

An Integrated Approach to Uncover Drivers of Cancer

Uri David Akavia,^{1,2,5} Oren Litvin,^{1,2,5} Jessica Kim,^{3,4} Felix Sanchez-Garcia,¹ Dylan Kotliar,¹ Helen C. Causton,¹ Panisa Pochanard,^{3,4} Eyal Mozes,¹ Levi A. Garraway,^{3,4} and Dana Pe'er^{1,2,*}

¹Department of Biological Sciences, Columbia University, 1212 Amsterdam Avenue, New York, NY 10027, USA

²Center for Computational Biology and Bioinformatics, Columbia University, 1130 St. Nicholas Avenue, New York, NY 10032, USA

³Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney Street, Boston, MA 02115, USA

⁴Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02142, USA

⁵These authors contributed equally to this work

Andrea Sboner

JC - 2011-01-19

Problem

Problem

- ~ Systematic characterization of cancer genomes has revealed a *staggering number* of diverse aberrations that differ among individuals, such that the functional importance and physiological impact of most tumor genetic alterations remain poorly defined

Goal

Goal

- ~ To identify “*driver* mutations”: genetic changes that promote cancer progression

Goal

- ~ To identify “*driver* mutations”: genetic changes that promote cancer progression
- ~ and distinguish them from “*passenger* mutations”: with no selective advantage

Approach

~ Integration of:

1. Chromosomal Copy Number Alterations (CNAs)

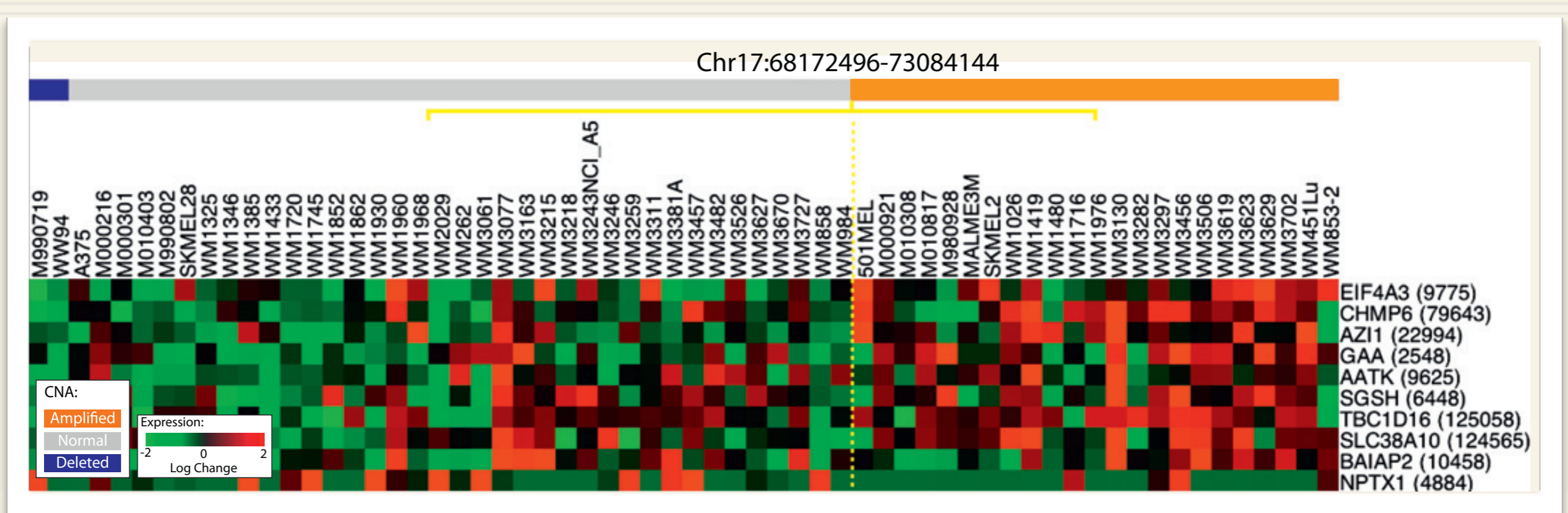
Approach

~ Integration of:

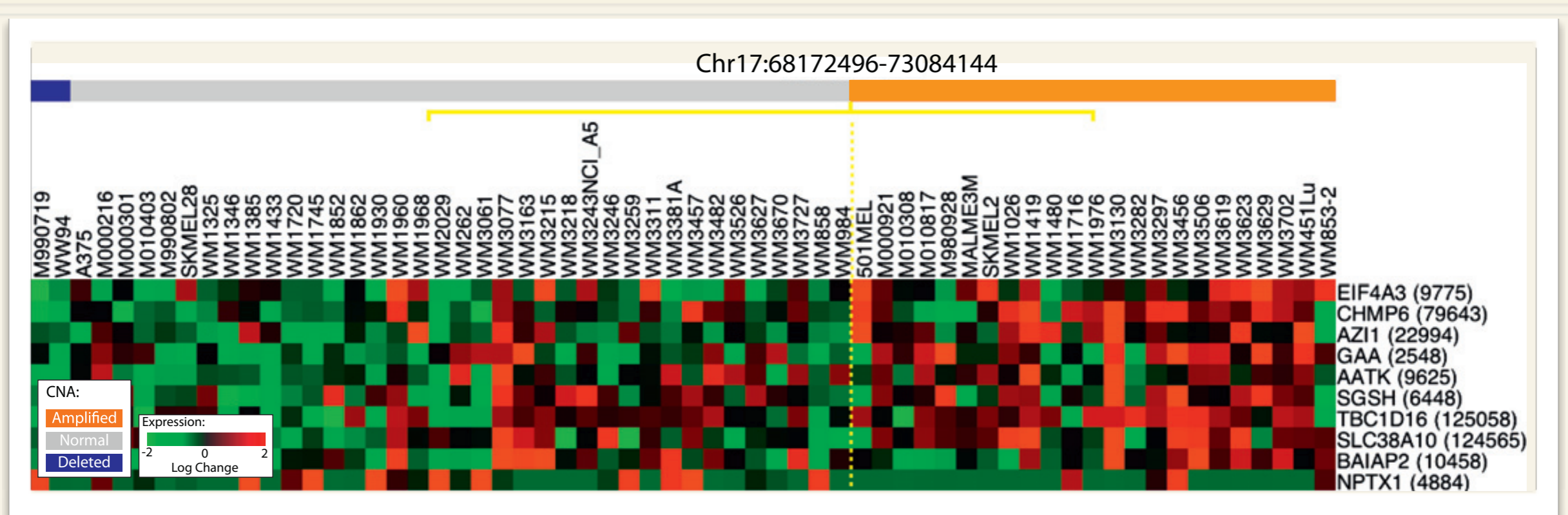
1. Chromosomal Copy Number Alterations (CNAs)
2. *Gene Expression*

Hypothesis

- ~ Driver mutations have a “*genomic footprint*”: a gene expression signature



Genes with identical amplification status,
but “slightly” different expression patterns



Genes with identical amplification status,
but “slightly” different expression patterns

“a driver mutation might be associated with a characteristic gene expression signature [...] representing a group of genes whose expression is modulated by the driver”

Limitations of CNAs only

Limitations of CNAs only

- ~ CNA regions include many genes
(most of them are passengers)

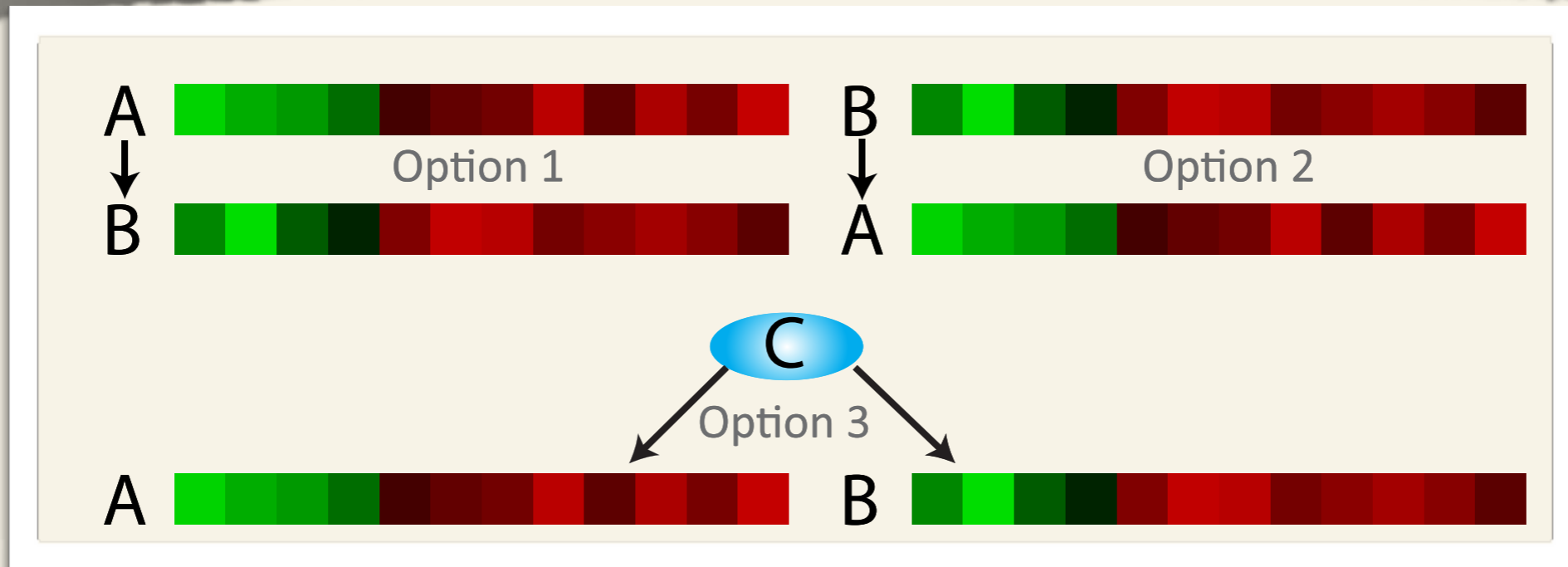
Limitations of CNAs only

- ~ CNA regions include many genes
(most of them are passengers)
- ~ Can detect the recurrent ones only

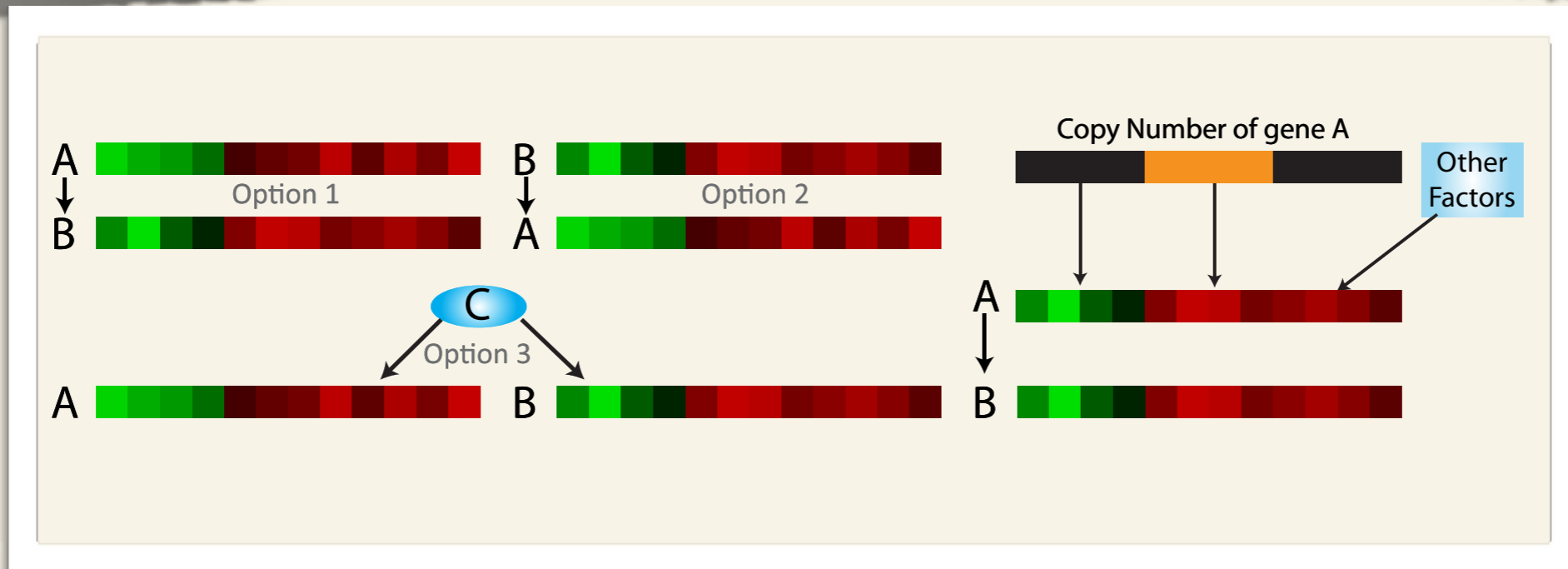
Limitations of CNAs only

- ~ CNA regions include many genes
(most of them are passengers)
- ~ Can detect the recurrent ones only
- ~ Do not provide functional insights

Establishing directionality



Establishing directionality



- ~ The use of expression and CNAs allows to establish the likelihood of influence

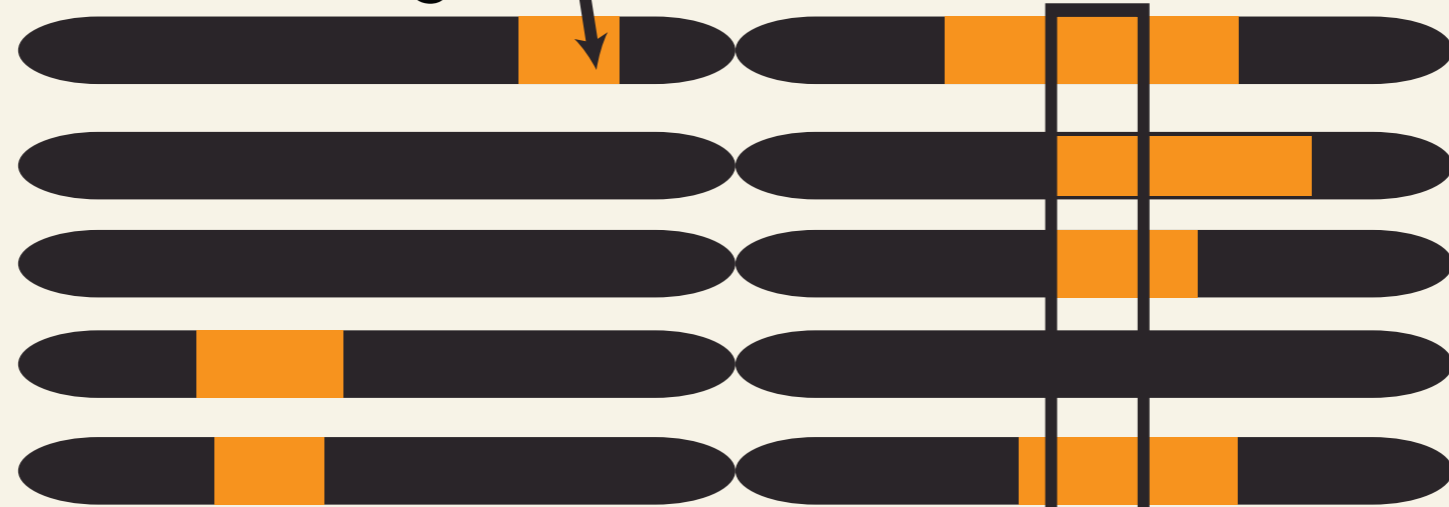
Three assumptions

1 Driver mutations enriched in multiple tumors

2 Driver expression correlated to a group of genes

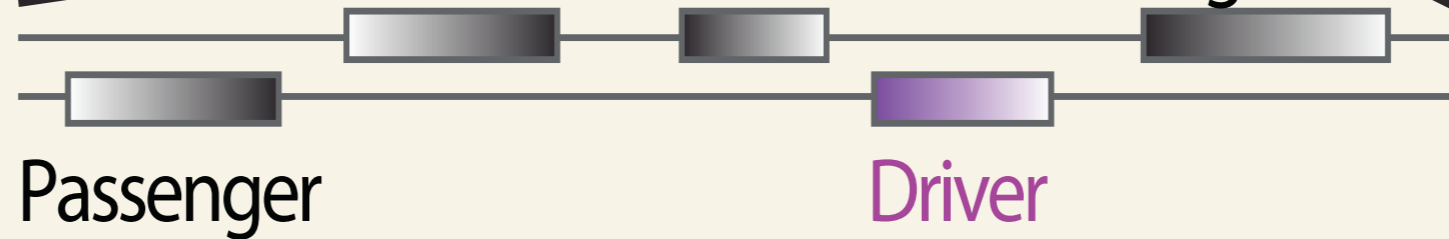
3 CNAs influence expression of genes in the module via the driver

Aberrant region



same chromosome
different samples

Genes in an aberrant region



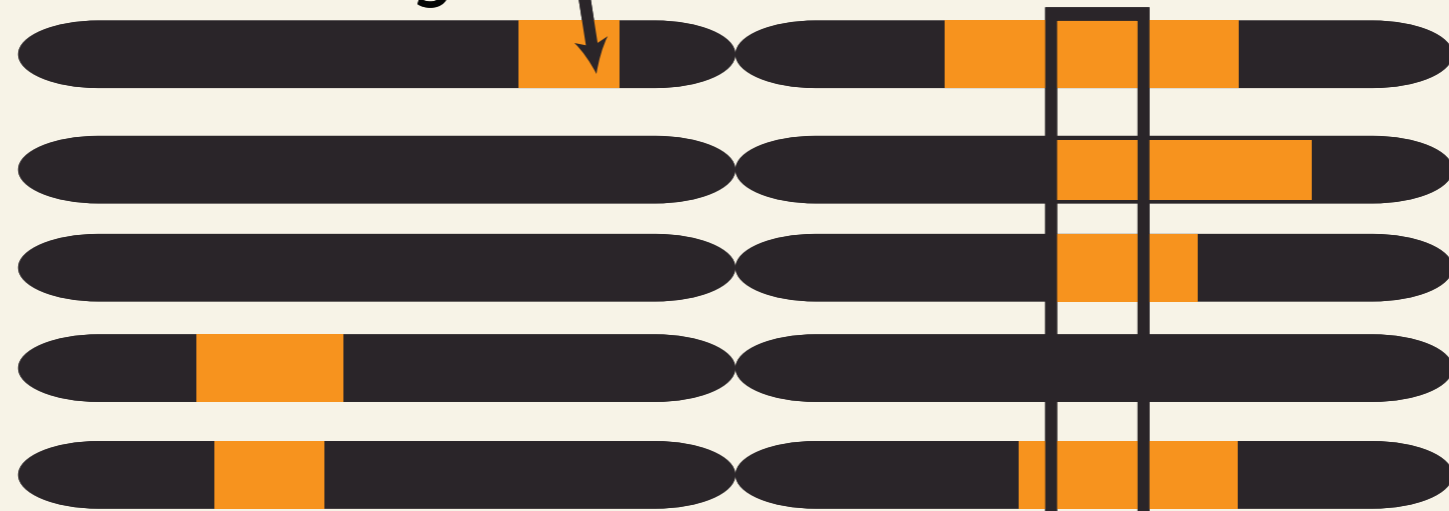
Three assumptions

1 Driver mutations enriched in multiple tumors

2 Driver expression correlated to a group of genes

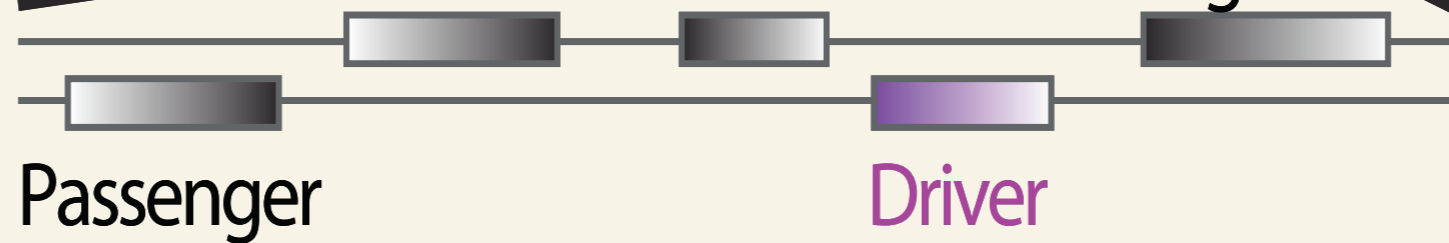
3 CNAs influence expression of genes in the module via the driver

