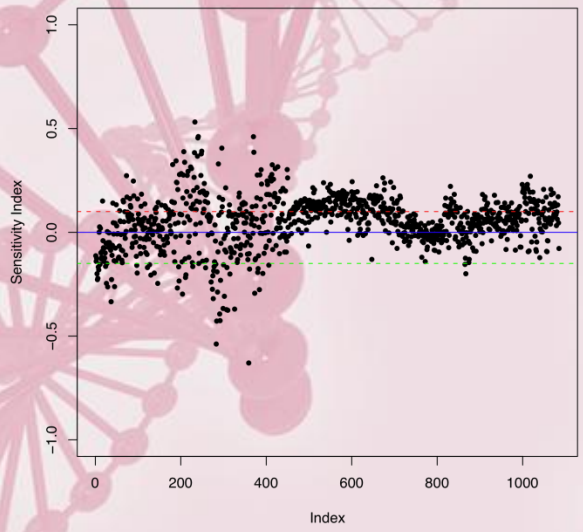
$$\%\text{cell growth} = \frac{\text{siRNA non-targeting}}{\text{siRNA for gene x}}$$

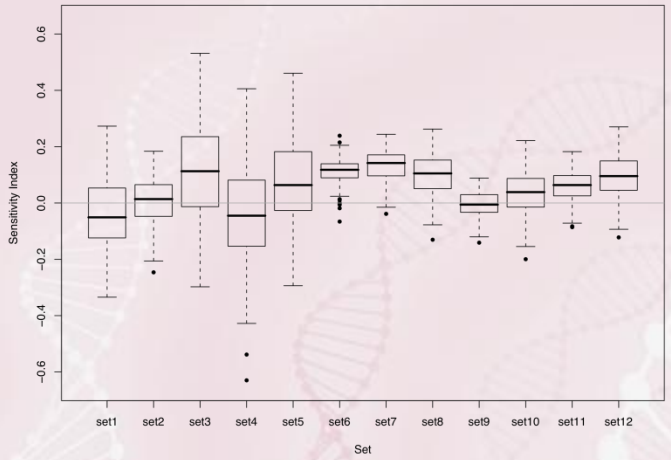
$\rho_{spearman} = 0.79$

Replicability?

2. Plate-plate variation



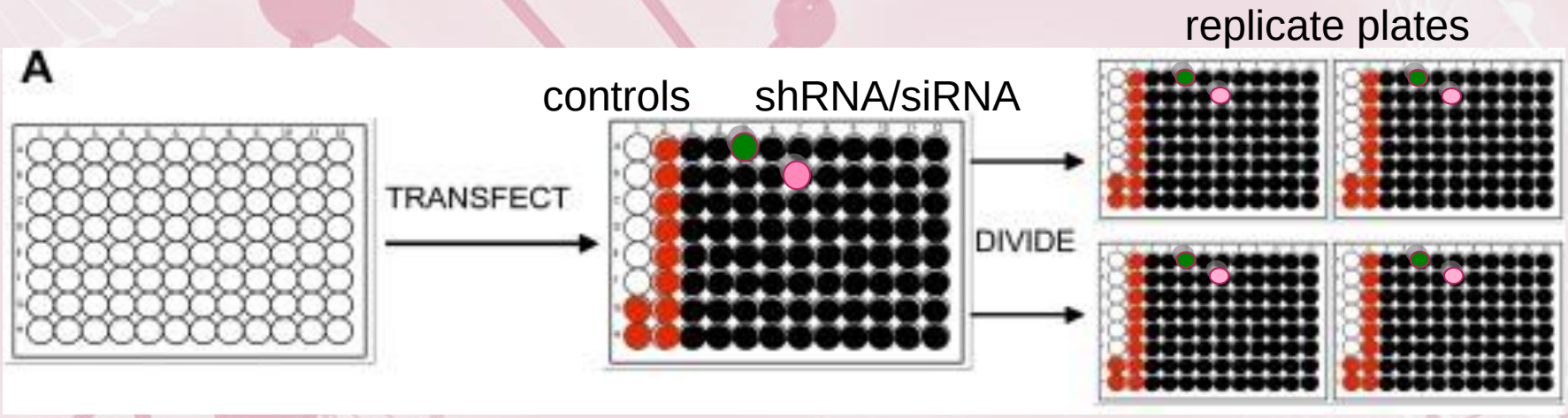




Global normalization?

Robustness?

