Module 4:

How do *unrealistic expectations* confound the results of our analyses

Case Studies in Bioinformatics

Giovanni Ciriello

giovanni.ciriello@unil.ch

Outline

- **Fundamentals of Cancer Genomics**
  - Types of genetic alterations in cancer
  - Most common alterations
  - Most commonly altered pathways
  - (Case studies)

- **Mutual exclusivity between alterations**
  - Why it occurs
  - Why it is important
  - How can we detect mutually exclusive alterations

- **The importance of null model designing**
  - 3 null models for testing mutual exclusivity

Cancer cells are associated with genetic abnormalities

Theodor Boveri (1862-1915)

Sea Urchin

Cancer cells are associated with genetic abnormalities

"A malignant tumour cell is [...] a cell with a specific abnormal chromosome constitution."

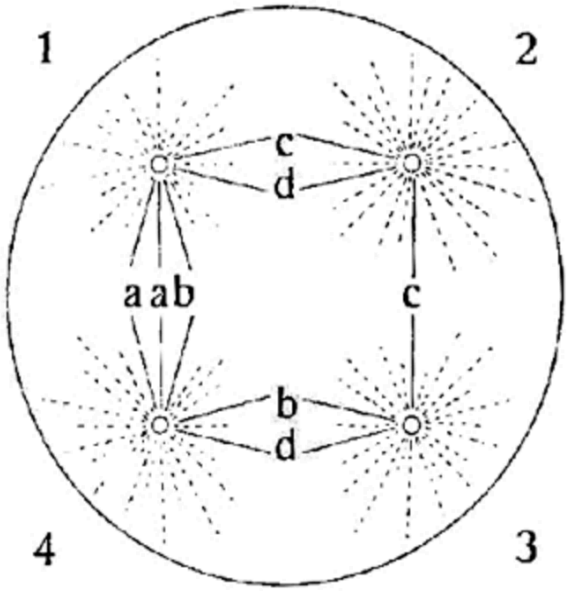


Fig. A.

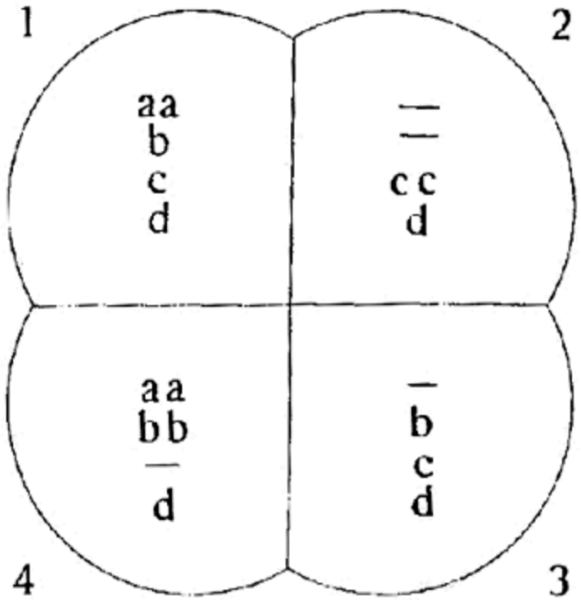
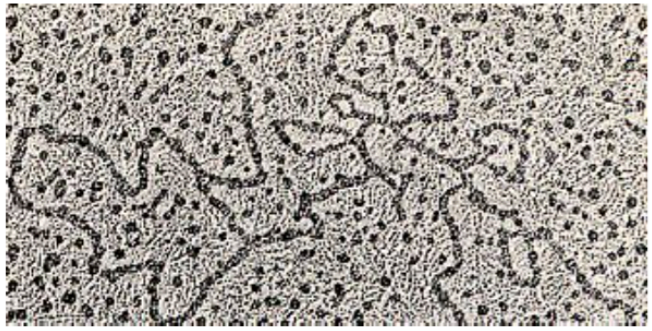


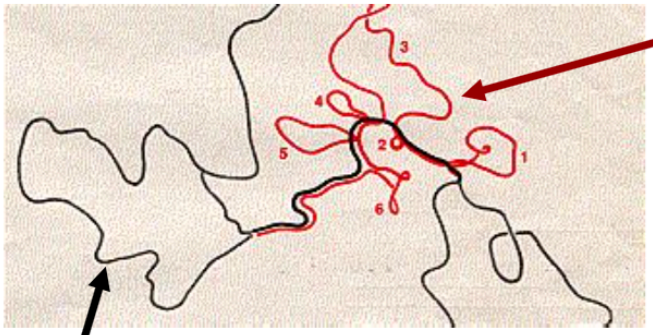
Fig. B.

Concerning the origin of malignant tumors. (1914) T. Boveri J Cell Sci. doi:10.1242/jcs.o25742

Cancer is a genetic disease

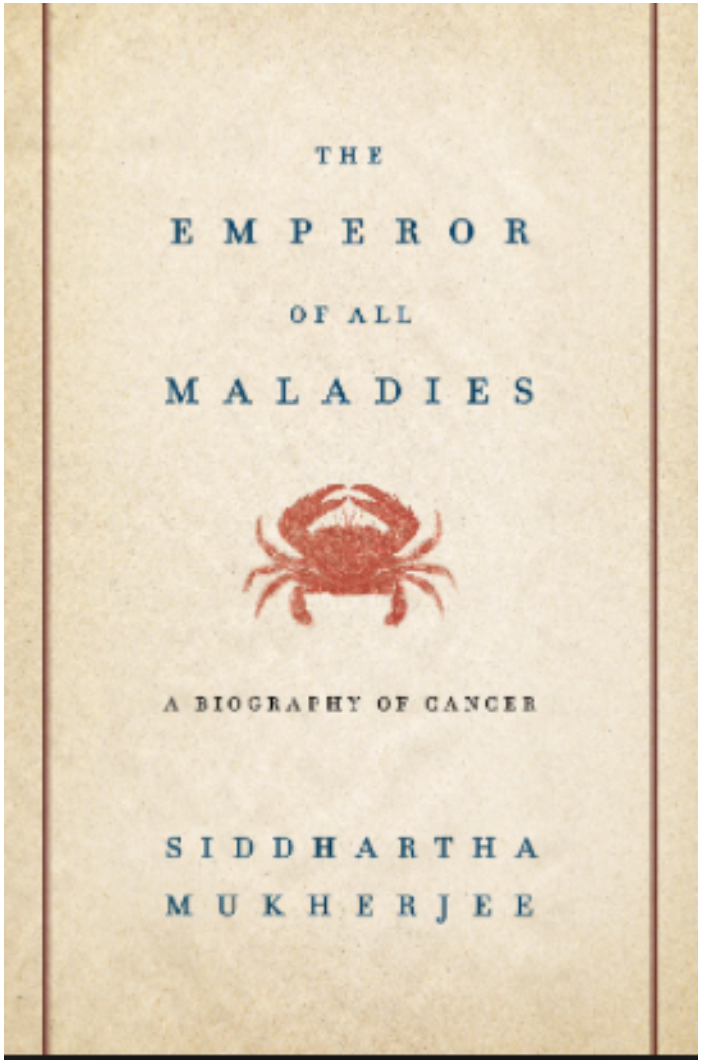


- Transforming src sequences from the Rous Sarcoma Virus are present in the DNA from normal cells.



Stehelin, Dominique, Varmus, Bishop, & Vogt, *Nature* 260, no. 5547 (1976): 170-173.

Cancer is a genetic disease



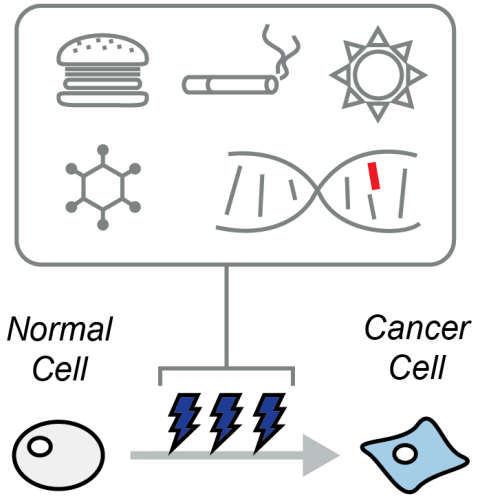
PBS Documentary

https://www.youtube.com/watch?v=iAbCa4k0Zfc

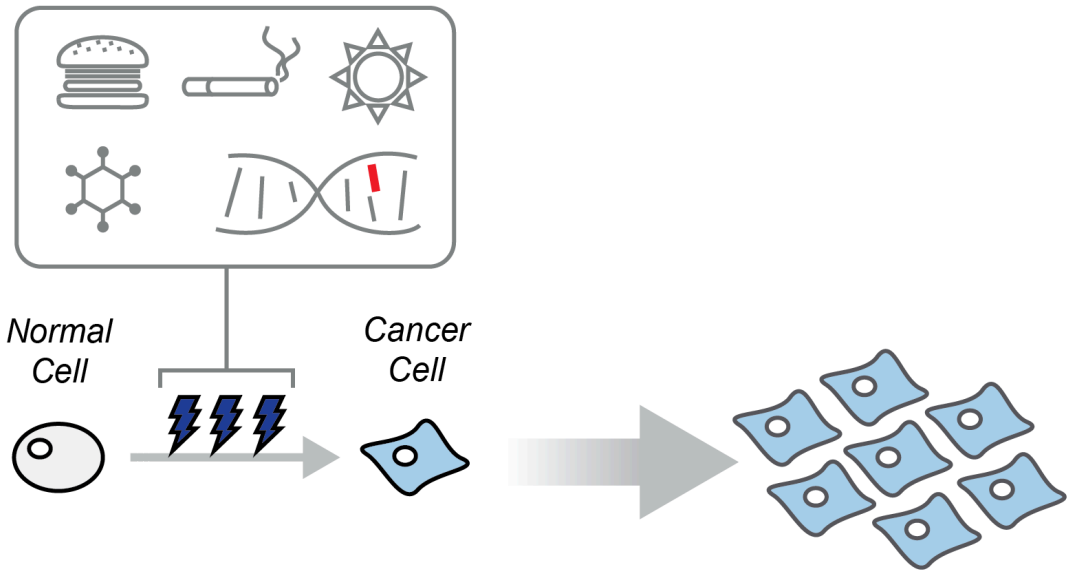
https://www.youtube.com/watch?v=gpjJIQK1QXA