Aberrant region



same chromosome
different samples

Genes in an aberrant region

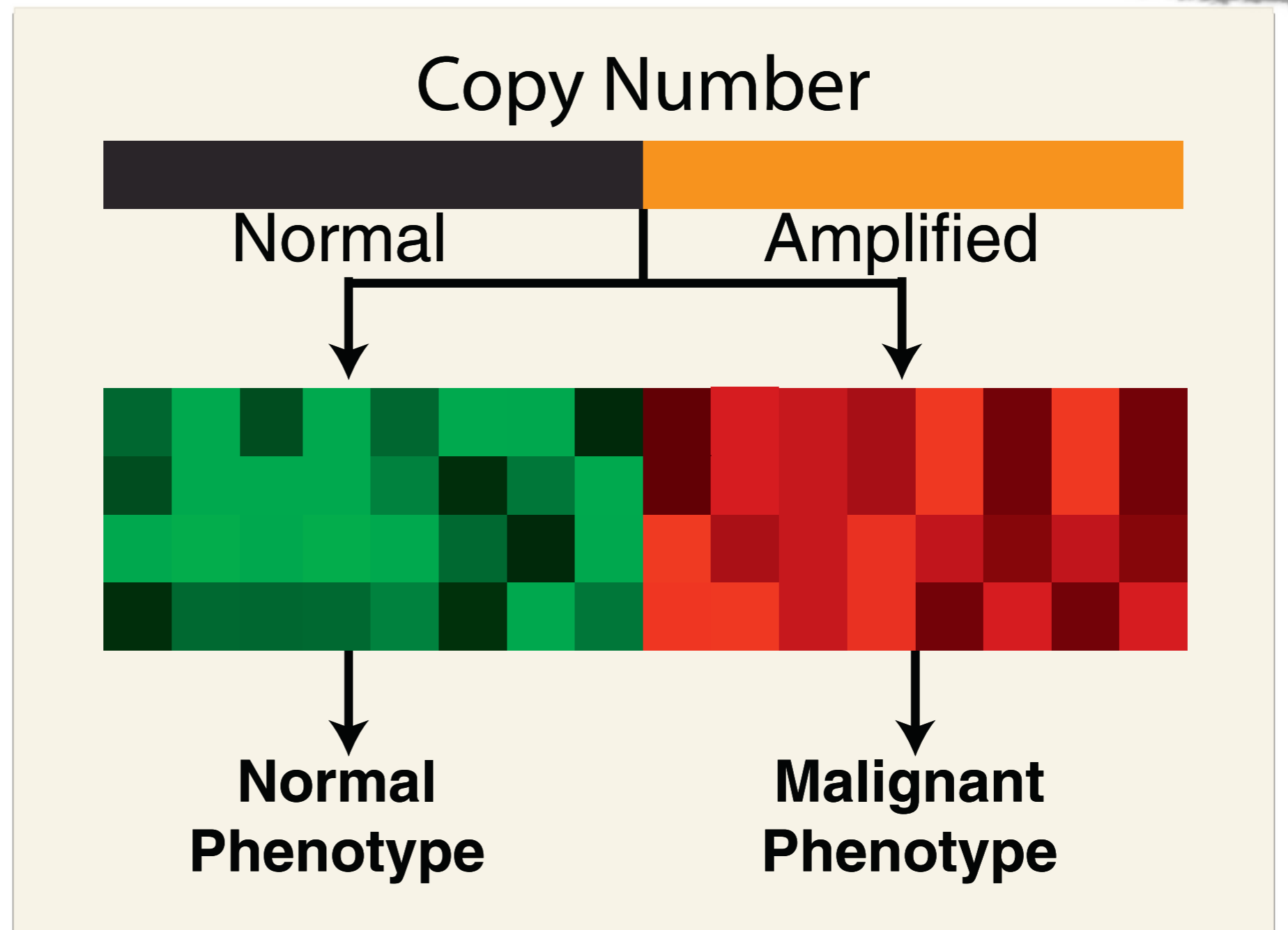


Three assumptions

1 Driver mutations enriched in multiple tumors

2 Driver expression correlated to a group of genes

3 CNAs influence expression of genes in the module via the driver

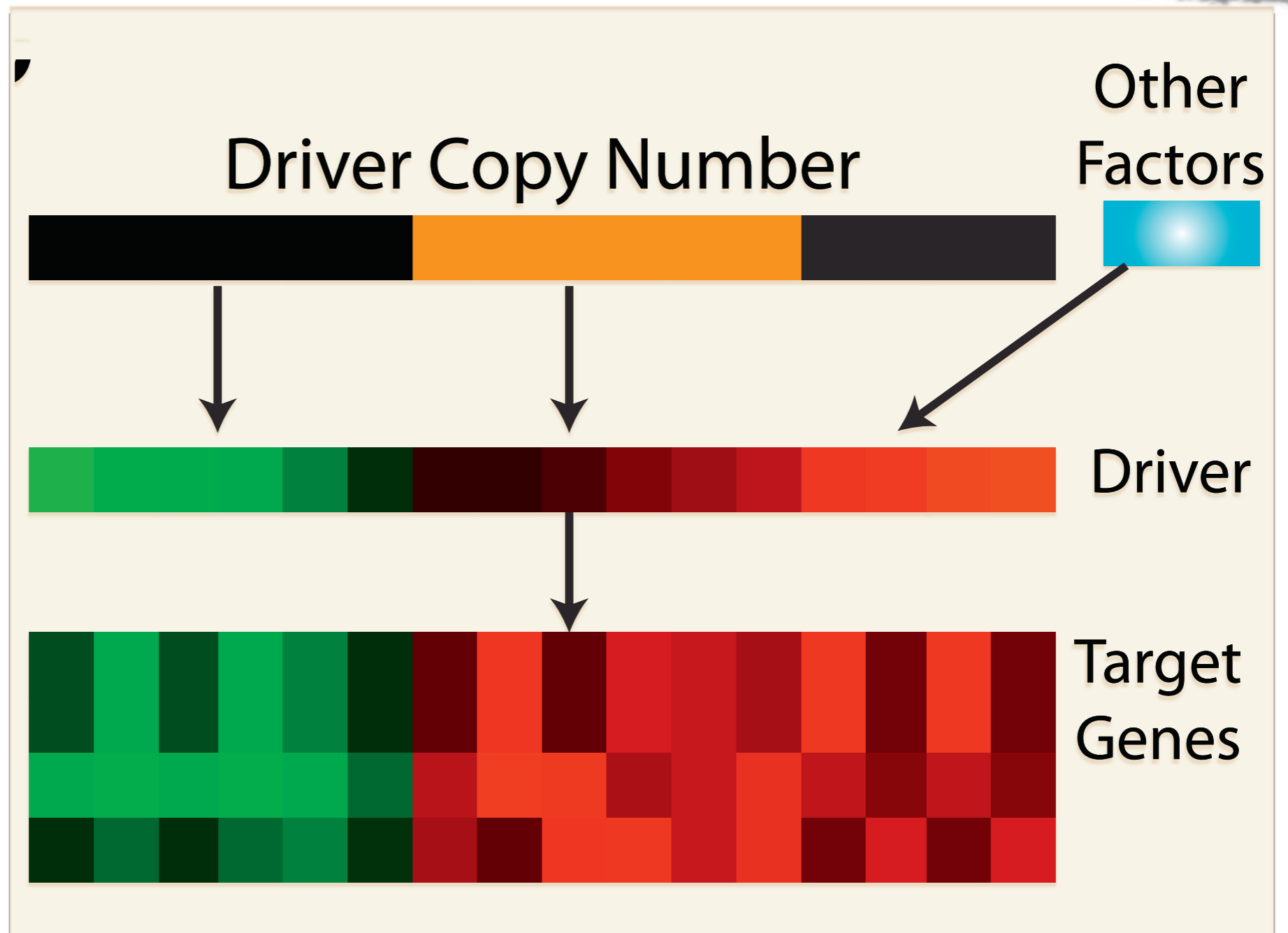


Three assumptions

1 Driver mutations enriched in multiple tumors

2 Driver expression correlated to a group of genes

3 CNAs influence expression of genes in the module via the driver



CONEXIC

~ **CO**py Number and **EX**pression In **C**ancer:

CONEXIC

- ~ **CO**py Number and **EX**pression In Cancer:
 - ~ integrates both copy number and gene expression data

CONEXIC

- ~ **COpy Number and EXpression In Cancer:**
 - ~ integrates both copy number and gene expression data
 - ~ identifies drivers and the processes they influence

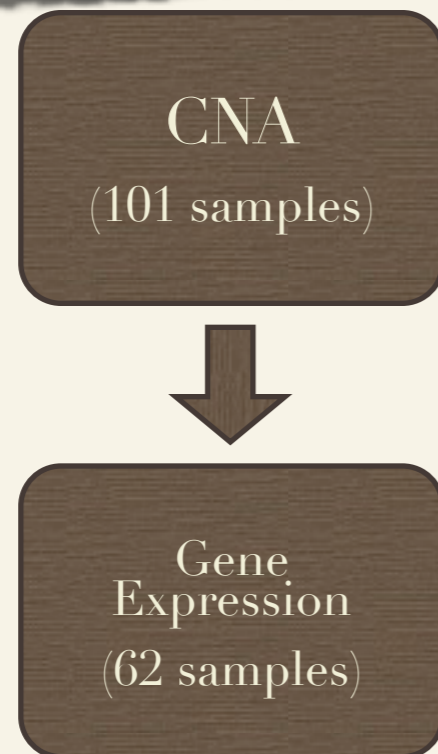
CONEXIC

- ~ **COpy Number and EXpression In Cancer:**
 - ~ integrates both copy number and gene expression data
 - ~ identifies drivers and the processes they influence
- ~ Applied to a melanoma data set

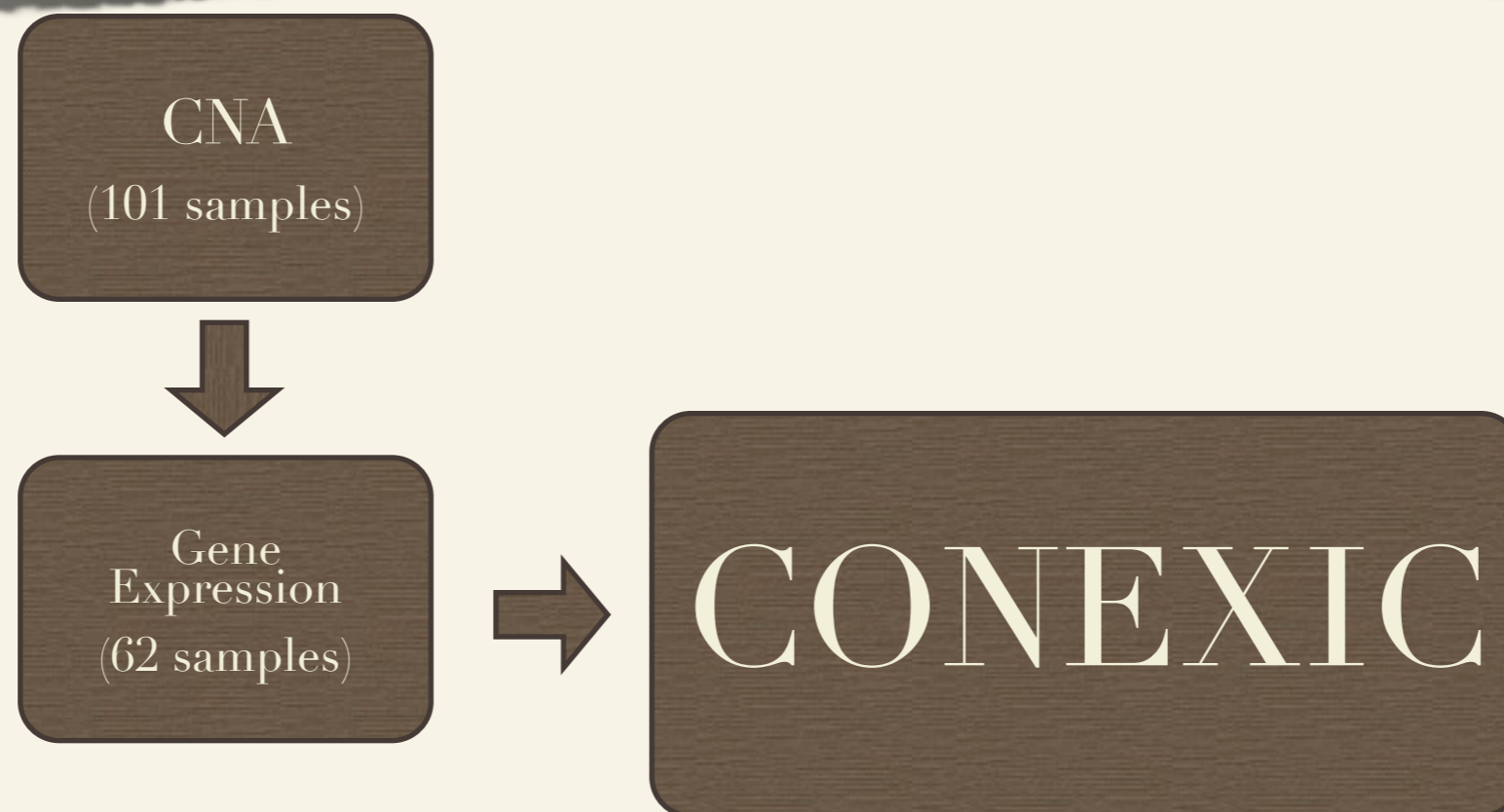
Overview of the study

CNA
(101 samples)

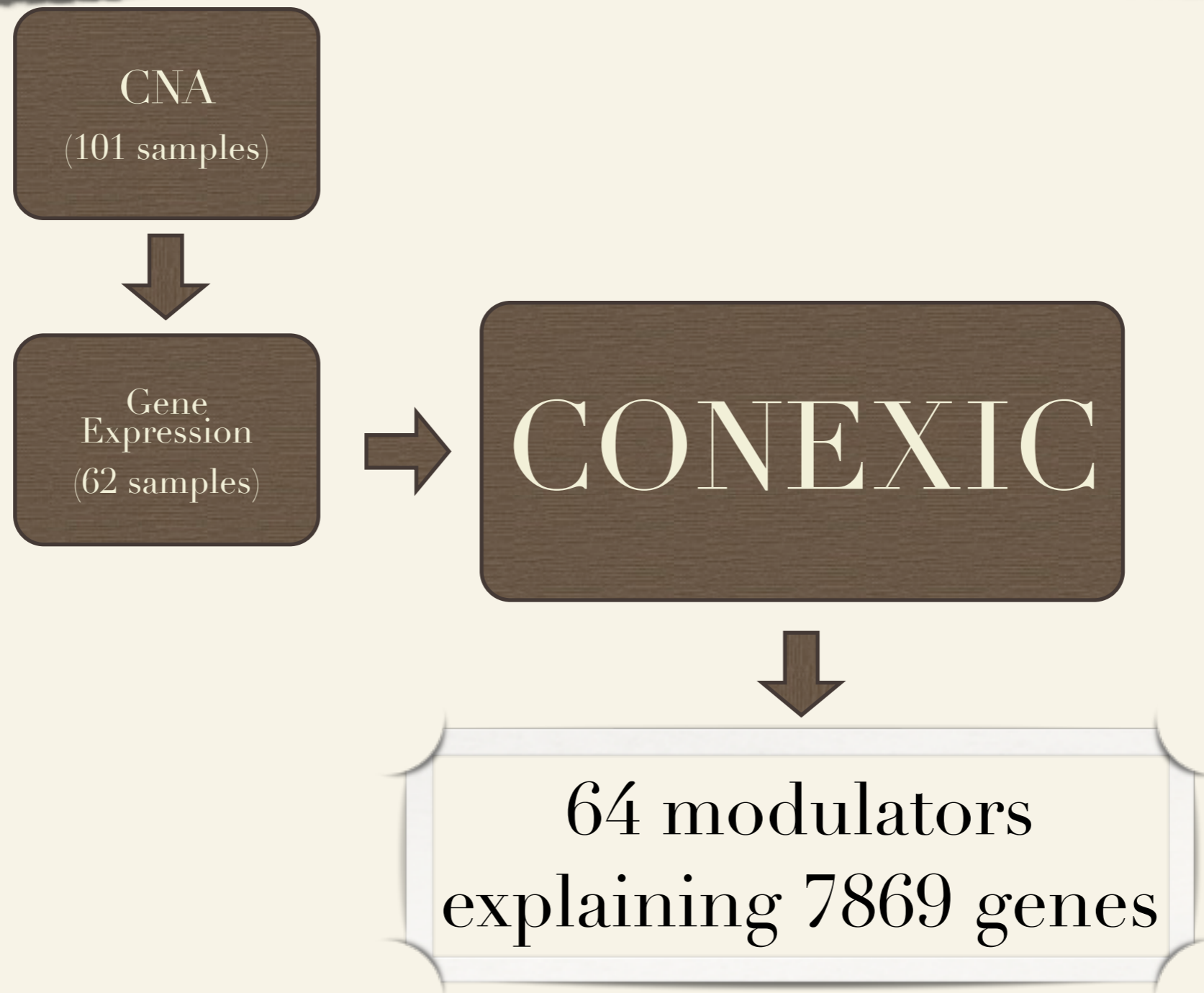
Overview of the study



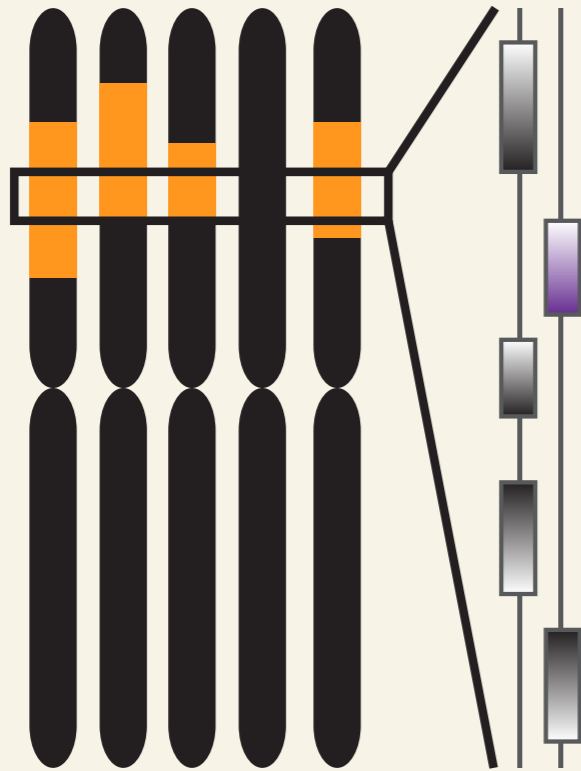
Overview of the study



Overview of the study



Details of the approach



Amplified Genes:

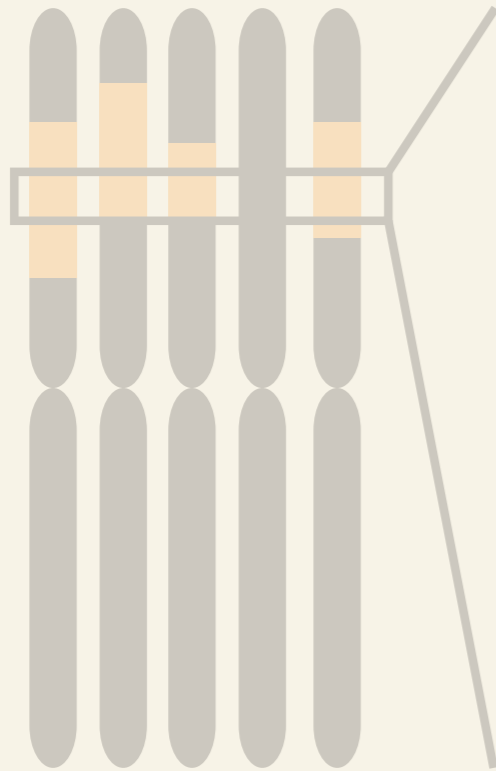
1. CCND1
2. MITF
- 3.....

