

Simulating pathways and mutual exclusivity

Blanca Lacruz Pleguezuelos

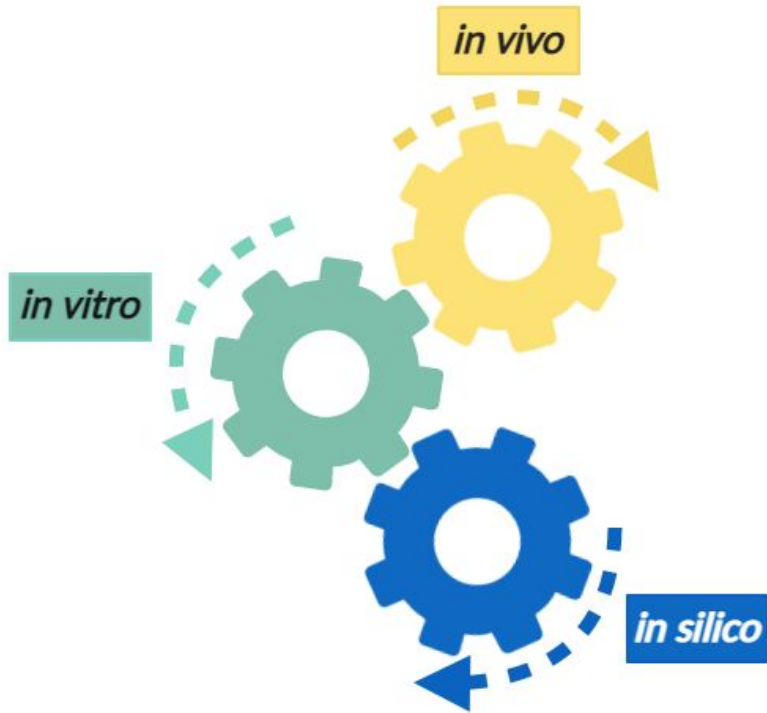
Víctor Mateo Cáceres

Manuel Moradiellos Corpus

Cancer is a **collection of complex diseases** difficult to tackle, many biological variables intertwined

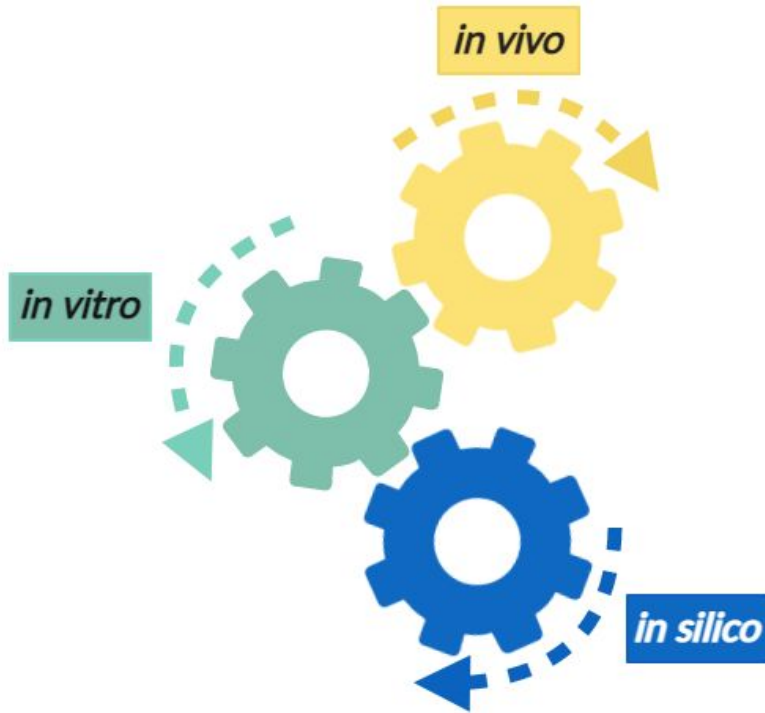
3

Cancer: Challenge and approaches



Cancer is a **collection of complex diseases** difficult to tackle, many biological variables intertwined

In silico **models** integrate data from experimental studies and analytical predictions



Cancer is a **collection of complex diseases** difficult to tackle, many biological variables intertwined

