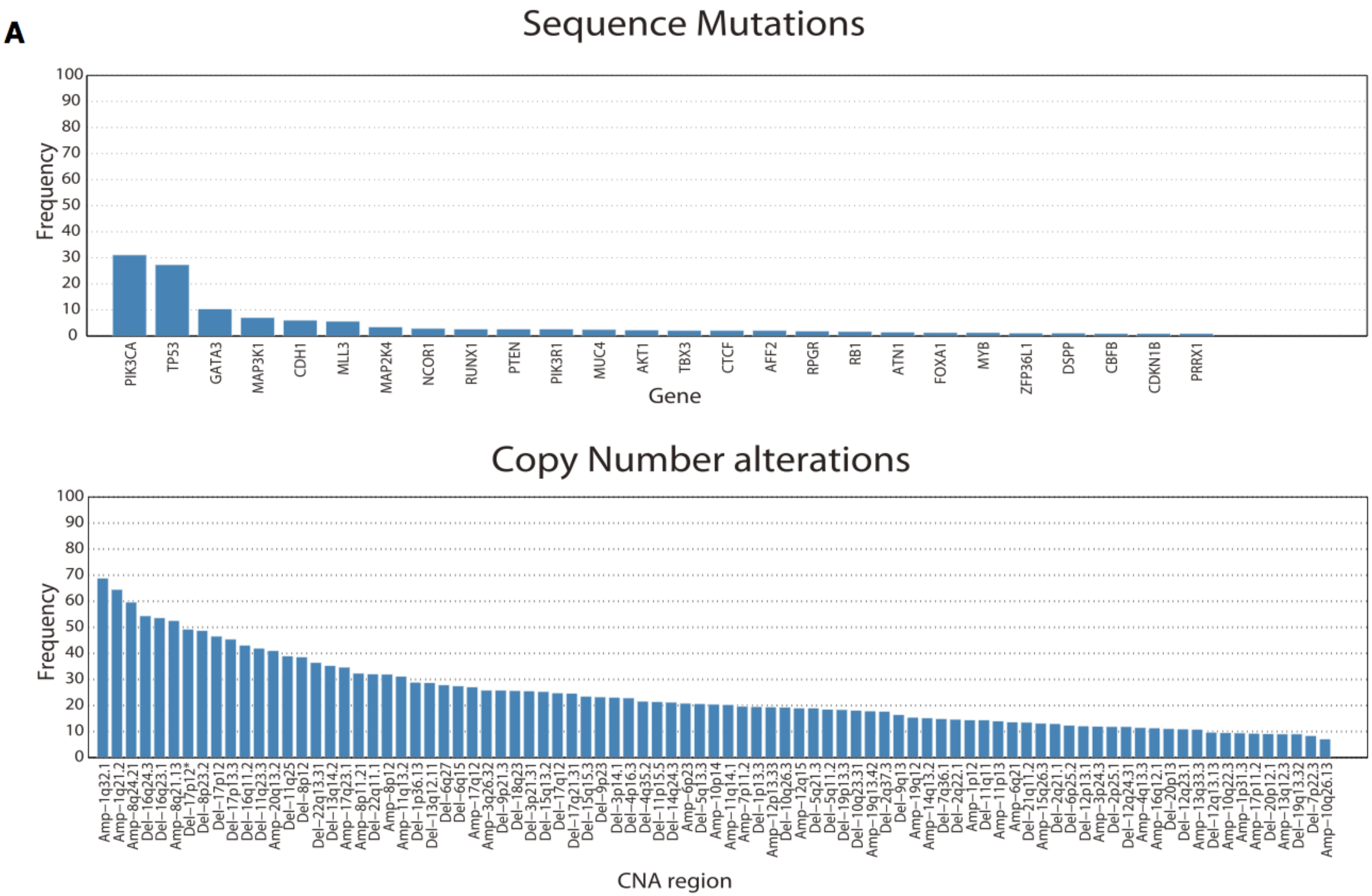
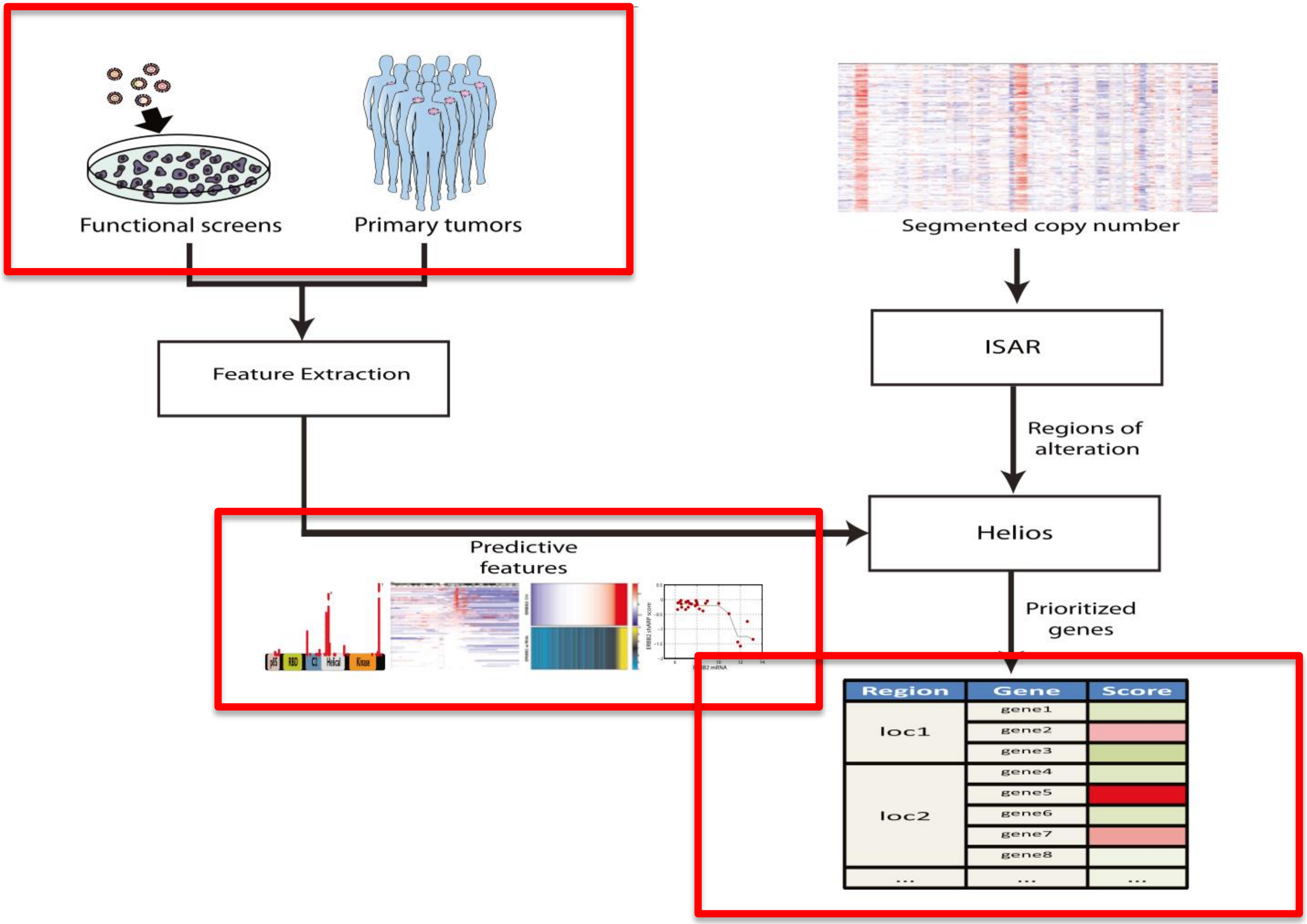
Helios