Deleted Genes:

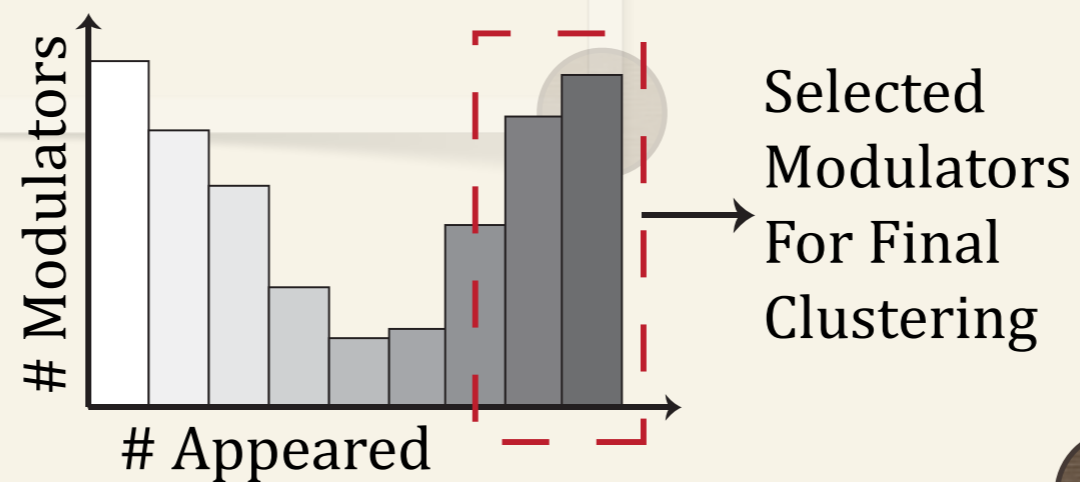
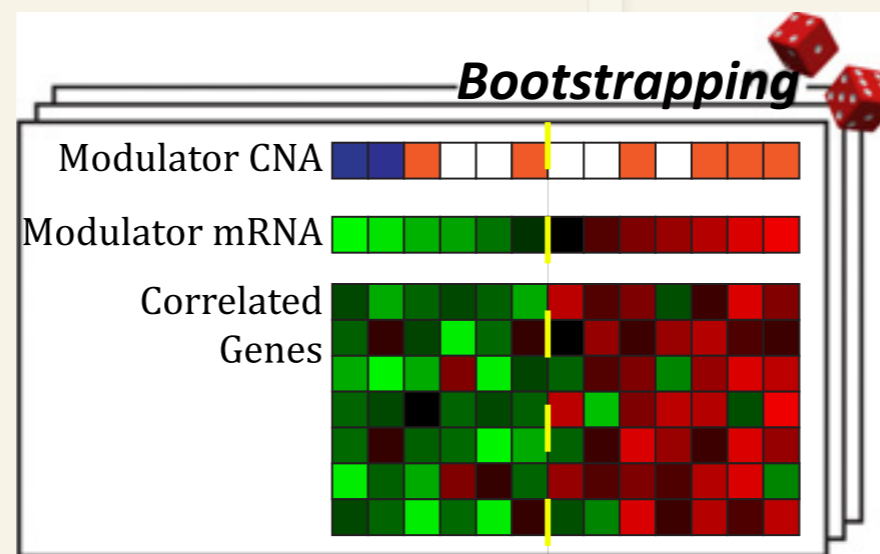
1. CDKN2A
2. KLF6
- 3.....

1

Details of the approach

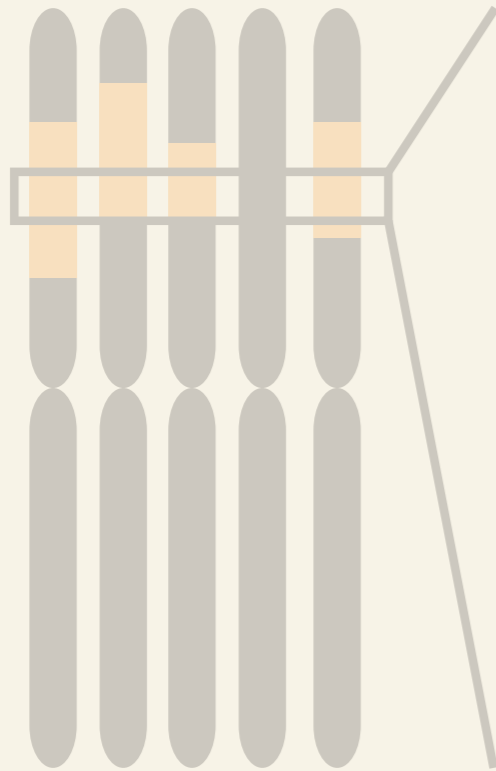


Amplified Genes:



2

Details of the approach



Amplified Genes:

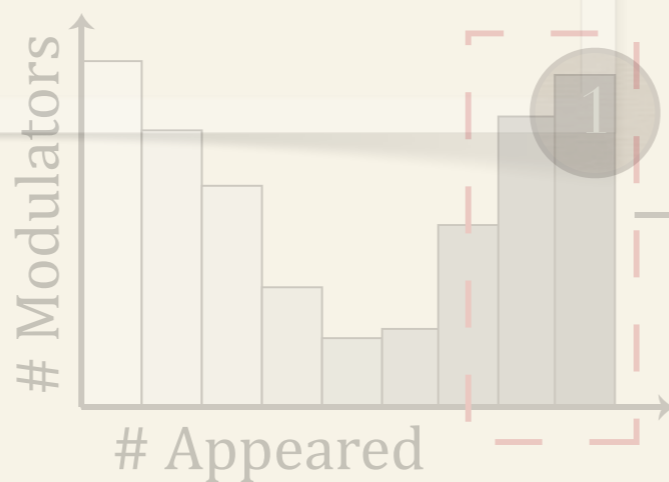
1. CCND1
2. MITF

3.....
Modulator CNA

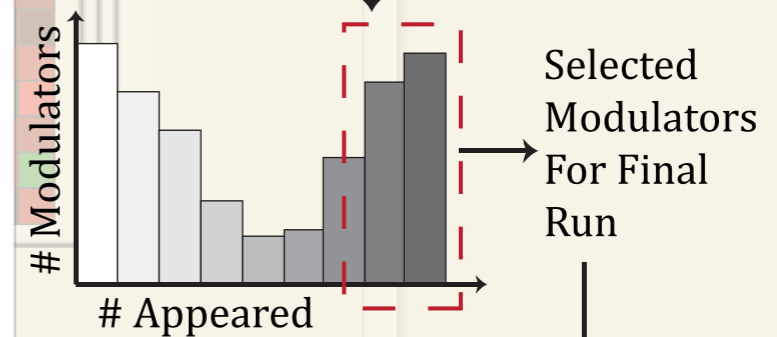
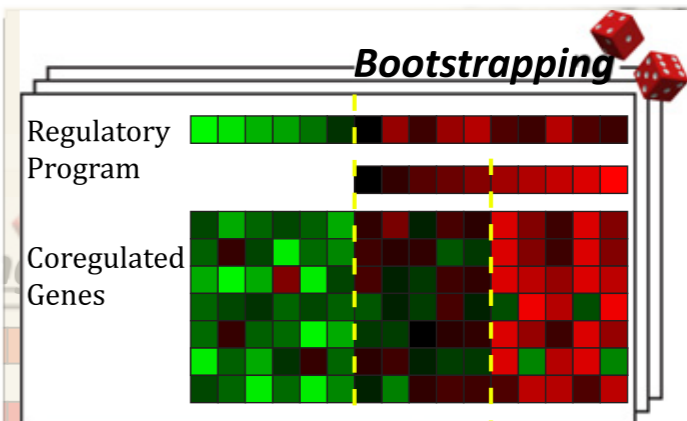
A horizontal bar chart showing copy number alterations (CNA) for various modulators. The bars are color-coded, with orange indicating amplification and blue indicating deletion.

Deleted Genes:

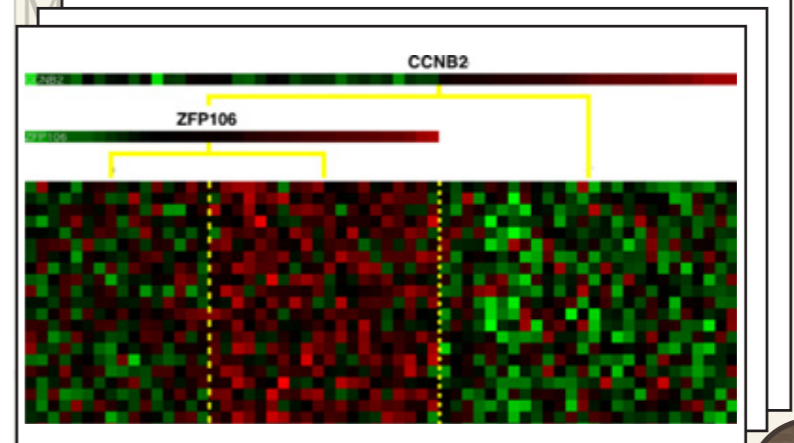
1. CDKN2A
2. KLF6
- 3.....



Bootstrapping



Final Results



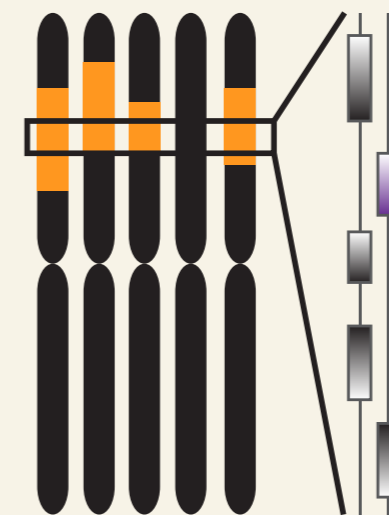
Selection of candidate drivers

~ Motivation:

- ~ identification of recurrently amplified/deleted regions in tumors

~ Method:

- ~ GISTIC (101 samples) - $qval=0.3$
- ~ Genes within $\pm 100Kb$ of each region
- ~ Expression filtering (62 samples) - $SD > 0.25$



Amplified Genes:

1. CCND1
2. MITF
- 3.....

Deleted Genes:

1. CDKN2A
2. KLF6
- 3.....

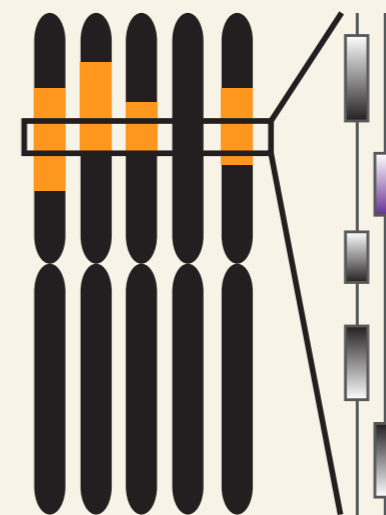
Selection of candidate drivers

~ Motivation:

- ~ identification of recurrently amplified/deleted regions in tumors

~ Method:

- ~ GISTIC (101 samples) - $qval=0.3$
- ~ Genes within $\pm 100Kb$ of each region
- ~ Expression filtering (62 samples) - $SD > 0.25$



Amplified Genes:

1. CCND1
2. MITF
- 3.....

Deleted Genes:

1. CDKN2A
2. KLF6
- 3.....

1

Result

- * 27 amplified regions
- * 23 deleted regions
- * 428 genes

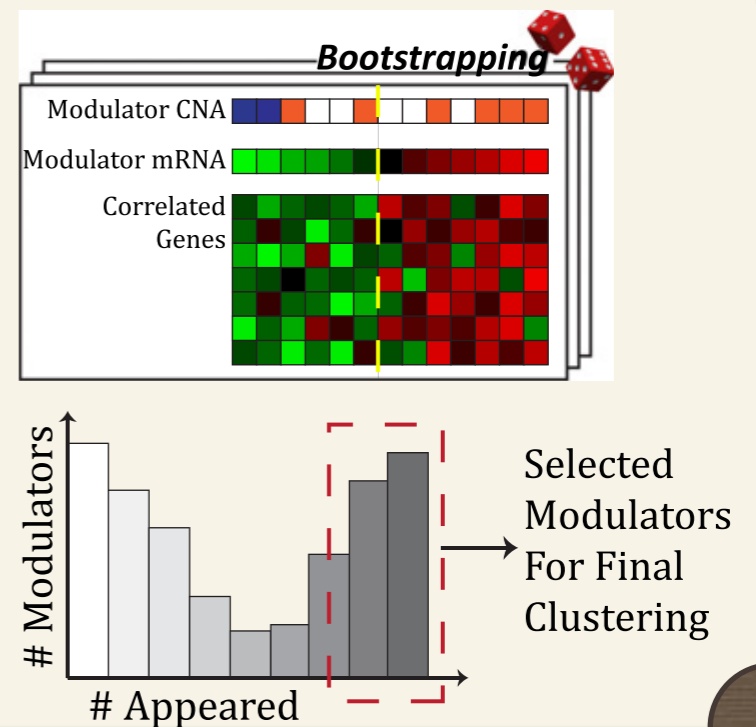
Single Modulator

~ Motivation:

- ~ to construct an initial model by associating each target gene with the single driver gene that fits it best

~ Method:

- ~ Based on Module Networks (Segal et al. 2003, 2005)



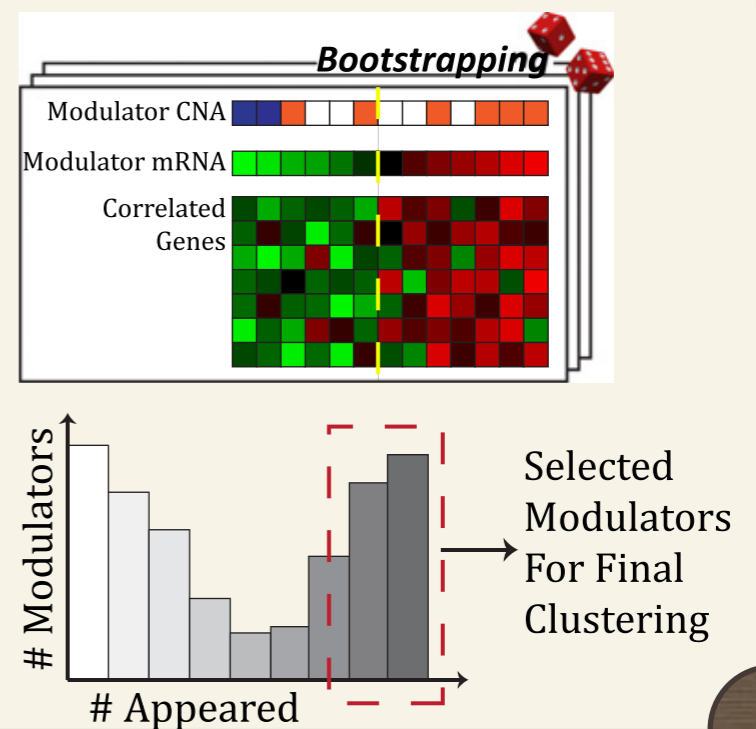
Single Modulator

~ Motivation:

- ~ to construct an initial model by associating each target gene with the single driver gene that fits it best

~ Method:

- ~ Based on Module Networks (Segal et al. 2003, 2005)
- ~ For each candidate driver gene (347):



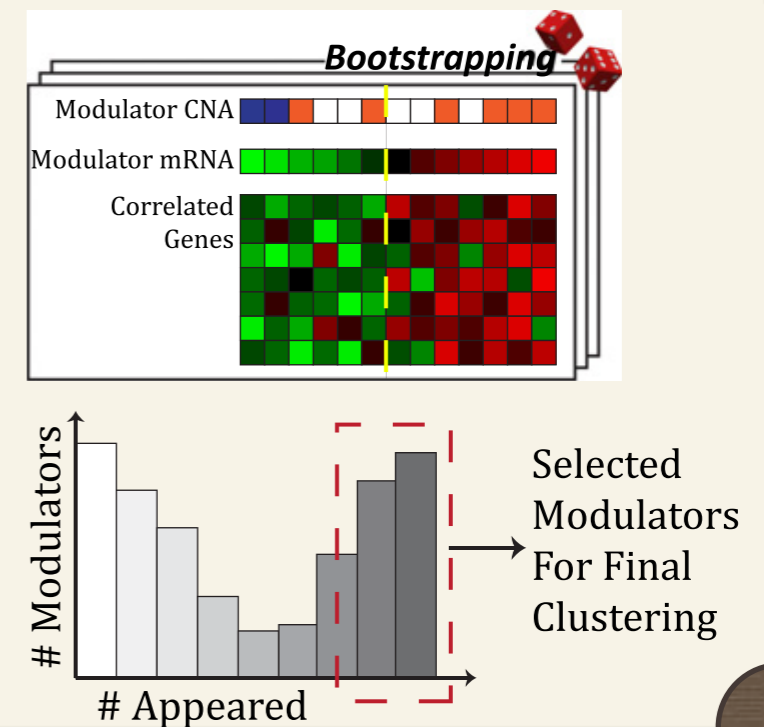
Single Modulator

~ Motivation:

- ~ to construct an initial model by associating each target gene with the single driver gene that fits it best

~ Method:

- ~ Based on Module Networks (Segal et al. 2003, 2005)
- ~ For each candidate driver gene (347):
 - ~ k-means clustering ($k=2$): normal vs amplified/deleted samples, to identify threshold



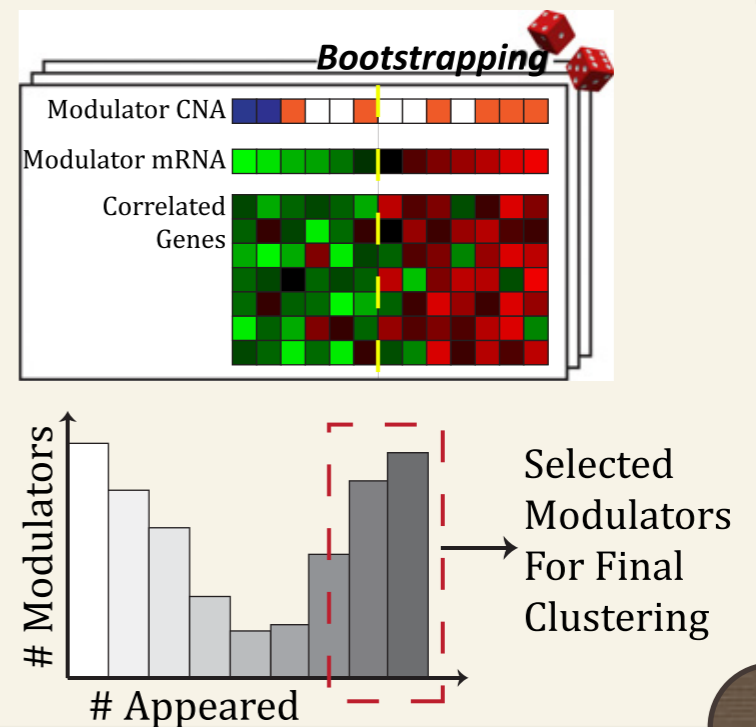
Single Modulator

~ Motivation:

- ~ to construct an initial model by associating each target gene with the single driver gene that fits it best

~ Method:

- ~ Based on Module Networks (Segal et al. 2003, 2005)
- ~ For each candidate driver gene (347):
 - ~ k-means clustering ($k=2$): normal vs amplified/deleted samples, to identify threshold
 - ~ each target gene associated to two sets: tumors where driver is below thresh and tumors where driver is above thresh, according to Normal Gamma



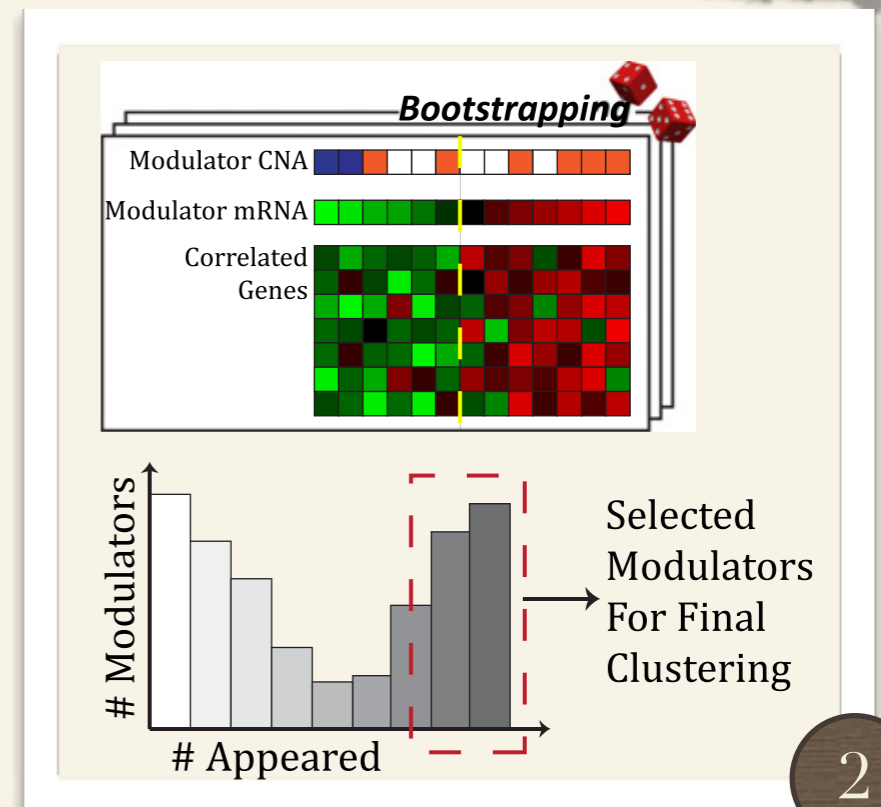
Single Modulator

~ Motivation:

- ~ to construct an initial model by associating each target gene with the single driver gene that fits it best

~ Method:

- ~ Based on Module Networks (Segal et al. 2003, 2005)
- ~ For each candidate driver gene (347):
 - ~ k-means clustering ($k=2$): normal vs amplified/deleted samples, to identify threshold
 - ~ each target gene associated to two sets: tumors where driver is below thresh and tumors where driver is above thresh, according to Normal Gamma
- ~ assignment of each gene to the *highest scoring modulator*



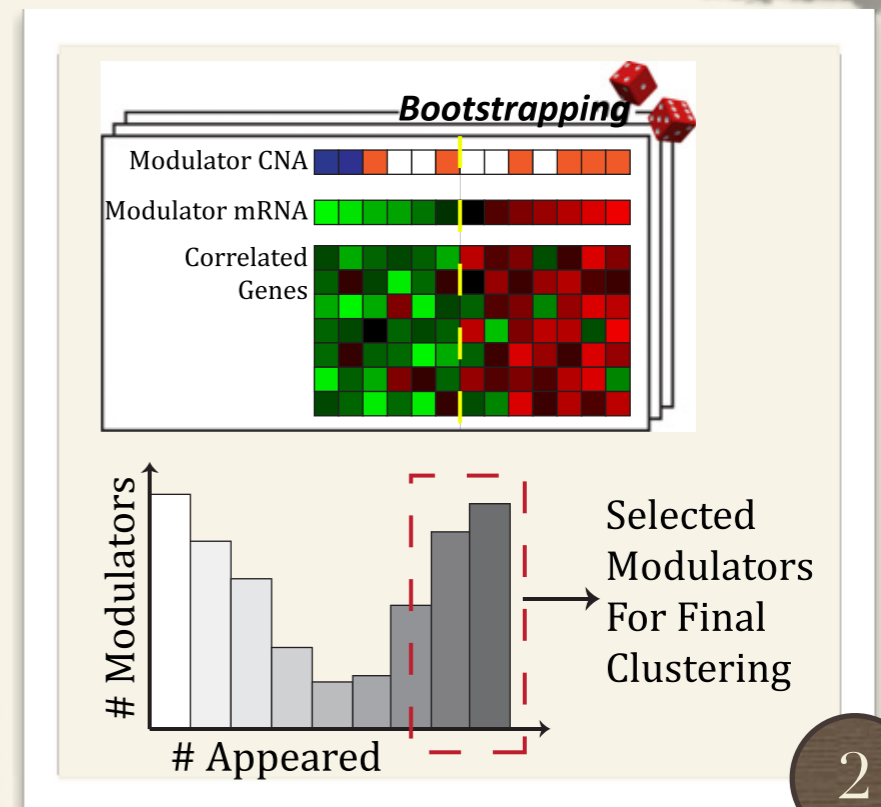
Single Modulator

~ Motivation:

- ~ to construct an initial model by associating each target gene with the single driver gene that fits it best

~ Method:

- ~ Based on Module Networks (Segal et al. 2003, 2005)
- ~ For each candidate driver gene (347):
 - ~ k-means clustering ($k=2$): normal vs amplified/deleted samples, to identify threshold
 - ~ each target gene associated to two sets: tumors where driver is below thresh and tumors where driver is above thresh, according to Normal Gamma
- ~ assignment of each gene to the *highest scoring modulator*
- ~ permutation to ensure robustness of modulators



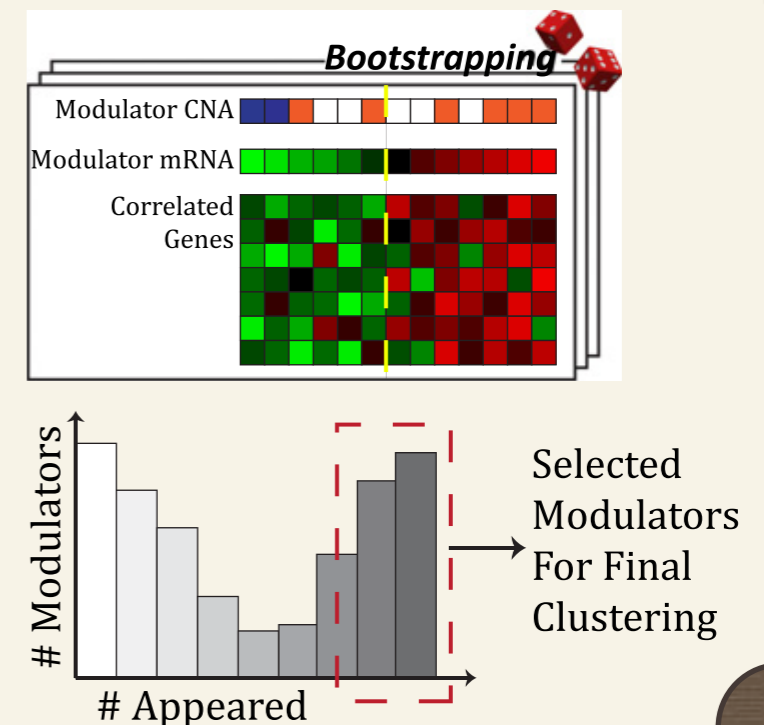
Single Modulator

~ Motivation:

- ~ to construct an initial model by associating each target gene with the single driver gene that fits it best

~ Method:

- ~ Based on Module Networks (Segal et al. 2003, 2005)
- ~ For each candidate driver gene (347):
 - ~ k-means clustering (k=2): normal vs amplified/deleted samples, to identify threshold
 - ~ each target gene associated to two sets: tumors where driver is below thresh and tumors where driver is above thresh, according to Normal Gamma
- ~ assignment of each gene to the *highest scoring modulator*
- ~ permutation to ensure robustness of modulators



2

Result

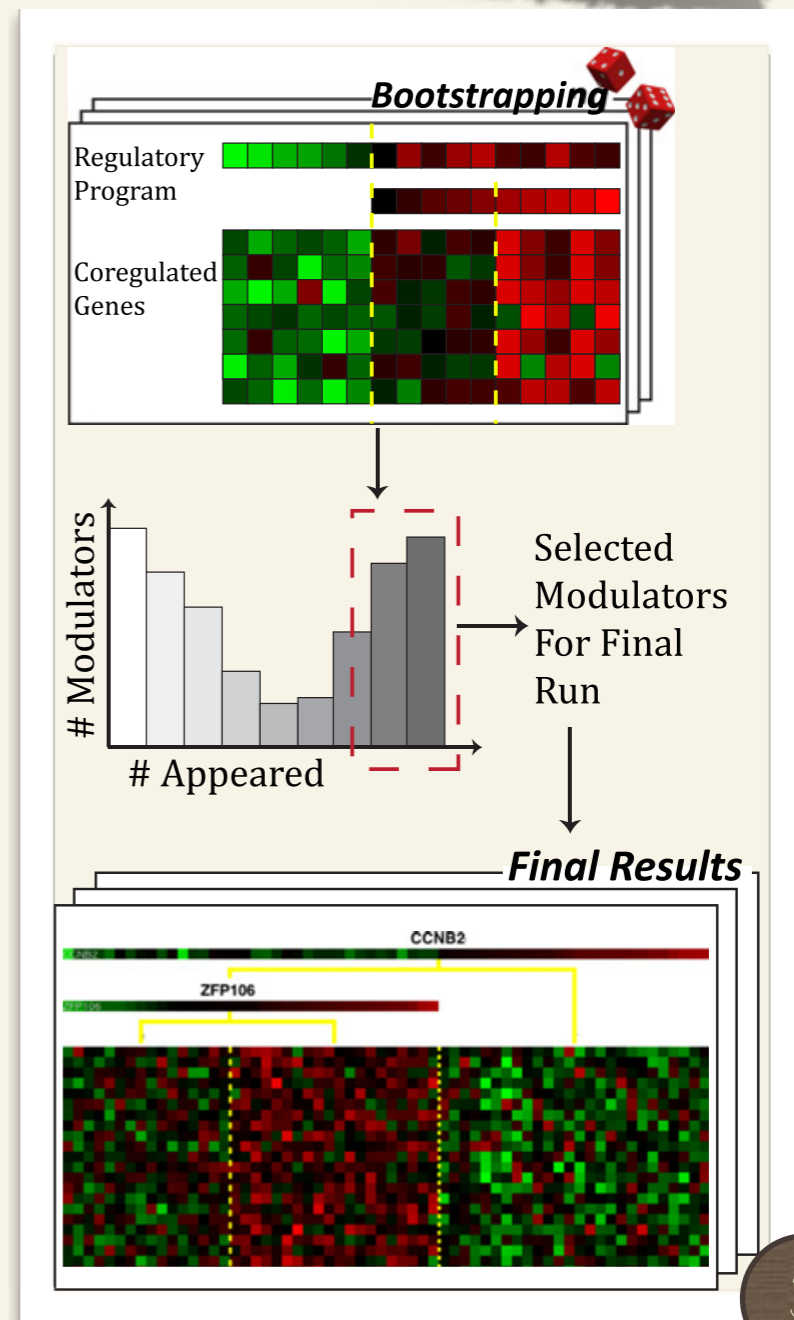
- * 78 modulators
- * each module > 20 genes
- * 4018 genes

Network learning

~ Motivation:

~ refinement of the modulators and modules by allowing more than one modulator for each module (regulation program)

~ Method:



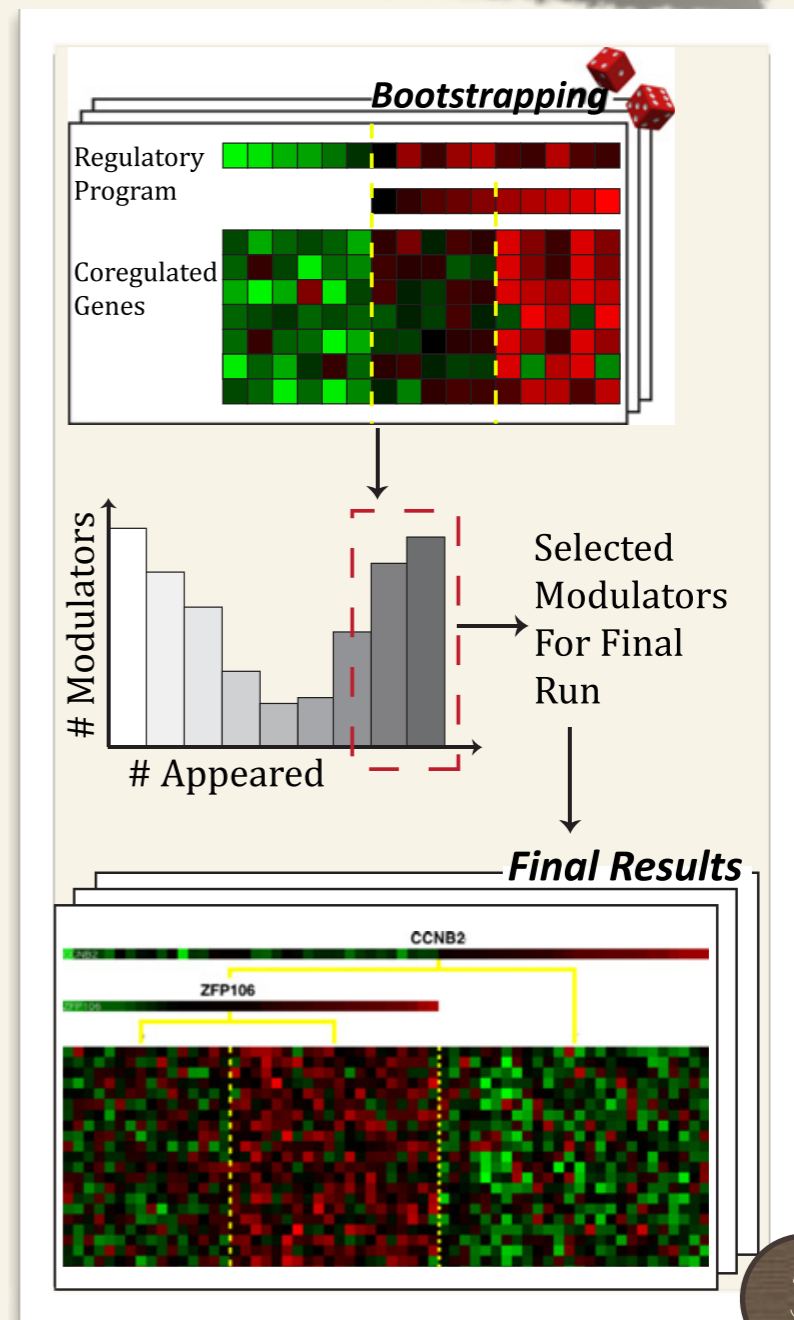
Network learning

~ Motivation:

~ refinement of the modulators and modules by allowing more than one modulator for each module (regulation program)

~ Method:

~ It performs two tasks, iteratively:



Network learning

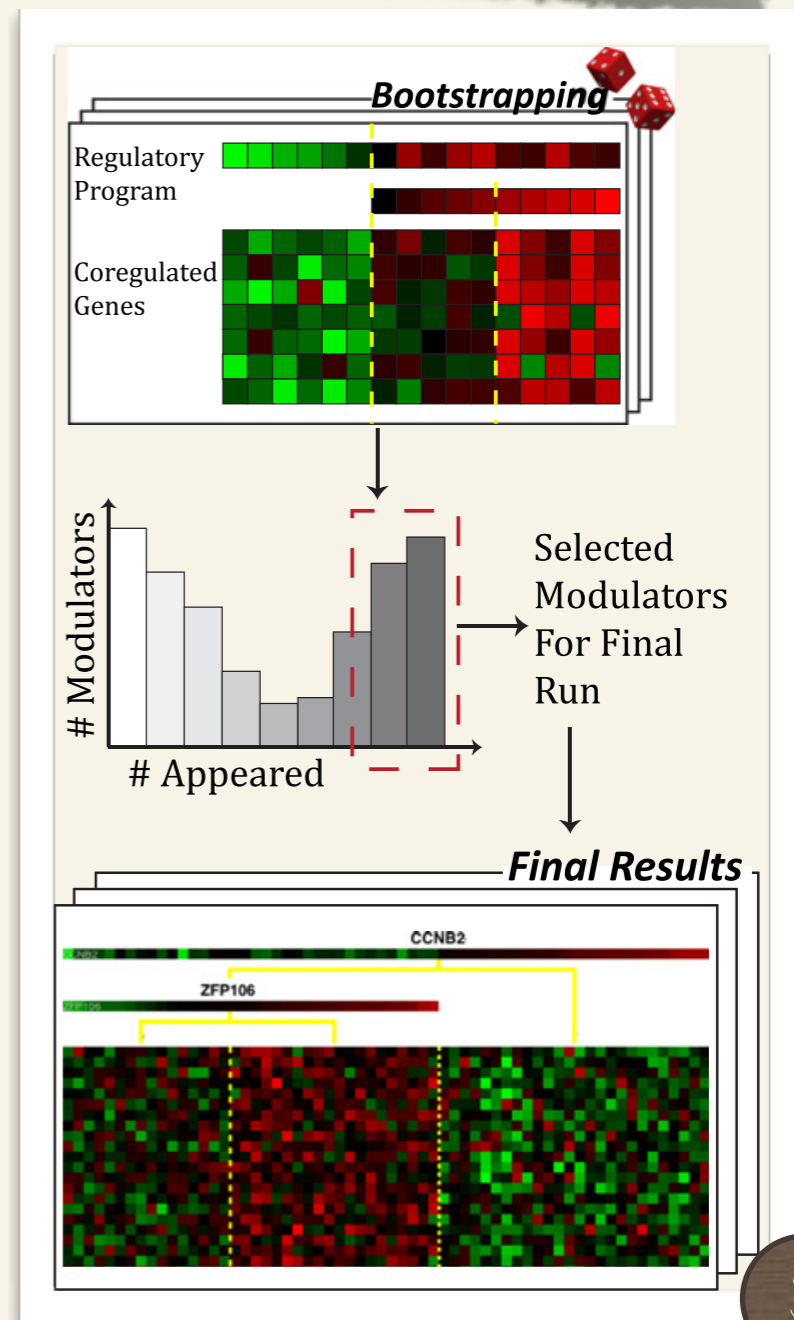
~ Motivation:

~ refinement of the modulators and modules by allowing more than one modulator for each module (regulation program)

~ Method:

~ It performs two tasks, iteratively:

1. learns the regulation program of each module



Network learning

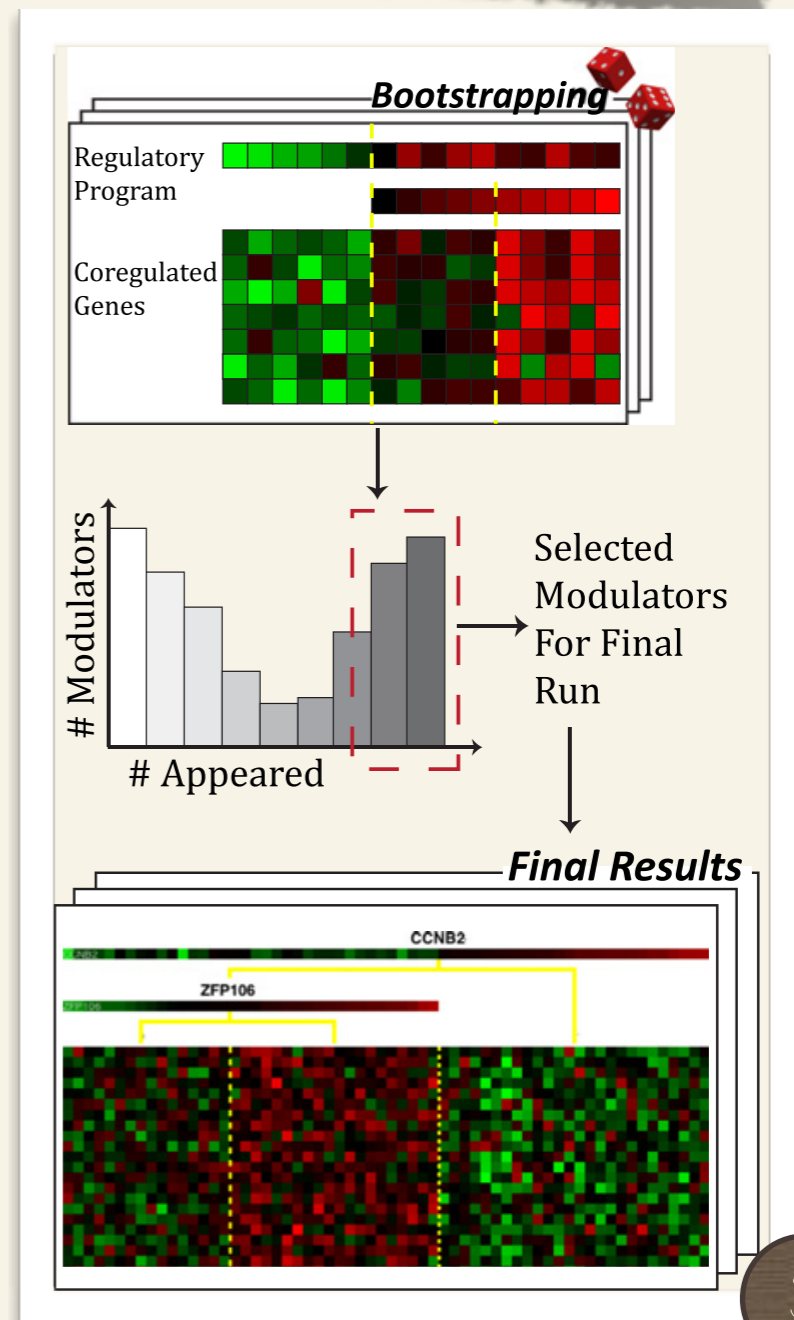
~ Motivation:

~ refinement of the modulators and modules by allowing more than one modulator for each module (regulation program)

~ Method:

~ It performs two tasks, iteratively:

1. learns the regulation program of each module
2. re-assigns the genes to the modules



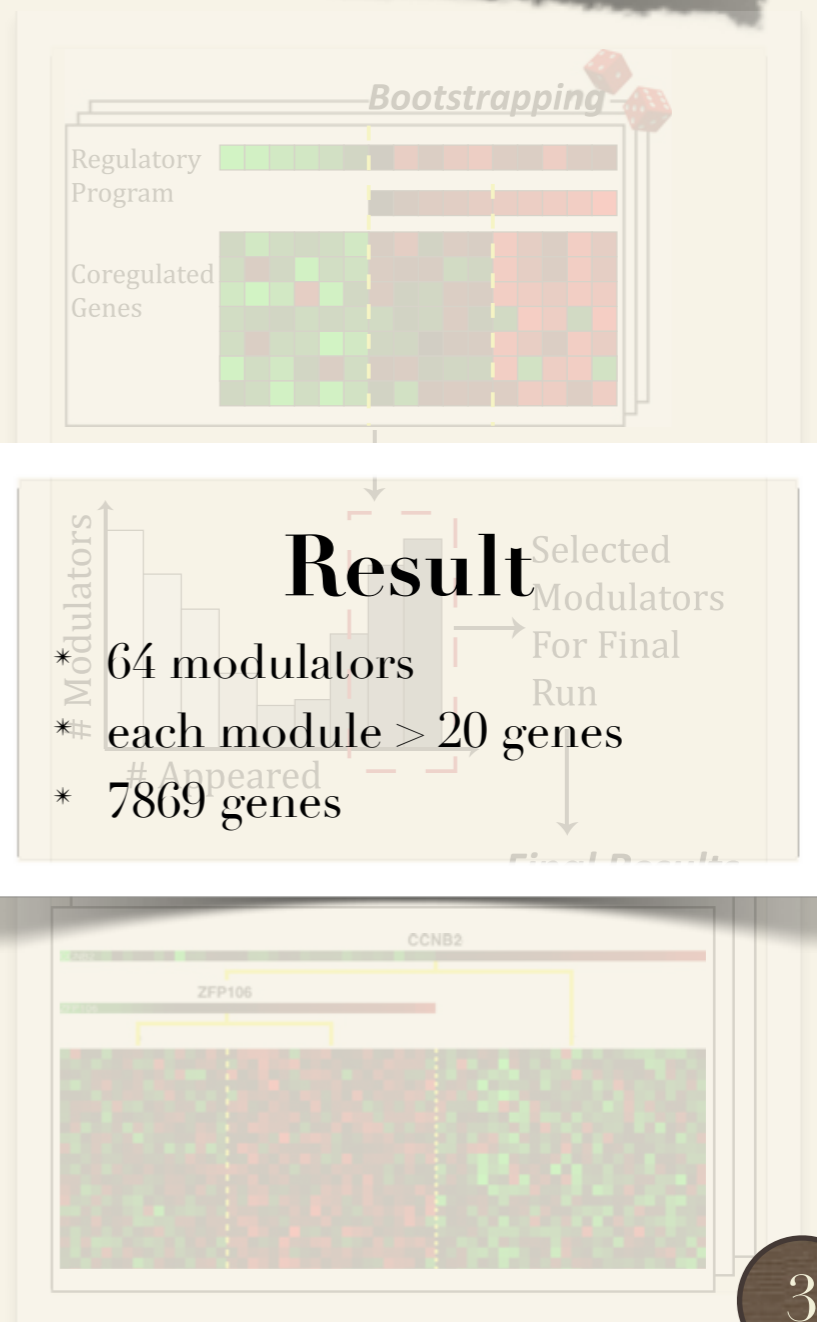
Network learning

~ Motivation:

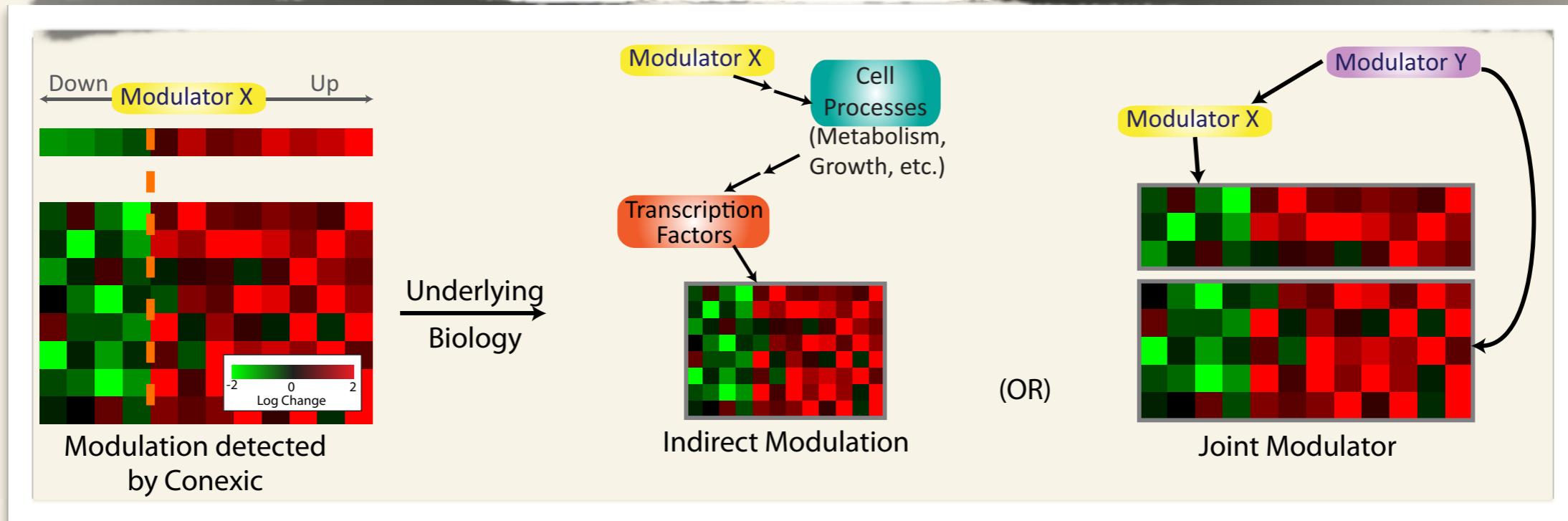
- ~ refinement of the modulators and modules by allowing more than one modulator for each module (regulation program)

~ Method:

- ~ It performs two tasks, iteratively:
 1. learns the regulation program of each module
 2. re-assigns the genes to the modules



Modules: Modulators & Genes

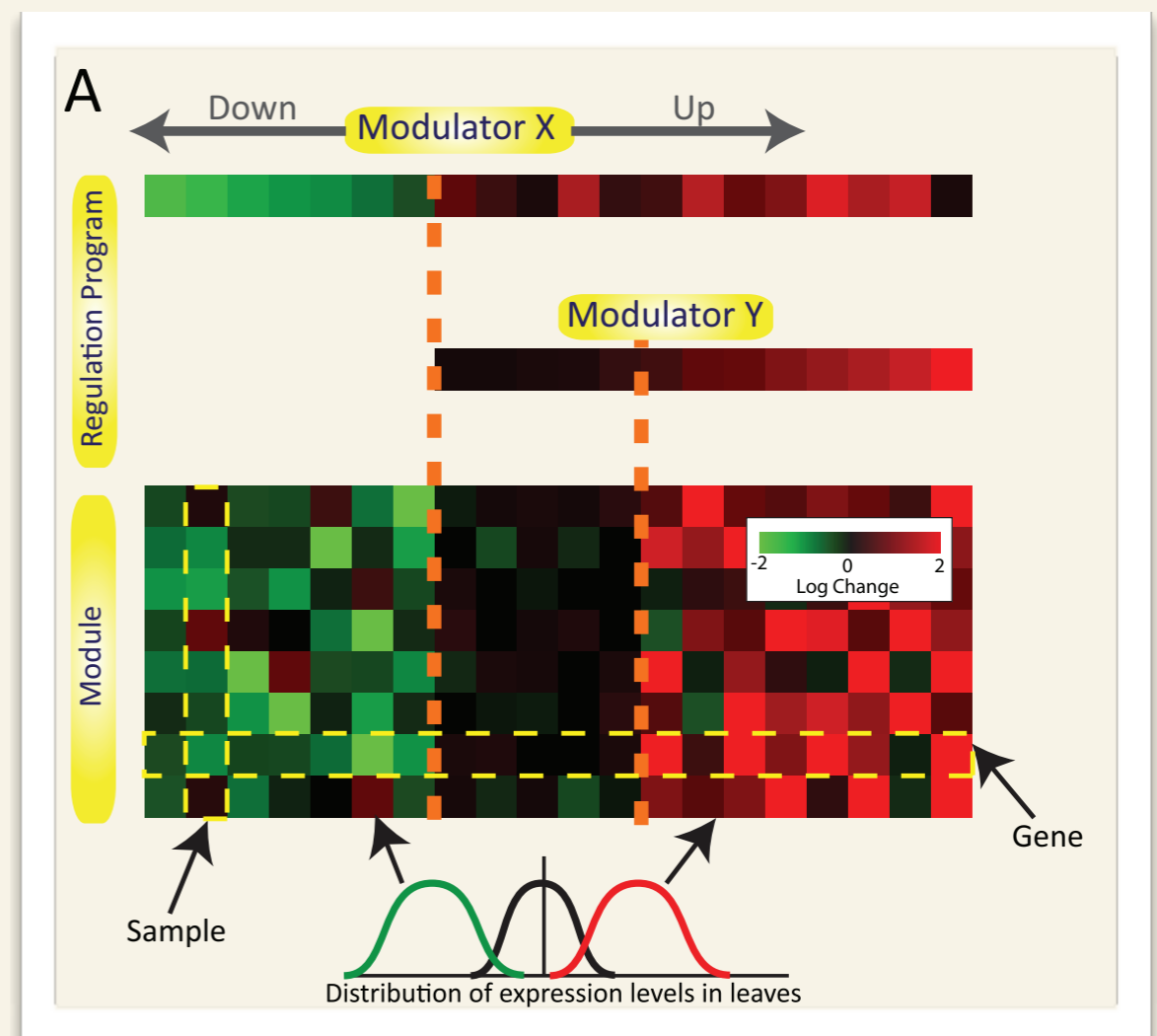


- ~ The influence of a modulator on genes can be indirect (through processes) or joint with another modulator in the module
- ~ This helps define the altered cellular physiology leading to the malignant phenotype

Learning the regulation program

Given a set of modules, a regulation program is learned for each module:

1. all candidate drivers are considered (428)
2. Multiple splits are allowed (penalty for multiple splits)
3. a tree is generated recursively:
 - 3.1. the best driver-split is selected according to the score
 - 3.2. outlier-removal test and linear influence test as robustness criteria
4. all genes are reassigned to the modules, and moved to the new one if the score improves



NormalGamma score

NormalGamma(Leaf, λ, α) :

$$N = \text{Size}(\text{Leaf})$$

$$\beta = \text{Max}\left(1, \frac{\lambda * (\alpha - 2)}{\lambda + 1}\right)$$

$$\beta^+ = \beta + \frac{\text{Var}(\text{Leaf}) * N}{2} + N * \lambda * \frac{\overline{\text{Leaf}^2}}{2 * (N + \lambda)}$$

$$\alpha^+ = \alpha + \frac{N}{2}$$

$$\text{Score} = -N * \ln(\sqrt{2\pi}) + \frac{\ln(\frac{\lambda}{\lambda + N})}{2} + \ln(\Gamma(\alpha^+)) - \ln(\Gamma(\alpha)) + \alpha * \ln(\beta) - \alpha^+ * \ln(\beta^+)$$

NormalGamma score

$NormalGamma(Leaf, \lambda, \alpha) :$

$$N = Size(Leaf)$$

$$\beta = Max\left(1, \frac{\lambda * (\alpha - 2)}{\lambda + 1}\right)$$

$$\beta^+ = \beta + \frac{Var(Leaf) * N}{2} + N * \lambda * \frac{\overline{Leaf^2}}{2 * (N + \lambda)}$$

$$\alpha^+ = \alpha + \frac{N}{2}$$

$$Score = -N * \ln(\sqrt{2\pi}) + \frac{\ln(\frac{\lambda}{\lambda + N})}{2} + \ln(\Gamma(\alpha^+)) - \ln(\Gamma(\alpha)) + \alpha * \ln(\beta) - \alpha^+ * \ln(\beta^+)$$

~ Leaf: vector of gene expression values

NormalGamma score

NormalGamma(Leaf, λ, α) :

$$N = \text{Size}(\text{Leaf})$$

$$\beta = \text{Max}\left(1, \frac{\lambda^*(\alpha - 2)}{\lambda + 1}\right)$$

$$\beta^+ = \beta + \frac{\text{Var}(\text{Leaf}) * N}{2} + N * \lambda * \frac{\overline{\text{Leaf}^2}}{2 * (N + \lambda)}$$

$$\alpha^+ = \alpha + \frac{N}{2}$$

$$\text{Score} = -N * \ln(\sqrt{2\pi}) + \frac{\ln(\frac{\lambda}{\lambda + N})}{2} + \ln(\Gamma(\alpha^+)) - \ln(\Gamma(\alpha)) + \alpha * \ln(\beta) - \alpha^+ * \ln(\beta^+)$$

- ~ Leaf: vector of gene expression values
- ~ lambda/alpha/beta: parameters (estimated by cross-validation)

NormalGamma score

$NormalGamma(Leaf, \lambda, \alpha) :$

$$N = Size(Leaf)$$

$$\beta = Max\left(1, \frac{\lambda * (\alpha - 2)}{\lambda + 1}\right)$$

$$\beta^+ = \beta + \frac{Var(Leaf) * N}{2} + N * \lambda * \frac{\overline{Leaf^2}}{2 * (N + \lambda)}$$

$$\alpha^+ = \alpha + \frac{N}{2}$$

$$Score = -N * \ln(\sqrt{2\pi}) + \frac{\ln(\frac{\lambda}{\lambda + N})}{2} + \ln(\Gamma(\alpha^+)) - \ln(\Gamma(\alpha)) + \alpha * \ln(\beta) - \alpha^+ * \ln(\beta^+)$$

- ~ Leaf: vector of gene expression values
- ~ lambda/alpha/beta: parameters (estimated by cross-validation)
- ~ Comparison with the entire data

$$NormalGamma(Left_Leaf) + NormalGamma(Right_Leaf) > = NormalGamma(Entire_data) + Penalty$$

NormalGamma score

$NormalGamma(Leaf, \lambda, \alpha) :$

$$N = Size(Leaf)$$

$$\beta = Max\left(1, \frac{\lambda^*(\alpha - 2)}{\lambda + 1}\right)$$

$$\beta^+ = \beta + \frac{Var(Leaf)*N}{2} + N*\lambda*\frac{\overline{Leaf^2}}{2*(N+\lambda)}$$

$$\alpha^+ = \alpha + \frac{N}{2}$$

$$Score = -N*\ln(\sqrt{2\pi}) + \frac{\ln(\frac{\lambda}{\lambda+N})}{2} + \ln(\Gamma(\alpha^+)) - \ln(\Gamma(\alpha)) + \alpha*\ln(\beta) - \alpha^+*\ln(\beta^+)$$

- ~ Leaf: vector of gene expression values
- ~ lambda/alpha/beta: parameters (estimated by cross-validation)
- ~ Comparison with the entire data

$$NormalGamma(Left_Leaf) + NormalGamma(Right_Leaf) > = NormalGamma(Entire_data) + Penalty$$

Penalty based on #leaves in regulation program (module specific) and #modulators (network-wide penalty)