https://www.youtube.com/watch?v=KYbxn1HtqFU

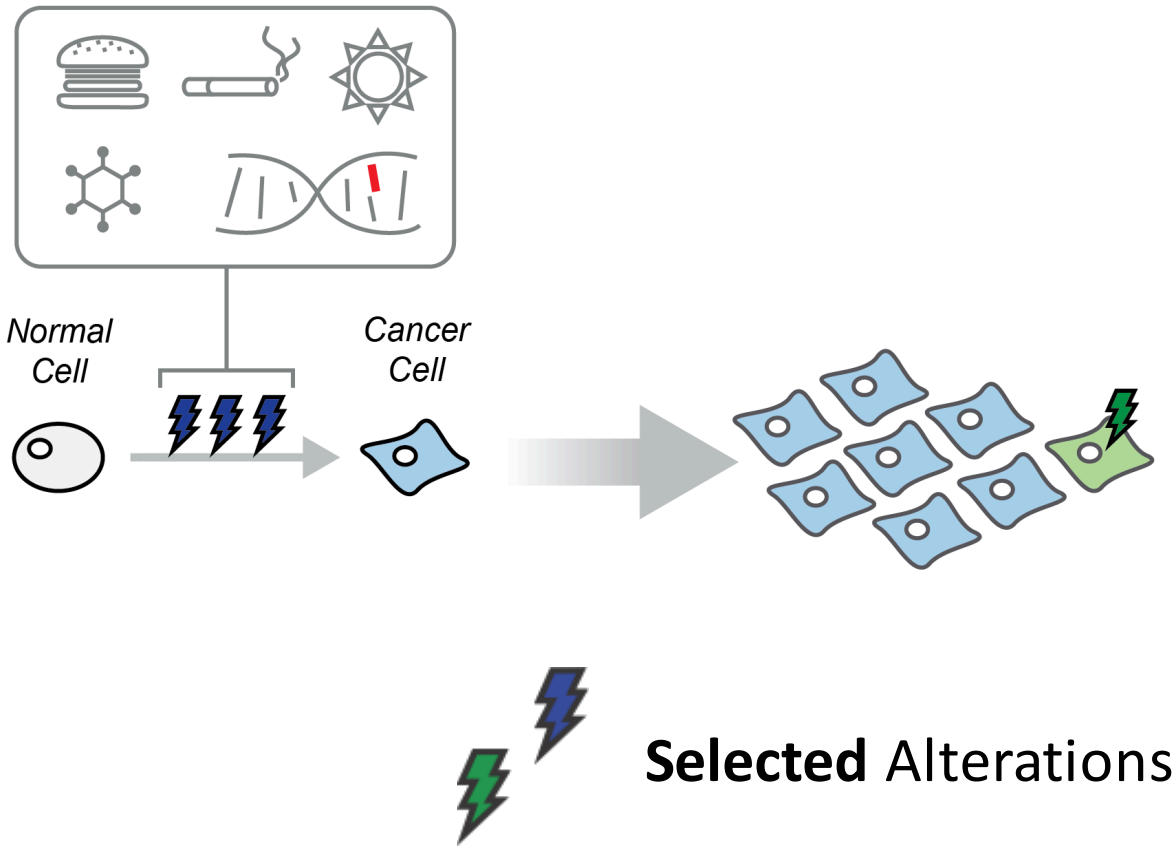
A simplified model of cancer evolution



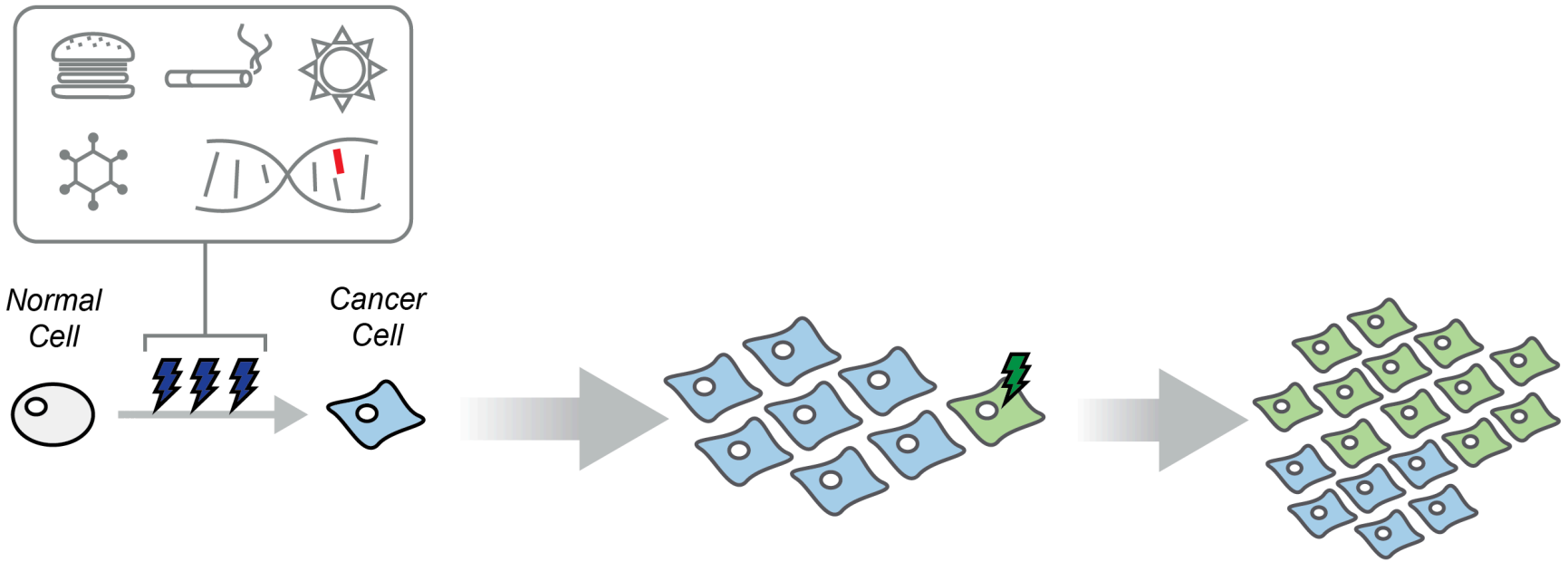
A simplified model of cancer evolution



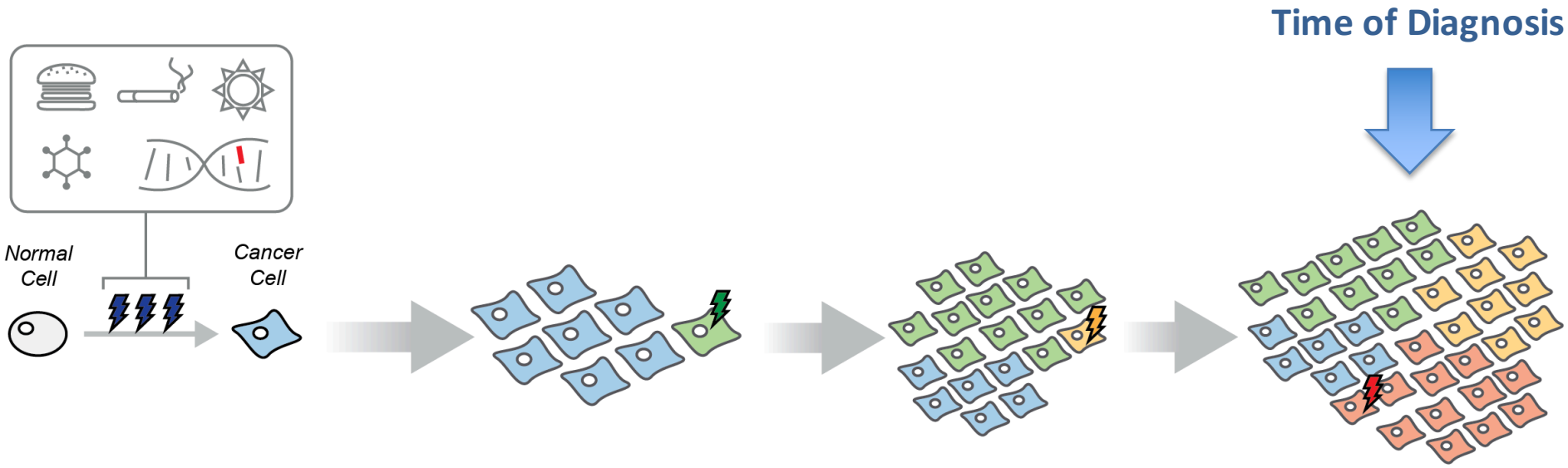
A simplified model of cancer evolution



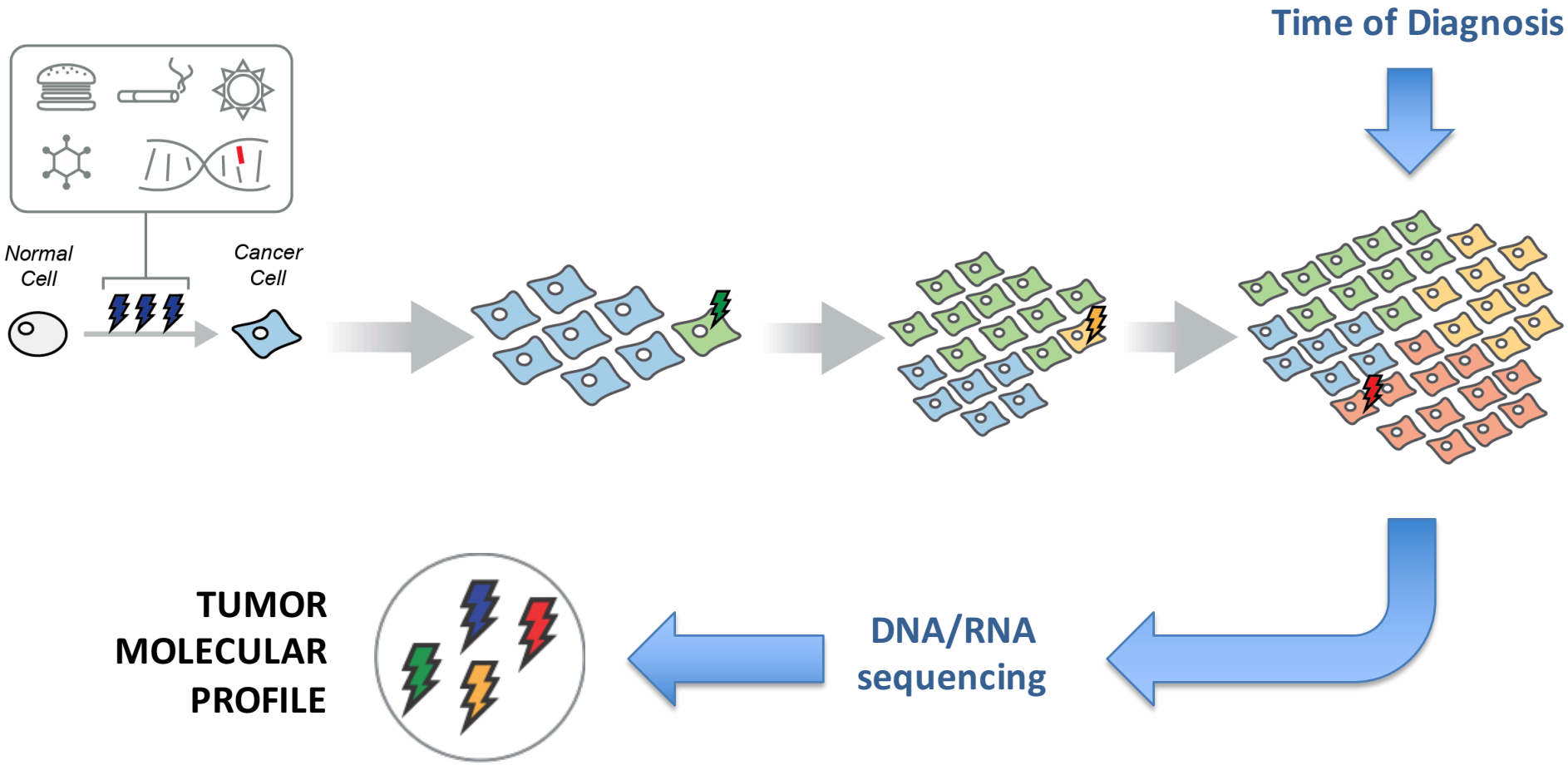
A simplified model of cancer evolution



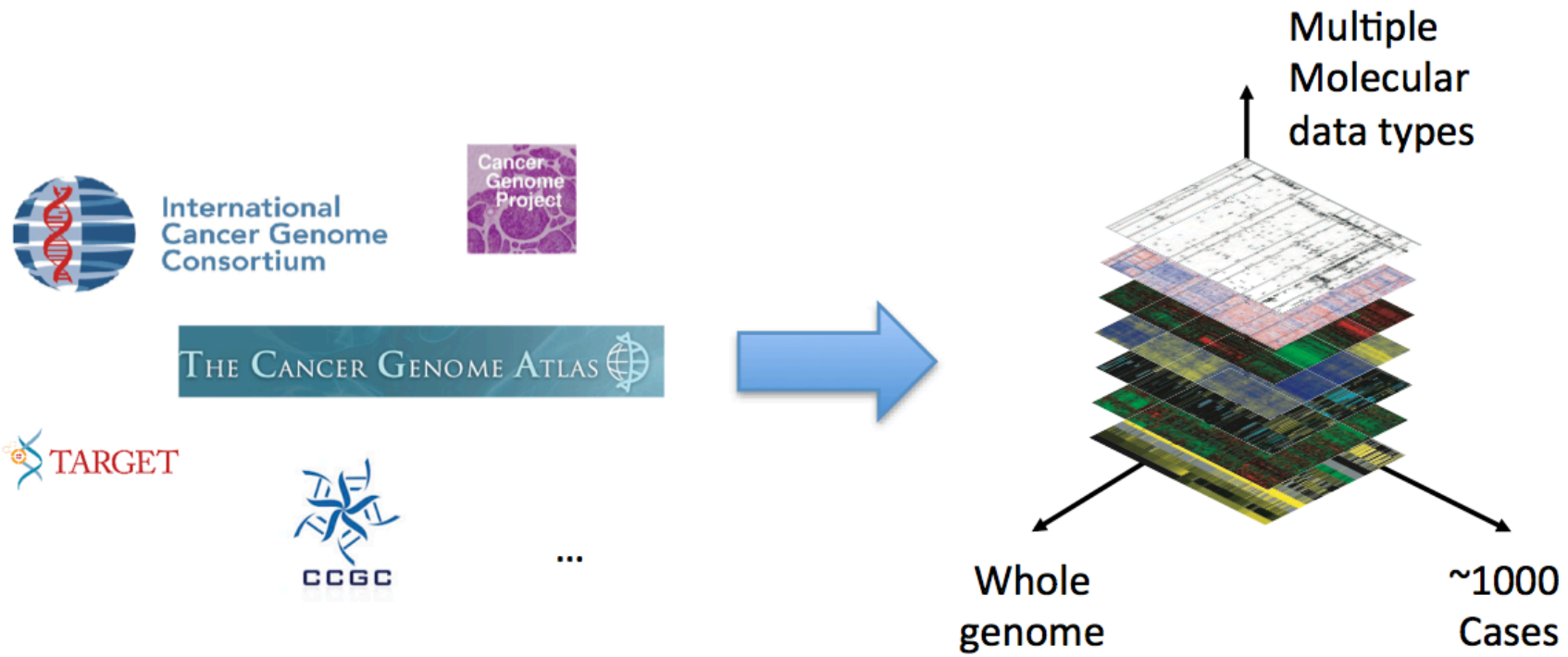
A simplified model of cancer evolution



A simplified model of cancer evolution

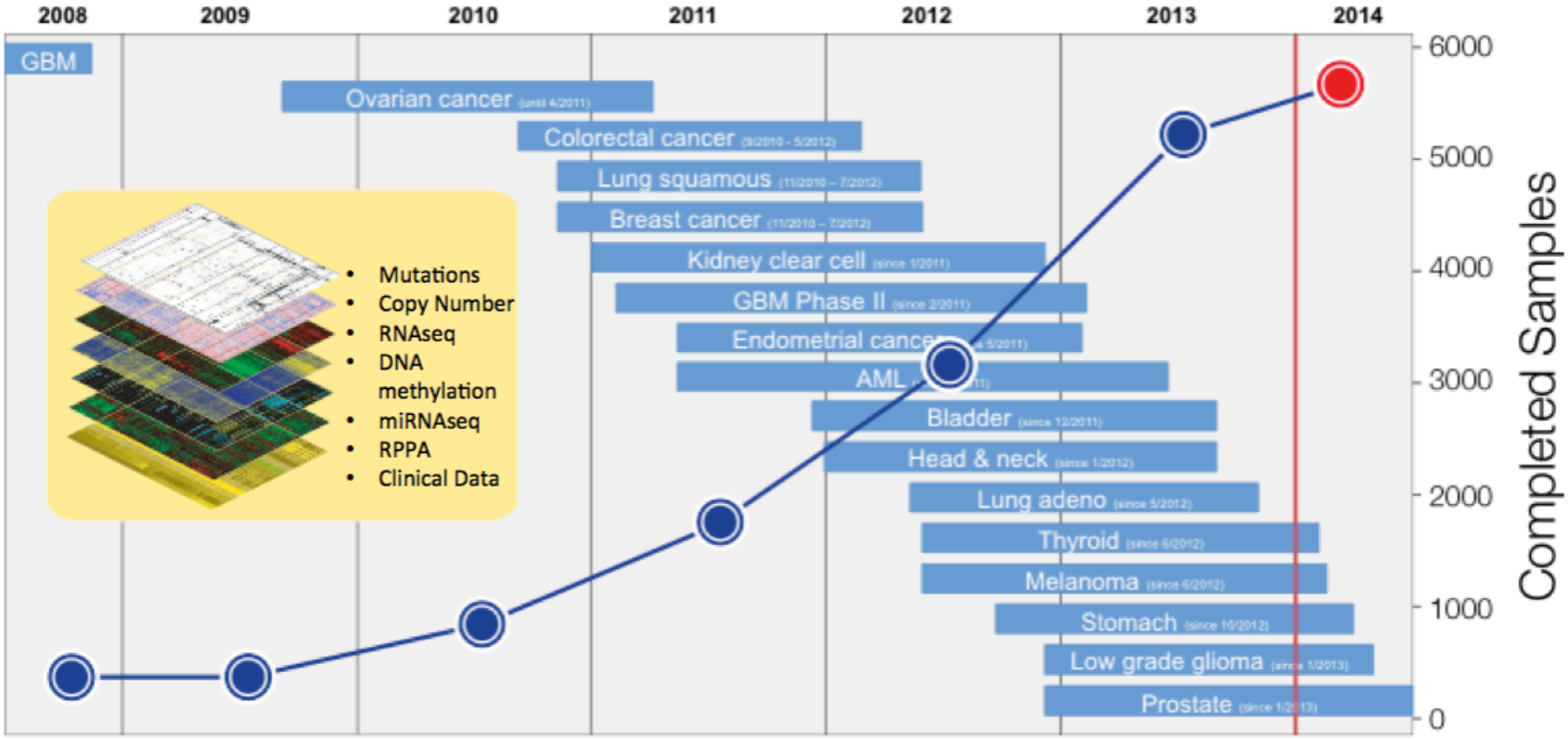


Cancer Genomics Projects



Cancer molecular landscape at unprecedented detail

The Cancer Genome Atlas



- **Recently Completed / ongoing:** Lobular breast cancer, chromophobe and papillary renal carcinoma, cervix carcinoma, adrenocortical carcinoma, sarcoma, hepatocellular carcinoma, DLBCL, pancreatic cancer, rare tumors, etc…

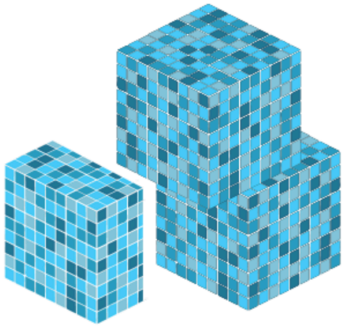
The Cancer Genome Atlas

TCGA produced over

2.5

PETABYTES

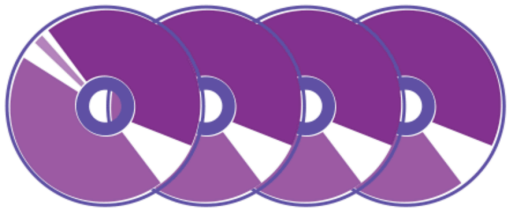
of data



To put this into perspective, **1 petabyte** of data is equal to

212,000

DVDs



TCGA data describes

33

DIFFERENT TUMOR TYPES

...including

10

RARE CANCERS

...based on paired tumor and normal tissue sets collected from

11,000

PATIENTS

...using

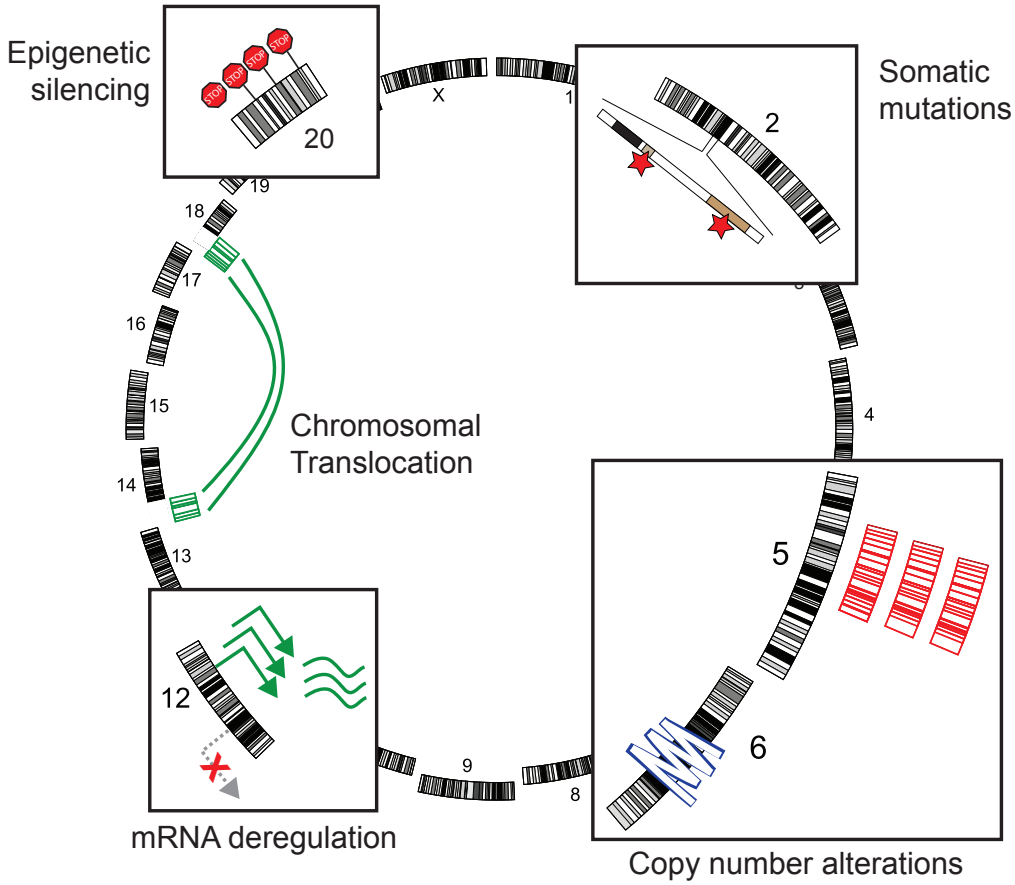
7

DIFFERENT DATA TYPES

Cancer molecular profiles

Alterations:

- **Mutations**
- **Copy number changes**
- Translocations
- Hyper/Hypo DNA Methylation
- Deregulation of transcription and translation



Gene Mutations

- **Single nucleotide changes**

	T			C			A			G		
T	TTT	Phe	F	TCT	Ser	S	TAT	Tyr	Y	TGT	Cys	C
	TTC	Phe	F	TCC	Ser	S	TAC	Tyr	Y	TGC	Cys	C
	TTA	Leu	L	TCA	Ser	S	TAA	stop	*	TGA	stop	*
	TTG	Leu	L	TCG	Ser	S	TAG	stop	*	TGG	Trp	W
C	CTT	Leu	L	CCT	Pro	P	CAT	His	H	CGT	Arg	R
	CTC	Leu	L	CCC	Pro	P	CAC	His	H	CGC	Arg	R
	CTA	Leu	L	CCA	Pro	P	CAA	Gln	Q	CGA	Arg	R
	CTG	Leu	L	CCG	Pro	P	CAG	Gln	Q	CGG	Arg	R
A	ATT	Ile	I	ACT	Thr	T	AAT	Asn	N	AGT	Ser	S
	ATC	Ile	I	ACC	Thr	T	AAC	Asn	N	AGC	Ser	S
	ATA	Ile	I	ACA	Thr	T	AAA	Lys	K	AGA	Arg	R
	ATG	Met	M	ACG	Thr	T	AAG	Lys	K	AGG	Arg	R
G	GTT	Val	V	GCT	Ala	A	GAT	Asp	D	GGT	Gly	G
	GTC	Val	V	GCC	Ala	A	GAC	Asp	D	GGC	Gly	G
	GTA	Val	V	GCA	Ala	A	GAA	Glu	E	GGA	Gly	G
	GTG	Val	V	GCG	Ala	A	GAG	Glu	E	GGG	Gly	G

Gene Mutations

- **Single nucleotide changes**

- **Silent mutations**: nucleotide change no amino acid change

TCT=Serine
TCC=Serine

	T			C			A			G		
T	TTT	Phe	F	TCT	Ser	S	TAT	Tyr	Y	TGT	Cys	C
	TTC	Phe	F	TCC	Ser	S	TAC	Tyr	Y	TGC	Cys	C
	TTA	Leu	L	TCA	Ser	S	TAA	stop	*	TGA	stop	*
	TTG	Leu	L	TCG	Ser	S	TAG	stop	*	TGG	Trp	W
C	CTT	Leu	L	CCT	Pro	P	CAT	His	H	CGT	Arg	R
	CTC	Leu	L	CCC	Pro	P	CAC	His	H	CGC	Arg	R
	CTA	Leu	L	CCA	Pro	P	CAA	Gln	Q	CGA	Arg	R
	CTG	Leu	L	CCG	Pro	P	CAG	Gln	Q	CGG	Arg	R
A	ATT	Ile	I	ACT	Thr	T	AAT	Asn	N	AGT	Ser	S
	ATC	Ile	I	ACC	Thr	T	AAC	Asn	N	AGC	Ser	S
	ATA	Ile	I	ACA	Thr	T	AAA	Lys	K	AGA	Arg	R
	ATG	Met	M	ACG	Thr	T	AAG	Lys	K	AGG	Arg	R
G	GTT	Val	V	GCT	Ala	A	GAT	Asp	D	GGT	Gly	G
	GTC	Val	V	GCC	Ala	A	GAC	Asp	D	GGC	Gly	G
	GTA	Val	V	GCA	Ala	A	GAA	Glu	E	GGA	Gly	G
	GTG	Val	V	GCG	Ala	A	GAG	Glu	E	GGG	Gly	G

Gene Mutations

- **Single nucleotide changes**

- **Missense**: change a nucleotide and encode for a different amino acid

  TCT= Serine
  CCT= Proline

- **Nonsense:** change a nucleotide and induce a stop codon

  TAT = Serine
  TAA = Stop Codon!

Gene Mutations

- **Frame-shift mutations** (change the reading frame)

  - **Deletion:** deletion of 1 or more nucleotide

ACC AGC TGC ACT
Thr Ser Cys Thr

ACC AGC TGA CT
Thr Ser STOP

  - **Insertion**: Add one or more extra-nucleotide to the DNA

ACC AGC TGC ACT
Thr Ser Cys Thr

ACC AGC TGC CAC CT
Thr Ser Cys His

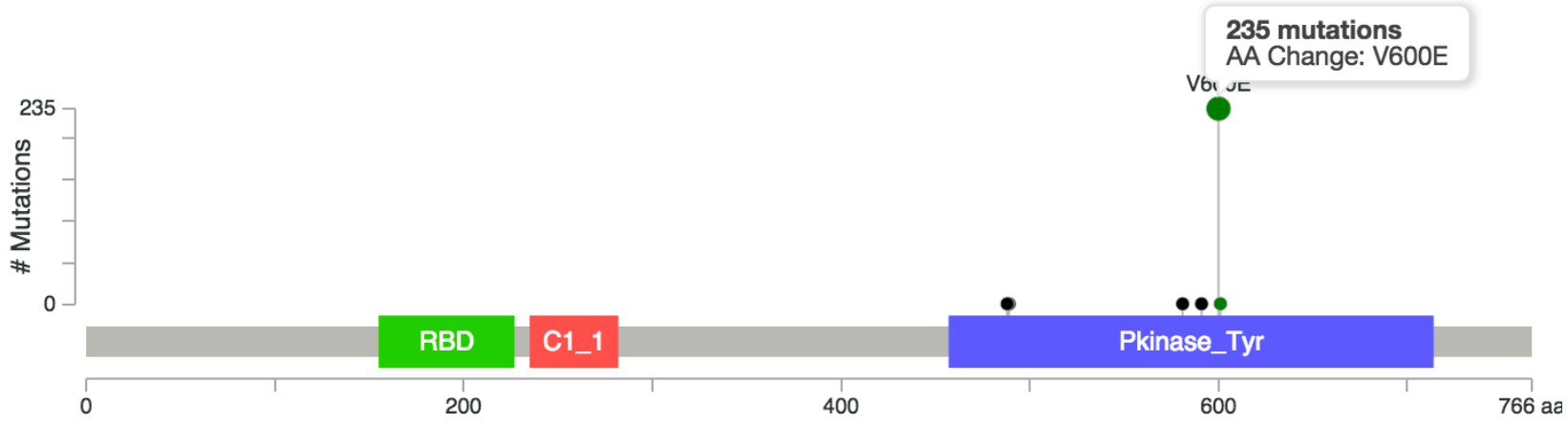
HOTSPOT mutations

(activating an **oncogene**)

BRAF V600E mutations in Thyroid Carcinoma (399 patients)

GTG = Valine (V)

GAG = Glutamate (E)

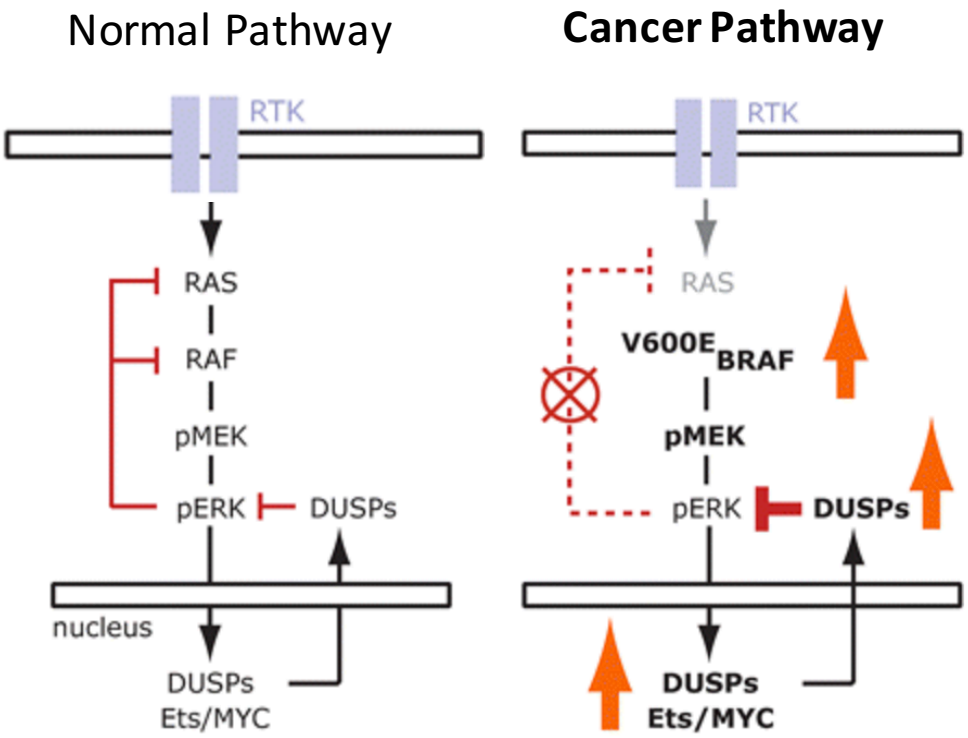


HOTSPOT mutations

(activating an **oncogene**)

BRAF V600E mutations in Thyroid Carcinoma (399 patients)

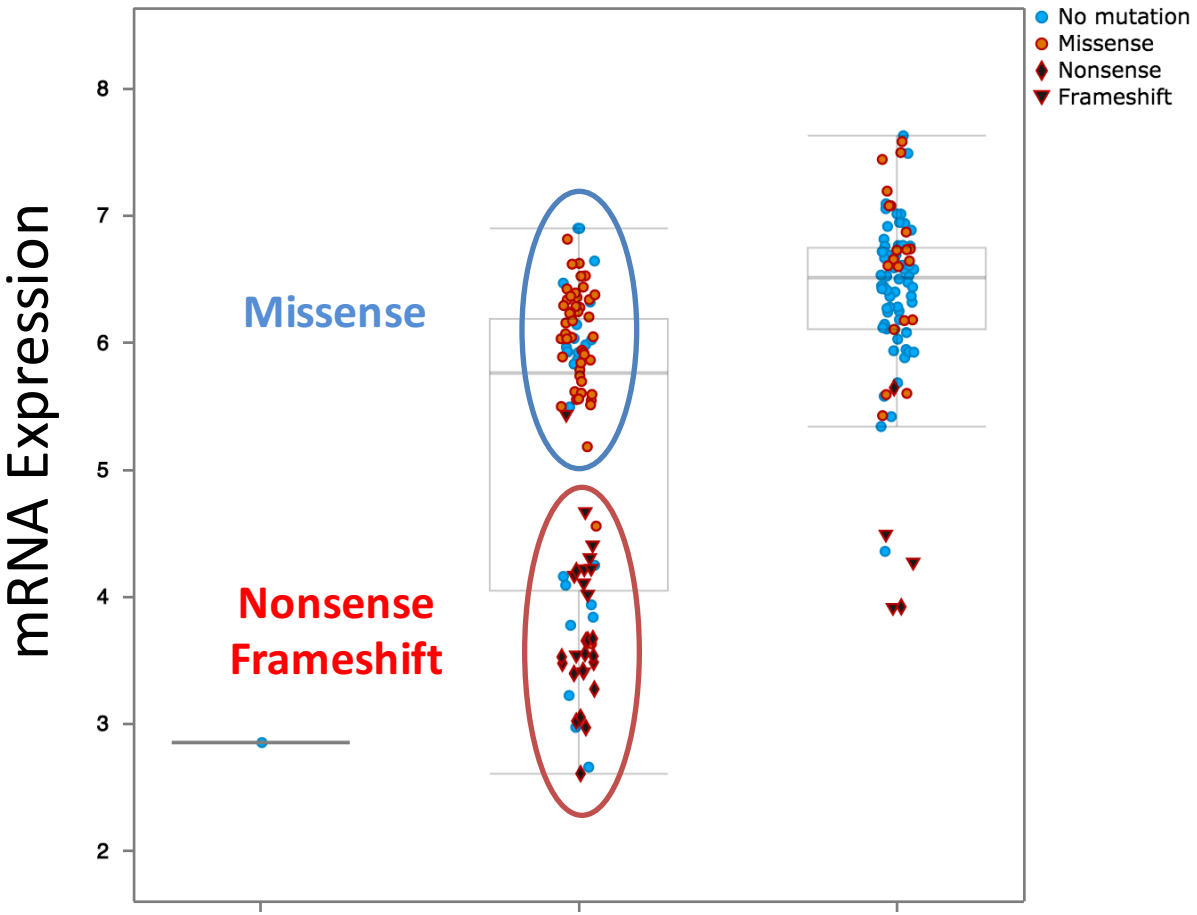
GTG = Valine (V)
GAG = Glutamate (E)