Motivation


A
Sequence Mutations
Frequency
100
90
80
70
60
50
40
30
20
10
0
PIK3CA
TP53
GATA3
MAP3K1
CDH1
MLL3
MAP2K4
NCOR1
RUNX1
PTEN
PIK3R1
MUC4
AKT1
TBX3
CTCF
AFF2
RPGR
RB1
ATN1
FOXA1
MYB
ZFP36L1
DSPP
CBFB
CDKN1B
PRRX1
Gene
Copy Number alterations
Frequency
100
90
80
70
60
50
40
30
20
10
0
Amp−1q32.1
Amp−1q21.2
Amp−8q24.21
Del−16q24.3
Del−16q23.1
Amp−8q21.13
Del−17p12*
Del−8p23.2
Del−17p12
Del−17p13.3
Del−16q11.2
Del−11q23.3
Amp−20q13.2
Del−11q25
Del−8p12
Del−22q13.31
Del−13q14.2
Amp−17q23.1
Amp−8p11.21
Del−22q11.1
Amp−8p12
Amp−11q13.2
Del−1p36.13
Del−13q12.11
Del−6q27
Del−6q15
Amp−17q12
Amp−3q26.32
Del−9p21.3
Del−18q23
Del−3p21.31
Del−15q13.2
Del−17q12
Del−17q21.31
Del−15q15.3
Del−9p23
Del−3p14.1
Del−4p16.3
Del−4q35.2
Del−11p15.5
Del−14q24.3
Amp−6p23
Del−5q13.3
Amp−10p14
Amp−11q14.1
Amp−7p11.2
Del−1p13.3
Amp−12p13.33
Del−10q26.3
Amp−12q15
Del−5q21.3
Del−5q11.2
Del−19p13.3
Del−10q23.31
Amp−19q13.42
Del−2q37.3
Del−9q13
Amp−19q12
Amp−14q13.2
Del−7q36.1
Del−2q22.1
Amp−1p12
Del−11q11
Amp−11p13
Amp−6q21
Del−21q11.2
Amp−15q26.3
Del−2q21.1
Del−6p25.2
Del−12p13.1
Amp−3p24.3
Del−2p25.1
Del−12q24.31
Amp−4q13.3
Amp−16q12.1
Del−20p13
Del−12q23.1
Amp−13q33.3
Del−12q13.13
Amp−10q22.3
Amp−1p31.3
Amp−17p11.2
Del−20p12.1
Amp−13q12.3
Del−19q13.32
Del−7p22.3
Amp−10q26.13
CNA region

Development of Helios



Functional screens

Primary tumors

Feature Extraction

Predictive features

Segmented copy number

ISAR

Regions of alteration

Helios

Prioritized genes

Region	Gene	Score
loc1	gene1	🟩
	gene2	🟥
	gene3	🟩
loc2	gene4	🟩
	gene5	🟥
	gene6	🟩
	gene7	🟥
	gene8	
...	...	...

ISAR

Gistic Method

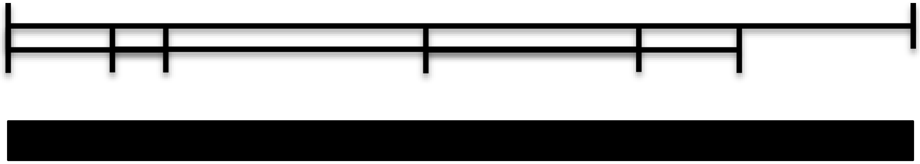
$$G^{AMP}(m) = \sum_{i=1}^{N} CN(m,i) \times I(CN(m,i) > \theta^{AMP})$$

CN(m,i) = copy # of marker m in sample I
I = indicator function
Θ = Aberration threshold