3. Well-well variation



Within-plate normalization?

siRNA Data Analysis

- Methods used to find the genes that are sensitive/resistant to the drug include
  - **fold-change**
  - ***t* test and its variants**
  - **Wilcoxon rank-sum**
  - **Sensitivity Index (SI)**
  - **…**
- It is unclear whether:
  (1) Drug effect, siRNA effect, and the interaction effect are all considered,
  (2) Variation among replicates is taken into account in the estimation,
  (3) Decision error rates (false-positive and false-negative) are appropriately controlled.

Sensitivity Index

-Swanton et al Cancer Cell 11, 498-512 2007

Effect of the siRNA: Rc/Cc

Effect of the drug: Cd/Cc

Expected Combined Effect (ECE): Rc/Cc X Cd/Cc

Observed Combined Effect: Rd/Cc

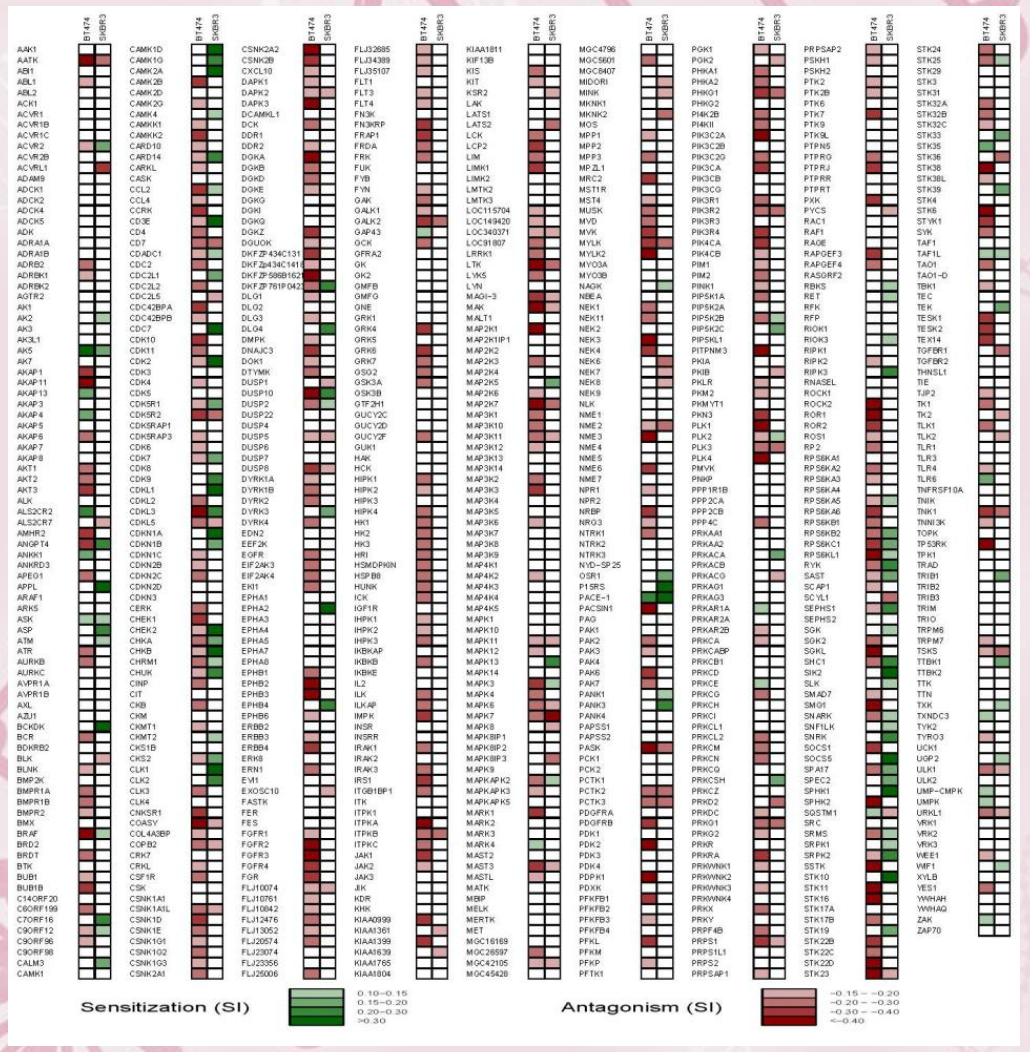
SI= ECE - OCE > 0 = sensitive
< 0 = resistant