Truncating Mutations

(inactivating a **tumor suppressor**)

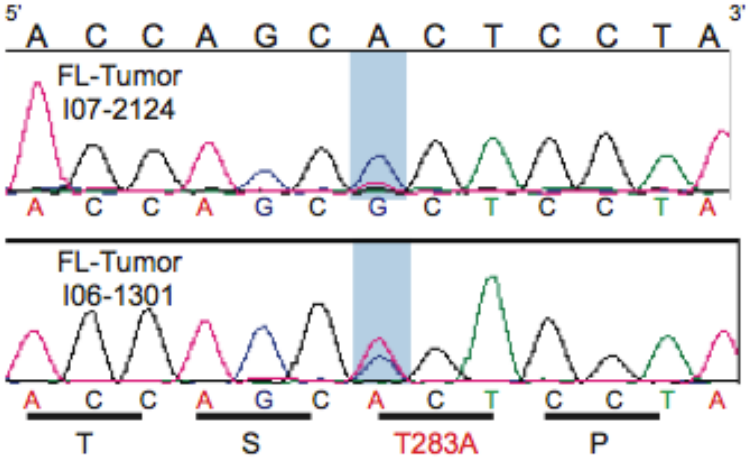
- **TP53 mutations in Colorectal cancer**

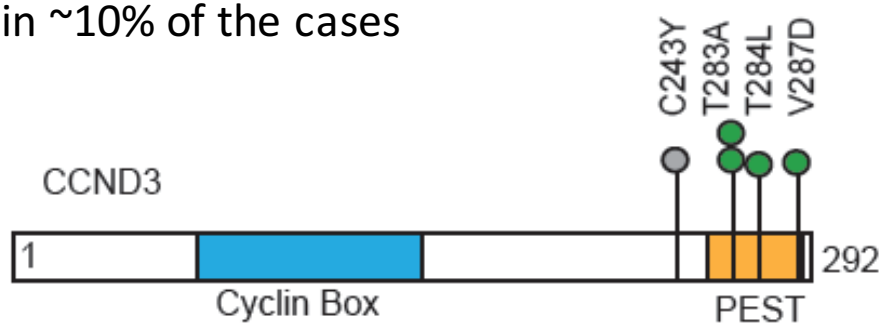


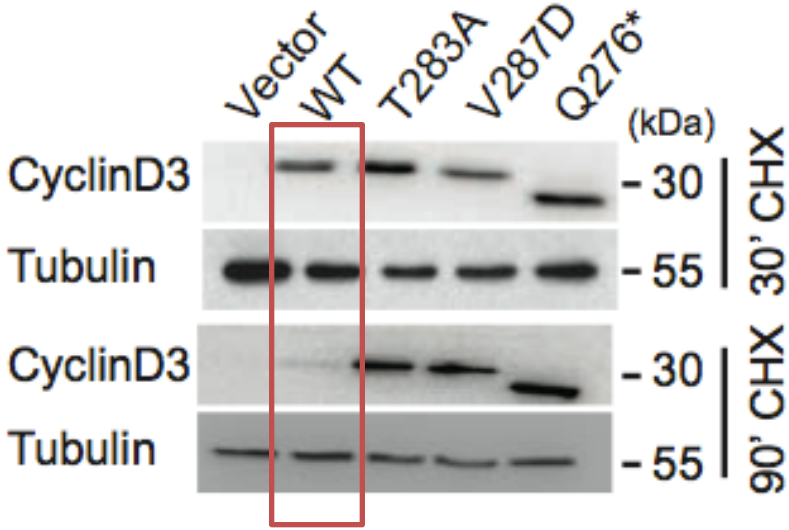
Truncating Mutations

(activating an **oncogene**)

In Lymphoma mutation in CyclinD3 occurs in ~10% of the cases







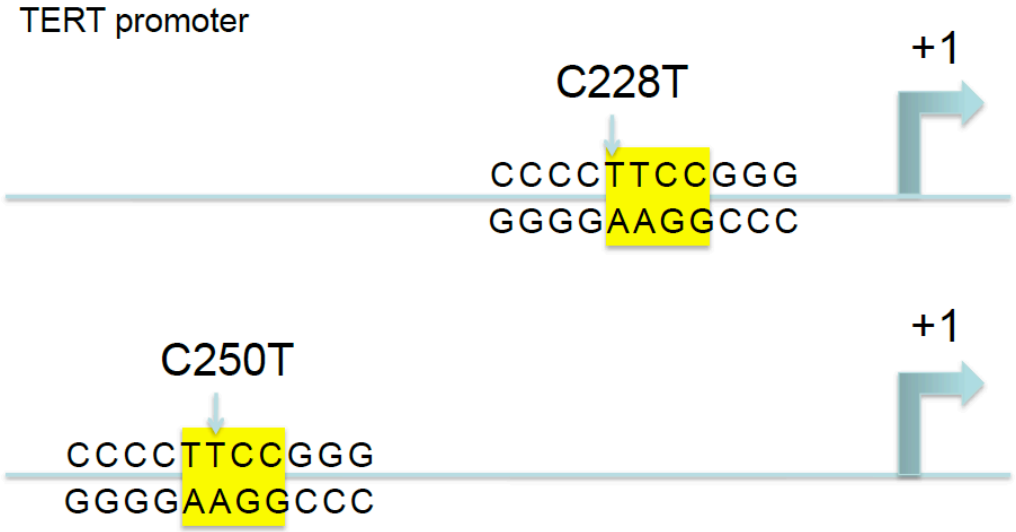
(Oricchio et al. JEM, 2014)

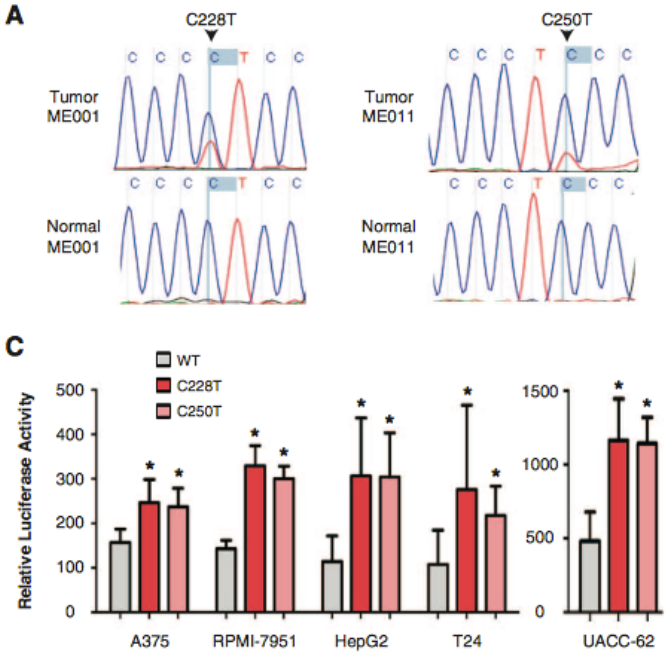
Non-coding Mutations

Highly Recurrent *TERT* Promoter Mutations in Human Melanoma

Franklin W. Huang,[1,2,3]* Eran Hodis,[1,3,4]* Mary Jue Xu,[1,3,4] Gregory V. Kryukov,[1] Lynda Chin,[5,6] Levi A. Garraway[1,2,3]†

(Science, 2013)

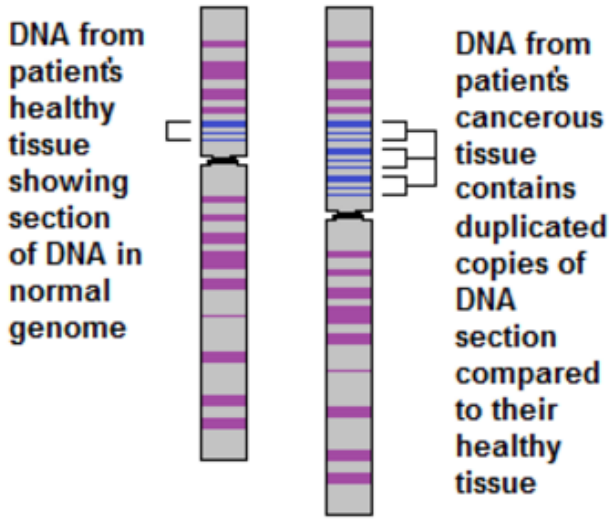




Copy Number Alterations

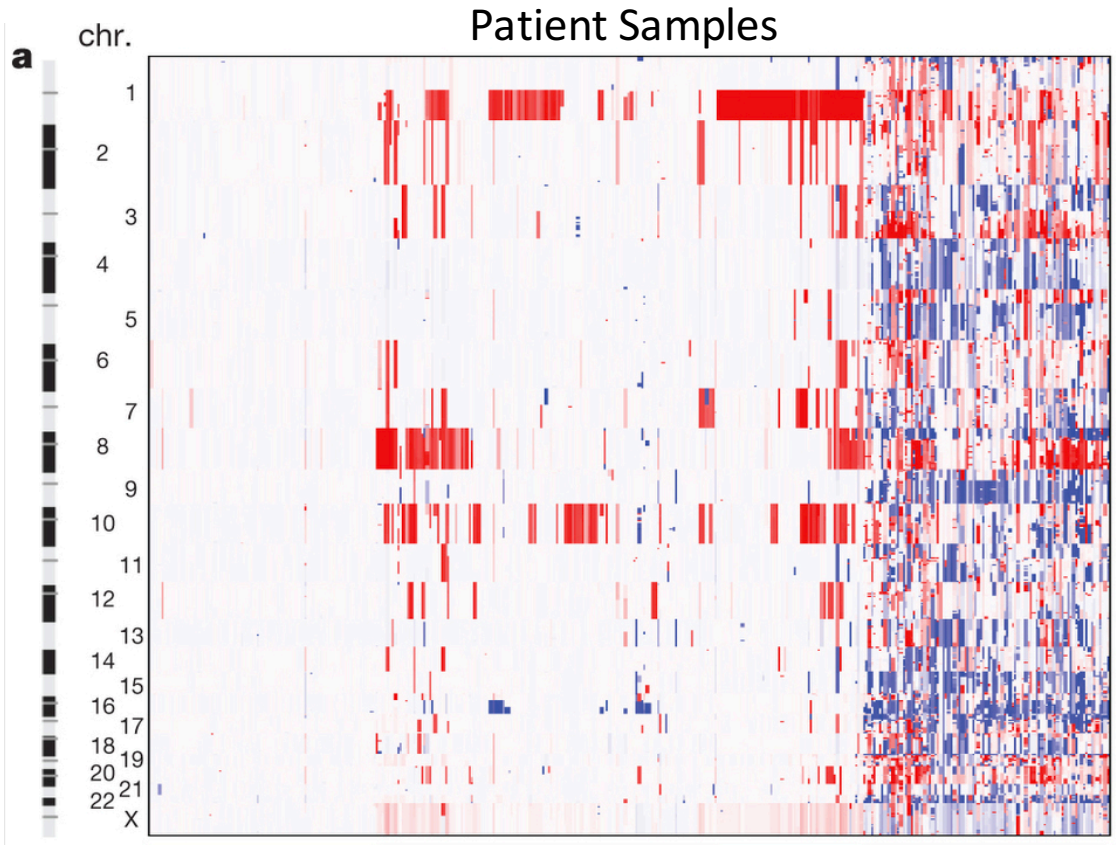
- ***Deletion***: Loss of chromosomal regions
  (Heterozygous or Homozygous)

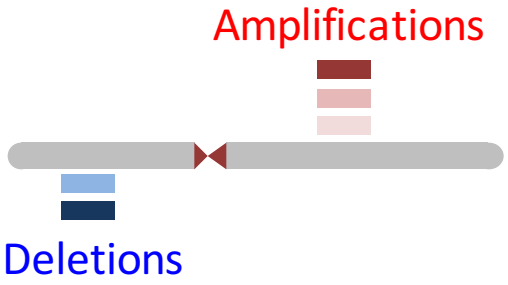
- ***Amplifications***: Acquire one or more copy of chromosomal regions (Duplication or Amplification)



Copy Number Alterations

- **Endometrial Carcinoma**



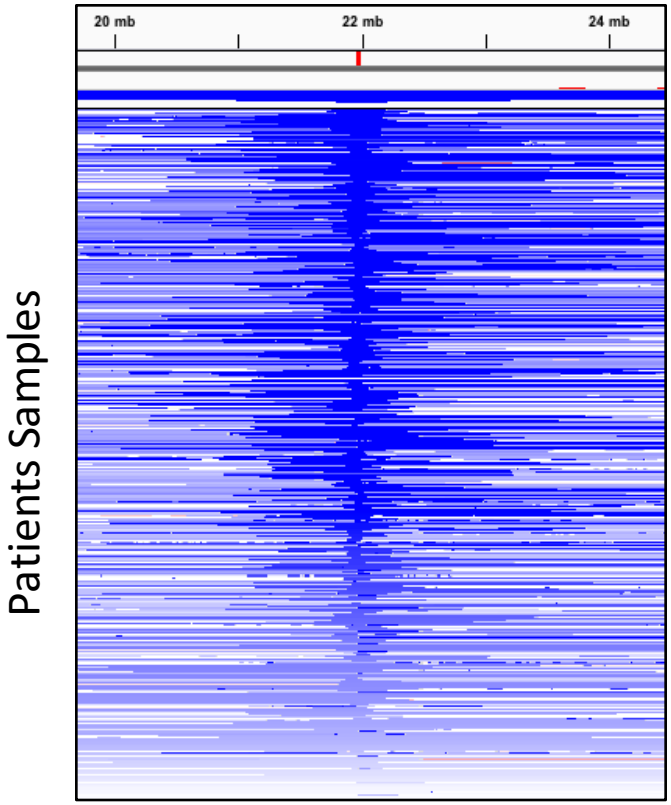


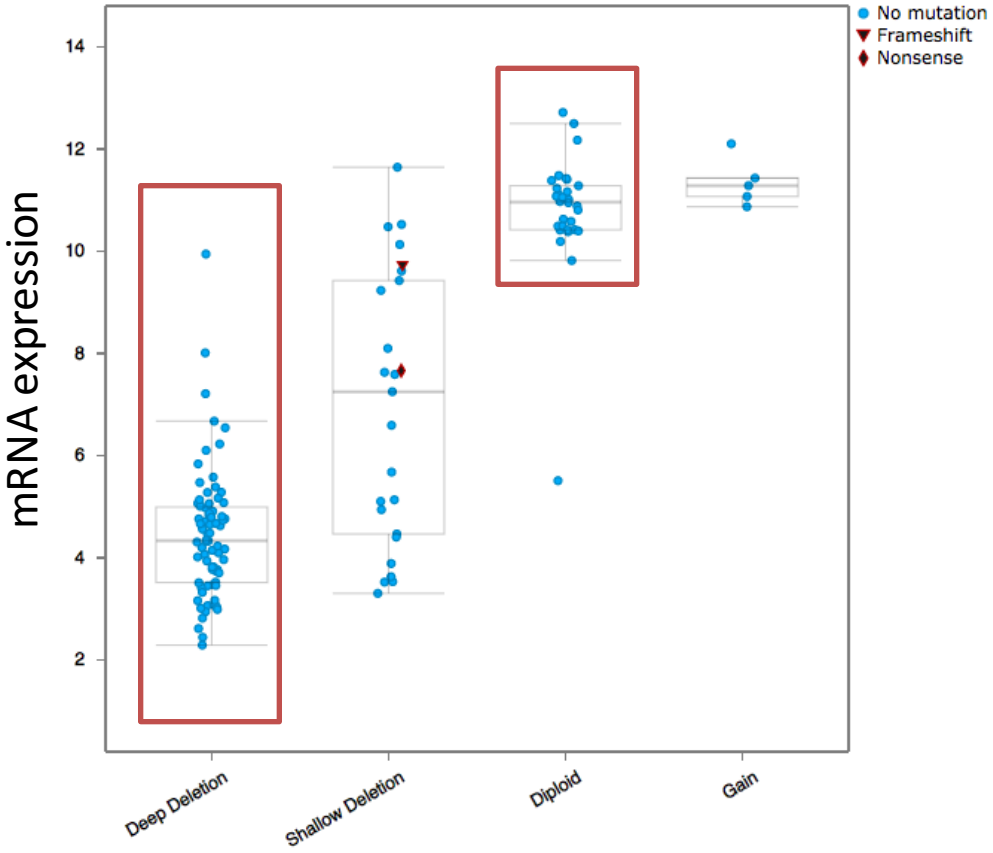
(TCGA, Nature 2013)

Focal Deletions

(inactivating a **tumor suppressor**)

- **Glioblastoma**

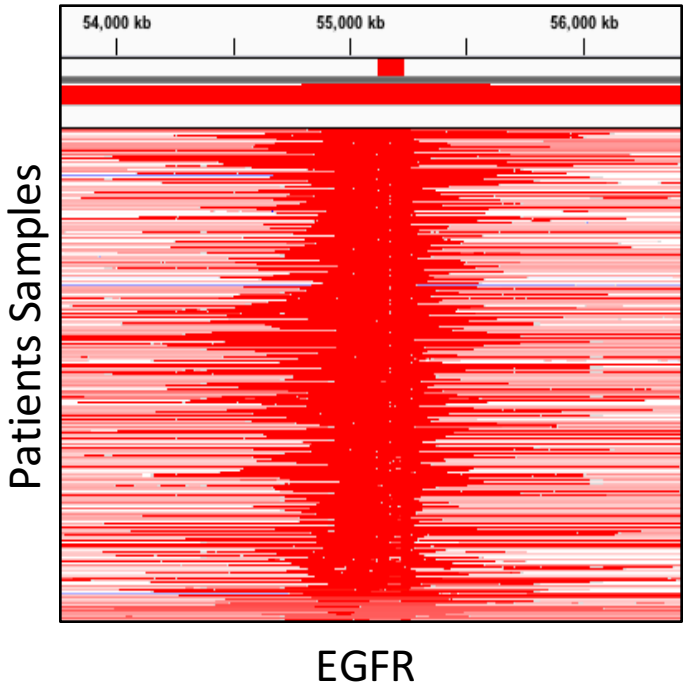


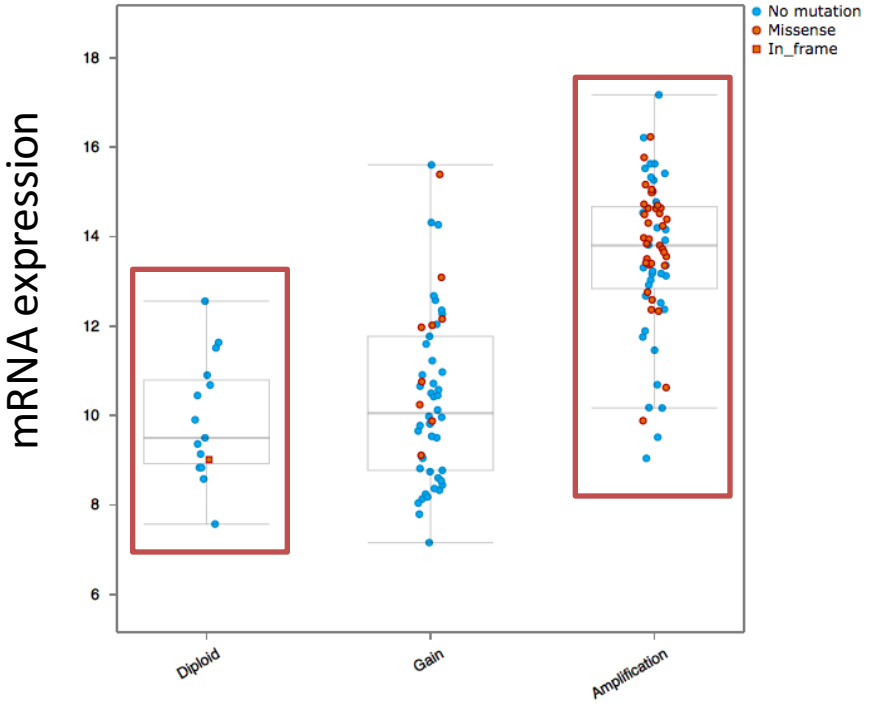


Focal Amplifications

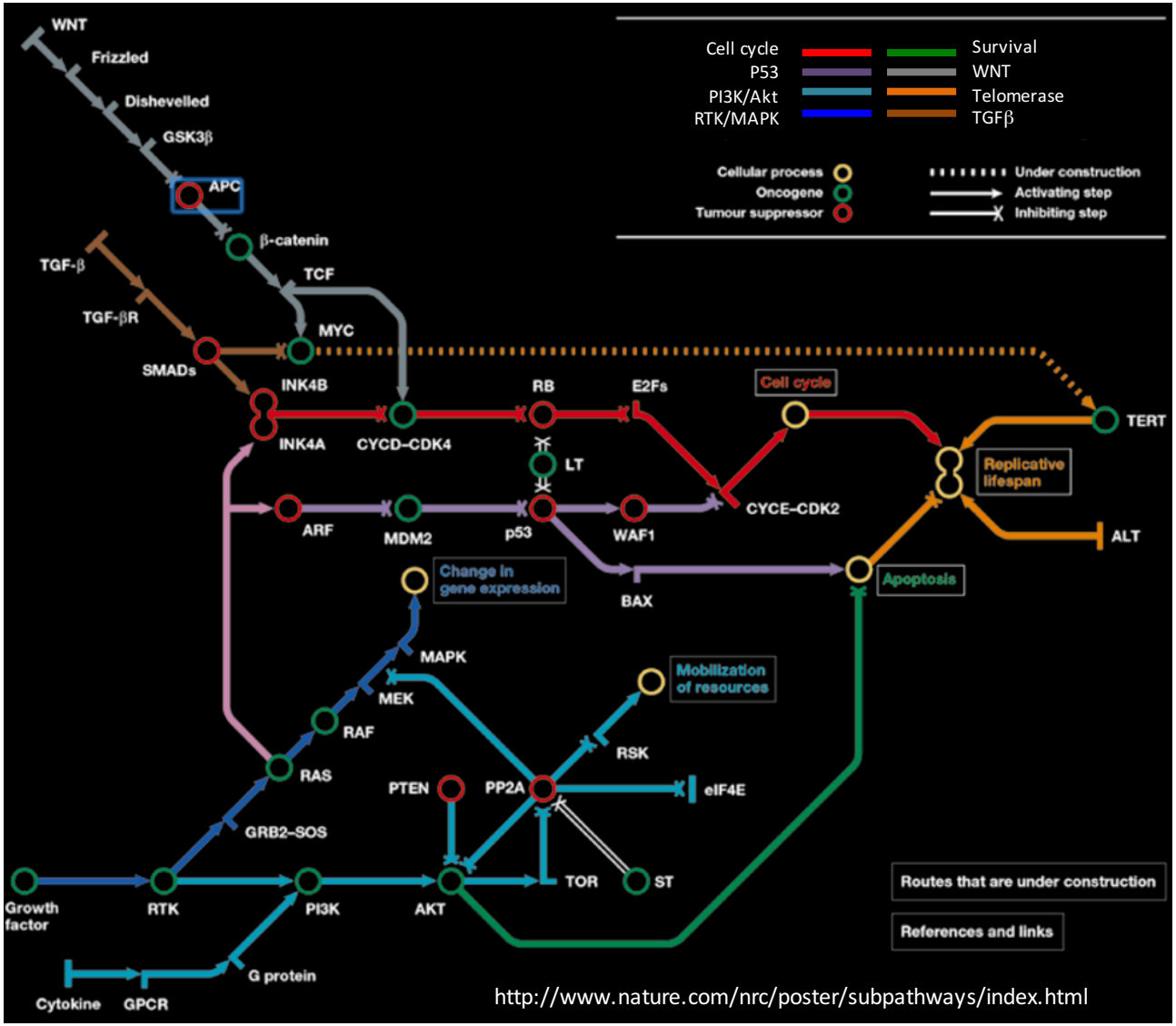
(activating an **oncogene**)