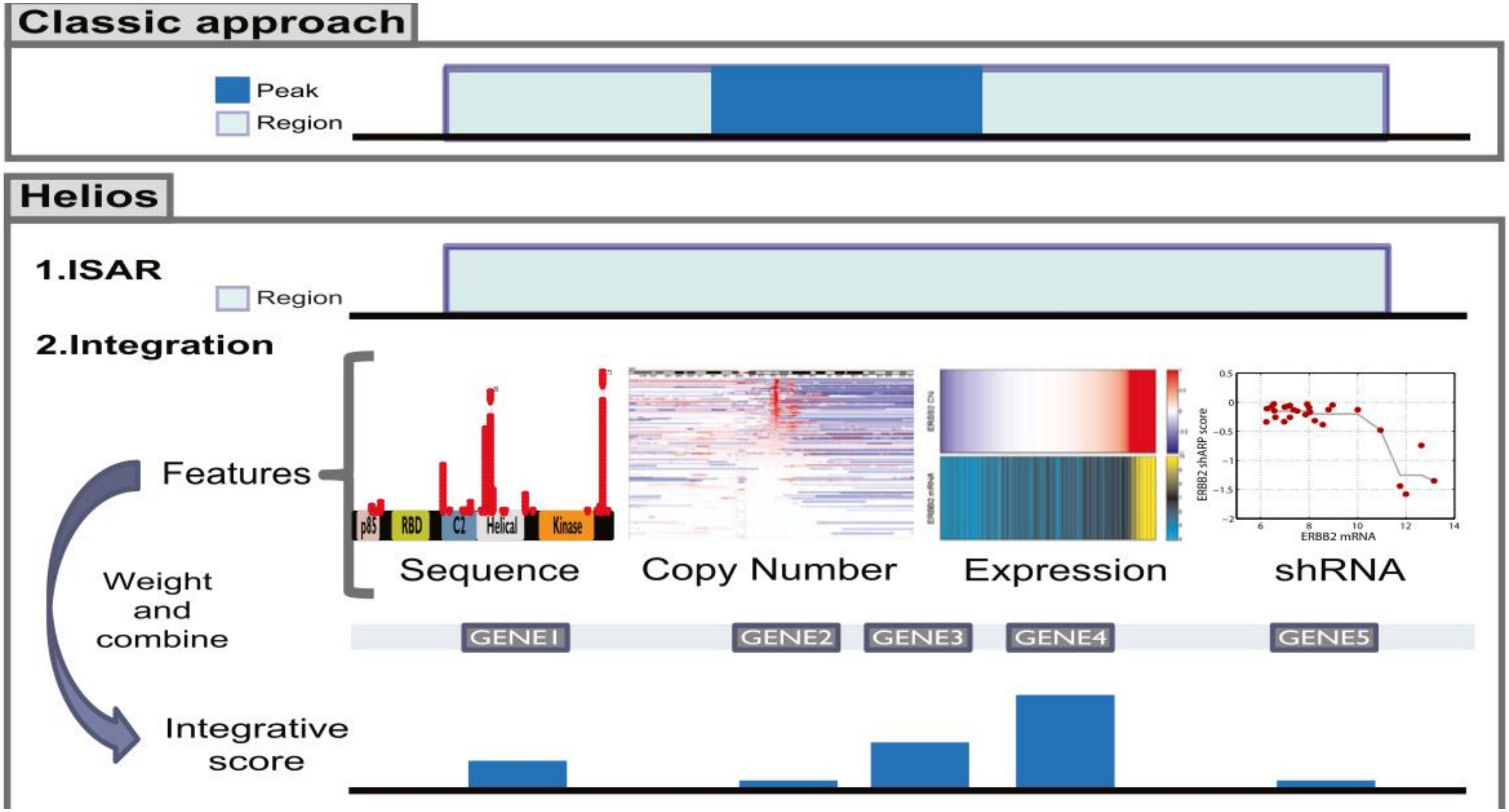
ISAR

$$S(m) = \max_{i \in W} -log10(qvalue_i(m))$$

W = set of window sizes
I = Window size used
Qvalue = based on local distribution
m = marker

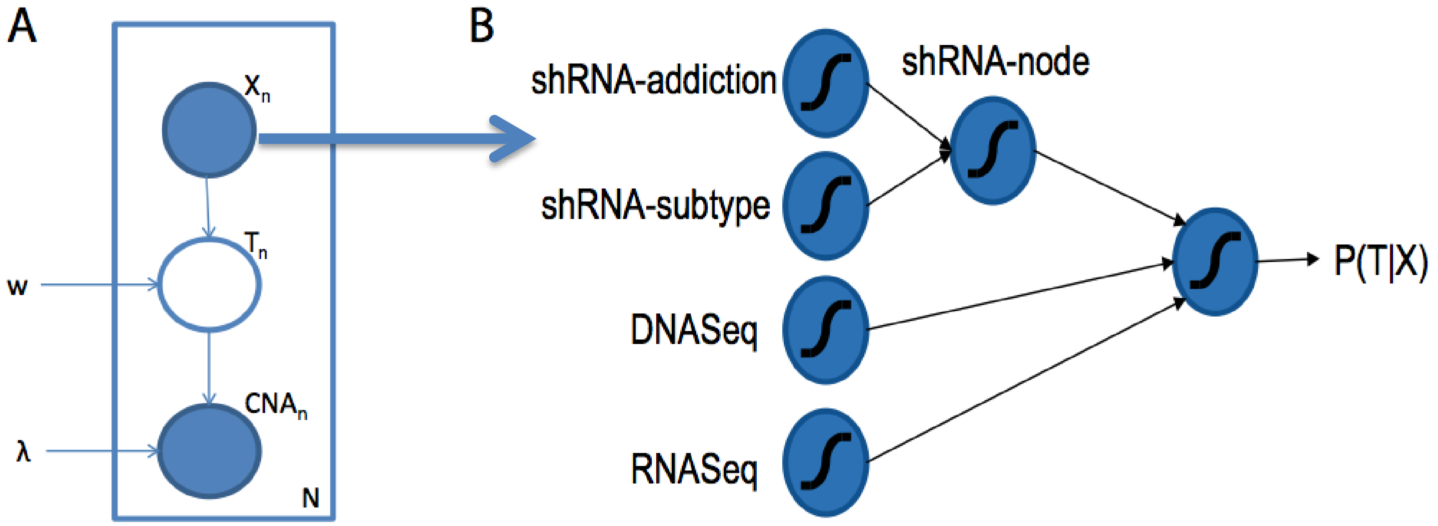


Helios


Classic approach
Peak
Region
Helios
1.ISAR
Region
2.Integration
Features
Weight and combine
Integrative score
p85
RBD
C2
Helical
Kinase
ERBB2 shARP score
ERBB2 mRNA
Sequence
Copy Number
Expression
shRNA
GENE1
GENE2
GENE3
GENE4
GENE5