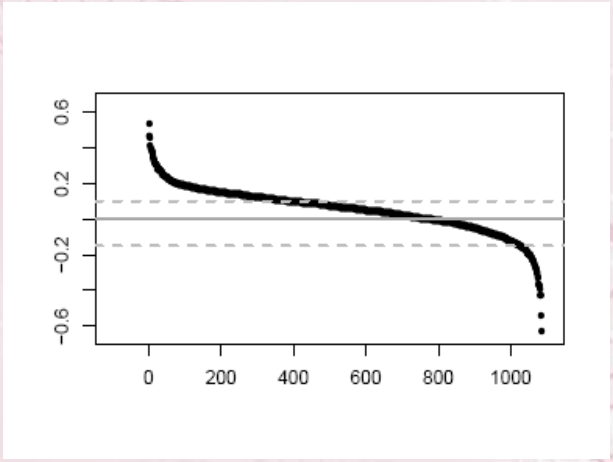
Range: from -1 to 1.

Rc: untreated RNA; Cc: untreated control;

Cd: drugged control; Rd: drugged RNA.

Results Based on SI





Top gene targets of paclitaxel sensitivity in breast cancer cells

	MDA-MB-231		MDA-MB-468			Previous combination with paclitaxel		
	% viab rel to		% viab rel to					
Gene	NS ctrl	Mean SI	NS ctrl	Mean SI	Drug/chemical inhibitor	Pre-clinical	Clinical	Ref
PPM1D	82.6	0.055	91.8	0.136	thioxanthen-9-one; CCT007093; anti-estrogens*	no	no	
CENPF	96.1	0.049	106.6	0.113	farnesyltransferase inhibitors*	yes	yes	
BCL2L1	88.4	0.041	82.7	0.093	ABT-737; AT-101	yes	yes	
FRAP1	78.2	0.037	82.1	0.078	rapamycin; RAD001	yes	yes	
IGF1	80.9	0.038			NVP-AEW541; 9-cis-retinoic acid*; raloxifene*	no	no	
MARK1	90.0	0.053				-	-	
EGFR			82.7	0.154	erlotinib; gefitinib; cetuximab	yes	yes	
ERK1			98.0	0.148	ERK/MEK inhibitors	yes	yes	
RPS6KB1			105.0	0.140	celecoxib*; staurosporine*	yes	no	
TGFB1			104.0	0.121	LY2109761; LY2157299; SD-208	no	no	
SP1			92.4	0.085	mithramycin; arsenic trioxide*	no	no	

* indirect inhibitors

Linear Model

- **Disadvantage of SI: it ignores the variation among replicates.**
- **Assuming normal distribution, a LM of cell viability (Y) for each siRNA can be constructed with the predictor variables treatment effect (x_1, yes/no), siRNA effect (x_2, yes/no), and their interaction term (x_1x_2):**

$$Y = a + b_1*x_1 + b_2*x_2 + b_3*x_1x_2 + err$$

- **Using the goodness-of-fit test, a Chi-square statistic can be computed, based on the difference between the deviance of the null model (without any predictors) D_0 and the residual deviance of the fitted model D_1 with 3 degrees of freedom.**