- **Glioblastoma**





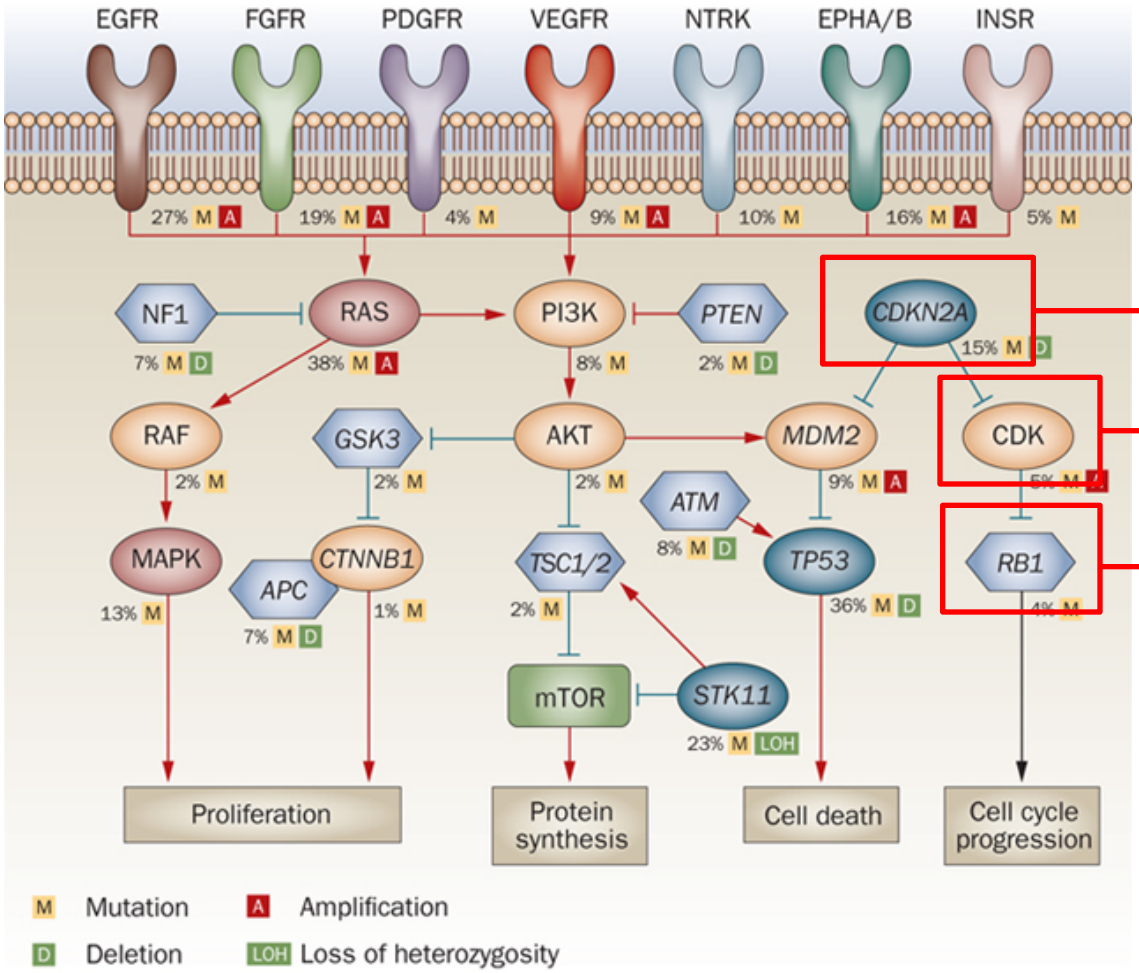
Cancer Pathways



http://www.nature.com/nrc/poster/subpathways/index.html

Rb Pathway

- **Cell cycle checkpoint G1/S phase**



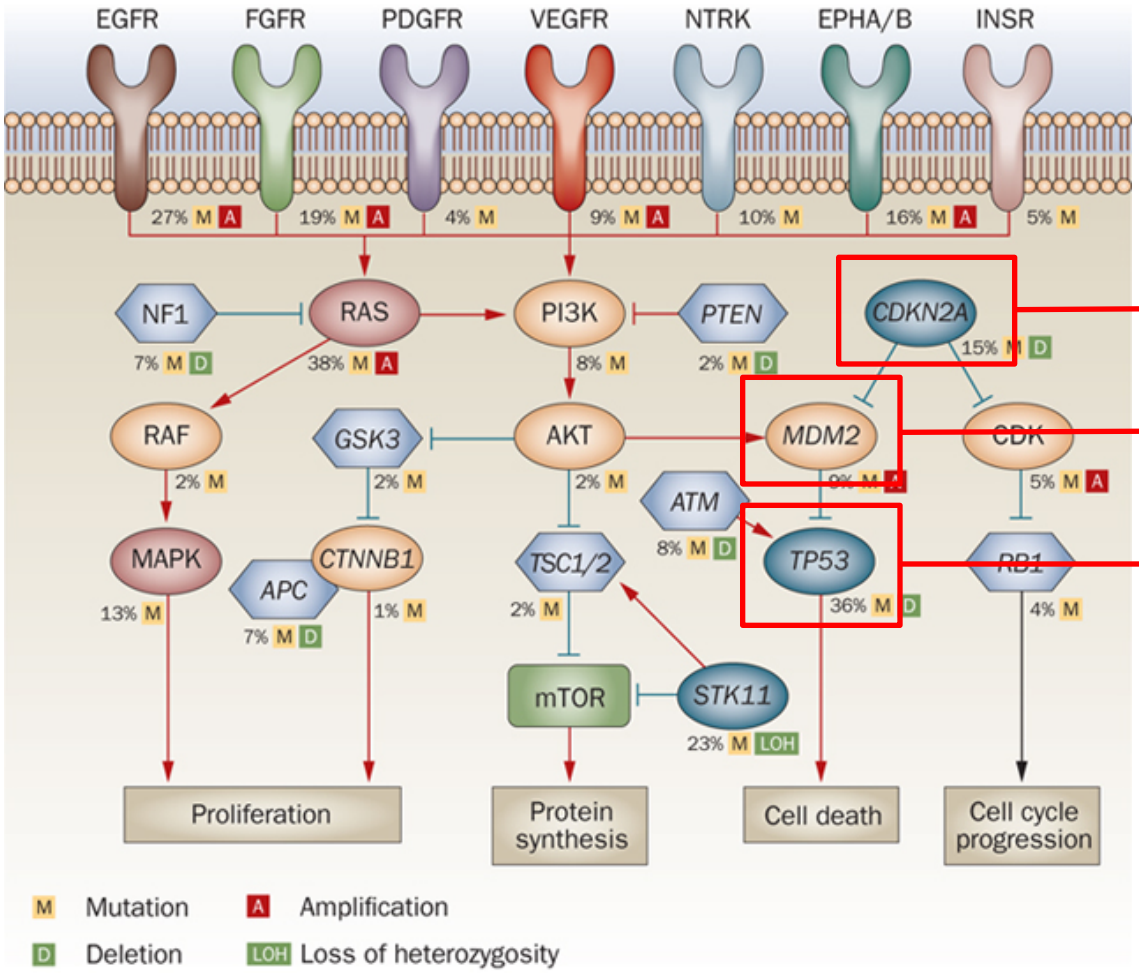
CDKN2A/CDKN2B

CCND1, CCND3, CCNE1
CDK2, CDK4, CDK6

RB1, E2F1

p53 pathway

- **Apoptosis**



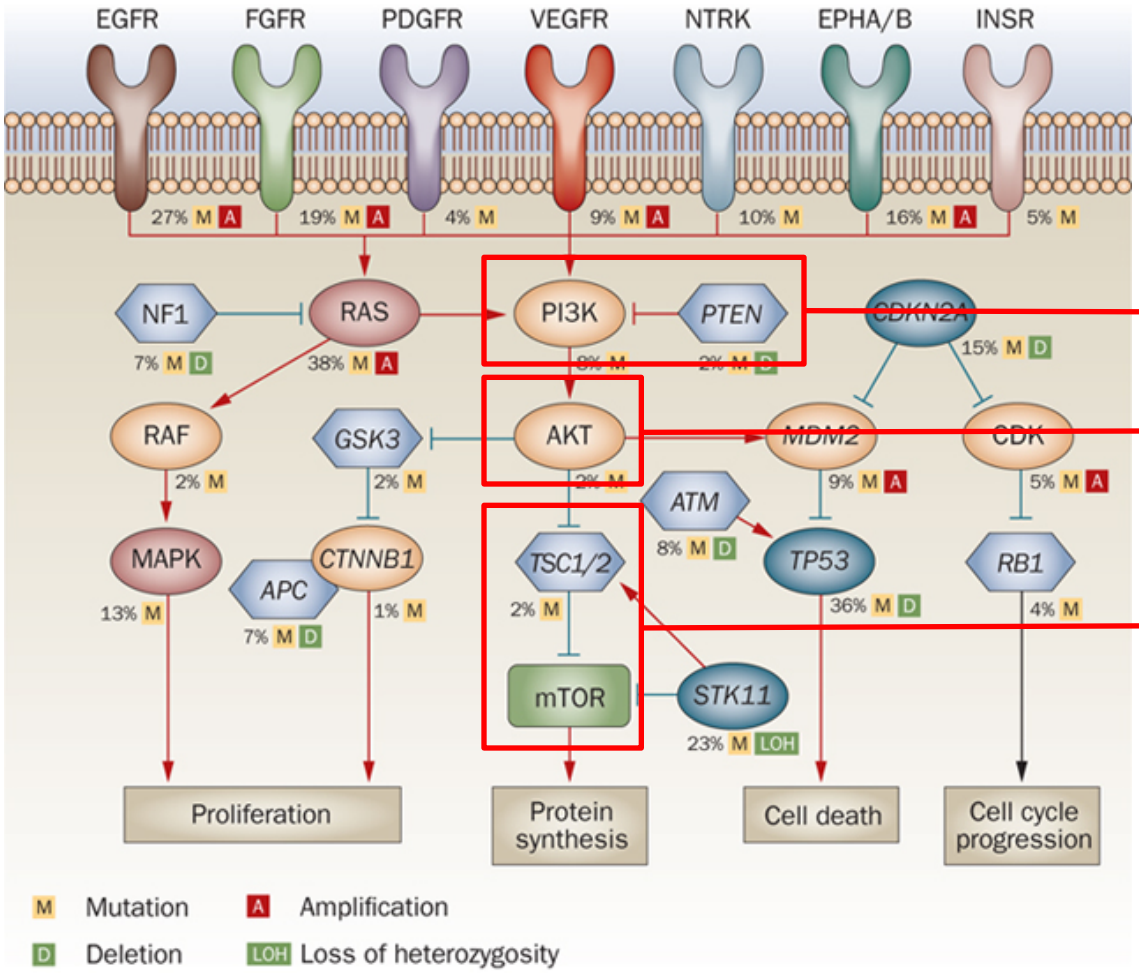
CDKN2A

MDM2, MDM4

TP53

PI3K/Akt pathway

- **Survival & Translation**



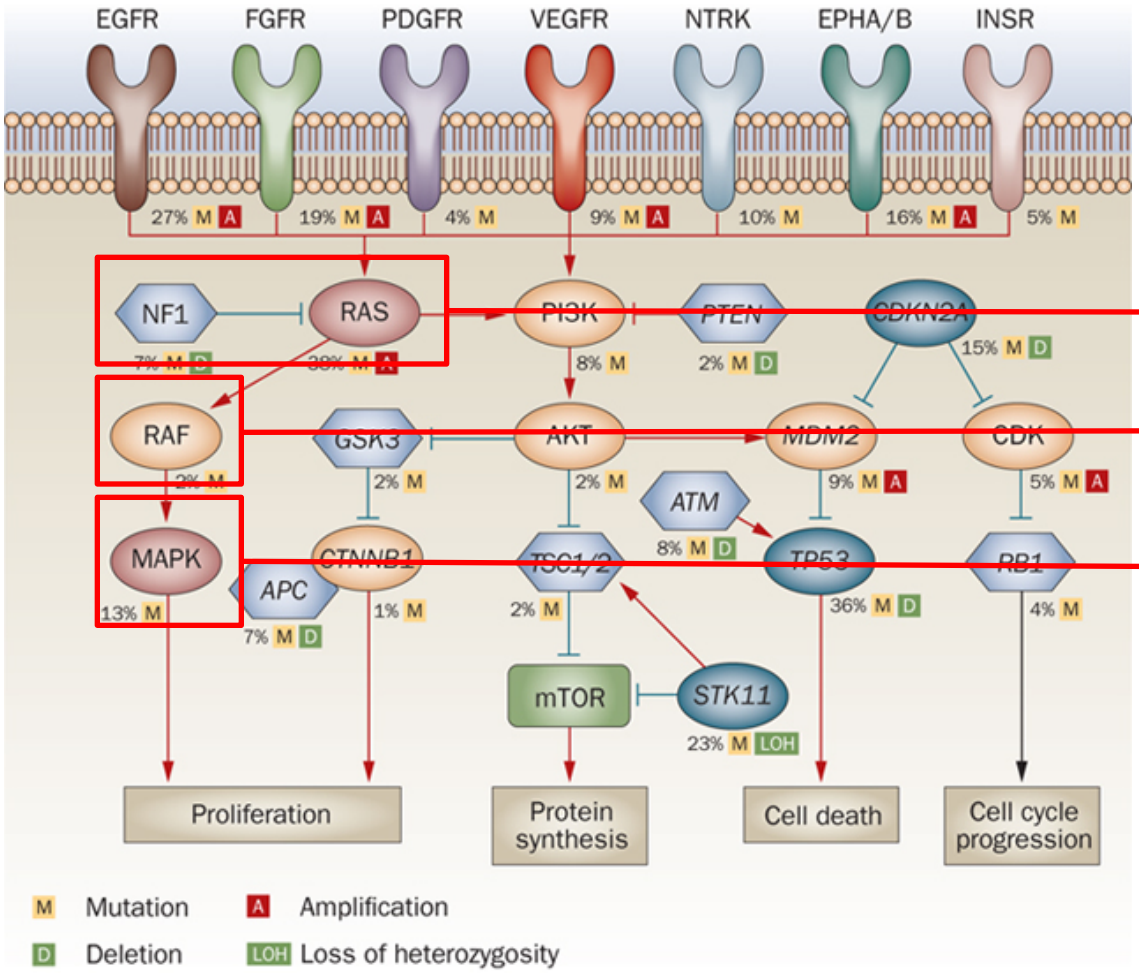
PIK3CA, PIK3R1, PTEN

AKT1

TSC1, TSC2, MTOR

MAPK Pathway

- **Cell growth**



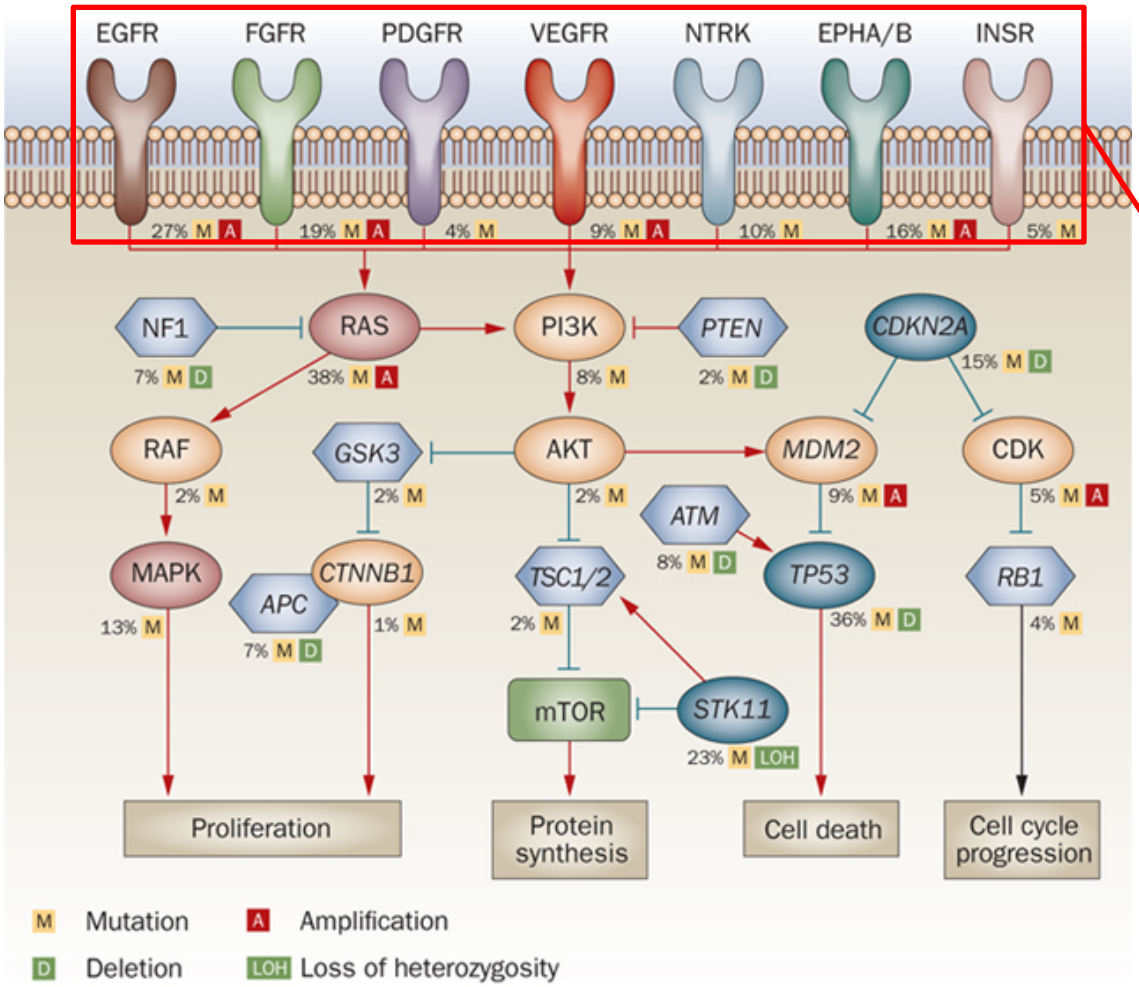
NF1, KRAS, HRAS, NRAS

BRAF

MAP2K1

Receptor Tyrosine Kinases

- **Cell growth**



EGFR, ERBB2, ERBB3
FGFR1
PDGFRA
KDR, KIT, MET
…

A Case Study

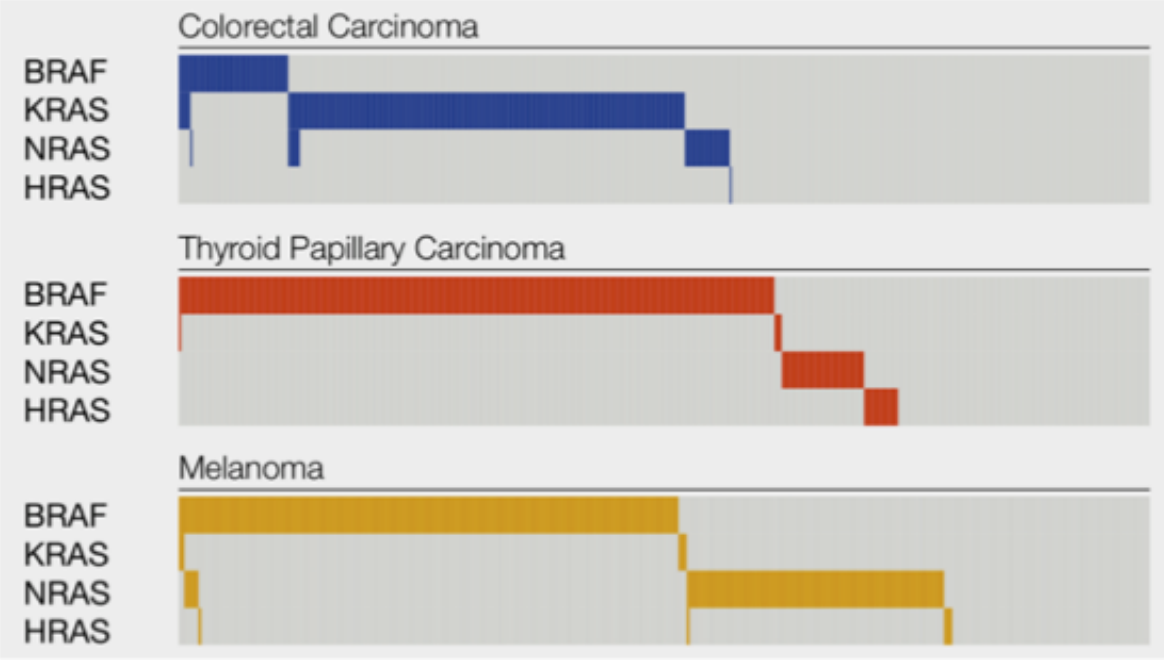
Comprehensive genomic characterization defines human glioblastoma genes and core pathways

The Cancer Genome Atlas Research Network*

http://www.nature.com/nature/journal/v455/n7216/pdf/nature07385.pdf

Mutual Exclusivity

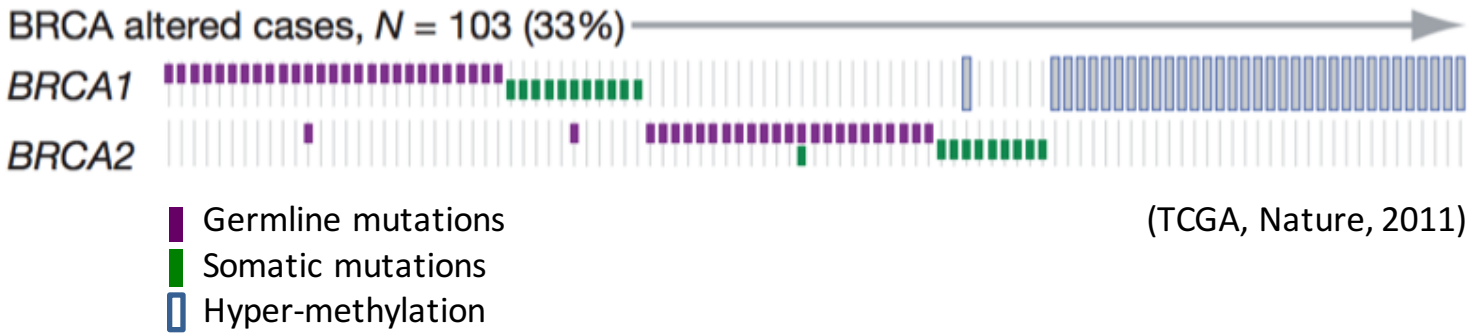
- Observations of mutually exclusive alterations



Patient Samples

Mutual Exclusivity

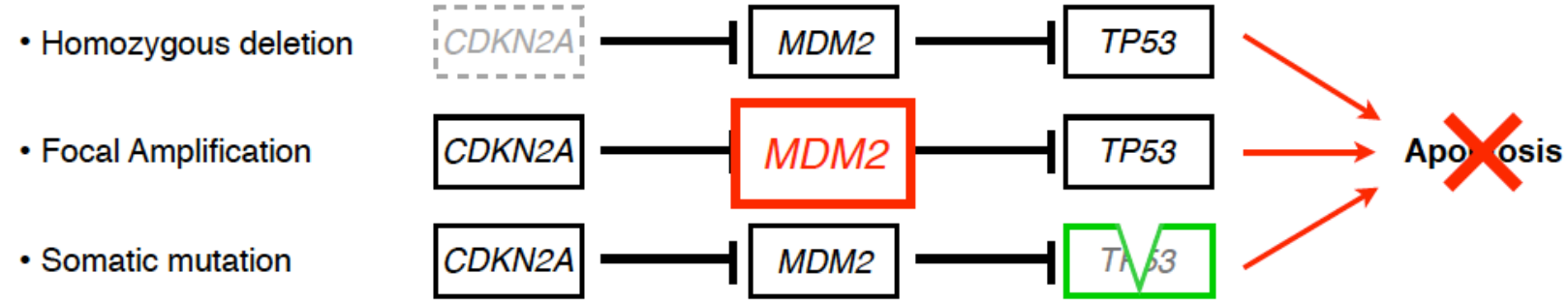
- Observations of mutually exclusive alterations



(TCGA, Nature, 2011)

Why Mutual Exclusivity?