Results

Gene Symbol	Pathway	Band	Genes in Region	Validation p-value
MITF	Melanoma	3p14.2-p14.1	1	<10 ⁻⁶
TBC1D16	Vesicular Trafficking	17q25.3	24	<10 ⁻⁶
ZFP106	Insulin/Ras	15q15.1	7	<10 ⁻⁶
DIXDC1	Wnt/JNK/PI3K	11q23.1	17	0.0001
OIP5	Cell Cycle	15q15.1	13	<10 ⁻⁶
TTBK2		15q15.2	7	0.0383
TRAF3	NFkappaB/JNK	14q32.32	19	0.0121
RAB27A	Vesicular Trafficking	15q15-q21.1	33	<10 ⁻⁶
C12orf35		12p11.21	45	<10 ⁻⁶
WBP2		17q25	92	0.0275
MOCS3		20q13.13	16	<10 ⁻⁶
NDUFB2		7q34	10	<10 ⁻⁶
ST6GALNAC2		17q25.1	92	<10 ⁻⁶
GRB2	EGFR/Ras	17q24-q25	92	0.1373
ECM1		1q21	55	0.0083
KCNG1		20q13	16	0.202
DPM1		20q13.13	16	0.097
PFKP	Metabolism	10p15.3-p15.2	3	0.0801
KLF6	Cell cycle, c-JUN (JNK)	10p15	3	<10 ⁻⁶
TIMM8B	Mitochondria	11q23.1-q23.2	17	0.7622
PI4KB		1q21	55	0.0003
PSMB4		1q21	55	0.0005
VPS72		1q21	55	<10 ⁻⁶
TARS2		1q21.3	55	0.0001
MNS1		15q21.3	33	0.0908
TDRD3	RNA processing	13q21.2	203	<10 ⁻⁶
CCNB2	Cell Cycle	15q22.2	33	<10 ⁻⁶
EIF5	Cell Cycle	14q32.32	19	0.1096
RAB7A	Vesicular Trafficking	3q21.3	16	<10 ⁻⁶
PIK3CB	PI3K signaling	3q22.3	15	<10 ⁻⁶

Results

Gene Symbol	Pathway	Band	Genes in Region	Validation p-value
MITF	Melanoma	3p14.2-p14.1	1	<10 ⁻⁶
TBC1D16	Vesicular Trafficking	17q25.3	24	<10 ⁻⁶
ZFP106	Insulin/Ras	15q15.1	7	<10 ⁻⁶
DIXDC1	Wnt/JNK/PI3K	11q23.1	17	0.0001
OIP5	Cell Cycle	15q15.1	13	<10 ⁻⁶
TTBK2		15q15.2	7	0.0383
TRAF3	NFkappaB/JNK	14q32.32	19	0.0121
RAB27A	Vesicular Trafficking	15q15-q21.1	33	<10 ⁻⁶
C12orf35		12p11.21	45	<10 ⁻⁶
WBP2		17q25	92	0.0275
MOCS3		20q13.13	16	<10 ⁻⁶
NDUFB2		7q34	10	<10 ⁻⁶
ST6GALNAC2		17q25.1	92	<10 ⁻⁶
GRB2	EGFR/Ras	17q24-q25	92	0.1373
ECM1		1q21	55	0.0083
KCNG1		20q13	16	0.202
DPM1		20q13.13	16	0.097
PFKP	Metabolism	10p15.3-p15.2	3	0.0801
KLF6	Cell cycle, c-JUN (JNK)	10p15	3	<10 ⁻⁶
TIMM8B	Mitochondria	11q23.1-q23.2	17	0.7622
PI4KB		1q21	55	0.0003
PSMB4		1q21	55	0.0005
VPS72		1q21	55	<10 ⁻⁶
TARS2		1q21.3	55	0.0001
MNS1		15q21.3	33	0.0908
TDRD3	RNA processing	13q21.2	203	<10 ⁻⁶
CCNB2	Cell Cycle	15q22.2	33	<10 ⁻⁶
EIF5	Cell Cycle	14q32.32	19	0.1096
RAB7A	Vesicular Trafficking	3q21.3	16	<10 ⁻⁶
PIK3CB	PI3K signaling	3q22.3	15	<10 ⁻⁶

Top 30 modulators include 10 known oncogenes and tumor suppressors

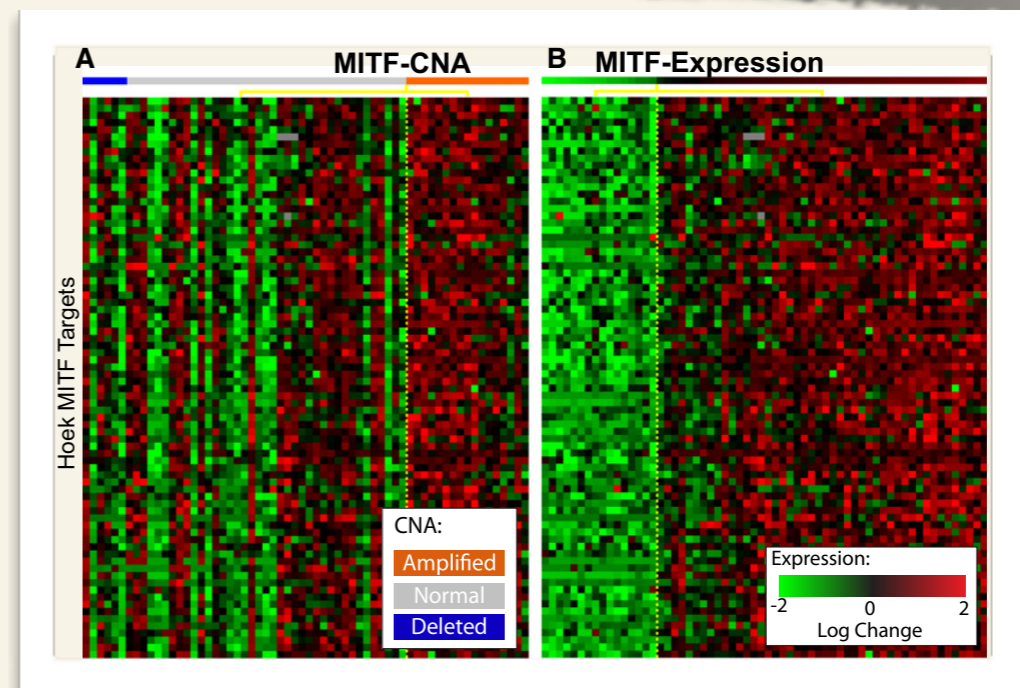
Results

Gene Symbol	Pathway	Band	Genes in Region	Validation p-value
MITF	Melanoma	3p14.2-p14.1	1	<10 ⁻⁶
TBC1D16	Vesicular Trafficking	17q25.3	24	<10 ⁻⁶
ZFP106	Insulin/Ras	15q15.1	7	<10 ⁻⁶
DIXDC1	Wnt/JNK/PI3K	11q23.1	17	0.0001
OIP5	Cell Cycle	15q15.1	13	<10 ⁻⁶
TTBK2		15q15.2	7	0.0383
TRAF3	NFkappaB/JNK	14q32.32	19	0.0121
RAB27A	Vesicular Trafficking	15q15-q21.1	33	<10 ⁻⁶
C12orf35		12p11.21	45	<10 ⁻⁶
WBP2		17q25	92	0.0275
MOCS3		20q13.13	16	<10 ⁻⁶
NDUFB2		7q34	10	<10 ⁻⁶
ST6GALNAC2		17q25.1	92	<10 ⁻⁶
GRB2	EGFR/Ras	17q24-q25	92	0.1373
ECM1		1q21	55	0.0083
KCNG1		20q13	16	0.202
DPM1		20q13.13	16	0.097
PFKP	Metabolism	10p15.3-p15.2	3	0.0801
KLF6	Cell cycle, c-JUN (JNK)	10p15	3	<10 ⁻⁶
TIMM8B	Mitochondria	11q23.1-q23.2	17	0.7622
PI4KB		1q21	55	0.0003
PSMB4		1q21	55	0.0005
VPS72		1q21	55	<10 ⁻⁶
TARS2		1q21.3	55	0.0001
MNS1		15q21.3	33	0.0908
TDRD3	RNA processing	13q21.2	203	<10 ⁻⁶
CCNB2	Cell Cycle	15q22.2	33	<10 ⁻⁶
EIF5	Cell Cycle	14q32.32	19	0.1096
RAB7A	Vesicular Trafficking	3q21.3	16	<10 ⁻⁶
PIK3CB	PI3K signaling	3q22.3	15	<10 ⁻⁶

Top 30 modulators include 10 known oncogenes and tumor suppressors

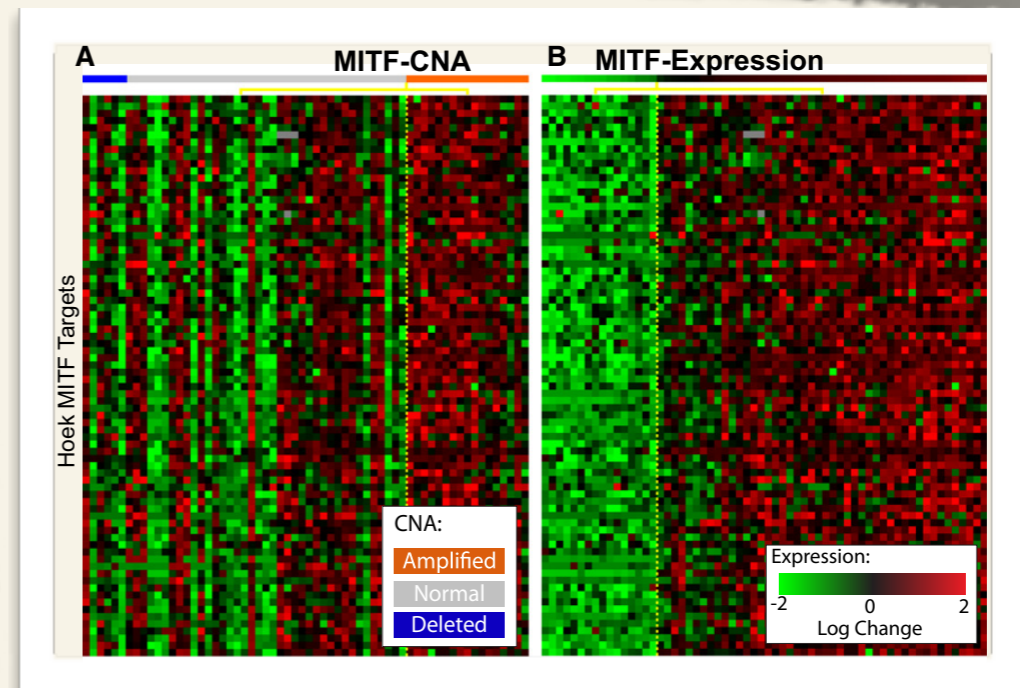
Literature search (using LitVan) suggests PI3K, MAPK, cyclin, RAB (vesicular trafficking)

MITF: a known driver



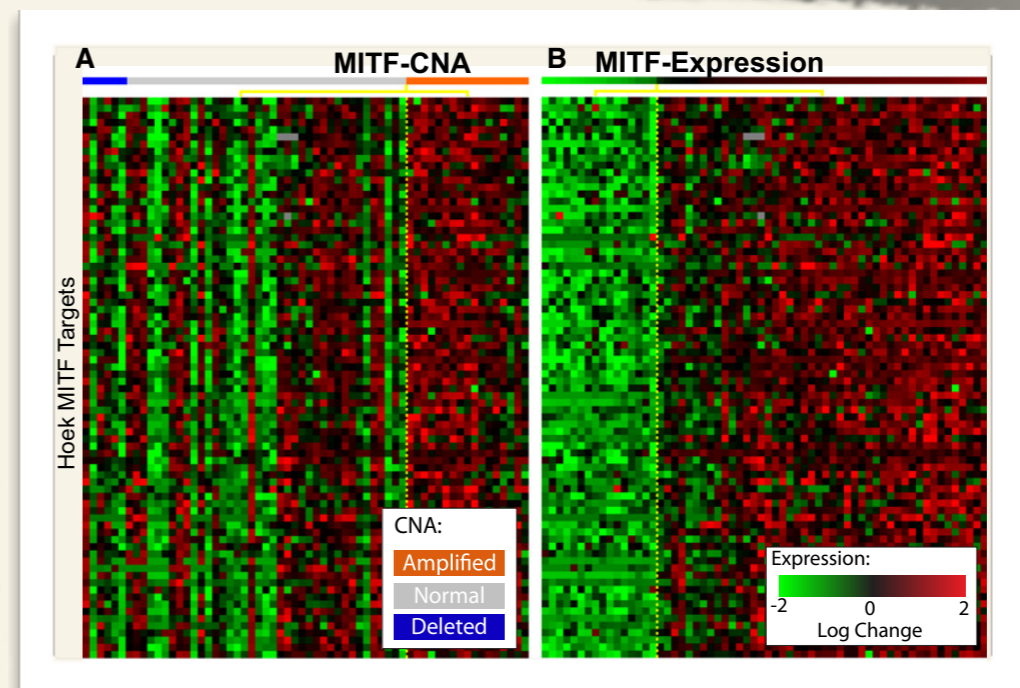
MITF: a known driver

~ Highest-scoring modulator



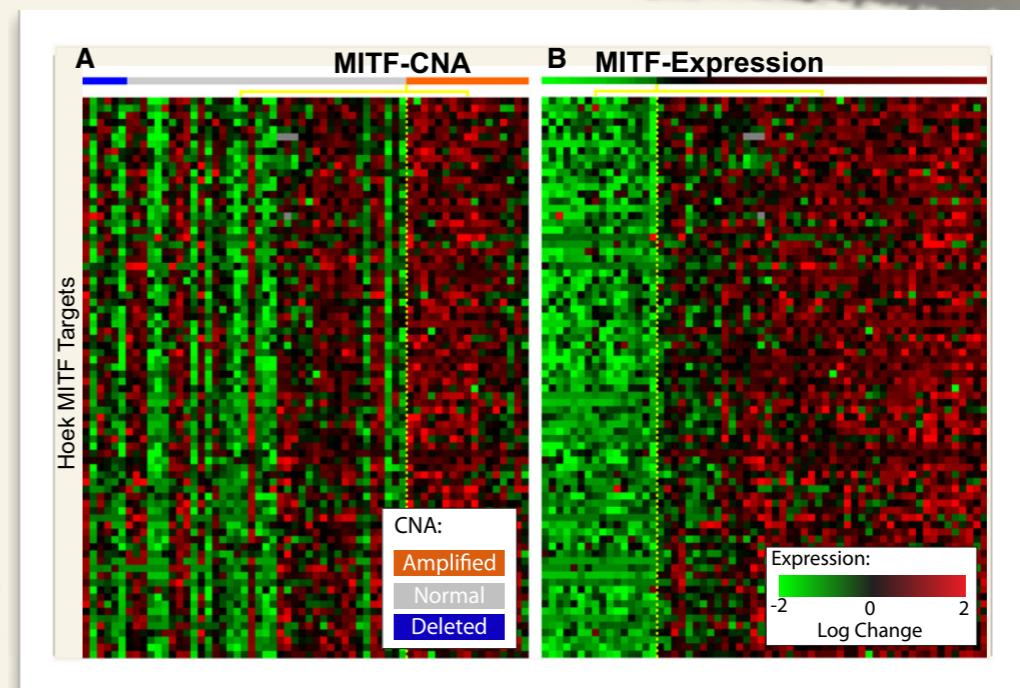
MITF: a known driver

- ~ Highest-scoring modulator
- ~ 45 of previously known 80 targets:
(1 gene \Leftrightarrow 1 module)



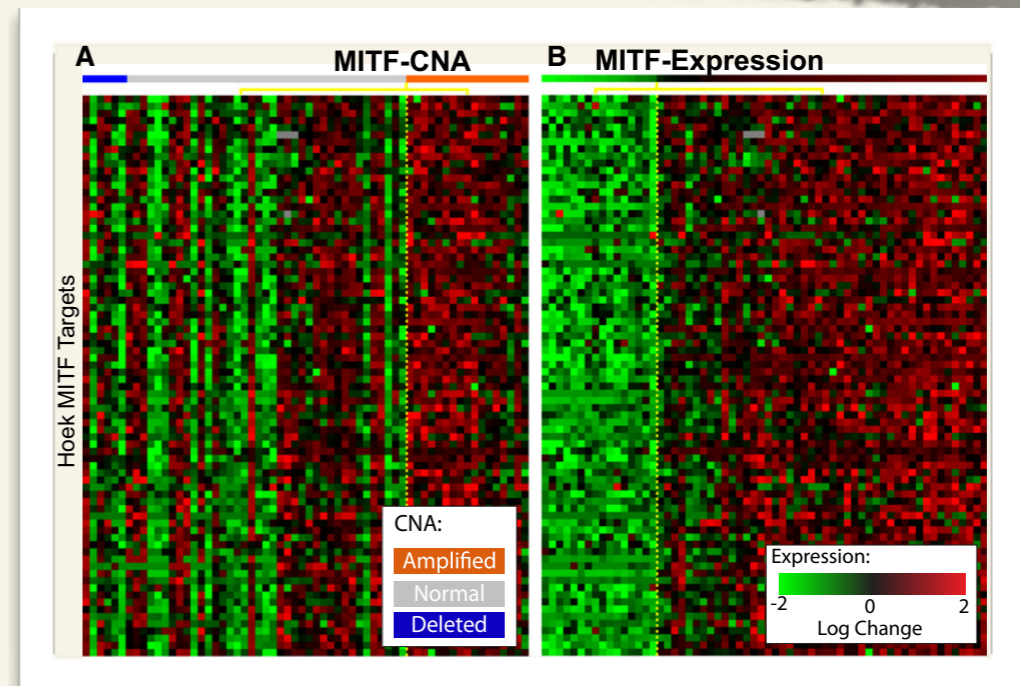
MITF: a known driver

- ~ Highest-scoring modulator
- ~ 45 of previously known 80 targets:
(1 gene \Leftrightarrow 1 module)
- ~ 76 out of 80 (if indirect associations)



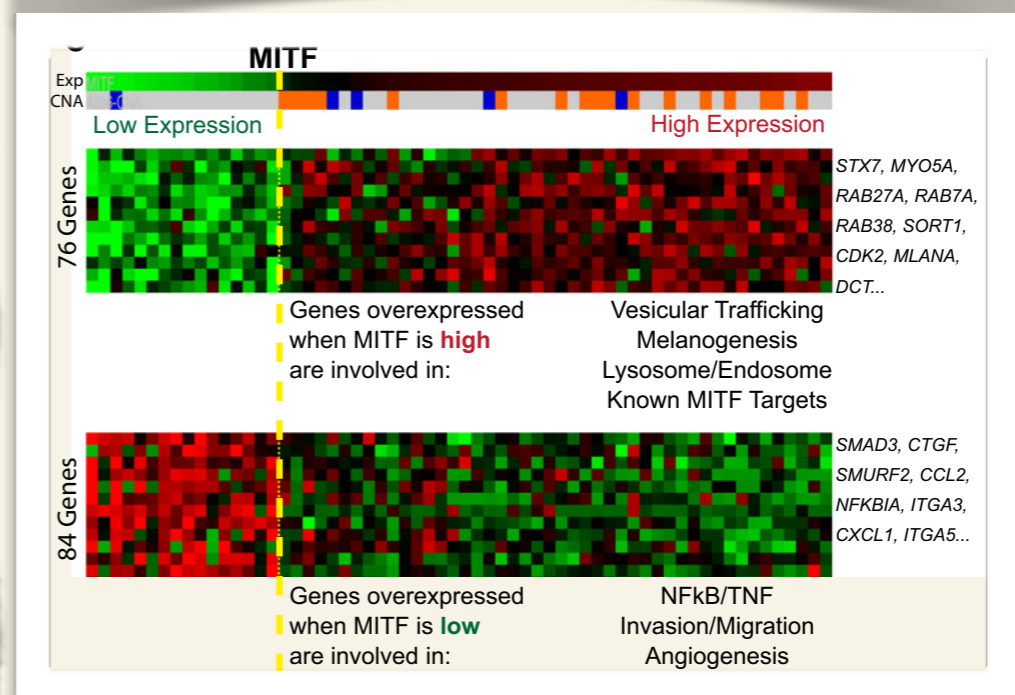
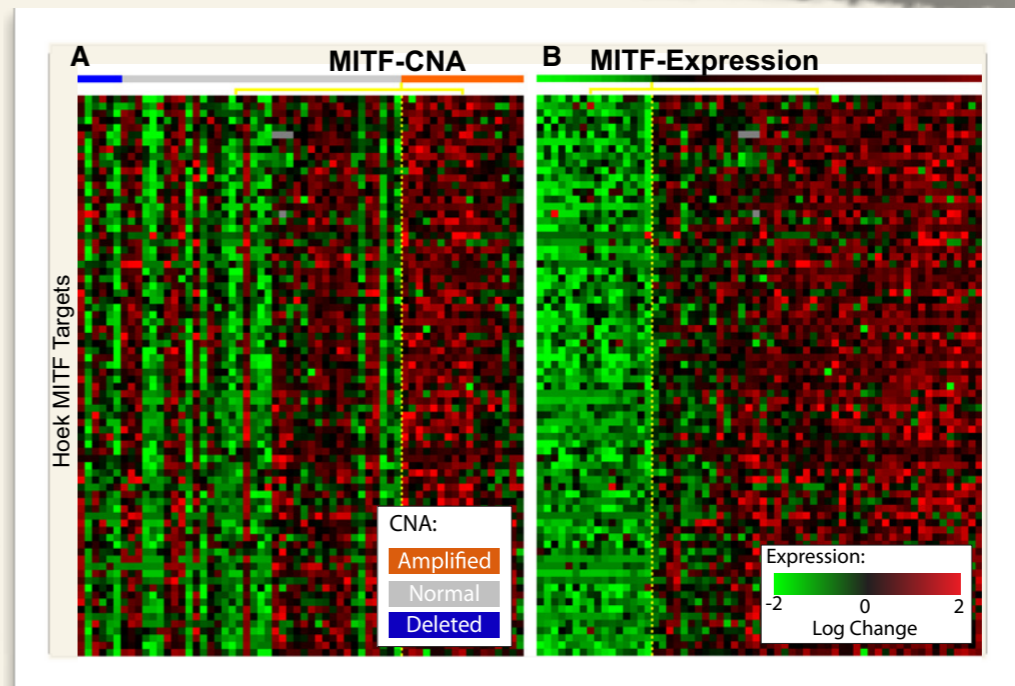
MITF: a known driver