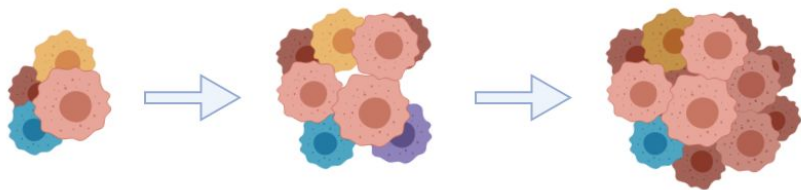
In silico **models** integrate data from experimental studies and analytical predictions

Model **tumor progression** to identify candidate genes, valuable information for:

- Diagnosis
- Treatment
- Prognostic

5 Modeling cancer: CPMs and DAGs

Cancer Progression Models (CPMs)

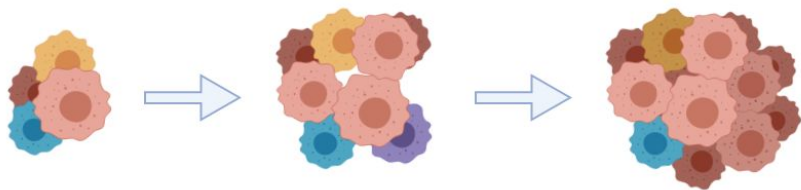


Genotype frequency data from
cross-sectional studies

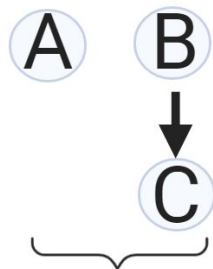
Cancer Progression Models (CPMs)



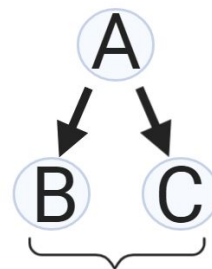
Directed Acyclic Graphs (DAGs)



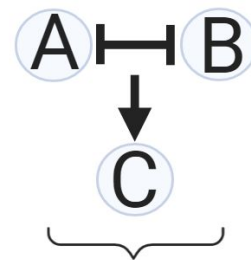
Genotype frequency data from
cross-sectional studies



Some genes with
no dependency



Temporal
constraint: one
mutation precedes



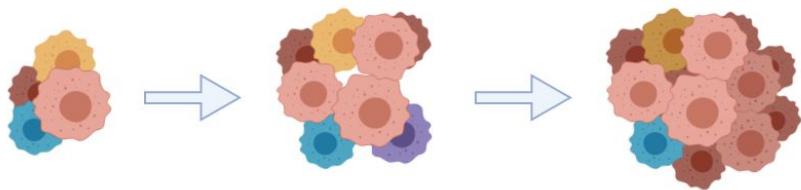
Inhibition and/or
mutual exclusivity

7 Modeling cancer: CPMs and DAGs

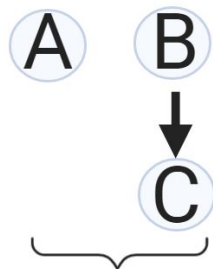
Cancer Progression Models (CPMs)



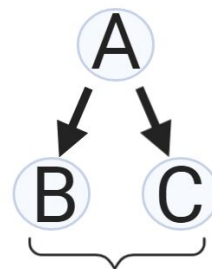
Directed Acyclic Graphs (DAGs)



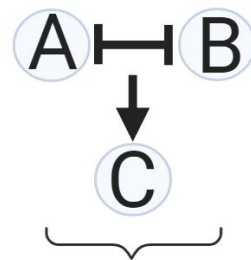
Genotype frequency data from cross-sectional studies



Some genes with no dependency



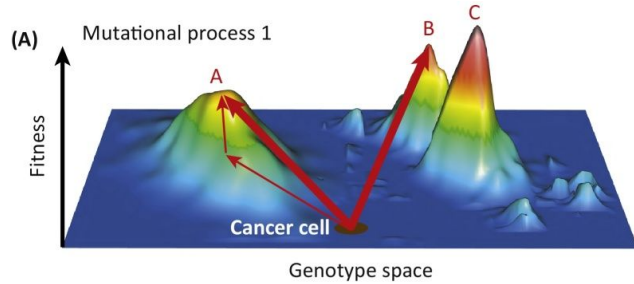
Temporal constraint: one mutation precedes