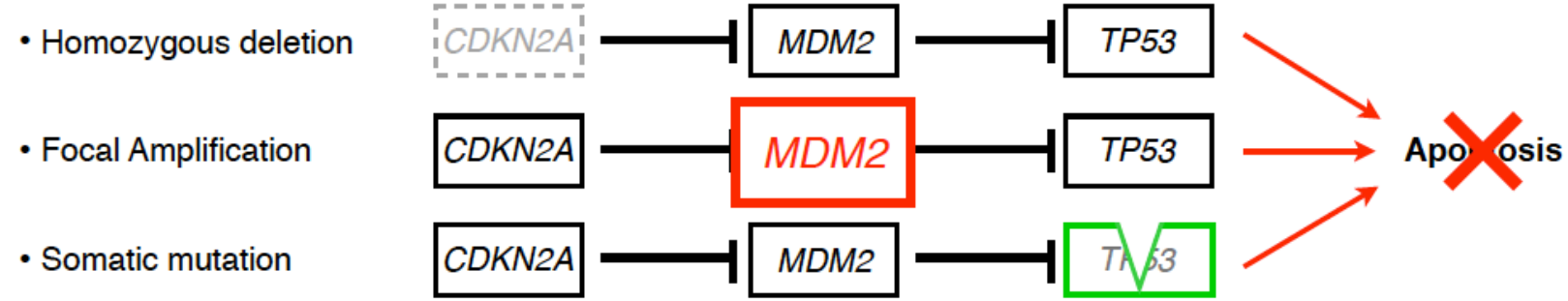
1) Selective Advantage



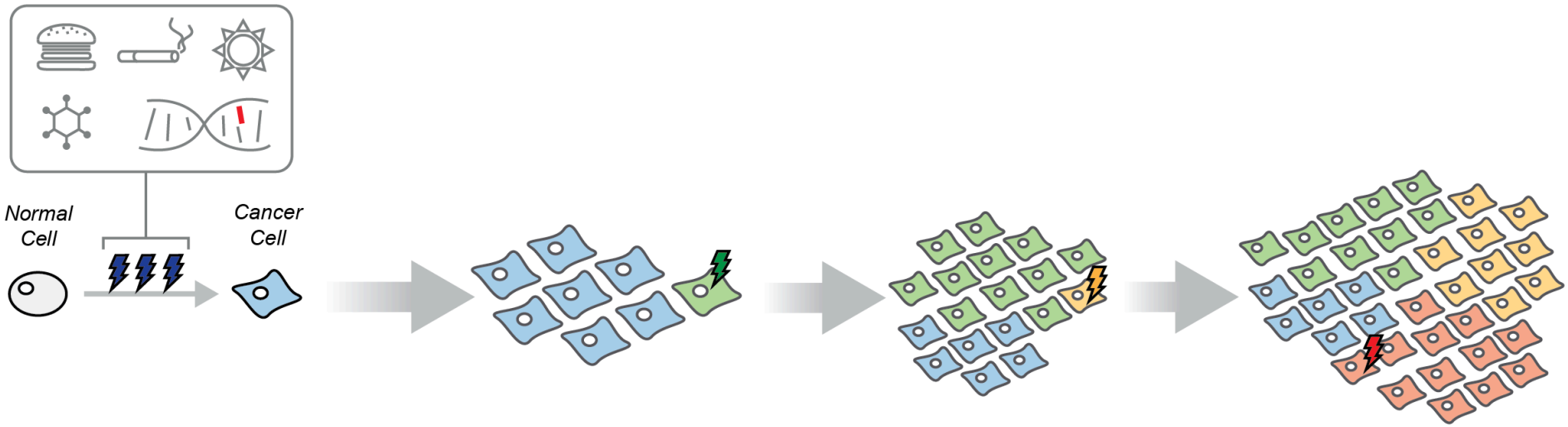
A second hit in the same pathway doesn't offer a further selective advantage

Why Mutual Exclusivity?

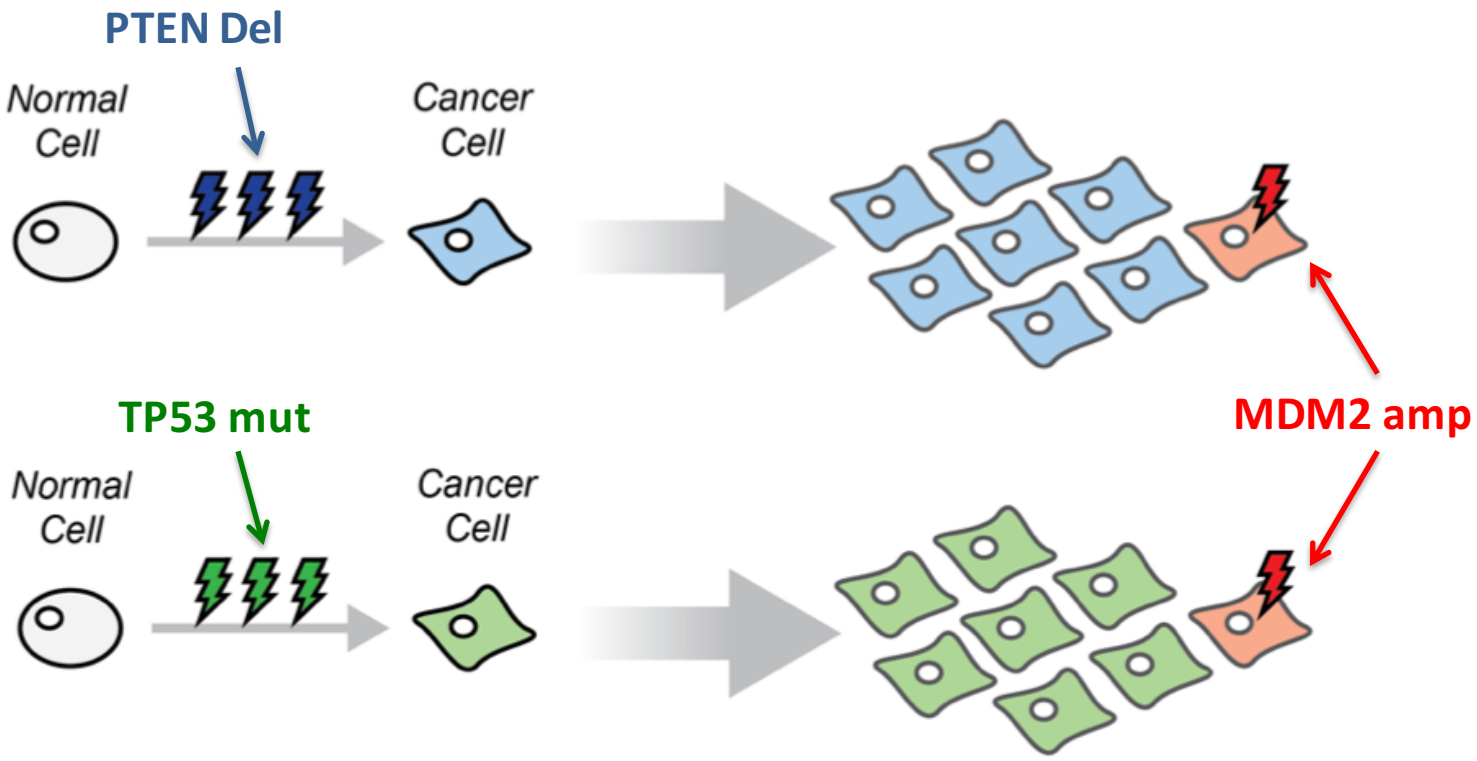
1) Selective Advantage



A second hit in the same pathway doesn't offer a further selective advantage



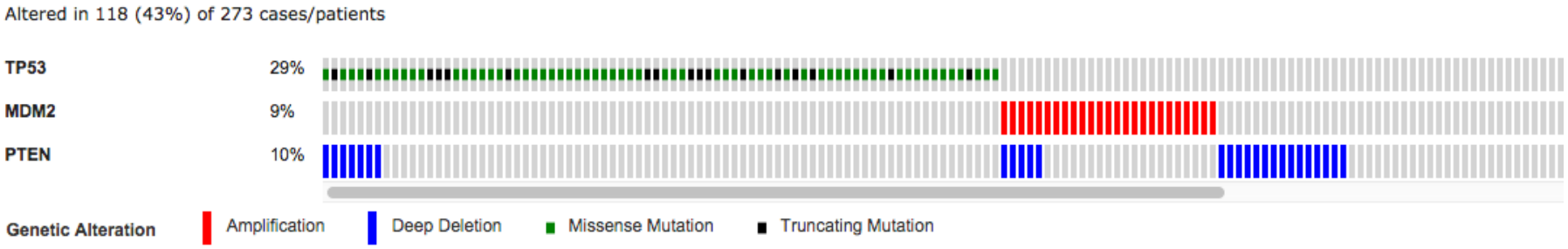
Mutual Exclusivity reflects Selection



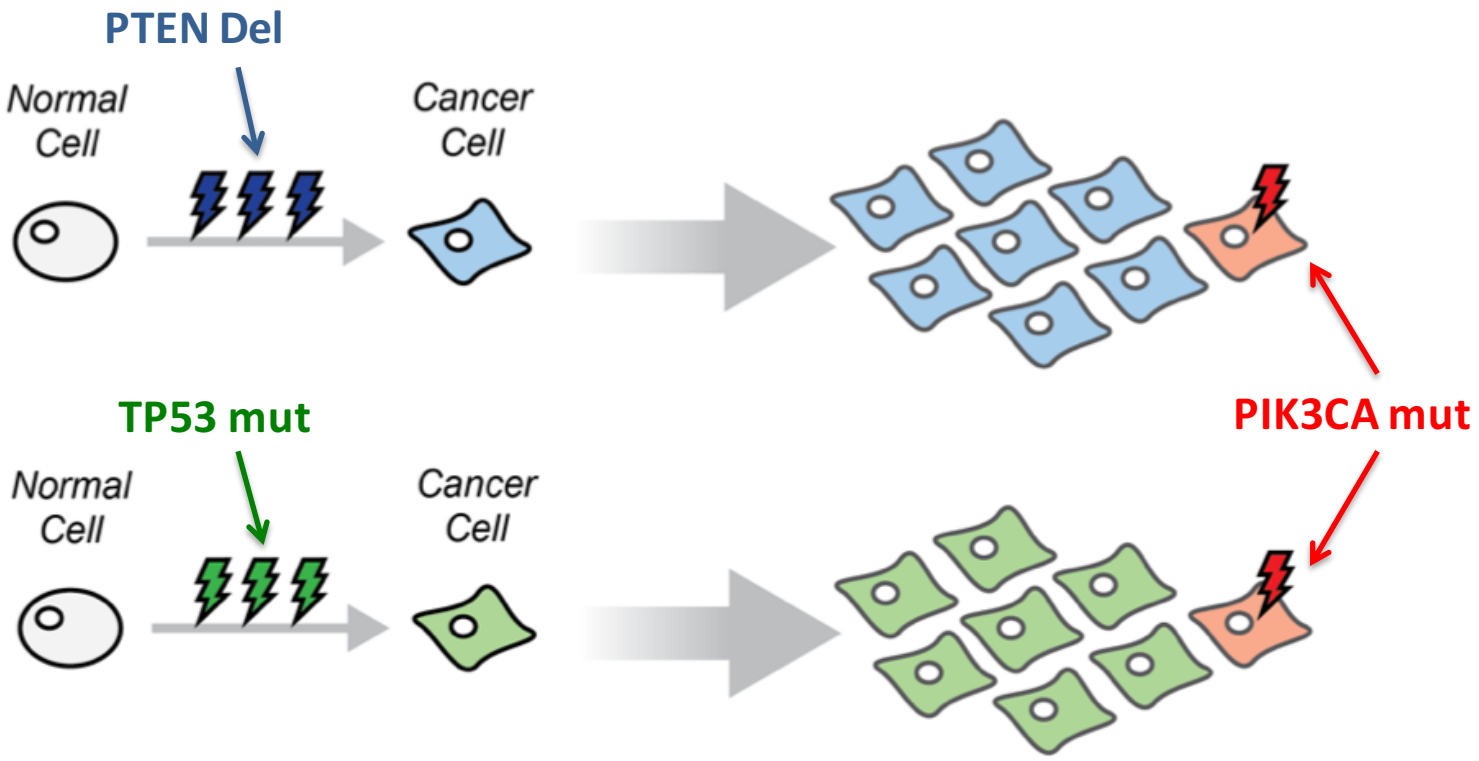
Is MDM2 amplification giving the same advantage in the 2 cases?

Mutual Exclusivity reflects Selection

TCGA Glioblastoma Dataset (source cBioPortal)



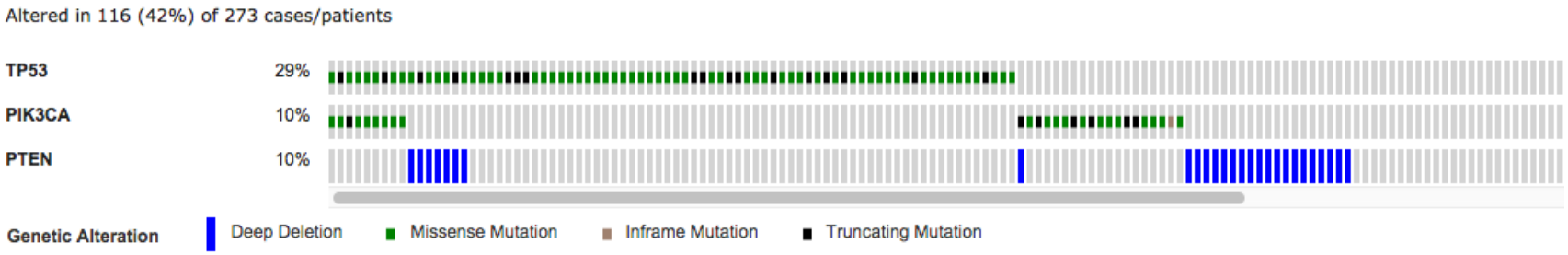
Mutual Exclusivity reflects Selection



Is PIK3CA mutation giving the same advantage in the 2 cases?

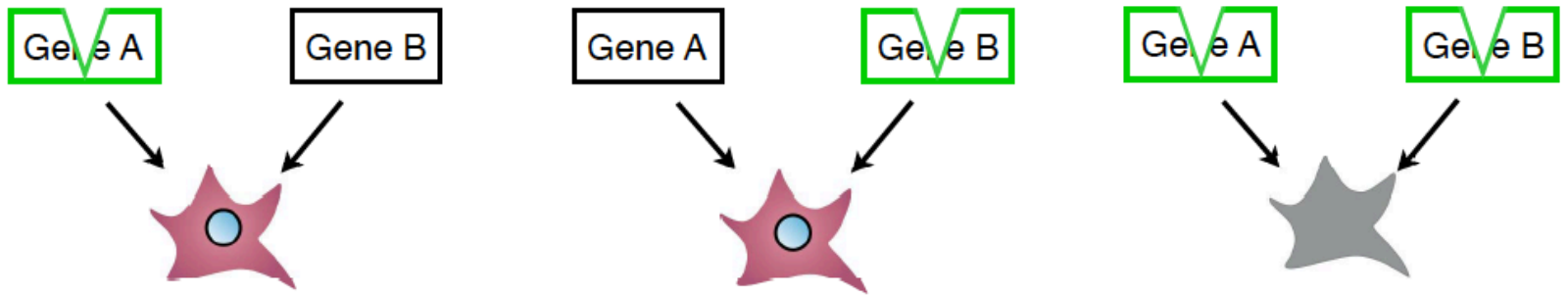
Mutual Exclusivity reflects Selection

TCGA Glioblastoma Dataset (source cBioPortal)



Why mutual exclusivity?

2) Synthetic Lethality

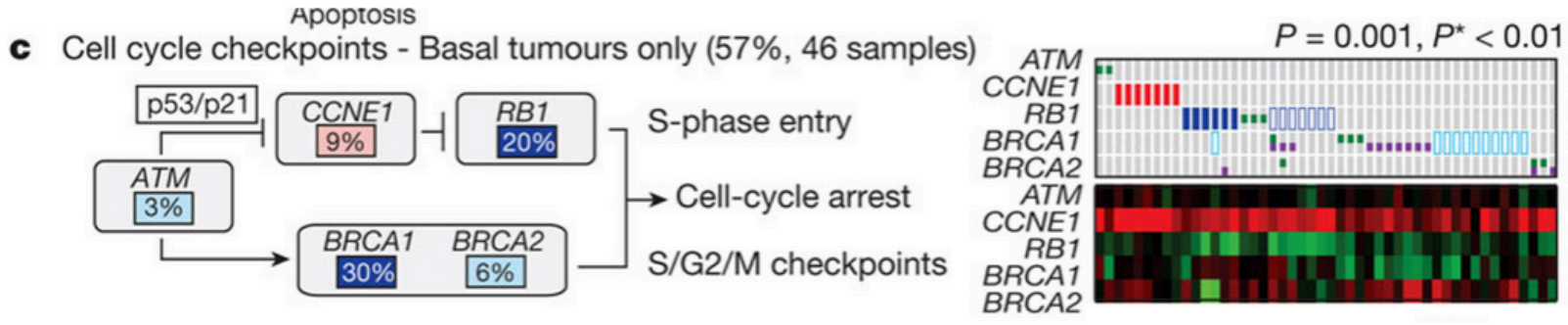


A second hit actually confers a disadvantage!

Synthetic Lethal interactions

Mutual exclusivity between alterations in DNA repair genes BRCA1/2 and cell cycle regulators CCNE1 and RB1 in **ovarian cancer** and **Basal breast cancer**

(Ciriello et al. Genome Res. 2012)

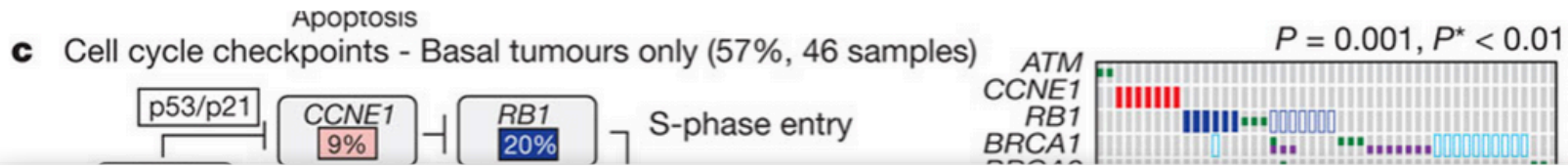


(TCGA, Nature 2012)

Synthetic Lethal interactions

Mutual exclusivity between alterations in DNA repair genes BRCA1/2 and cell cycle regulators CCNE1 and RB1 in **ovarian cancer** and **Basal breast cancer**

(Ciriello et al. Genome Res. 2012)



Synthetic lethality between *CCNE1* amplification and loss of *BRCA1*

(PNAS, 2013)

Dariush Etemadmoghadam[a,b,c], Barbara A. Weir[d,e], George Au-Yeung[a,f], Kathryn Alsop[a,f], Gillian Mitchell[a,b], Joshy George[a,f], Australian Ovarian Cancer Study Group[a,g,h,i,1], Sally Davis[a,c], Alan D. D'Andrea[d], Kaylene Simpson[b,c,j], William C. Hahn[d,e], and David D. L. Bowtell[a,b,c,f,2]

Why it is important?

If one alteration is **functional** and **sufficient** to deregulate a pathway activity, then a second alteration in the same pathway won't provide a further selective advantage

Why it is important?

If one alteration is **functional** and **sufficient** to deregulate a pathway activity, then a second alteration in the same pathway won't provide a further selective advantage

- Critical players of specific cellular processes
- Put alterations in a functional context
- Identify most relevant pathways in a tumor

Why it is important?

If one alteration is **functional** and **sufficient** to deregulate a pathway activity, then a second alteration in the same pathway won't provide a further selective advantage

If one alteration is **functional** and **sufficient** to deregulate a pathway activity, then therapeutically targeting that alteration will be enough to restore the pathway activity

Why it is important?