Helios

$$P(CNA) = \sum_{t \in 0,1} P(SCNA|T = t)P(T = t|X).$$

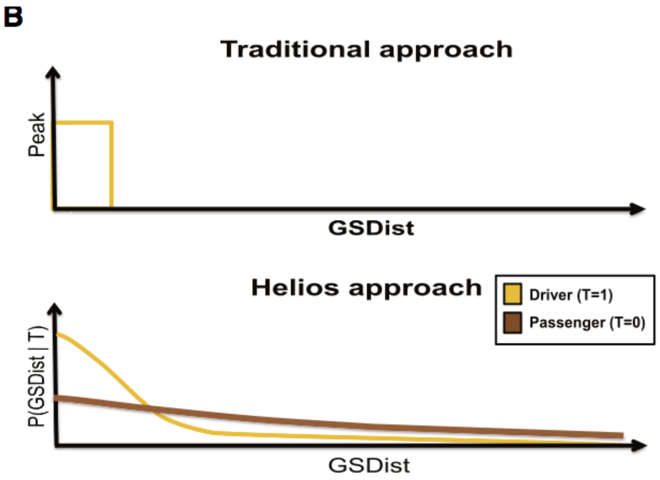


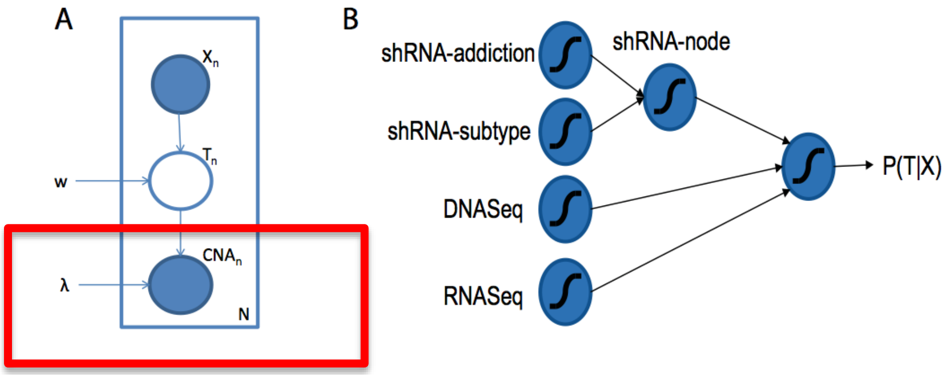
P(SCNA|T=t): Modeling of Copy Number

- Find Peak genes Independent of chromosomal region because distribution between regions is much larger than in region

$$GSDist(g) = \max_{j \in region(g)} (Gscore(j) - Gscore(g))$$

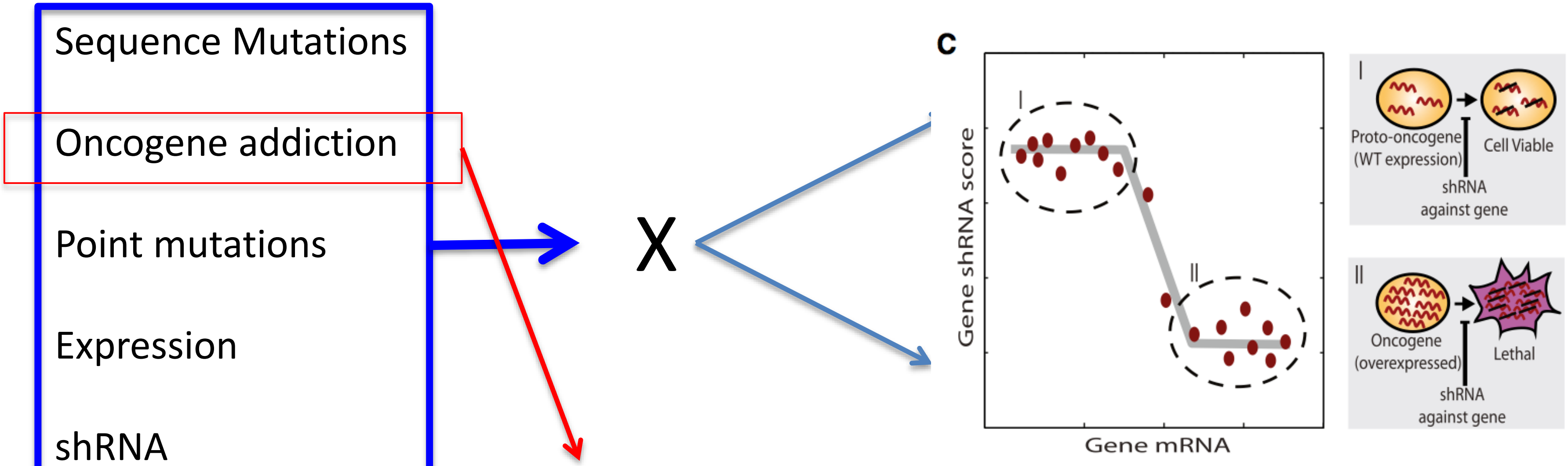
j = max in region
g = gene





P(T|X) – Modeling Additional Info

- Unified function made up of cues from all data



Sequence Mutations

Oncogene addiction

Point mutations

Expression

shRNA

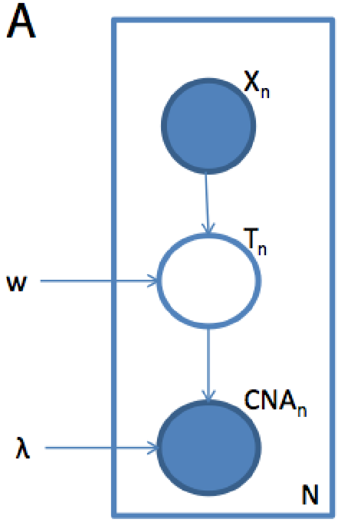
X

$$OA_h = -\log\Big(P\big(S_h \mid Exp_{target(h)}\big)\Big) = -\log(P(\epsilon))$$

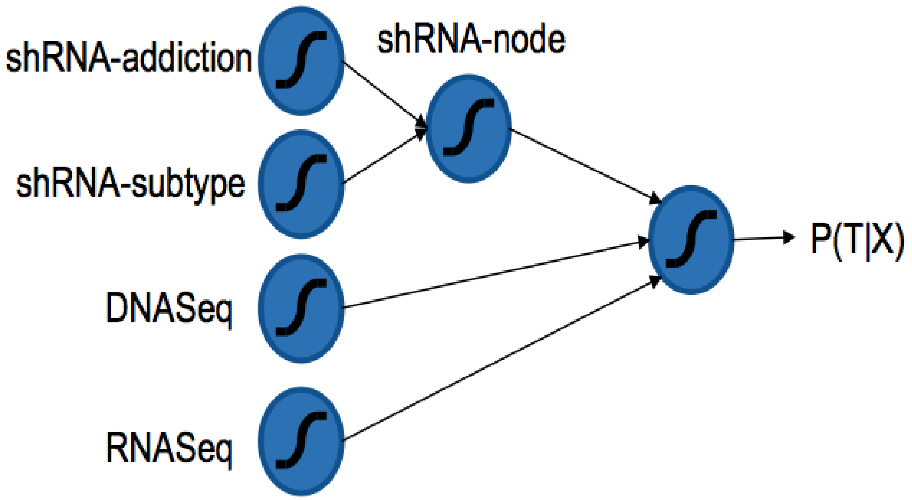
$$\epsilon = S_h - \widehat{S_h} = S_h - f(Exp_{target(h)})$$