Simulation Study

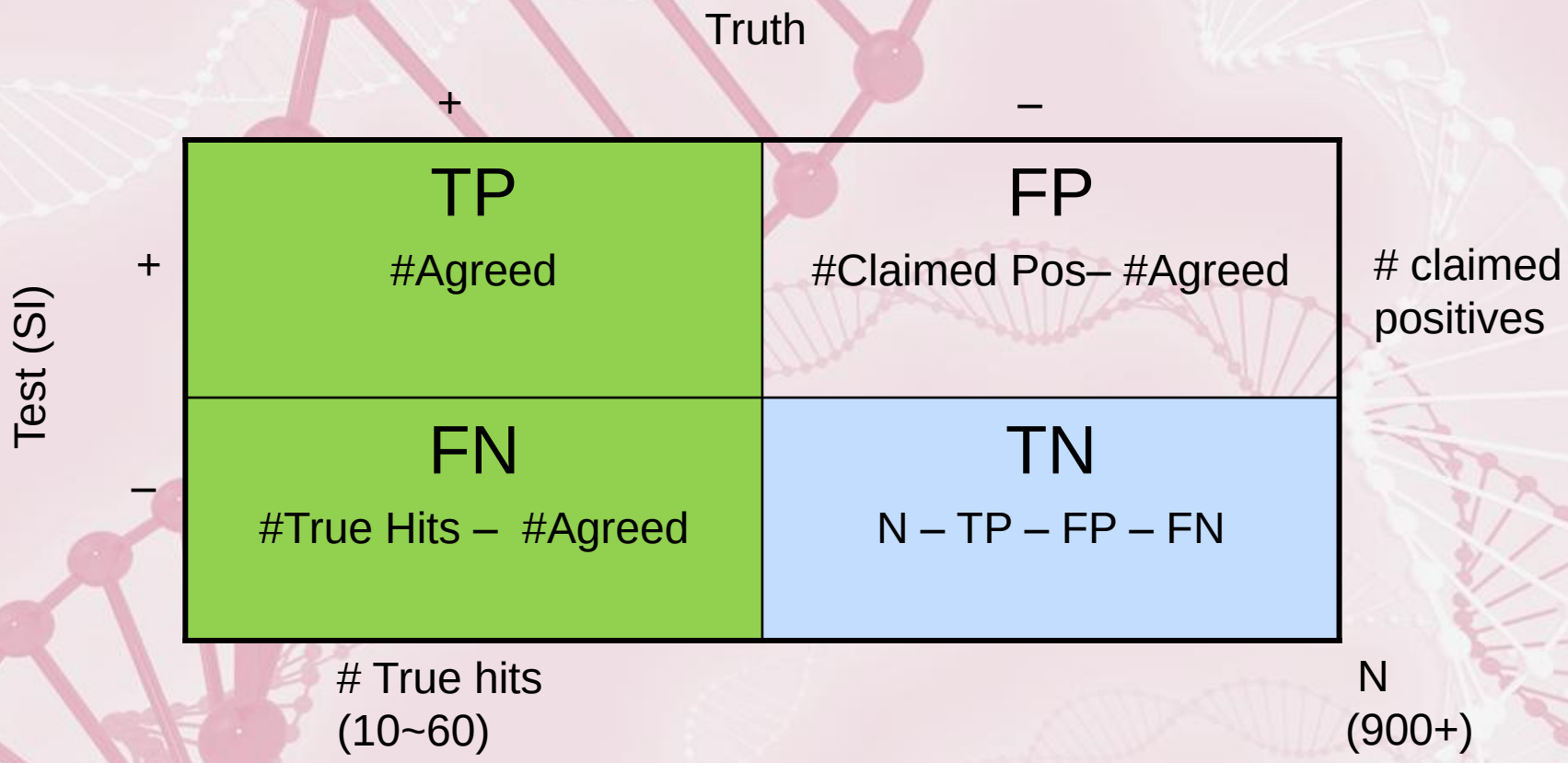
- **Number of true hits**: Uniform{10, 11, …, 60} out of 900 to 930 siRNAs
- The **viability** measurements of the samples transfected with **non-hits**:

  $N(\mu_{NH}, \sigma^2)$, with σ = c (0.2, 0.4, 0.6, 0.8).
- The **distribution of true hits** with a shifted mean relative to the non-hits:

  $N(\mu_{NH^*}C, \sigma^2)$

  C>1 for an ***antagonizing*** effect, C<1 for a ***sensitizing*** effect.
- The parameter D was used to tune the strength of the **drug effect** (or any other type of treatment effect)
- parameter K (>1) is defined such that **scrambled control** wells have a distribution with mean $\mu_{ctl} = \mu_{rna}*K$, where $\mu_{rna} = \mu_{NH} / \mu_{NH^*}D$.