If one alteration is **functional** and **sufficient** to deregulate a pathway activity, then a second alteration in the same pathway won't provide a further selective advantage

If one alteration is **functional** and **sufficient** to deregulate a pathway activity, then therapeutically targeting that alteration will be enough to restore the pathway activity

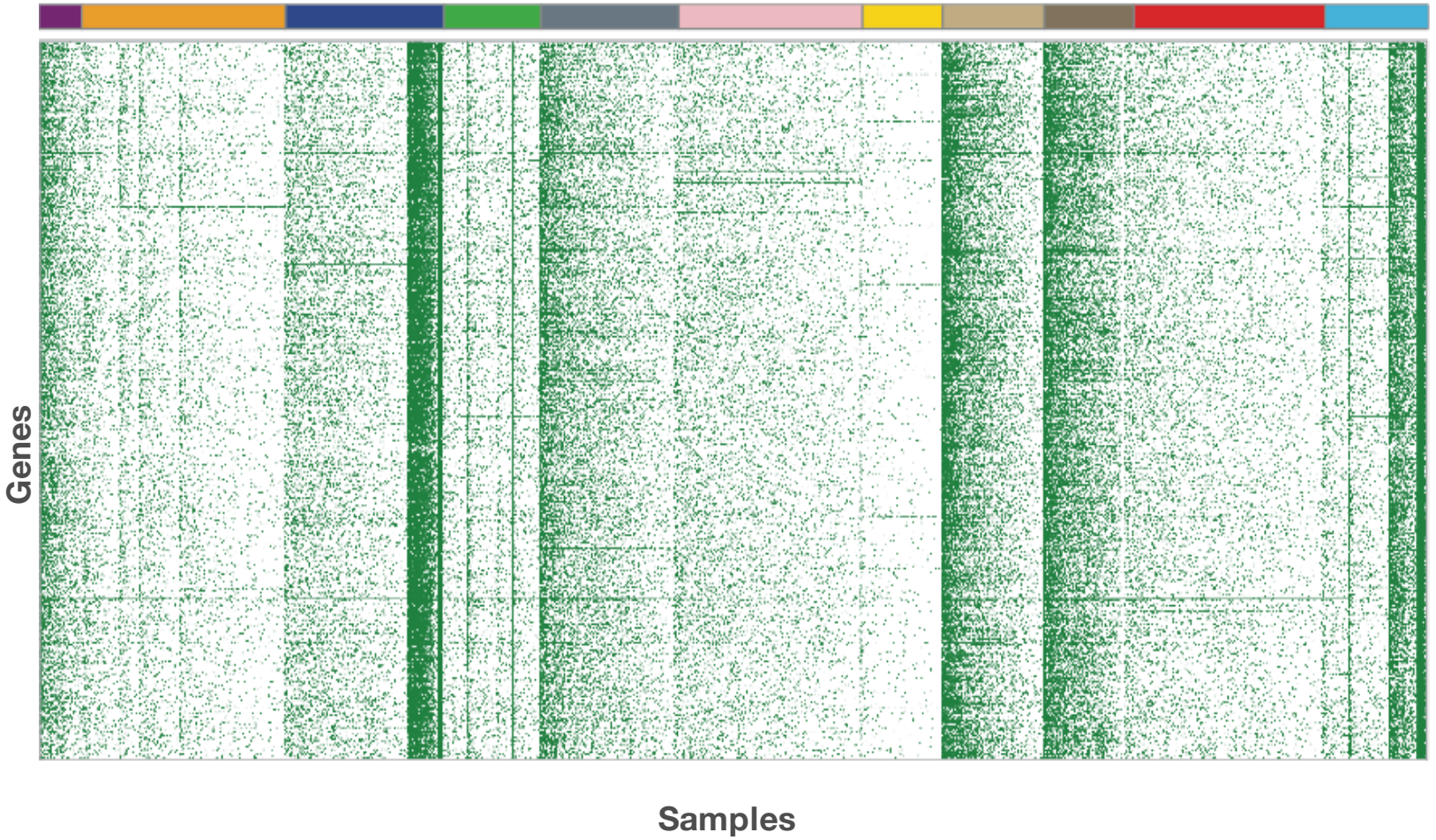
How do we identify significantly mutually exclusive patterns of alterations?

Key Steps:

- Identify *selected* alterations
- Determine which are *functionally related*
- Statistically evaluate their *mutual exclusivity*

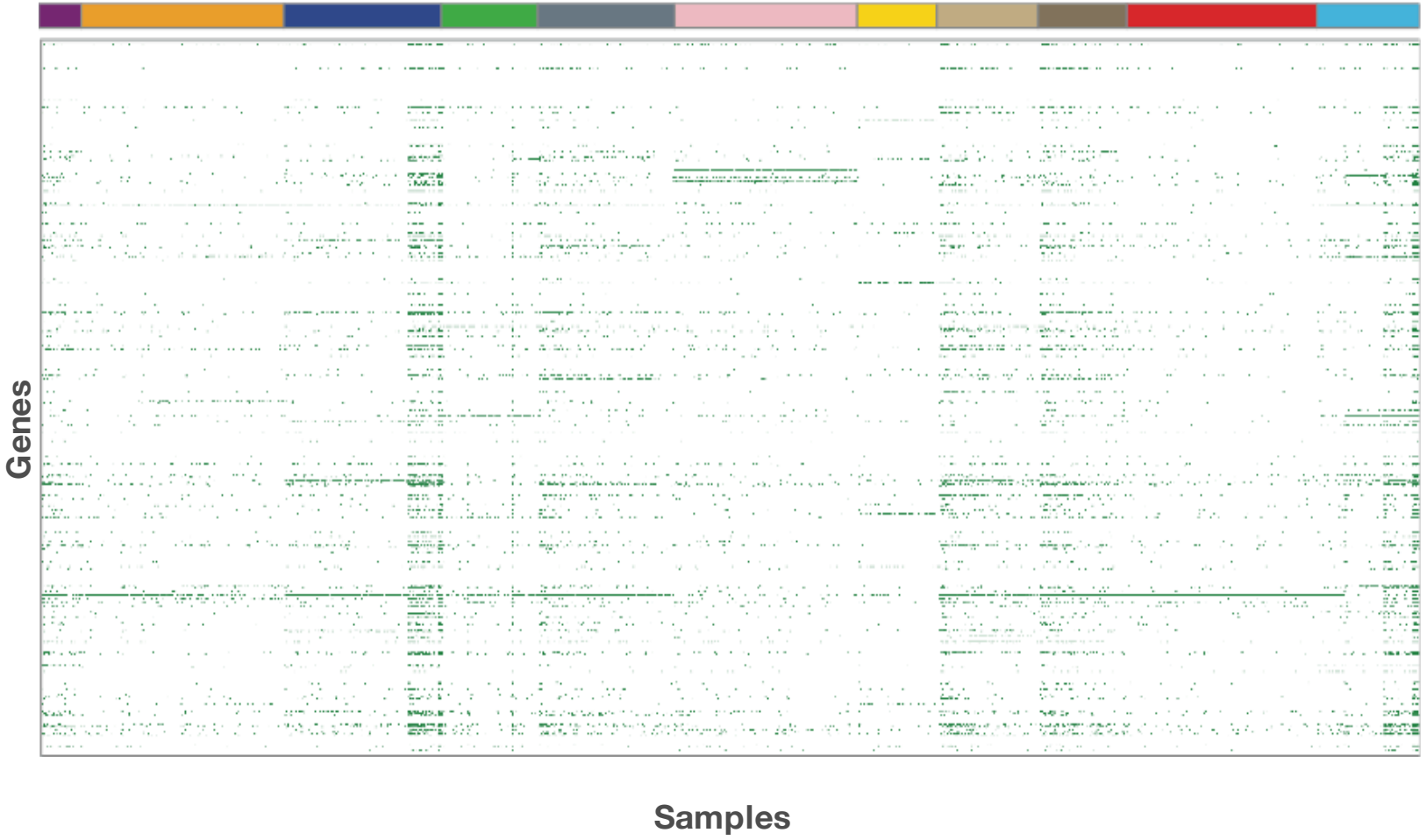
Tumor Molecular Profiles

Somatic mutations across 12 tumor types



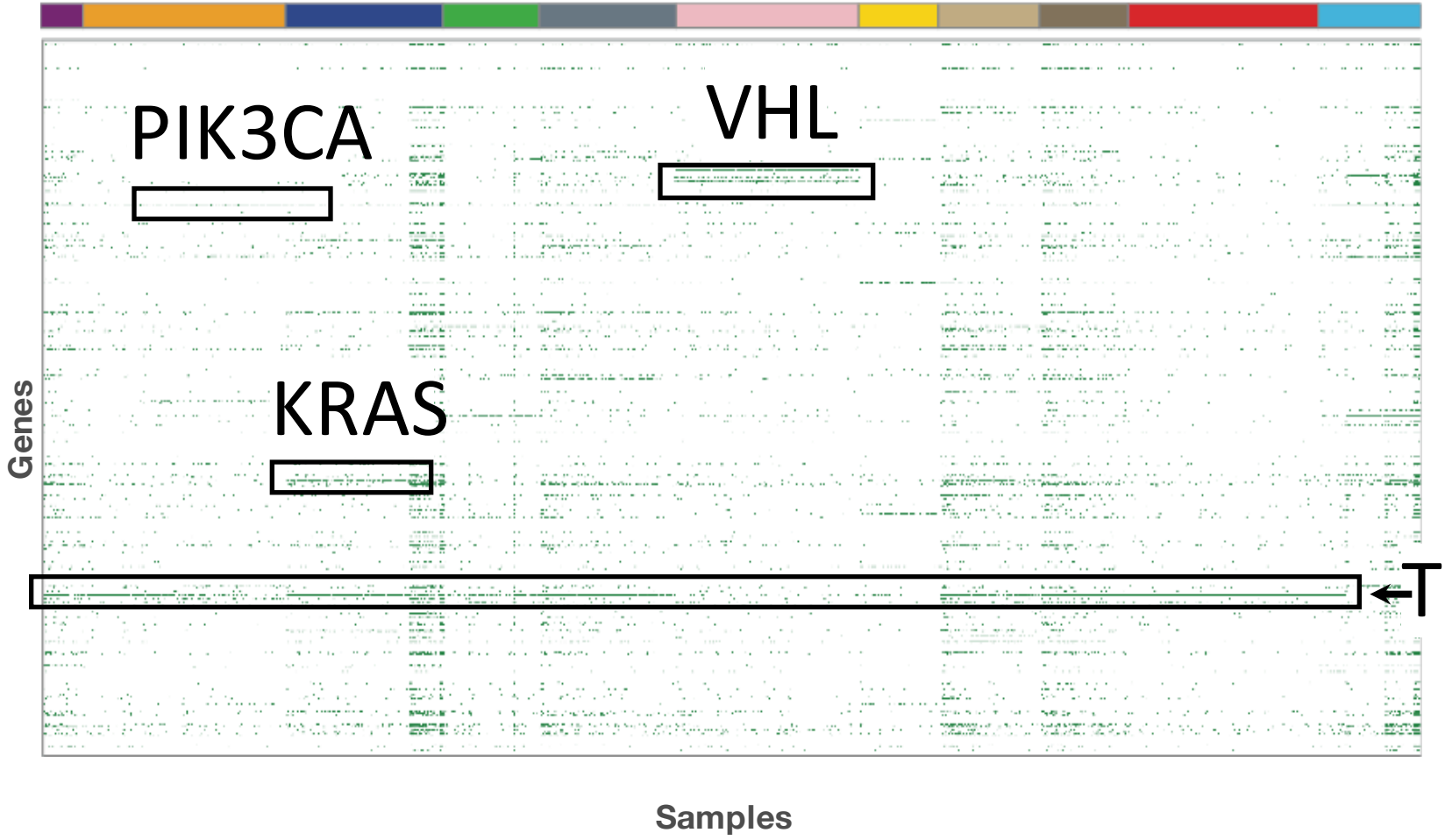
Tumor Molecular Profiles

Candidate driver mutations across 12 tumor types



Tumor Molecular Profiles

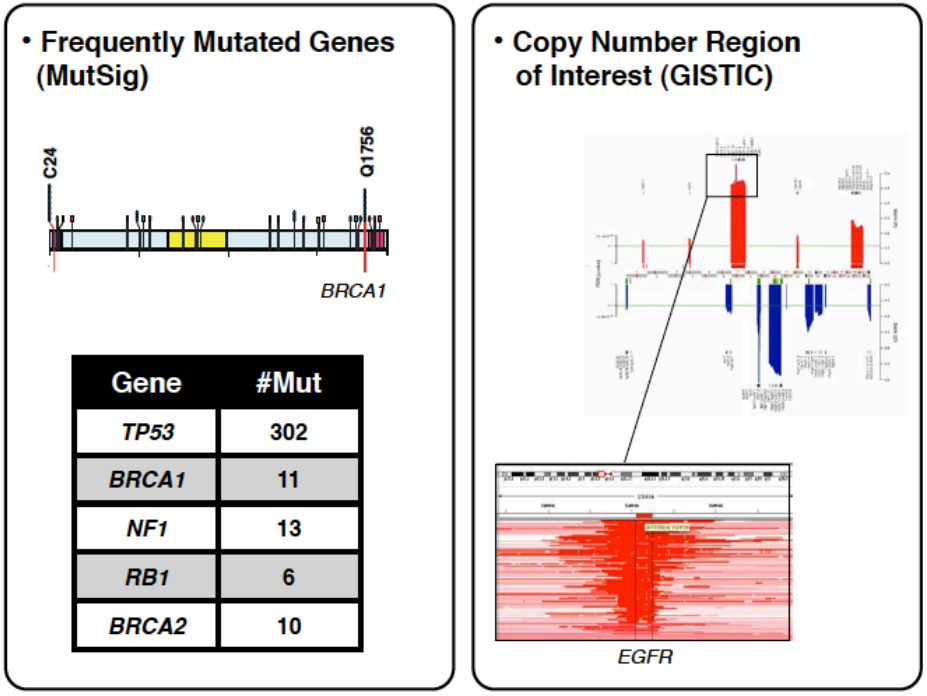
Candidate driver mutations across 12 tumor types



MEMo

1. Identify *selected* alterations

- **MutSig / MuSiC**
  - Recurrent mutations in cancer
- **GISTIC**
  - Recurrent Copy Number Alterations

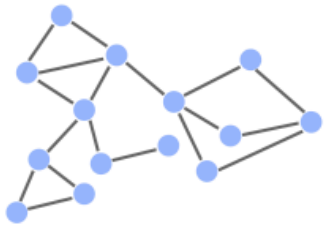


MEMo

2. Determine which are *functionally related*

STEP 1

Gene Network

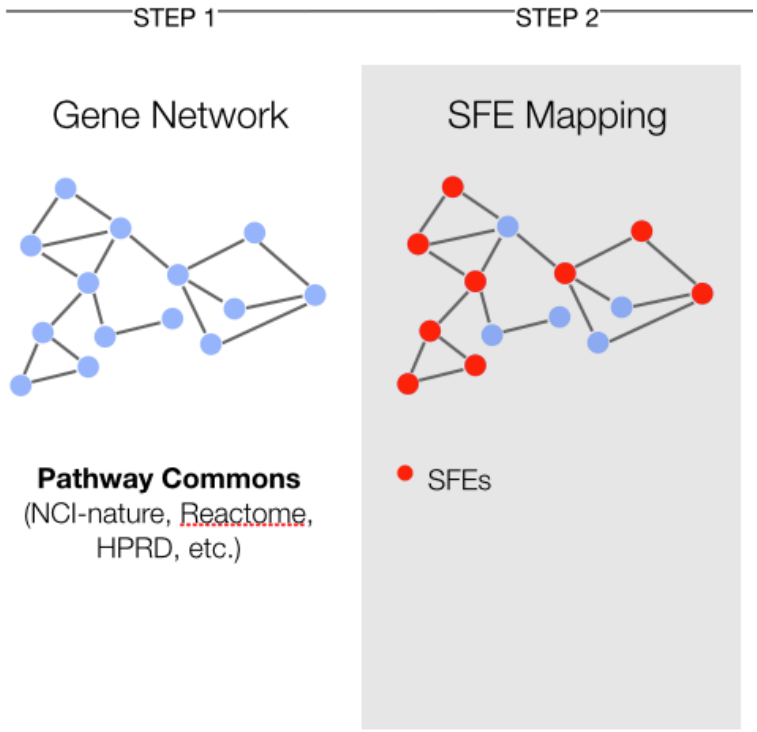


Pathway Commons
(NCI-nature, Reactome, HPRD, etc.)

- Pathway Commons
- National Cancer Institute nature
- Reactome
- Human Protein Reference Database
- The Cancer Cell Map — Memorial Sloan-Kettering Cancer Center

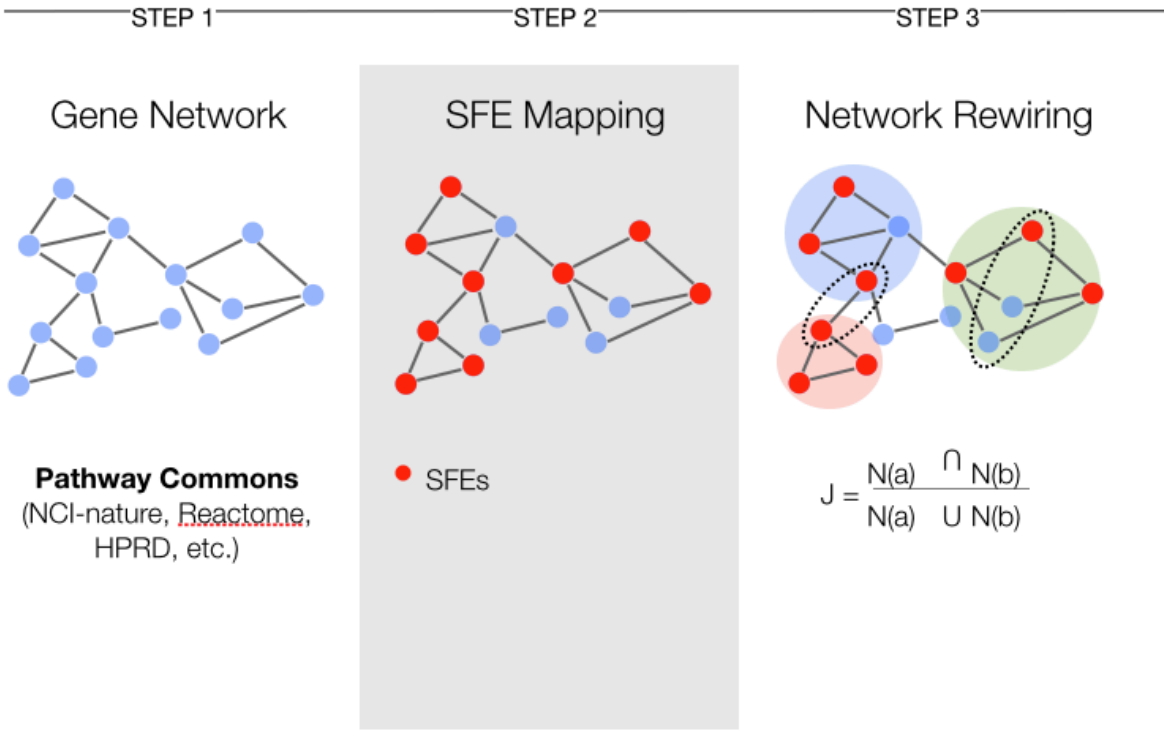
MEMo

2. Determine which are *functionally related*



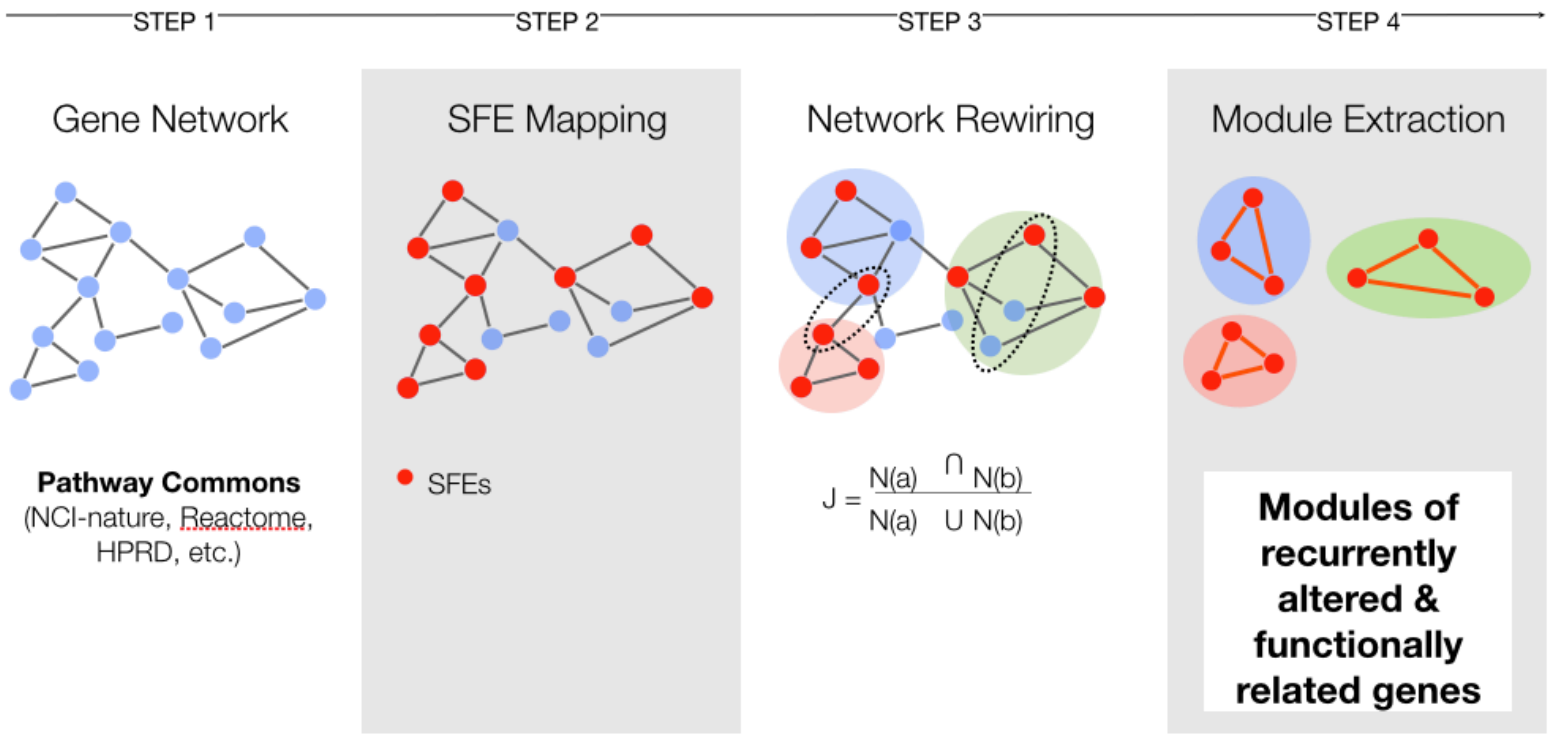
MEMo

2. Determine which are *functionally related*



MEMo

2. Determine which are *functionally related*



MEMo

3. Test the alterations in the module for mutual exclusivity

Alterations are “significantly”
mutually exclusive
if they occur together less frequently
than expected.

What do you expect?