$P(T|X)$

A

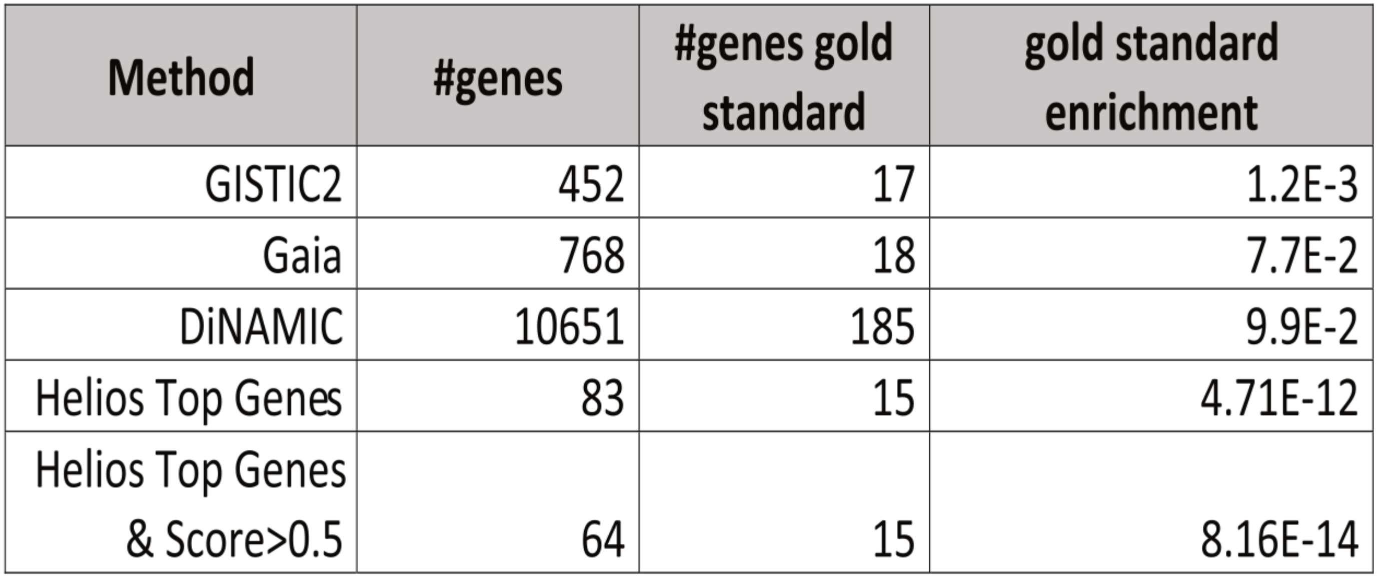

X_n
T_n
w
CNA_n
λ
N

B


shRNA-addiction
shRNA-node
shRNA-subtype
DNASeq
RNASeq
P(T|X)