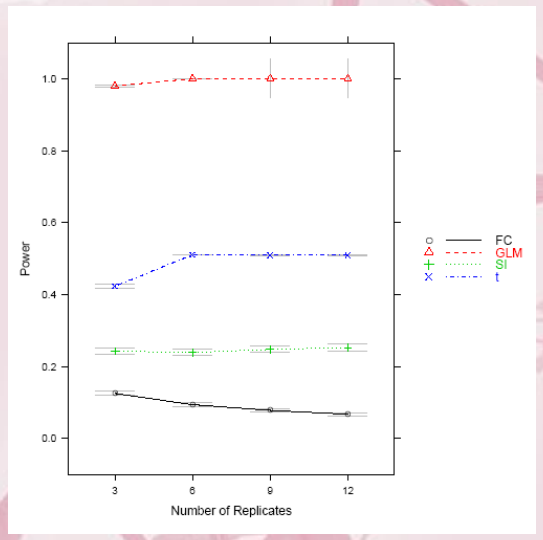
FPR & FNR



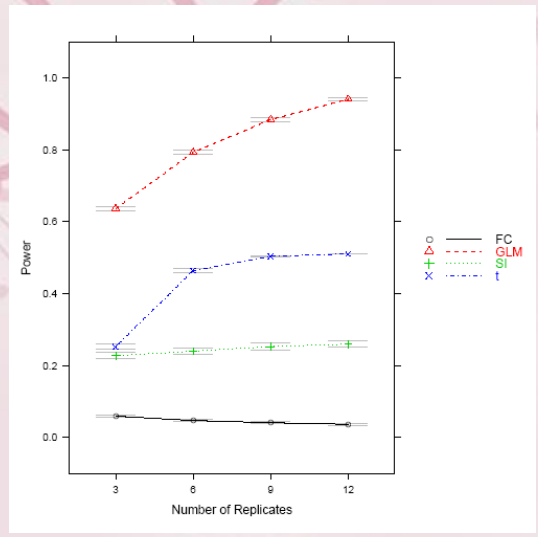
FNR = FN/(**TP**+FN) = 1-sensitivity

FPR = FP/(FP+TN) = 1-specificity

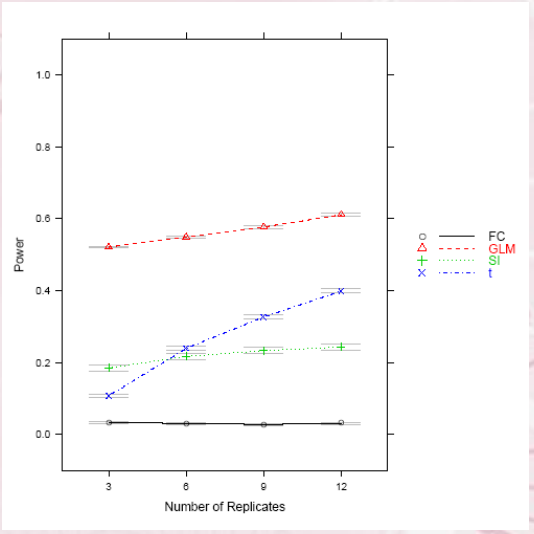
Power Analysis: weak drug effect (low concentration)



low σ, low D, high C

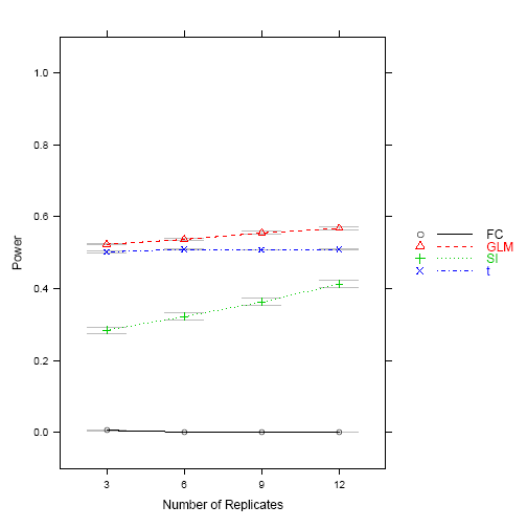


moderate σ, low D, high C

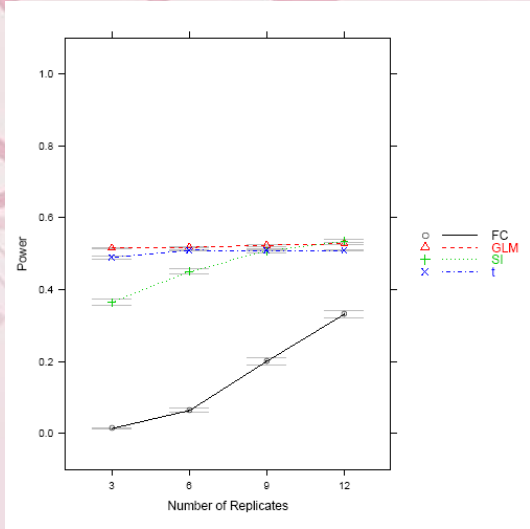


high σ, low D, high C

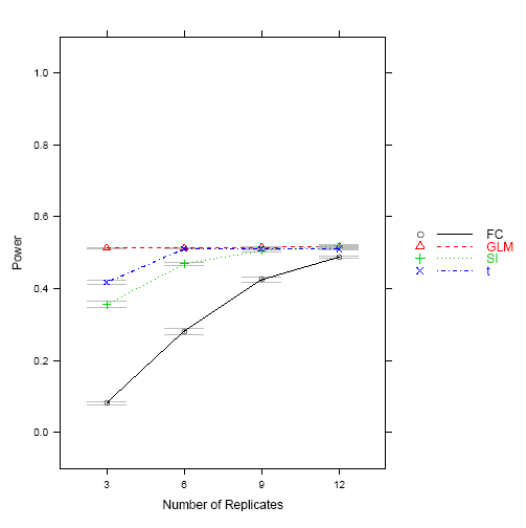
Power Analysis: strong drug effect (high concentration)