- ~ Highest-scoring modulator
- ~ 45 of previously known 80 targets:
(1 gene \Leftrightarrow 1 module)
- ~ 76 out of 80 (if indirect associations)
- ~ Unidirectional correlation of MITF with CNA

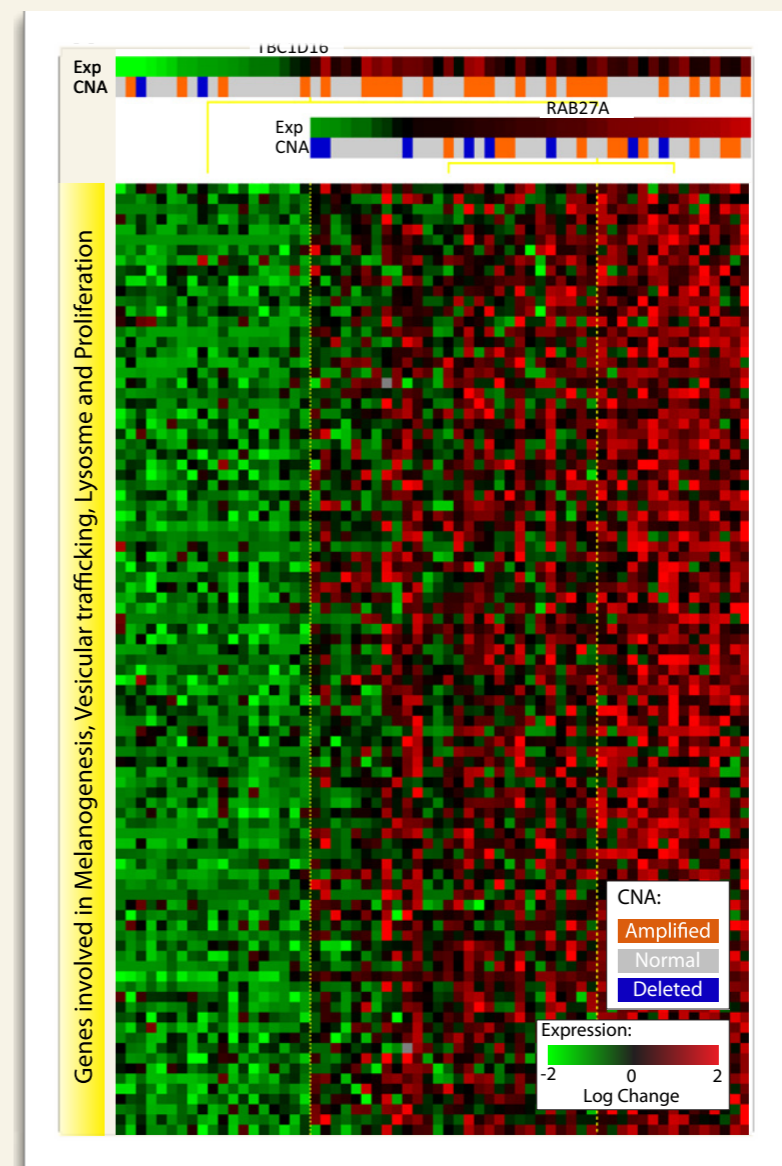


MITF: a known driver

- ~ Highest-scoring modulator
- ~ 45 of previously known 80 targets:
(1 gene \Leftrightarrow 1 module)
- ~ 76 out of 80 (if indirect associations)
- ~ Unidirectional correlation of MITF with CNA
- ~ The module explains MITF double role:
 - * high expression \rightarrow proliferation
 - * low expression \rightarrow invasion

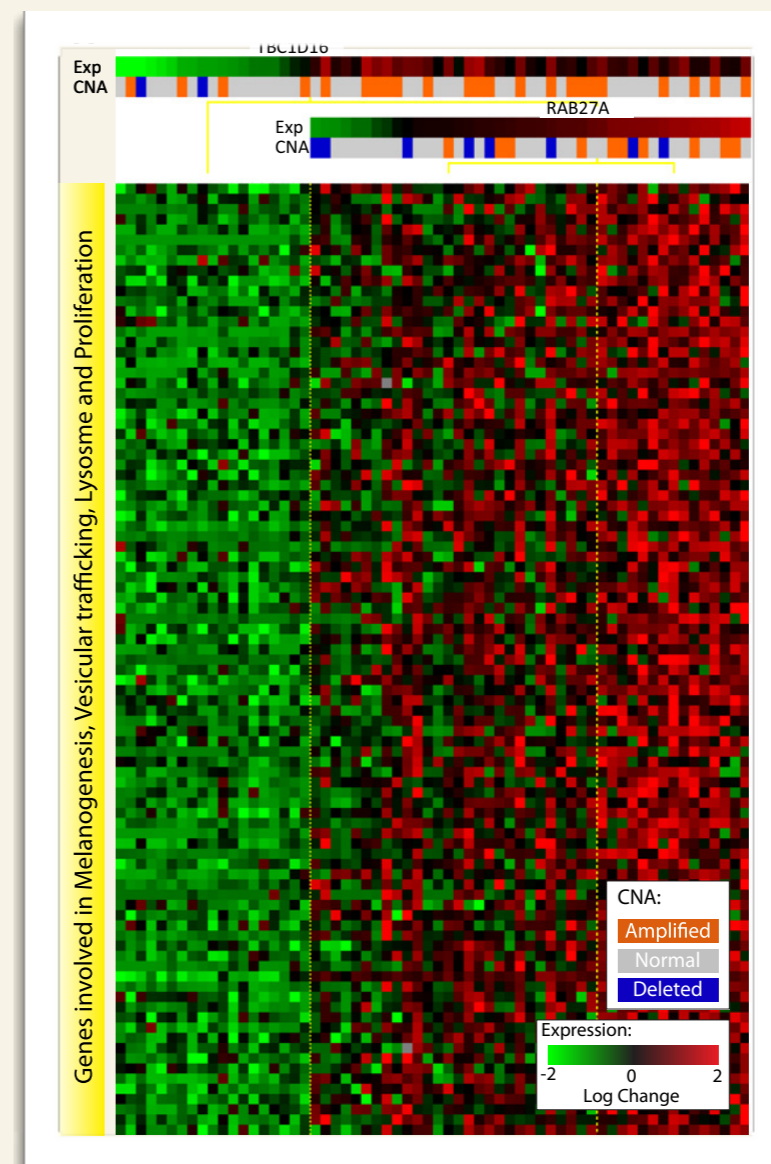


TBC1D16: unknown



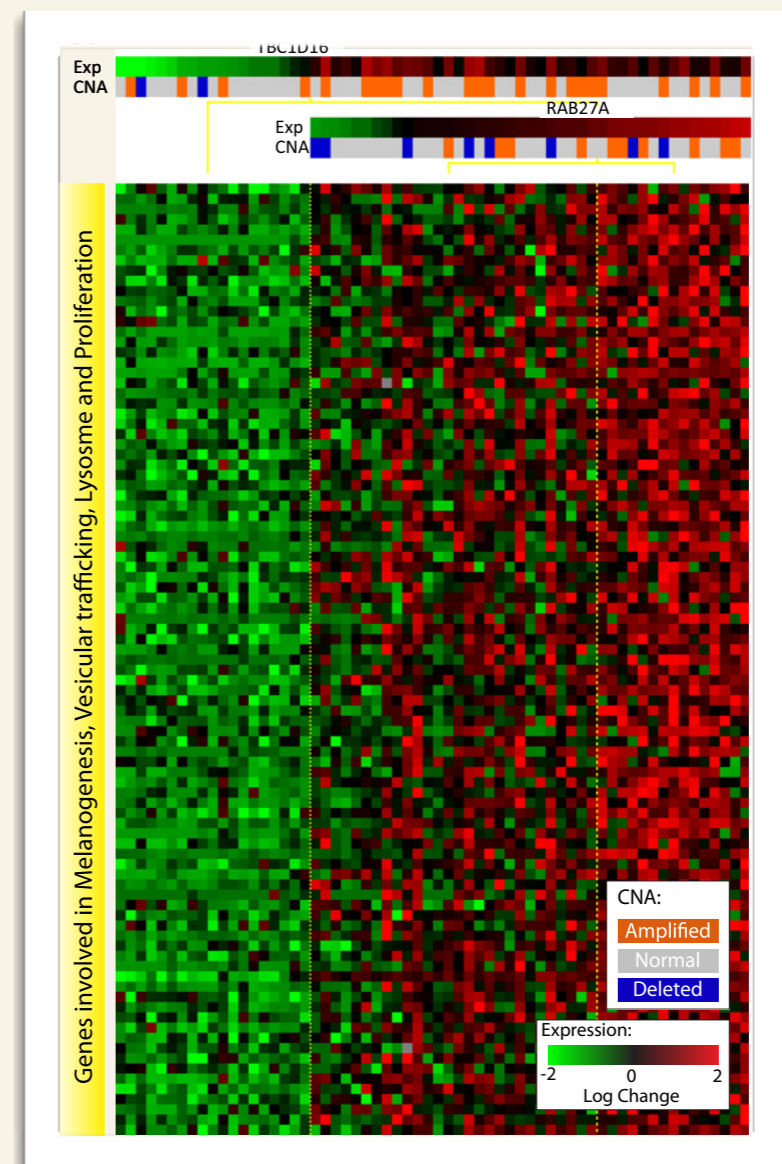
TBC1D16: unknown

- ~ RAB GTPase-activating protein, involved in trafficking and membrane transport



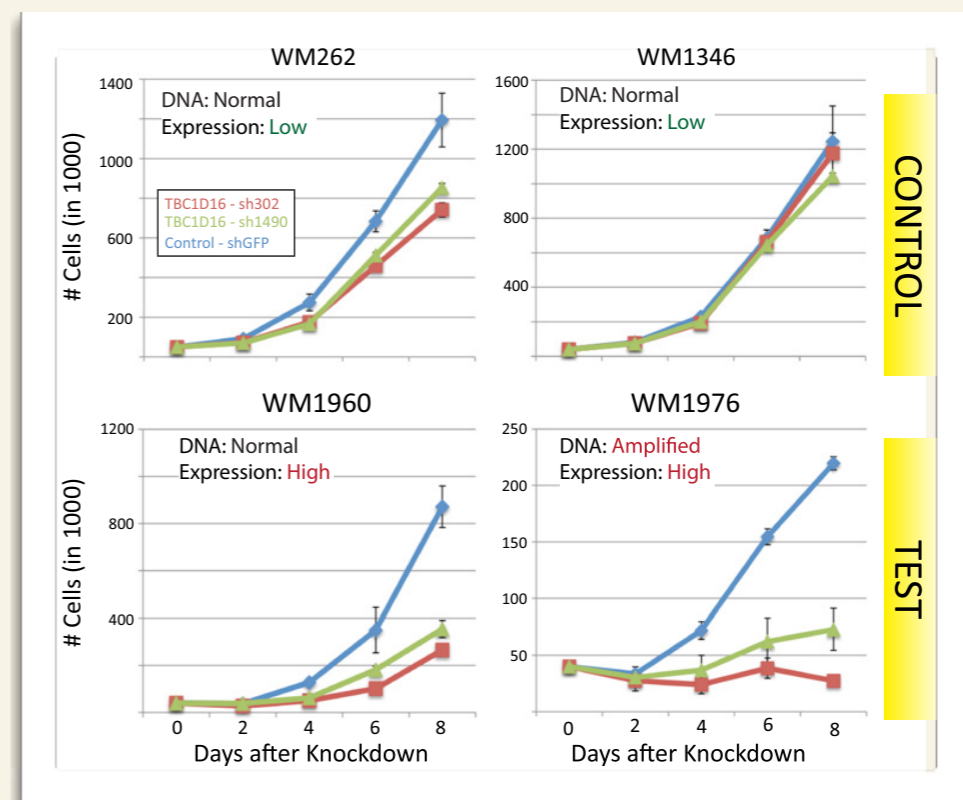
TBC1D16: unknown

- ~ RAB GTPase-activating protein, involved in trafficking and membrane transport
- ~ Involved in several processes, including: melanogenesis, vesicular trafficking, survival/proliferation

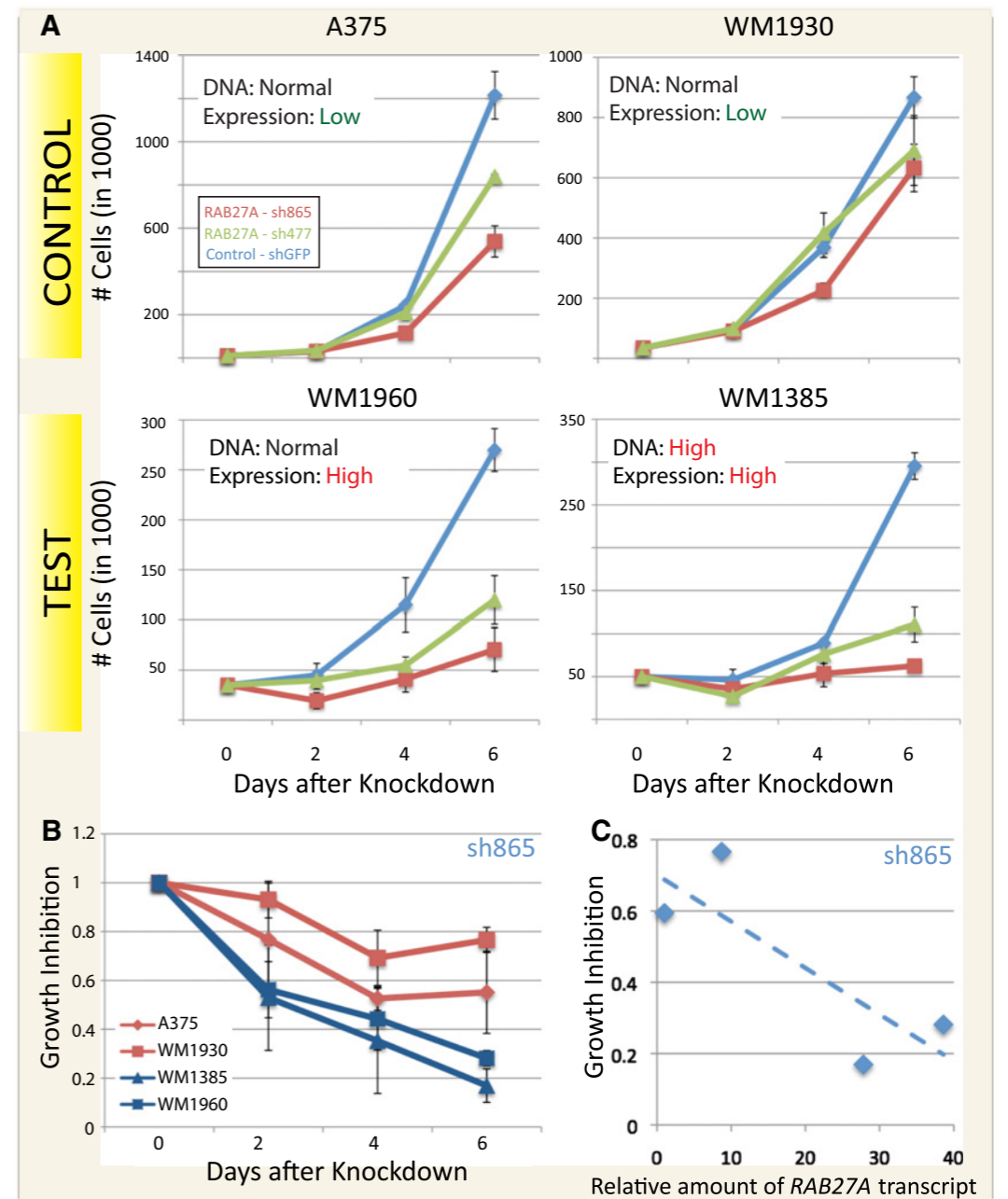


TBC1D16: unknown

- ~ RAB GTPase-activating protein, involved in trafficking and membrane transport
- ~ Involved in several processes, including: melanogenesis, vesicular trafficking, survival/proliferation
- ~ Knock-down experiment to confirm involvement in proliferation