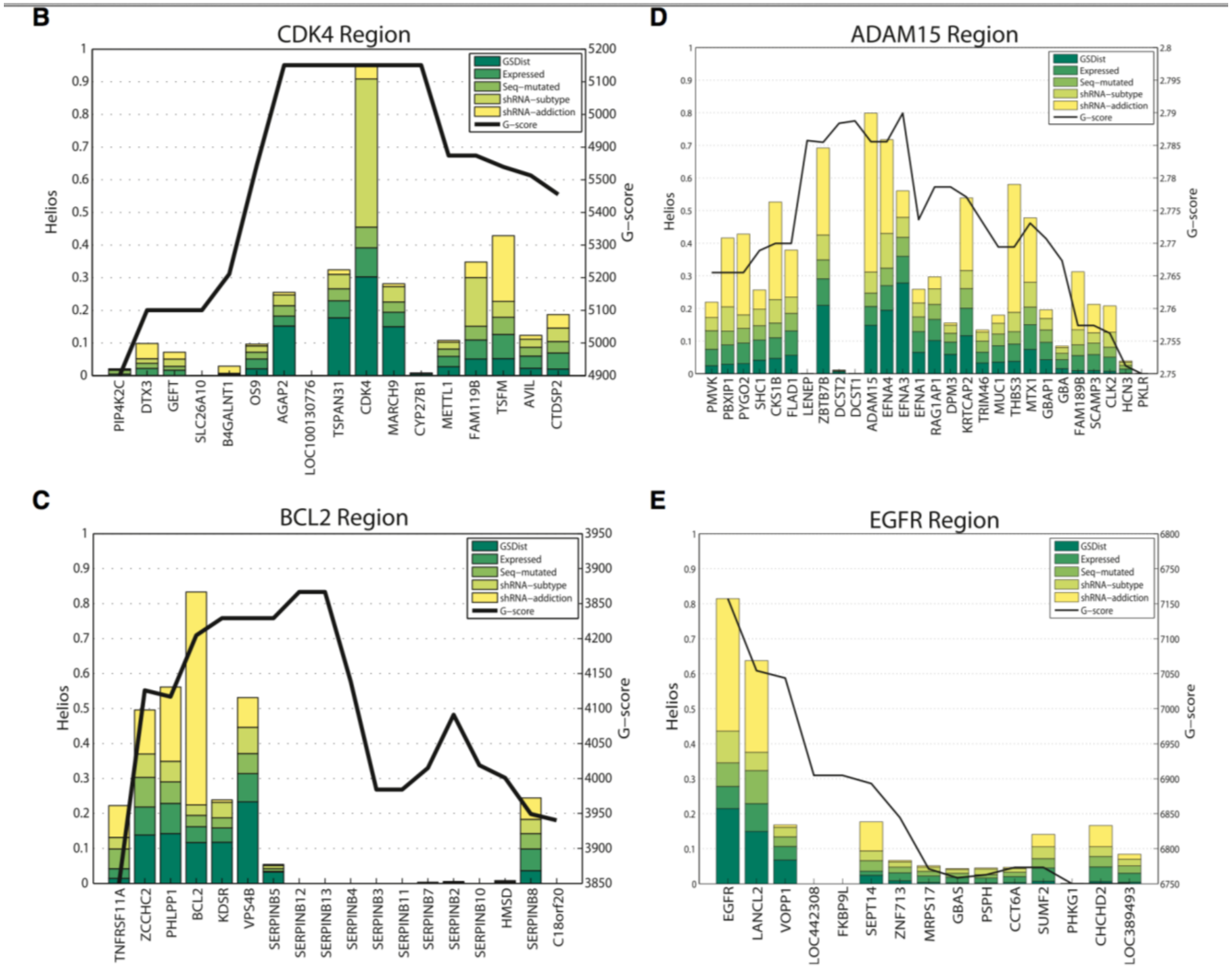
Results

- Using Helios, identified 64 candidate drivers using primary and cell line



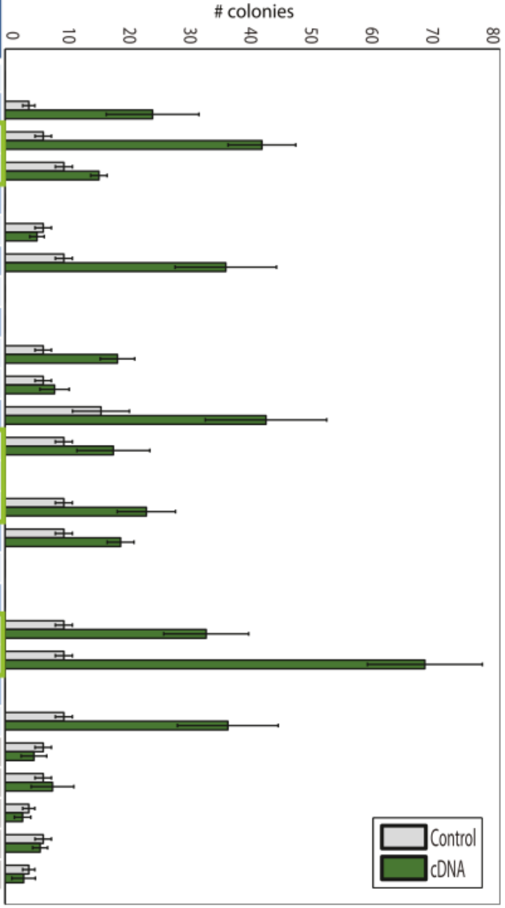
Method	#genes	#genes gold standard	gold standard enrichment
GISTIC2	452	17	1.2E-3
Gaia	768	18	7.7E-2
DiNAMIC	10651	185	9.9E-2
Helios Top Genes	83	15	4.71E-12
Helios Top Genes & Score>0.5	64	15	8.16E-14

Results

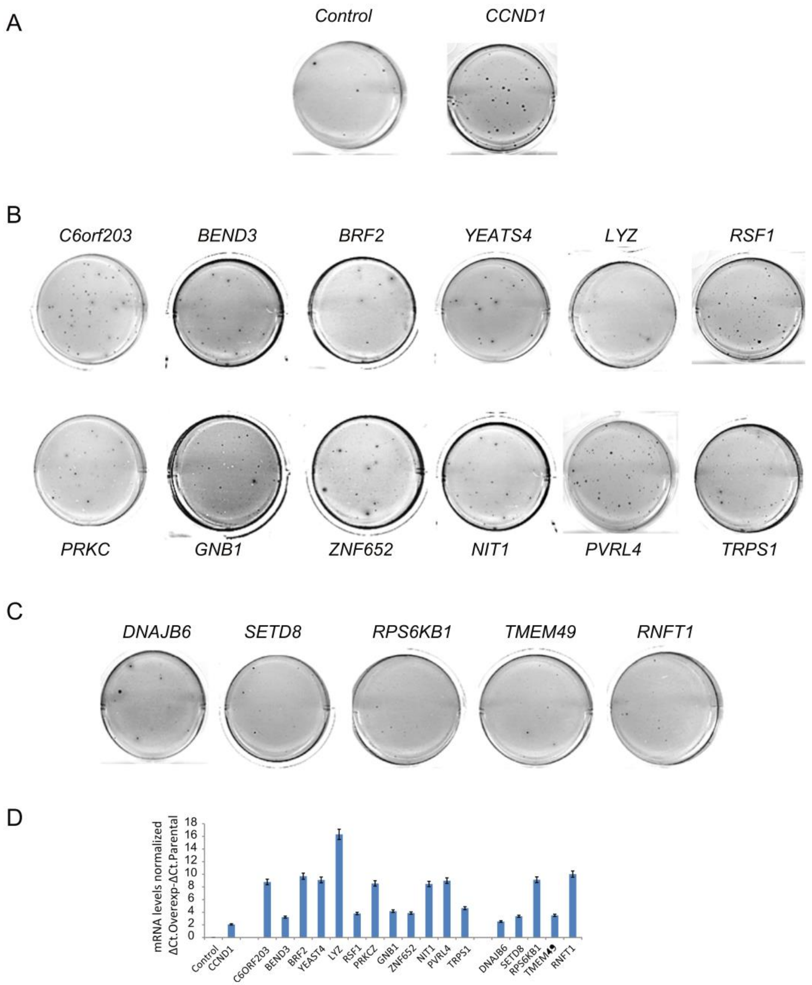

B
CDK4 Region
Helios
G–score
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
4900
5000
5100
5200
5300
5400
5500
4900
5000
5100
5200
GSDist
Expressed
Seq–mutated
shRNA–subtype
shRNA–addiction
G–score
PIP4K2C
DTX3
GEFT
SLC26A10
B4GALNT1
OS9
AGAP2
LOC100130776
TSPAN31
CDK4
MARCH9
CYP27B1
METTL1
FAM119B
TSFM
AVIL
CTDSP2
D
ADAM15 Region
Helios
G–score
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2.75
2.755
2.76
2.765
2.77
2.775
2.78
2.785
2.79
2.795
2.8
GSDist
Expressed
Seq–mutated
shRNA–subtype
shRNA–addiction
G–score
PMVK
PBXIP1
PYGO2
SHC1
CKS1B
FLAD1
LENEP
ZBTB7B
DCST2
DCST1
ADAM15
EFNA4
EFNA3
EFNA1
RAG1AP1
DPM3
KRTCAP2
TRIM46
MUC1
THBS3
MTX1
GBAP1
GBA
FAM189B
SCAMP3
CLK2
HCN3
PKLR
C
BCL2 Region
Helios
G–score
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
3850
3900
3950
4000
4050
4100
4150
4200
3850
3900
3950
GSDist
Expressed
Seq–mutated
shRNA–subtype
shRNA–addiction
G–score
TNFRSF11A
ZCCHC2
PHLPP1
BCL2
KDSR
VPS4B
SERPINB5
SERPINB12
SERPINB13
SERPINB4
SERPINB3
SERPINB11
SERPINB7
SERPINB2
SERPINB10
HMSD
SERPINB8
C18orf20
E
EGFR Region
Helios
G–score
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
6750
6800
6850
6900
6950
7000
7050
7100
7150
6750
6800
GSDist
Expressed
Seq–mutated
shRNA–subtype
shRNA–addiction
G–score
EGFR
LANCL2
VOPP1
LOC442308
FKBP9L
SEPT14
ZNF713
MRPS17
GBAS
PSPH
CCT6A
SUMF2
PHKG1
CHCHD2
LOC389493