low σ, high D, high C



moderate σ, high D, high C



high σ, high D, high C

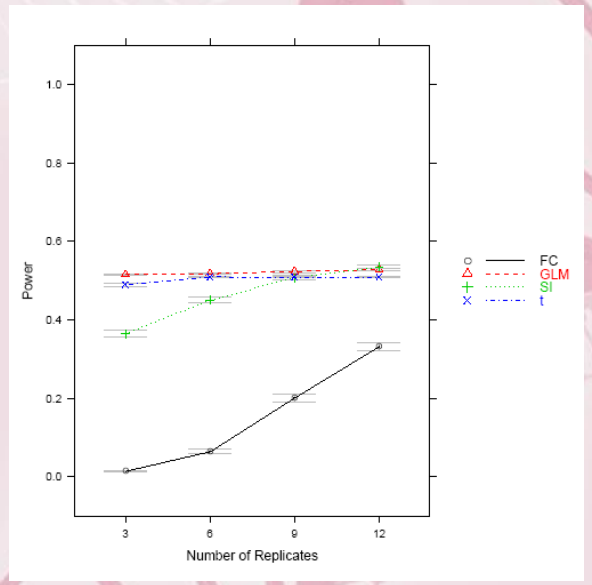
WHY?

In the Case of Skewed Data

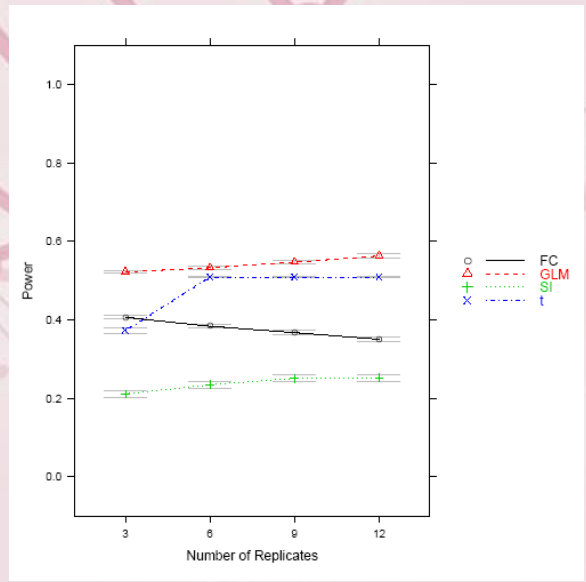
- Gamma distributions *Ga(r,λ)* used instead of Normal.
- The shape (*r*) and scale (*λ*) parameters of gamma distributions were calculated by solving $\mu = r\lambda$ *and* $\sigma^2 = r\lambda^2$.
- The skewness value ($2/\sqrt{r}$) is taken to be (0.5, 1, 1.5, 2).