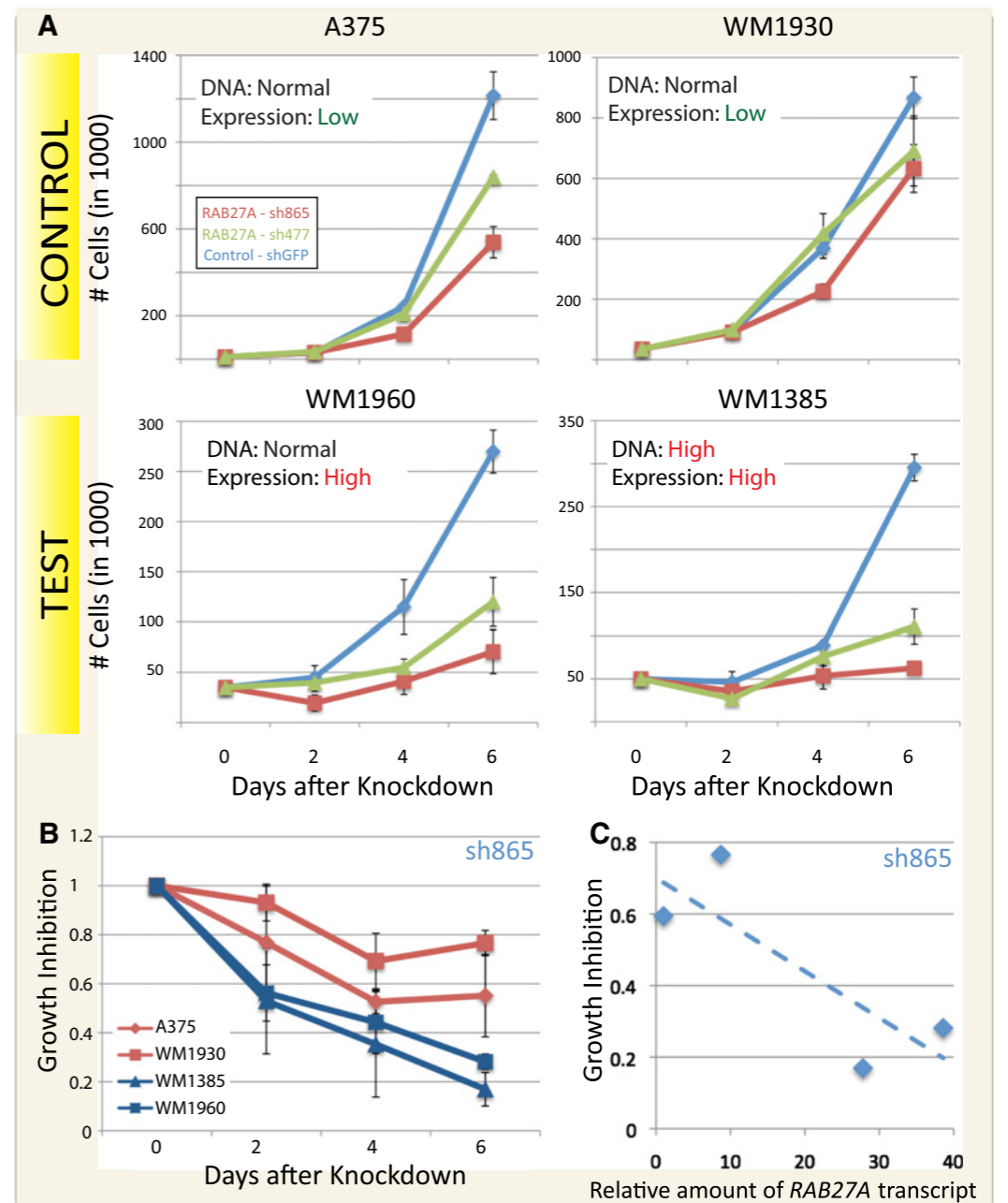


RAB27A



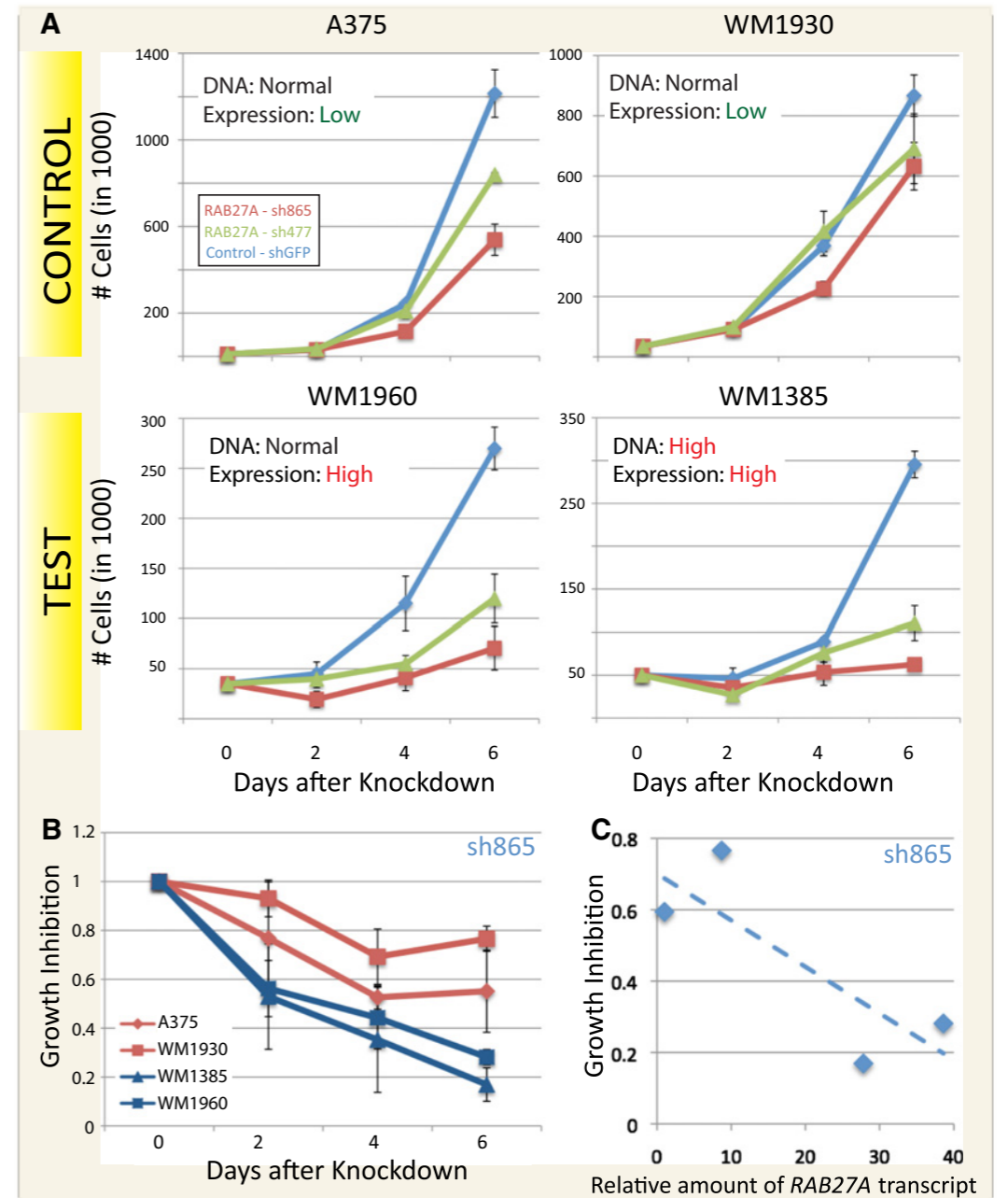
RAB27A

Involved in vesicular trafficking



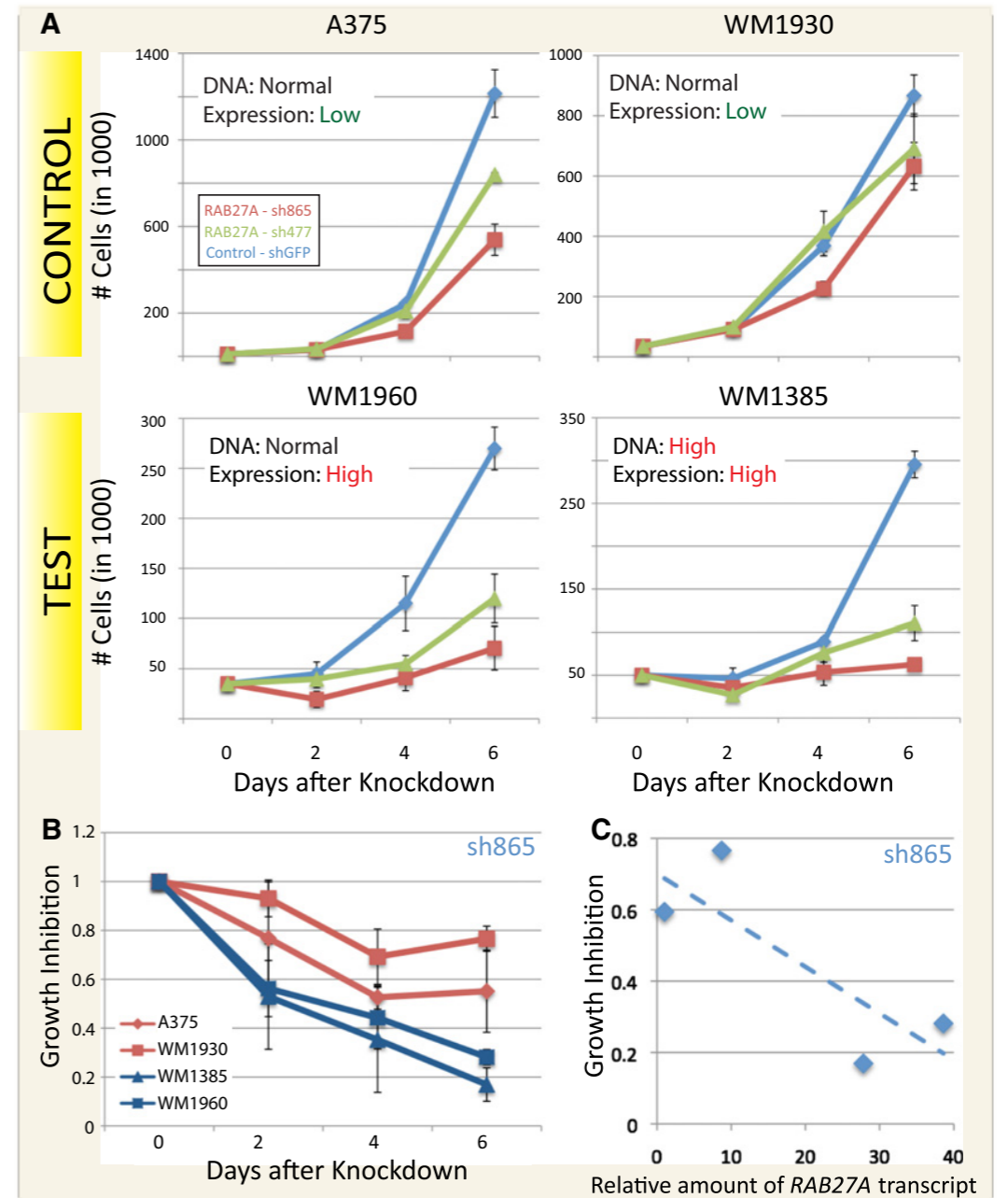
RAB27A

- Involved in vesicular trafficking
- Gene expression NOT correlated with CNA, but ranked at the top by CONEXIC



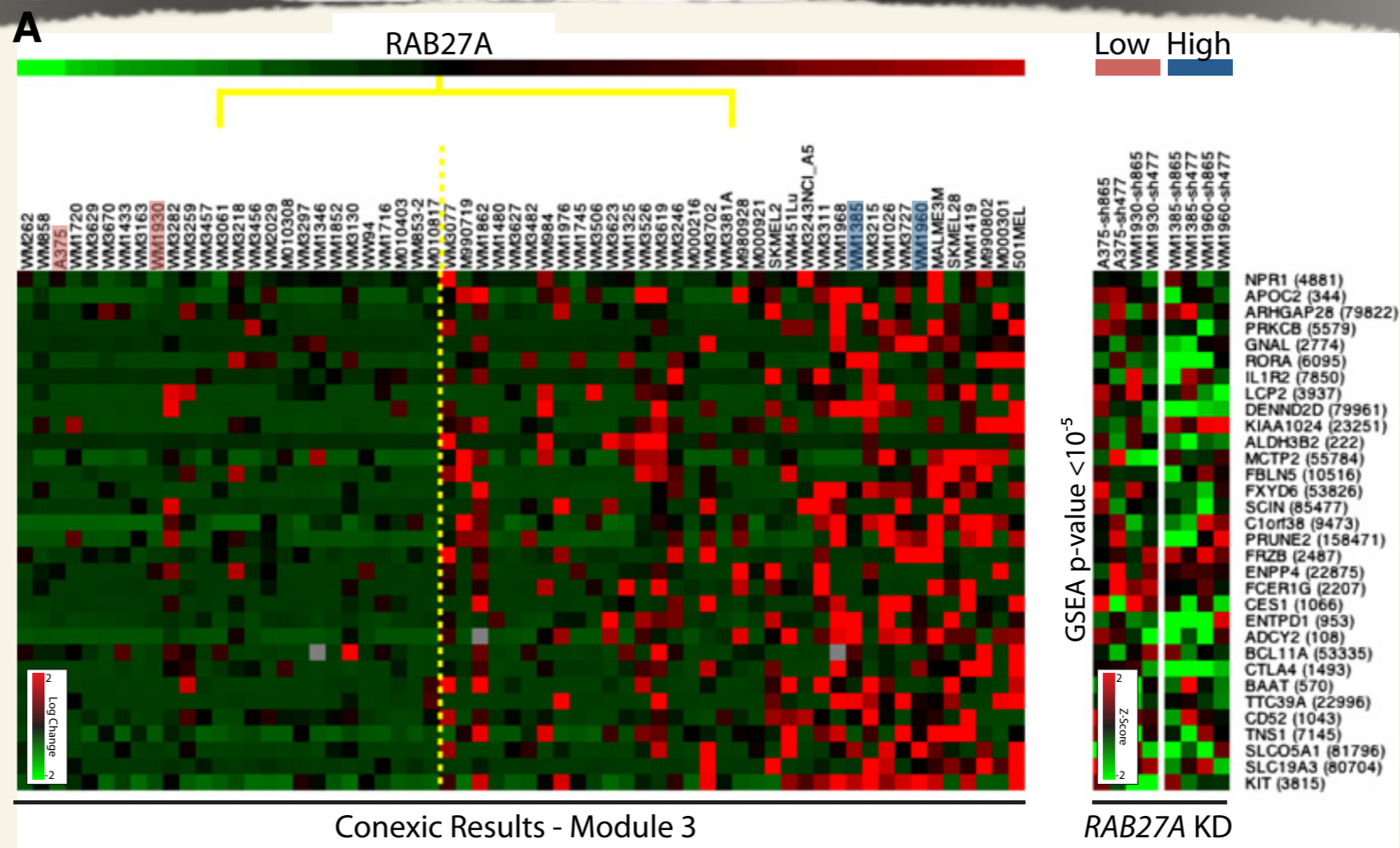
RAB27A

- Involved in vesicular trafficking
- Gene expression NOT correlated with CNA, but ranked at the top by CONEXIC
- Knock-down experiments



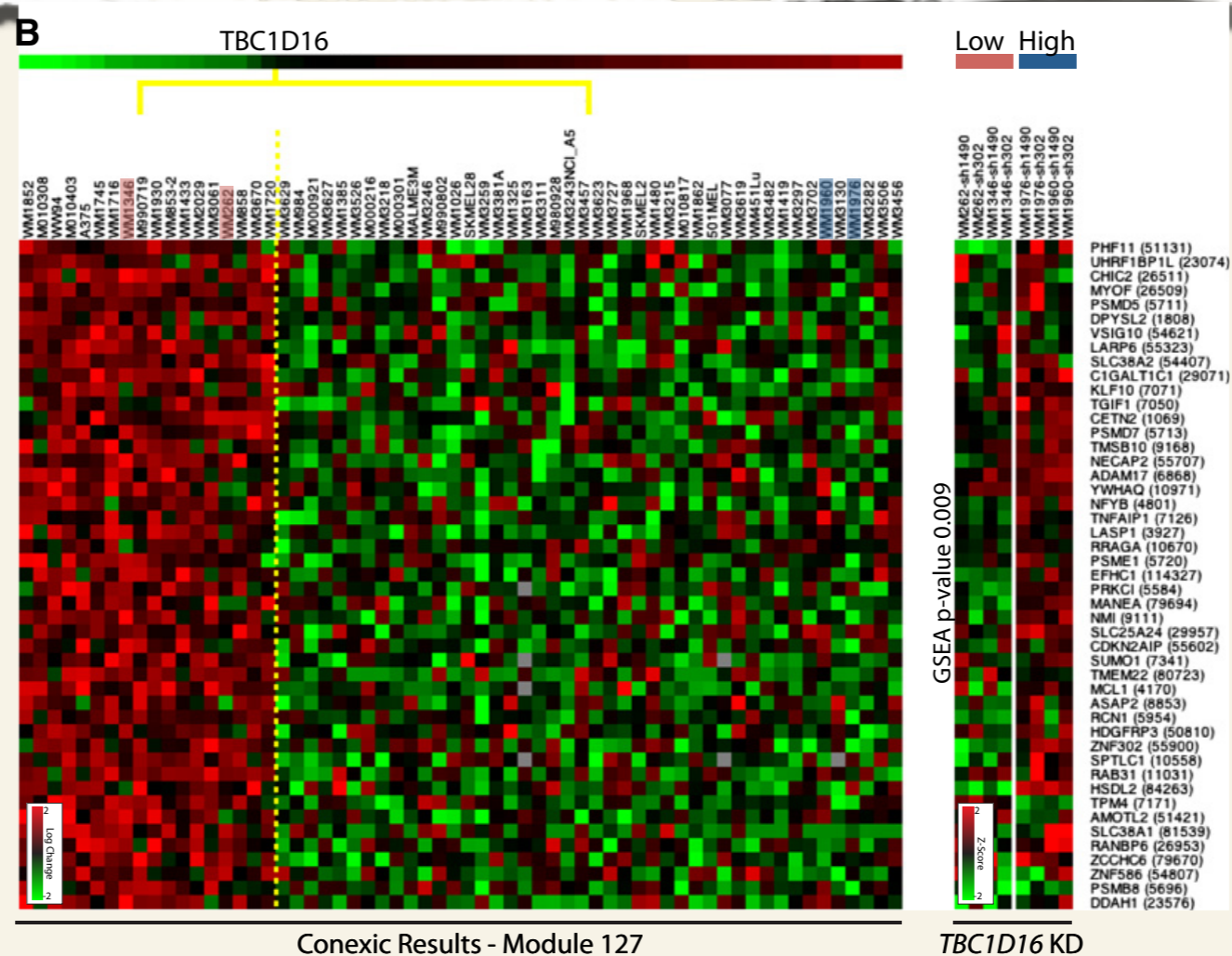
Do drivers influence the expression of their target genes?

- Microarray expression after knock-down: RAB27A; TBC1D16
- GSEA: to test the enrichment



Do drivers influence the expression of their target genes?

- Microarray expression after knock-down: RAB27A; TBC1D16
- GSEA: to test the enrichment



Conclusions

Conclusions

- ~ *Expression* of drivers, not copy number, drives phenotype

Conclusions

- ~ *Expression* of drivers, not copy number, drives phenotype
- ~ Correlation of drivers to copy number *lower* than correlation of passengers (selective pressure)

Conclusions

- ~ *Expression* of drivers, not copy number, drives phenotype
- ~ Correlation of drivers to copy number *lower* than correlation of passengers (selective pressure)
- ~ CONEXIC provides insights into physiological roles of drivers and their modules



Dana Pe'er Lab of Computational Systems Biology

Home

Publications

Research

People

Software

Biolearn

Genatomy

Camelot

JISTIC

LitVAn

CONEXIC

Collaborations

Openings

Contact

CONEXIC

COpy Number and EXpression In Cancer (CONEXIC) is an algorithm that integrates matched copy number (amplifications and deletions) and gene expression data from tumor samples to identify driving mutations and the processes they influence. CONEXIC is inspired by Module Networks (Segal et al, 2003), but has been augmented by a number of critical modifications that make it suitable for identifying drivers.

CONEXIC uses a score-guided search to identify the combination of modulators that best explains the behavior of a gene expression module across tumor samples and searches for those with the highest score within the amplified or deleted region.

When using CONEXIC, please cite the following article:

Akavia, U.D.*, Litvin O.*, Kim J., Sanchez-Garcia F., Kotliar D., Causton H.C., Pochanard P., Mozes E, Garraway L.A., Pe'er D. An Integrated Approach to Uncover Drivers of Cancer. Cell 2010; 143:1005-1017

*Equal Contribution

You can read detailed usage instructions [here](#). In order to run CONEXIC, please [download](#) it as well as the [math commons library](#) (see <http://commons.apache.org/math/> for more detail about this library).

For any questions, please contact conexic@gmail.com.