Results: Validation

ISAR Peak	Location	Gene	Helios Score	Ratio Next	Validation p-value	Supported Driver
1	17q12	ERBB2	0.88	2.32		Bona fide
2	11p13	CCND1	0.95	2.96	3.53E-05	Bona fide
3	6q21	C6orf203	0.44	1.00	1.37E-08	Validated
		BEND3	0.44	2.41	1.38E-05	Validated
4	20q13	ZNF217	0.67	667.71		Bona fide
5	8p11	BRF2	0.74	1.31	0.902	Negative
6	8q24	MYC	0.52	97.19	8.05E-06	Bona fide
7	17q23	MIR21	(*)			Bona fide
8	17q11	MYO18A	0.57	1.54		
9	12q13	YEATS4	0.60	1.04	1.25E-06	Validated
		LYZ	0.58	1.30	0.0645	Negative
10	11q14	RSF1	0.53	1.20	6.33E-05	Validated
11	1p36	PRKCZ	0.79	1.23	0.00444	Validated
		SKI	0.64	1.01		
		GNB1	0.63	1.73	2.83E-05	Validated
12	17q21	ZNF652	0.36	1.36	2.63E-06	Validated
13	10q26	FGFR2	0.74	14.65		Bona fide
14	4q13	ANKRD17	0.57	1.27		
15	1q21	NIT1	0.68	1.08	5.58E-06	Validated
		PVRL4	0.63	1.16	1.51E-08	Validated
16	15q26	IGF1R	0.84	4.77		Bona fide
17	8q24	TRPS1	0.70	1.24	6.91E-06	Validated
Negative Controls		DNAJB6	0.26		0.917	
		SETD8	0.21		0.172	
		RPS6KB1	0.23		0.915	
		TMEM49	0.06		0.744	
		RNFT1	0.02		0.819	

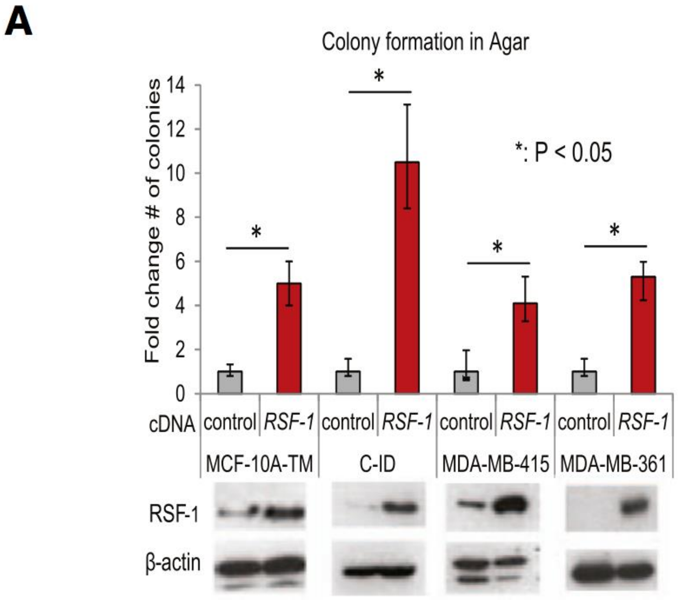


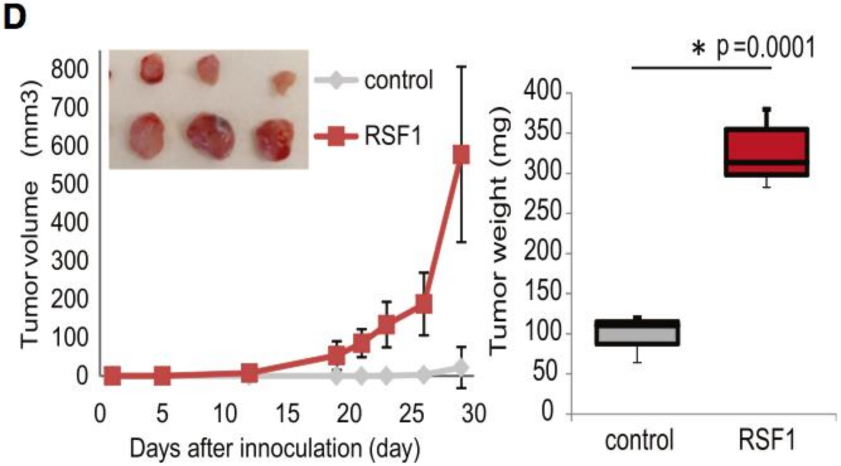
Results


A
Control
CCND1
B
C6orf203
BEND3
BRF2
YEATS4
LYZ
RSF1
PRKC
GNB1
ZNF652
NIT1
PVRL4
TRPS1
C
DNAJB6
SETD8
RPS6KB1
TMEM49
RNFT1
D
mRNA levels normalized
ΔCt.Overexp-ΔCt.Parental
0
2
4
6
8
10
12
14
16
18
Control
CCND1
C6ORF203
BEND3
BRF2
YEAST4
LYZ
RSF1
PRKCZ
GNB1
ZNF652
NIT1
PVRL4
TRPS1
DNAJB6
SETD8
RPS6KB1
TMEM49
RNFT1