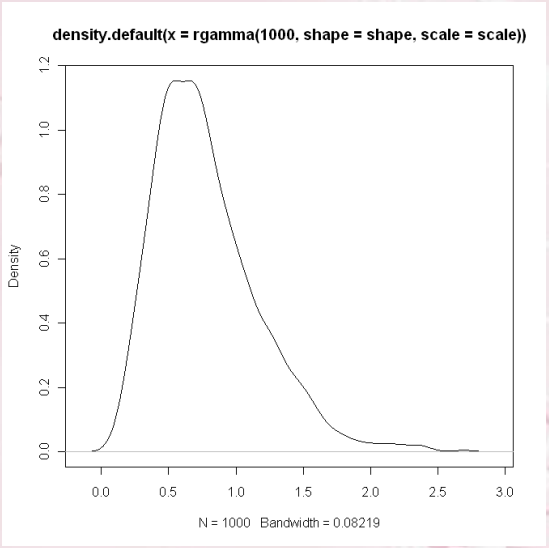
skewness =1

unskewed



skewed

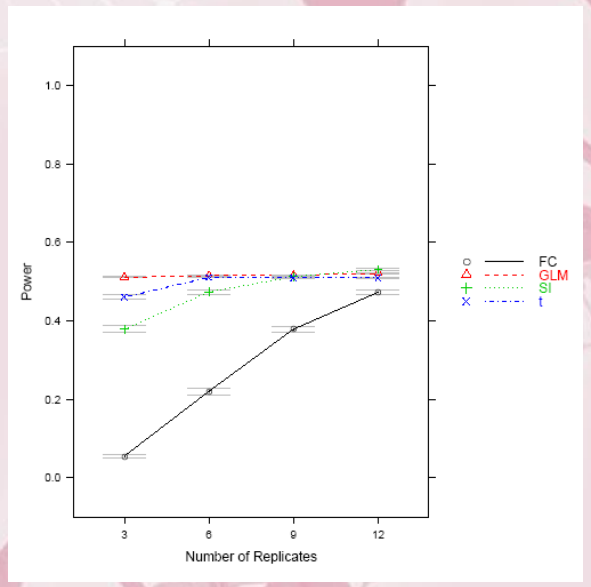




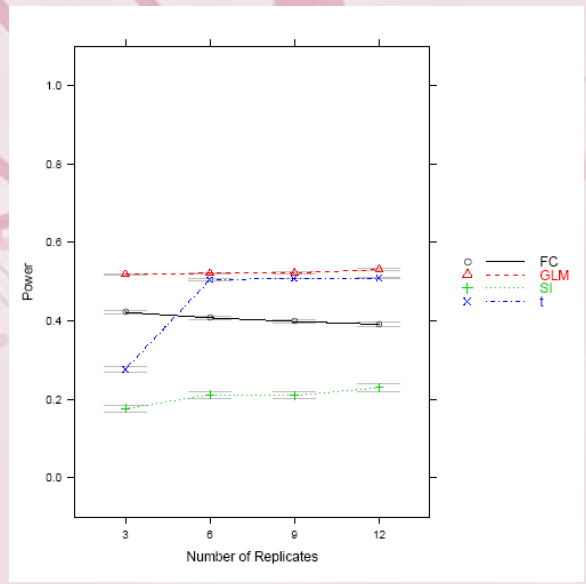
scenario=7; Scale=0.2; shape=4 ; **skewness =1** (for untreated non-hits)

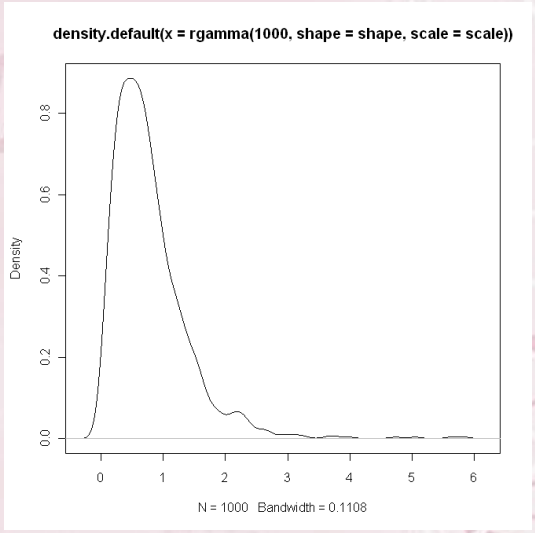
skewness=1.5

unskewed



skewed

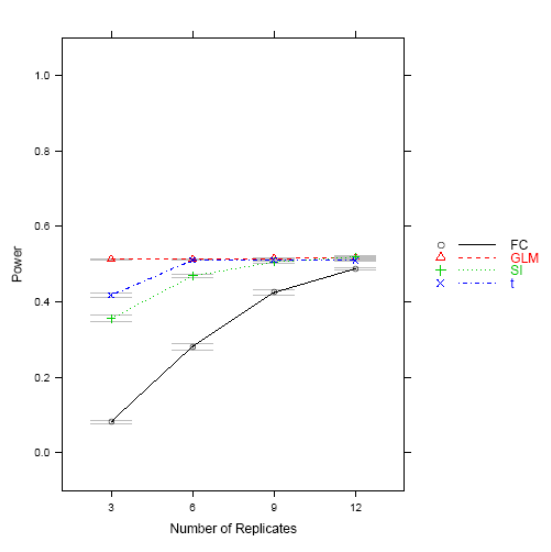




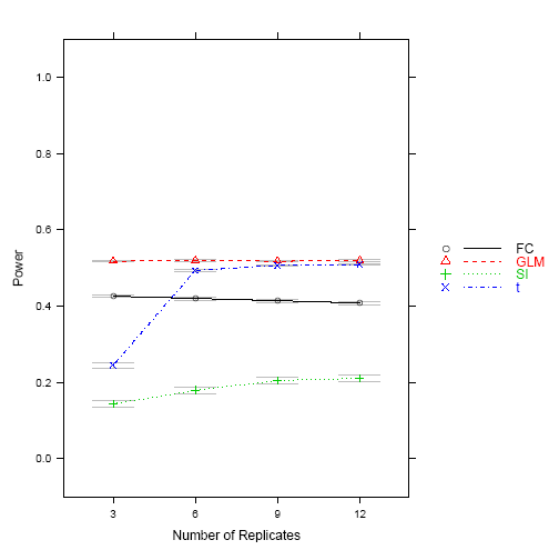
scenario=13; Scale=0.45; shape=1.78; **skewness=1.5** (for untreated non-hits)