Your *expectations* should preserve all the properties of the system
Except the one you're testing

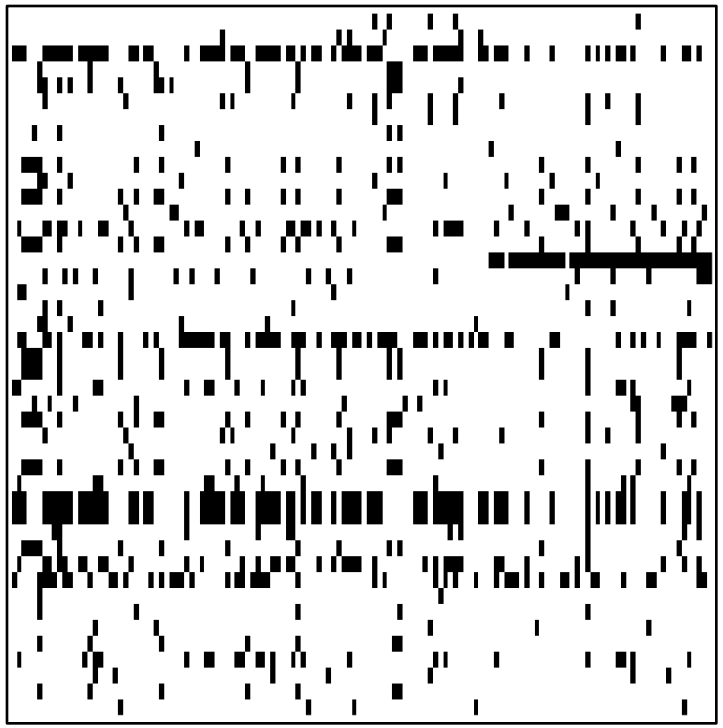
What do you expect?

Your *expectations* should preserve all the properties of the system
Except the one you're testing

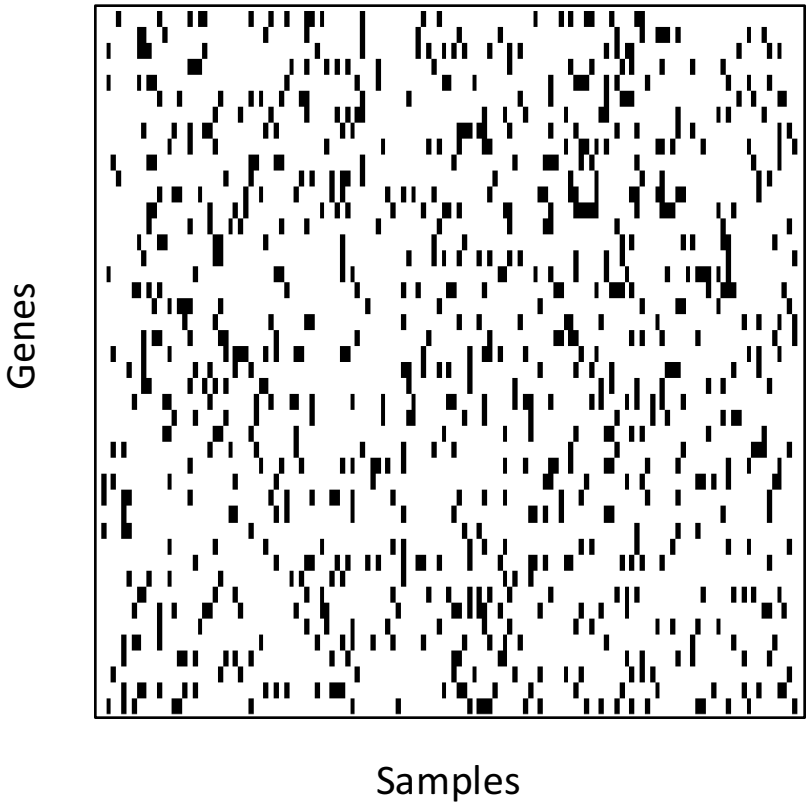
How do you test/model your expectations?

What do you expect?

Observed



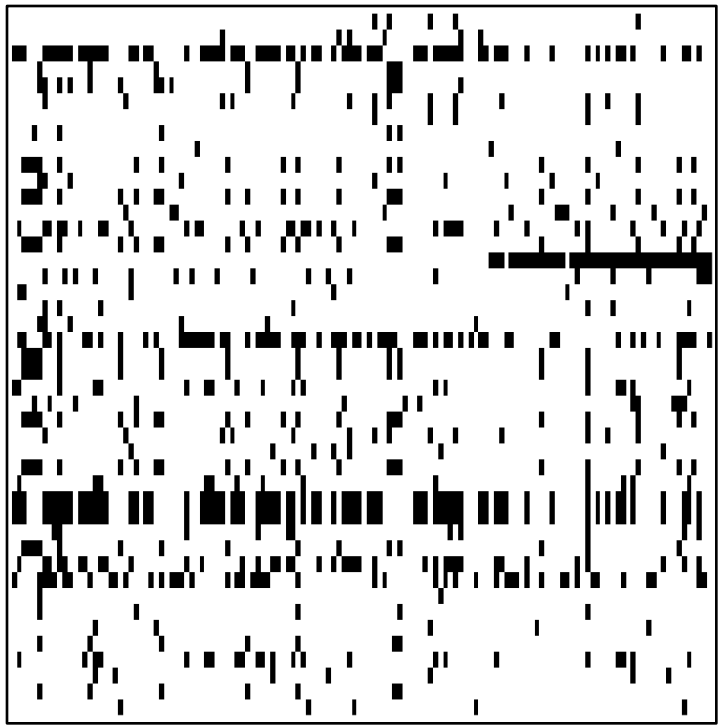
Random 1
"Complete Shuffling"



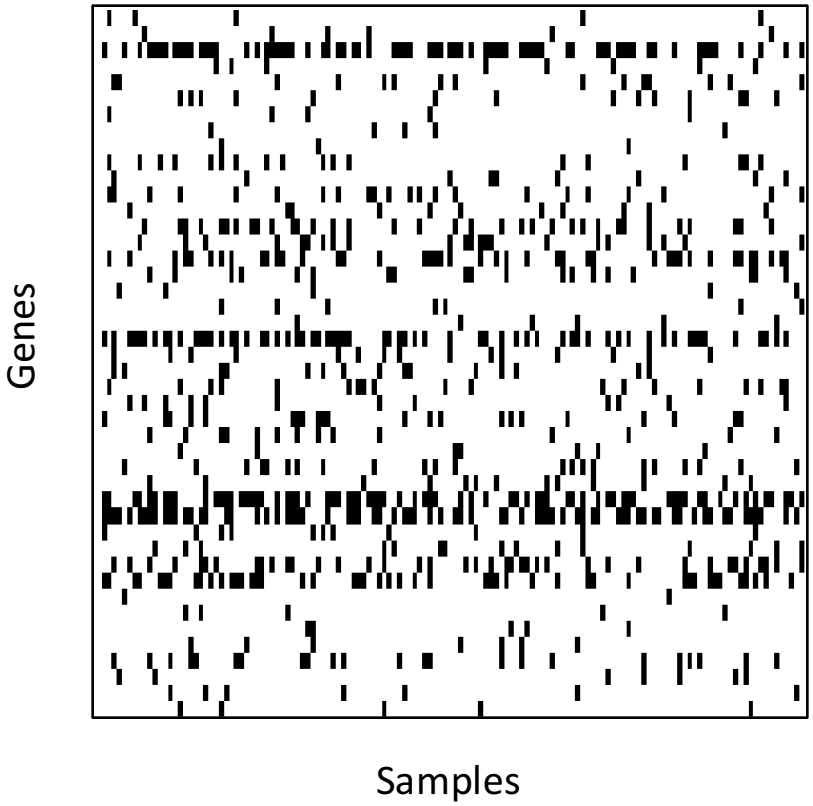
Both matrices have exactly 847 black cells

What do you expect?

Observed



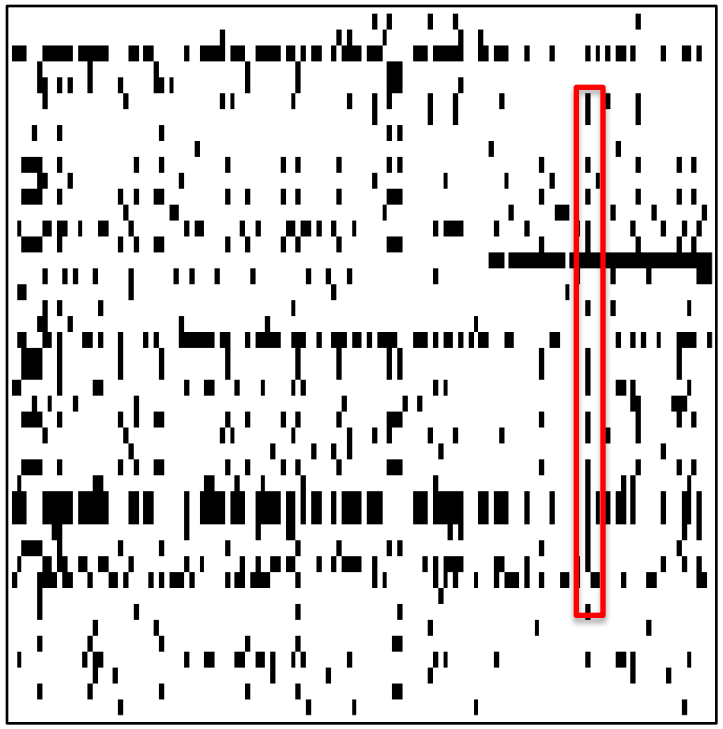
Random 2
"Rows Shuffling"



Here, I preserved the number of alterations on each row

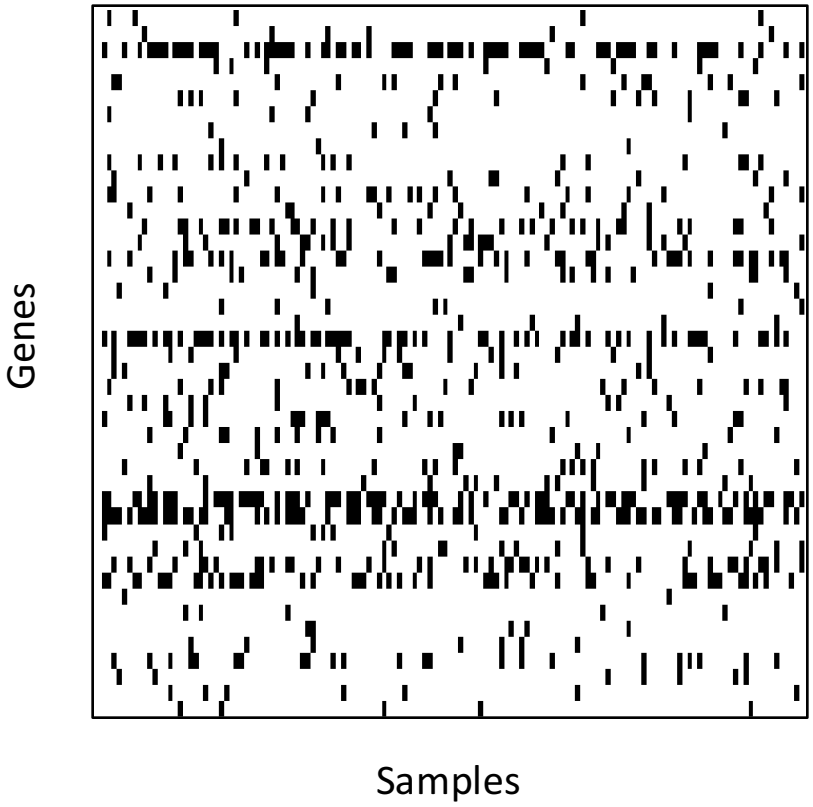
What do you expect?

Observed



Random 2

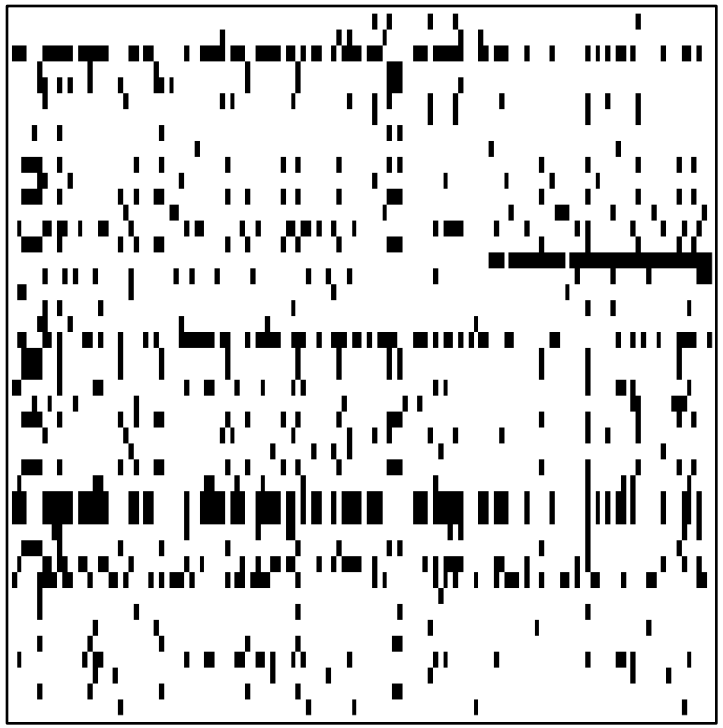
"Rows Shuffling"



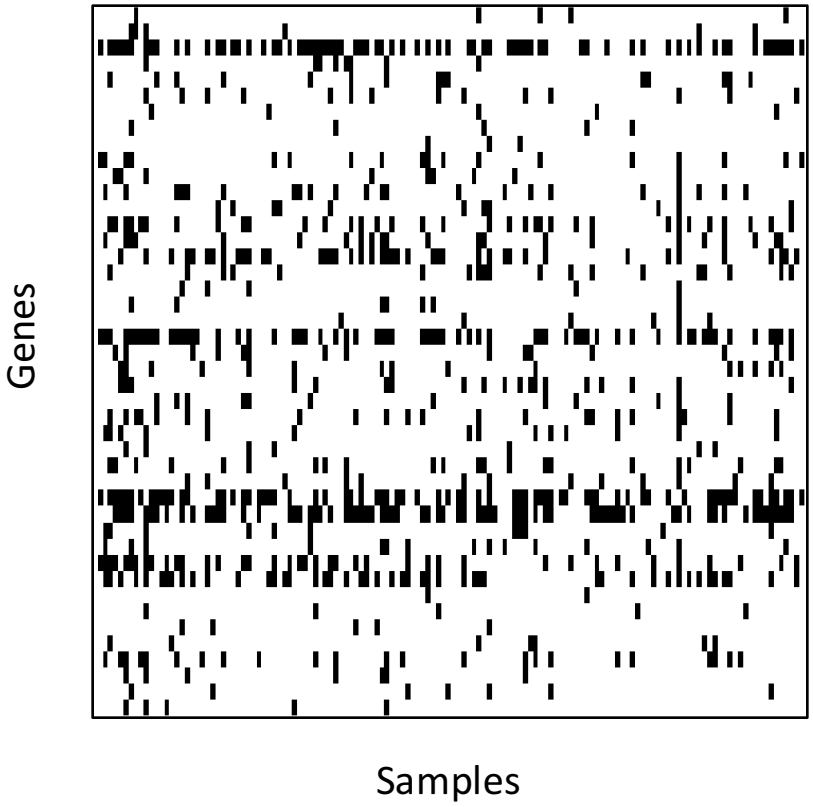
Here, I preserved the number of alterations on each row

What do you expect?

Observed



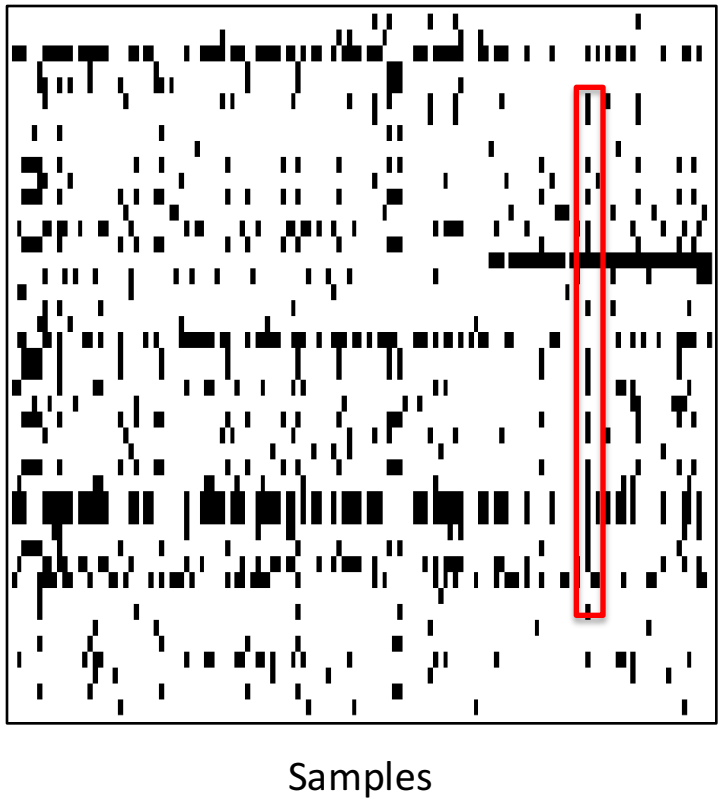
Random 3
"Switching Permutation"



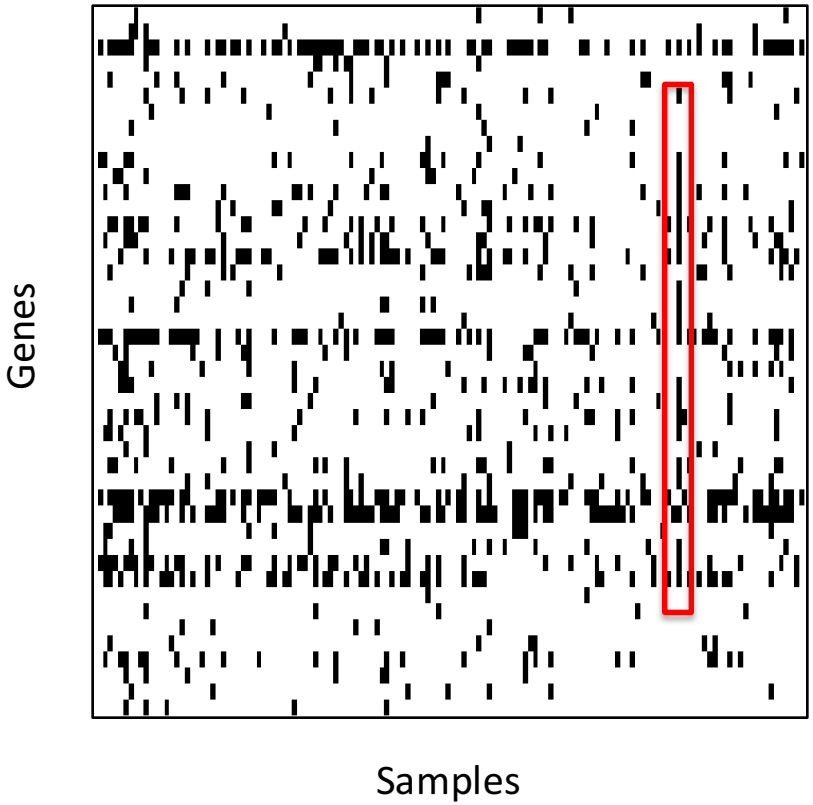
Here, I preserved the number of alterations on each row and column!

What do you expect?

Observed



Random 3
"Switching Permutation"



Both matrices have exactly 847 black cells

What do you expect?

3 null models

- Randomly shuffle the set of alterations with NO constrains
- Randomly shuffle the set of alterations such that the frequency of alteration per gene is identical to the observed
- Randomly shuffle the set of alterations such that the frequency of alteration per gene and per sample is identical to the observed

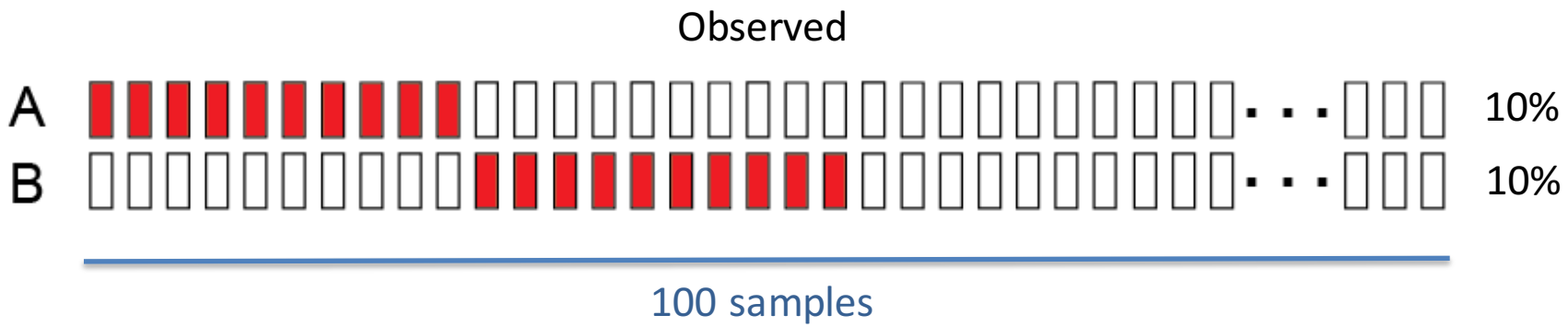
What do you expect?

3 null models

- Randomly shuffle the set of alterations with NO constrains
- Randomly shuffle the set of alterations such that the frequency of alteration per gene is identical to the observed
- Randomly shuffle the set of alterations such that the frequency of alteration per gene and per sample is identical to the observed

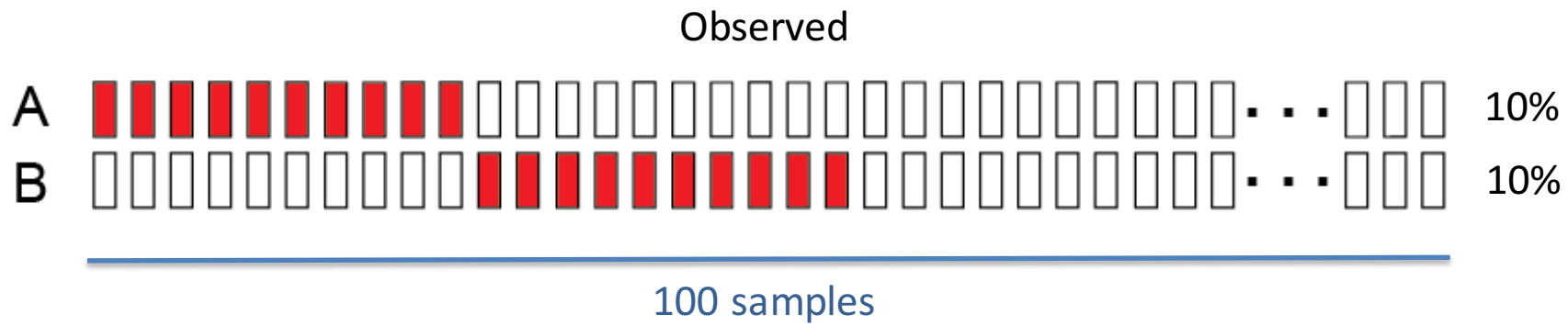
Does this matter when we test mutual exclusivity?