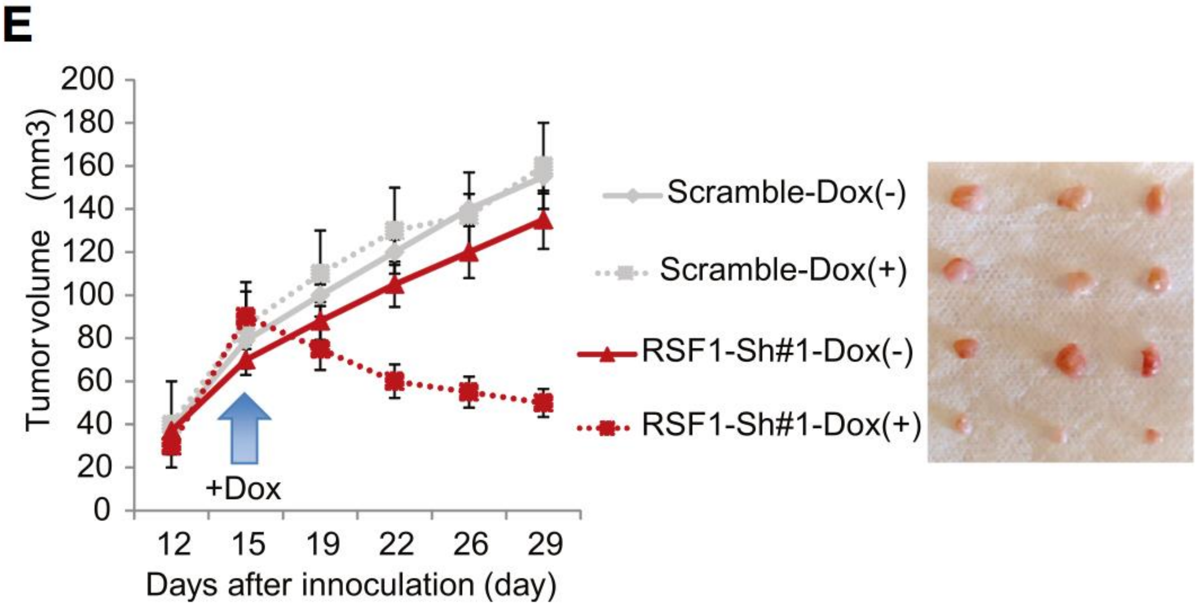
Results: RSF-1

Overexpression of RSF1 in CID cells

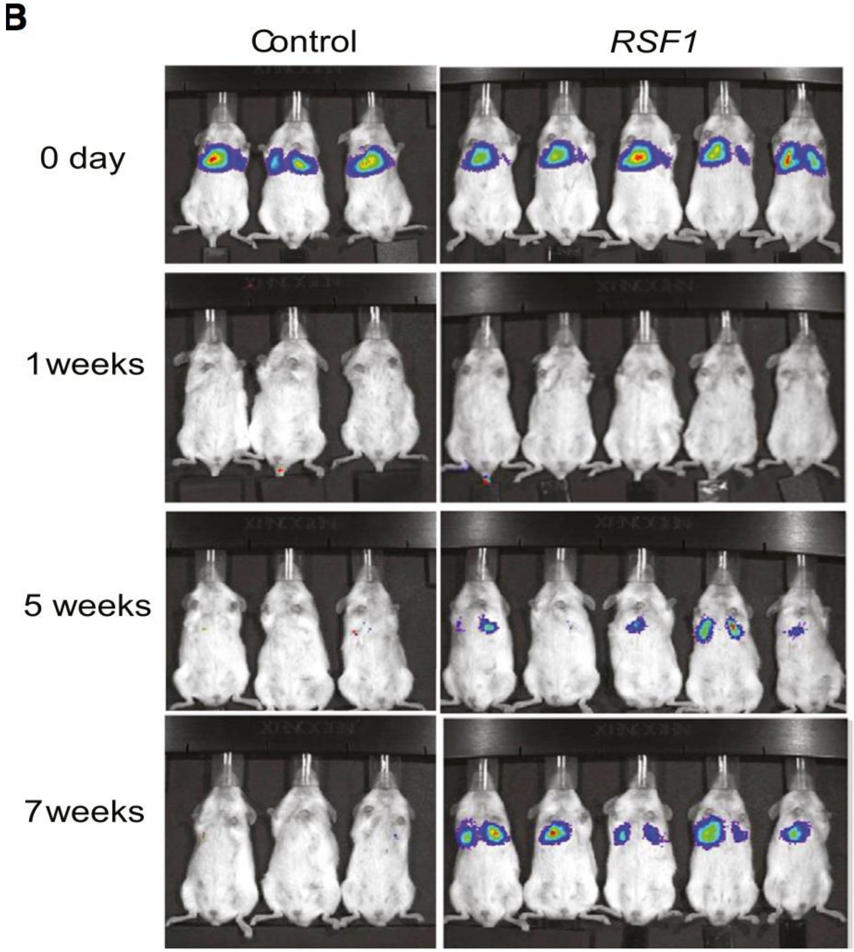




Results: Dox Inducible



Results: Xenograft



Results

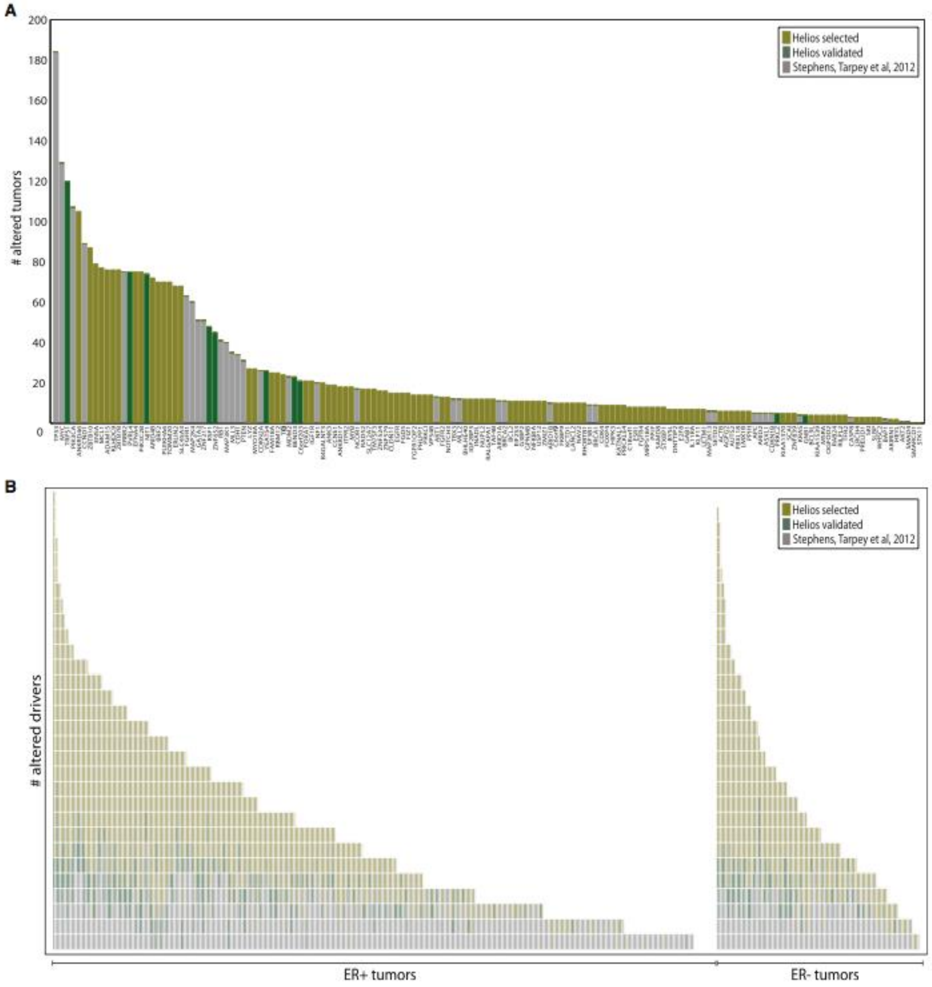


Figure 7. The Landscape of Driver Mutations in Breast Cancer

Take Aways

- A step forward in finding and validating driver genes
- First time (to my knowledge) of a computation→ *in vitro* →*in vivo* study
- Open up potential to incorporate therapeutic targets