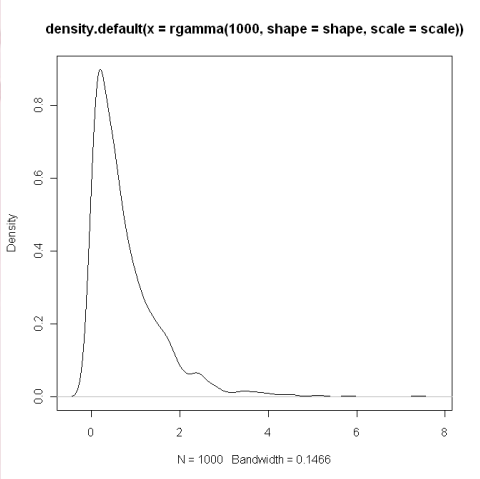
skewness=2

unskewed



skewed





scenario=8; Scale=0.8; shape=1; **skewness=2** (for untreated non-hits)

Summary Points of Skewed Data

- The SI method is affected by the skewness the most.
- The t-test has a lower power with skewed data when there is a small number of replicates.
- LM is quite stable.
- FC: very unstable.
- Do transformation if the data are heavily skewed.

Recommendation of Method

Number of Replicates	Noise[a]	Drug effect[b]	siRNA effect[c]	Recommended method(s)
3	Low	High	High	LM
	Moderate	High	High	LM
	High	High	High	LM
	Low	Moderate	High	LM
	Low	Low	High	LM
	Moderate	Low	High	LM
	High	Low	High	LM
	Low	High	Moderate	LM
	Low	High	Low	LM
6	Low	High	High	LM
	Moderate	High	High	LM
	High	High	High	LM
	Low	Moderate	High	LM
	Low	Low	High	LM
	Moderate	Low	High	LM
	High	Low	High	LM
	Low	High	Moderate	LM
	Low	High	Low	LM
9	Low	High	High	LM
	Moderate	High	High	LM, SI
	High	High	High	LM, SI
	Low	Moderate	High	LM
	Low	Low	High	LM
	Moderate	Low	High	LM
	High	Low	High	LM
	Low	High	Moderate	LM
	Low	High	Low	LM
12	Low	High	High	SI
	Moderate	High	High	SI
	High	High	High	SI
	Low	Moderate	High	LM
	Low	Low	High	LM
	Moderate	Low	High	LM
	High	Low	High	LM
	Low	High	Moderate	LM
	Low	High	Low	LM

[a] Noise can be measured by coefficient of variation (CV) or variance-to-mean ratio (VMR). VMR<0.2: low noise, 0.2≤VMR<0.5: moderate noise; VMR≥0.5 high noise.

[b] Drug effect can be estimated by Cd/Cc.

[c] siRNA effect can be estimated by Rc/Cc.

Increase the power of your study

Possible ways to improve the power of your study:

- Choose a statistical method that is most powerful for your study.
- Reduce the variability.
- Increase #replicates.
- Use lower concentrations of the drug if appropriate.

  "…Here we show that several of these targets sensitize lung cancer cells to paclitaxel concentrations ***1,000-fold lower than otherwise required for a significant response****, and we identify mechanistic relationships …"* (Whitehurst et al., Nature 2007)

Lovejoy CA, et al. “Functional genomic screens identify CINP as a genome maintenance protein” – PNAS. 2009 Nov 4. [Epub ahead of print].

Bauer JA, et al., “RNAi screening approach to identify paclitaxel sensitizers for use in breast cancer therapy – enhancing paclitaxel sensitivity by identifying novel drug combinations through RNAi screening in breast cancer cell lines” –to be submitted

Ye F, et al. “Statistical practice in high-throughput siRNA/shRNA screens identifying genes mediating sensitivity to chemotherapeutic drugs” – to be submitted

Acknowledgements

- Yu Shyr
- Jennifer Pietenpol
- William Wu
- Aixiang Jiang