Different expectations lead to different results



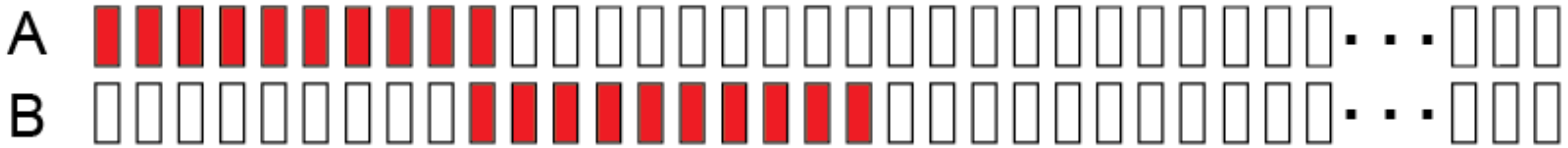
“The expected overlap should be 1, you observe 0, is that relevant?”

Different expectations lead to different results

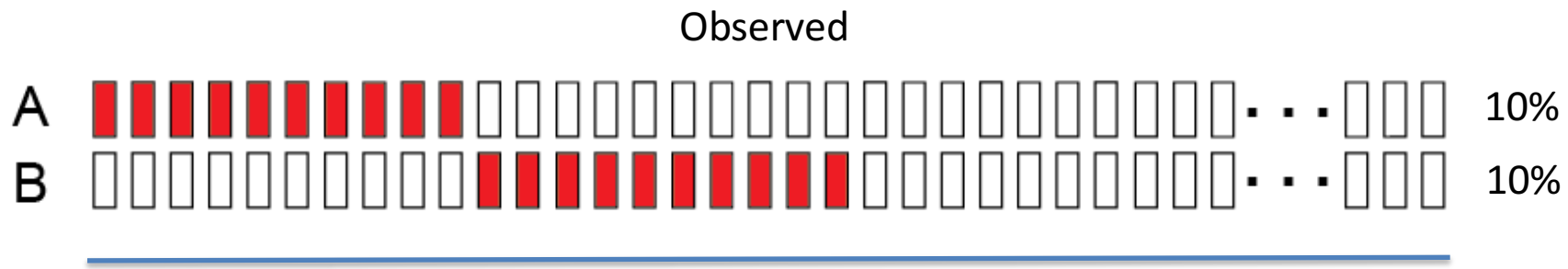


p(A) = 0.1
p(B) = 0.1

p(A,B) = 0.1*0.1 = 0.01 = 1%

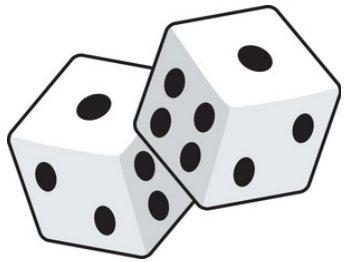


Different expectations lead to different results

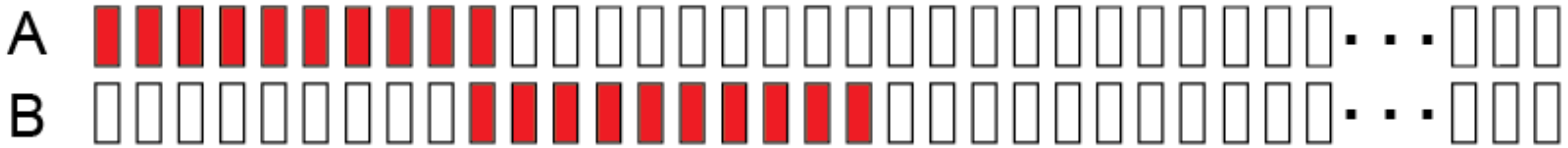


p(A) = 0.1
p(B) = 0.1

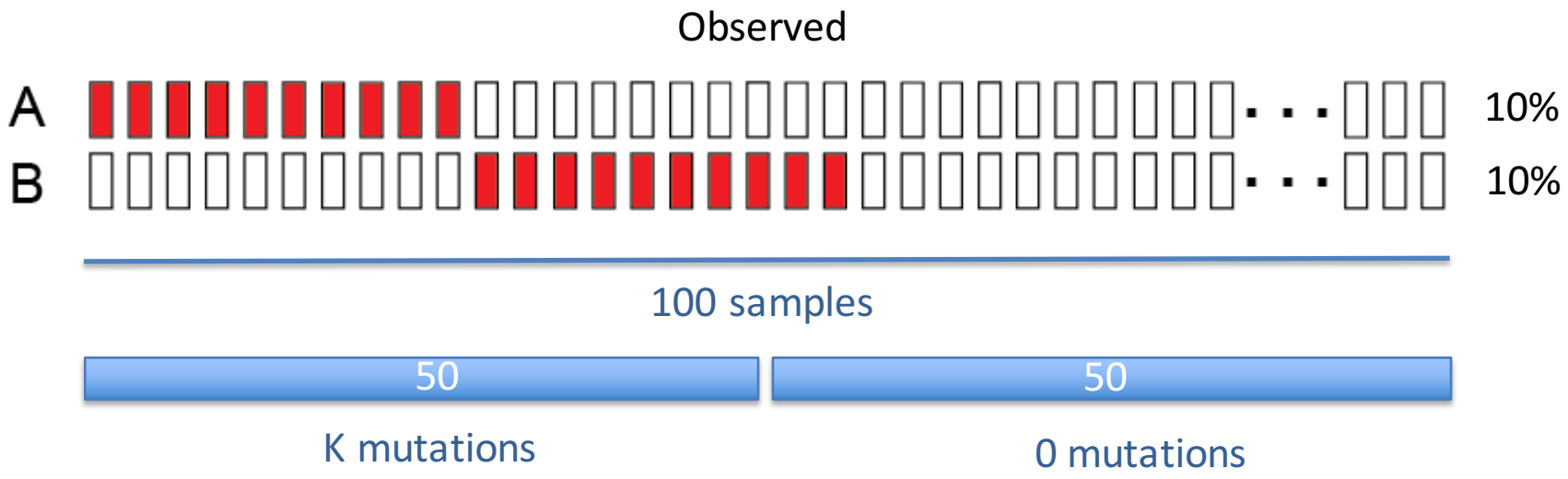
p(A,B) = 0.1*0.1 = 0.01 = 1%



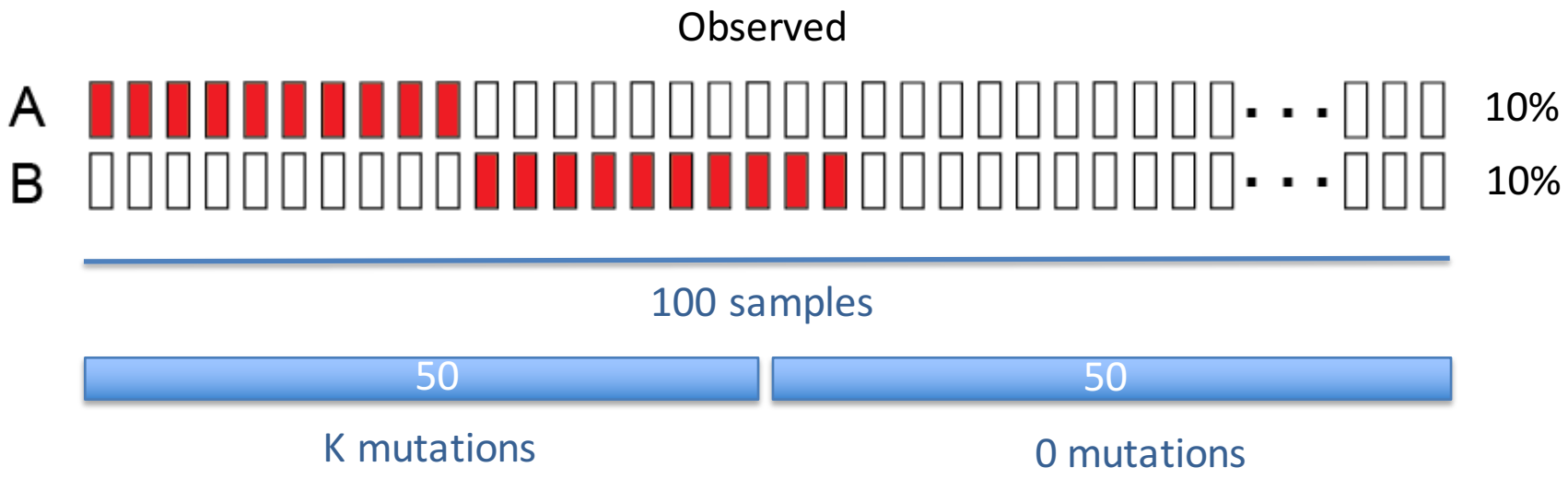
Is the dice fair?



Different expectations lead to different results



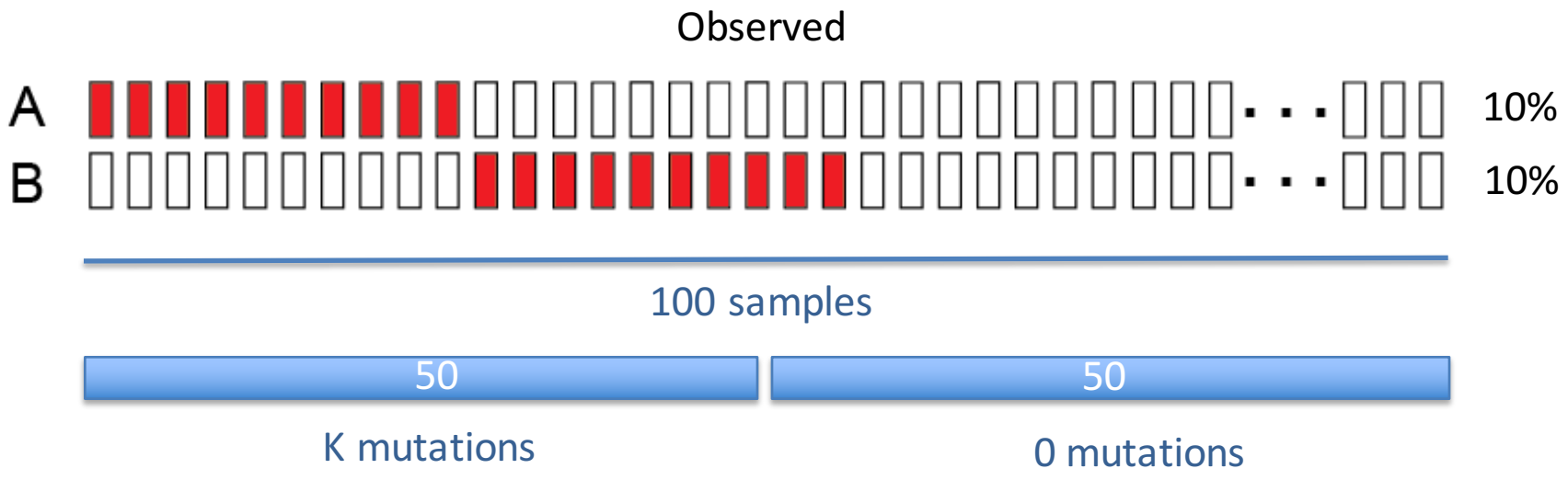
Different expectations lead to different results

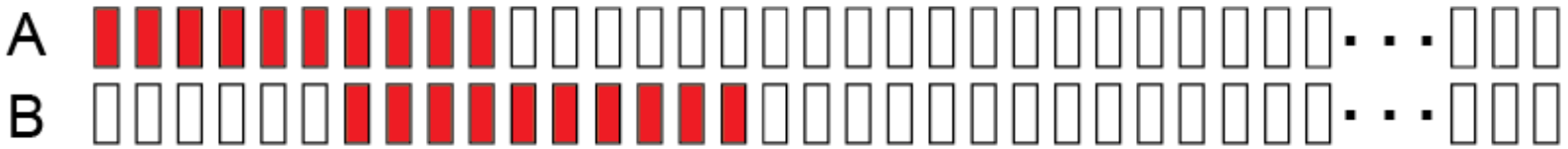


p(A) = 0.2
p(B) = 0.2

p(A,B) = 0.2*0.2 = 0.04 = 4%

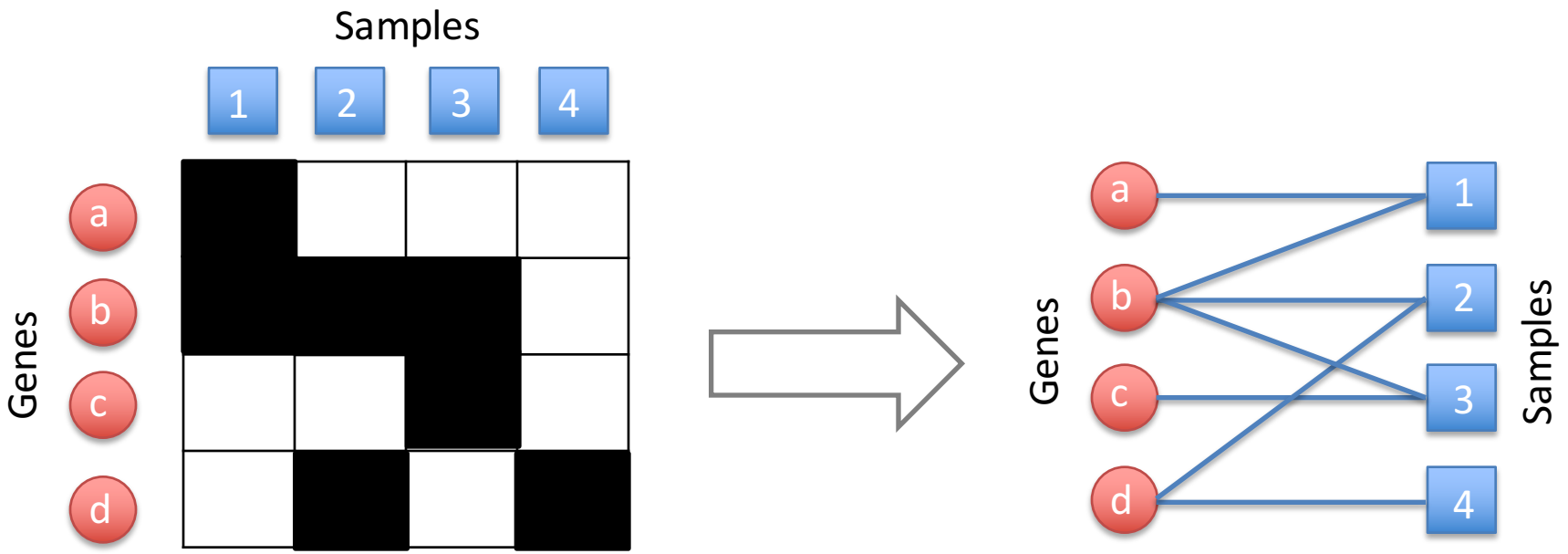
Different expectations lead to different results





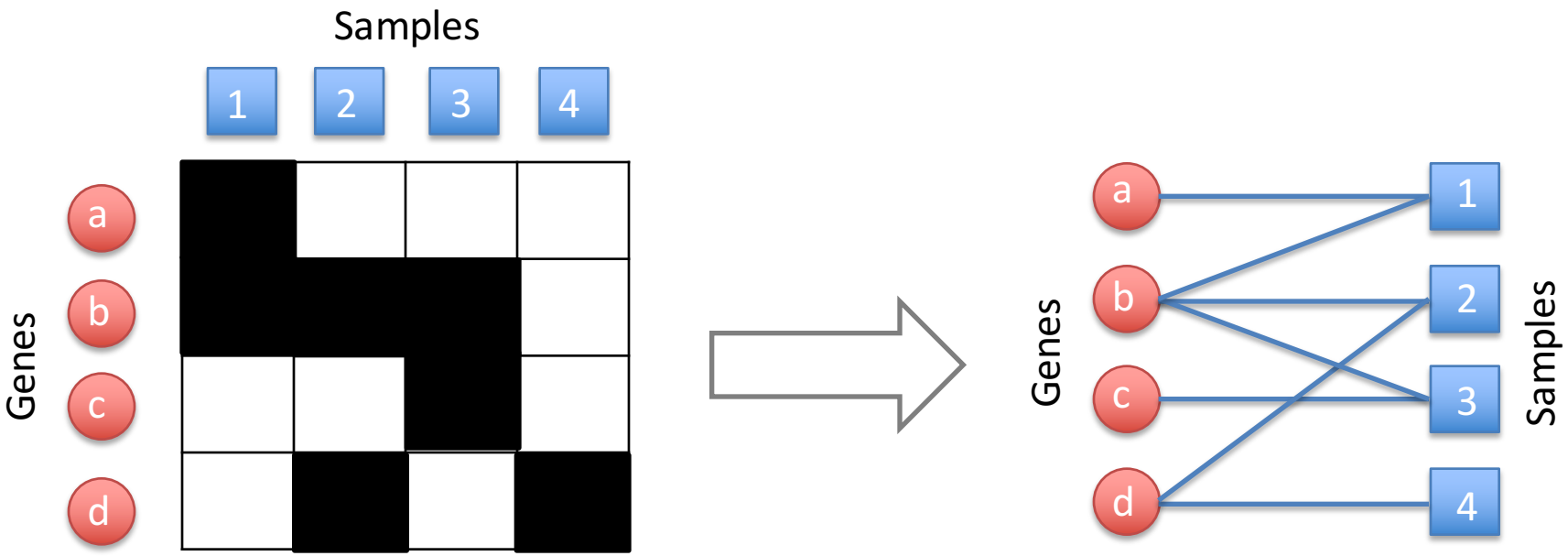
MEMo

3. Test the alterations in the module for mutual exclusivity



MEMo

3. Test the alterations in the module for mutual exclusivity

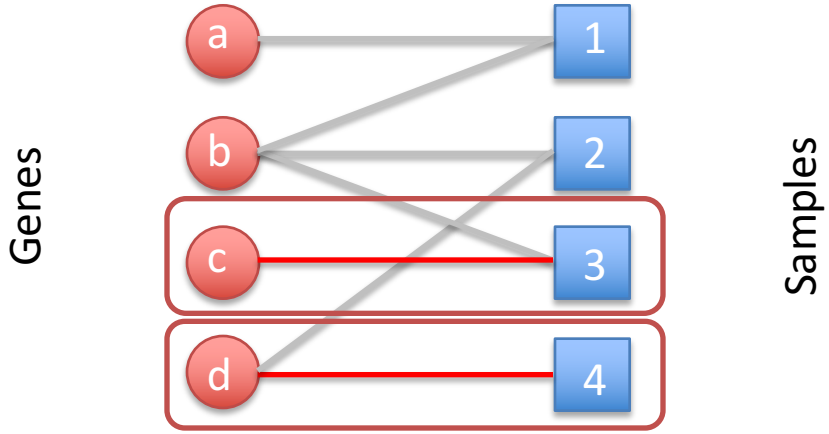


The frequencies of alteration of genes and samples correspond now to the number of edges connected to a node in the network (**degree**)

MEMo

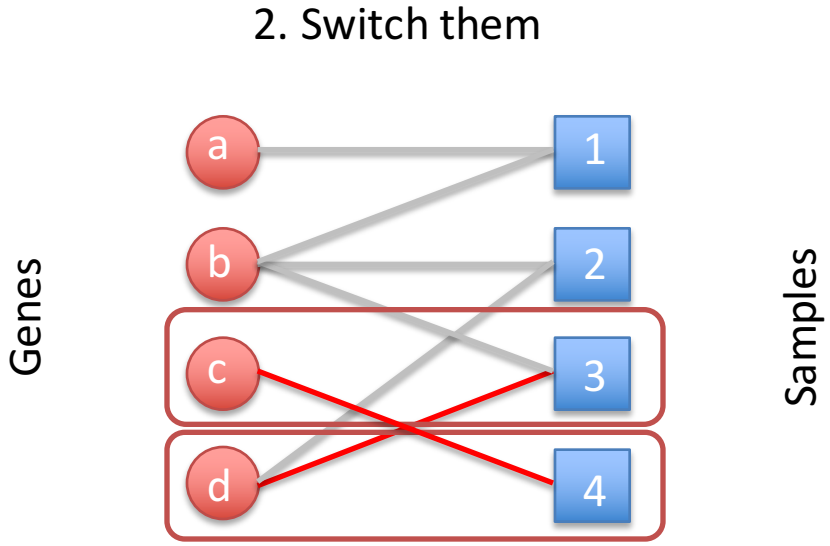
3. Test the alterations in the module for mutual exclusivity

1. Randomly select two edges



MEMo

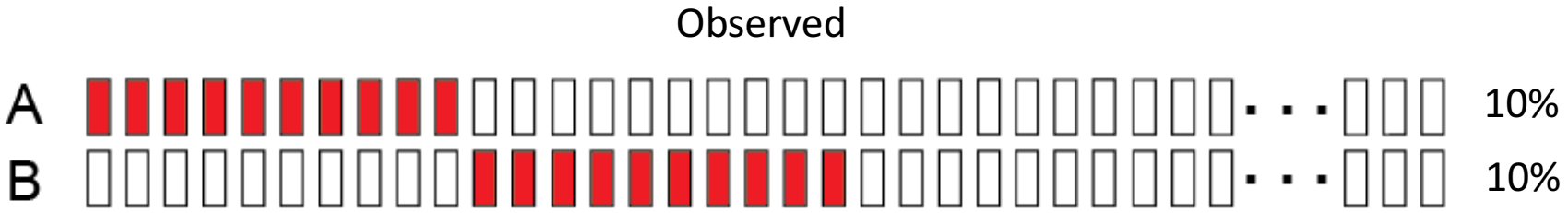
3. Test the alterations in the module for mutual exclusivity



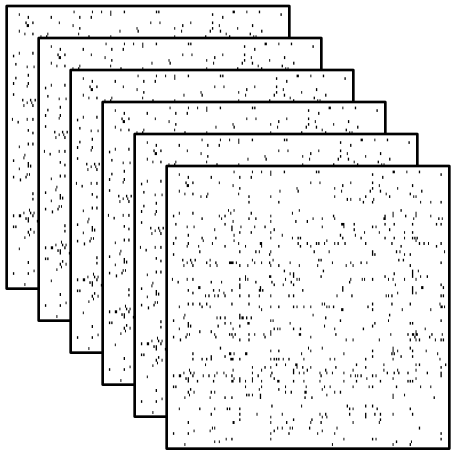
The degree of c, d, 3, and 4 has not changed!

(Switch is valid ONLY if it does not create “double” edges)

Empirical p-value



observedAltered: 20



Look for alterations in A and B across all random matrices.

Count how many times you find A+B > or = to observedAltered

Let's say this is 2 times out of 1000 matrices, then:
p = 2/1000 = 0.002

If p is smaller than a chosen threshold, your result is statistically significant
Typical threshold = 0.05 (THIS IS NOT GOD GIVEN!)

Exercise

- **Dec 14**
  - Load example of genomic data in R
  - Determine the distributions of alterations (genes/samples)
  - Compare the distributions against 3 possible null models
  - Test for mutual exclusivity specific set of modules (from the paper) using 3 null models
- **Dec 15**
  - Select TCGA cancer study (out of 4 proposed)
  - Determine alteration distributions
  - Based on the paper findings, select modules to test
  - Test for mutual exclusivity the modules you select and verify dependence of your results to the null model

Exercise

- **Required R packages**
  - **igraph** (from CRAN)
  - **BiRewire** (from Bioconductor)
    - Install all dependencies