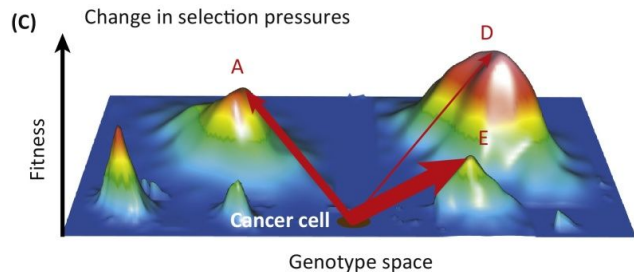
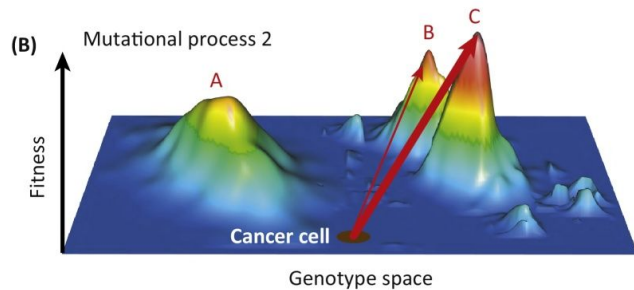
Inhibition and/or mutual exclusivity

These delimit possible **relationships between genes** and **possible mutational trajectories**, useful to identify candidate genes to block relevant pathways

8 Modeling cancer: Fitness landscapes

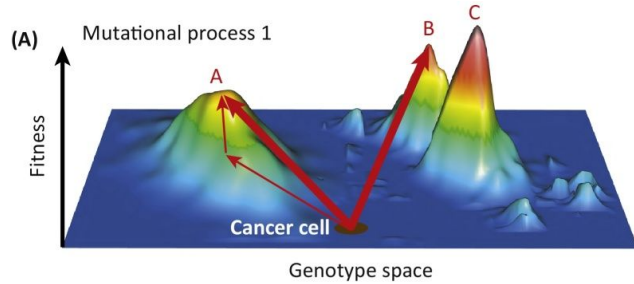


Maps of genotypes and their fitness, delimit **multiple paths** for accumulation of mutations

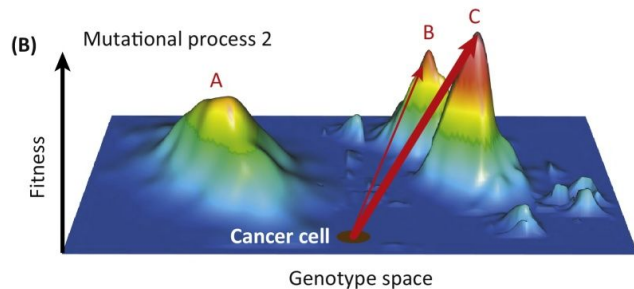


Lipinski et al. (2016)

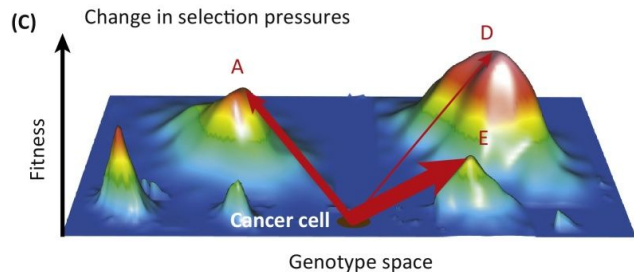
9 Modeling cancer: Fitness landscapes



Maps of genotypes and their fitness, delimit **multiple paths** for accumulation of mutations

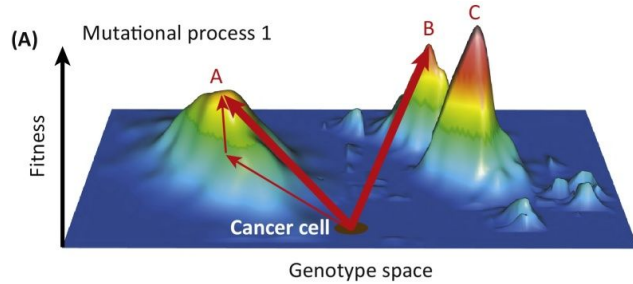


Accessible genotypes: Mutational pathways along different genotypes where each one is separated by a single mutational step and fitness increases

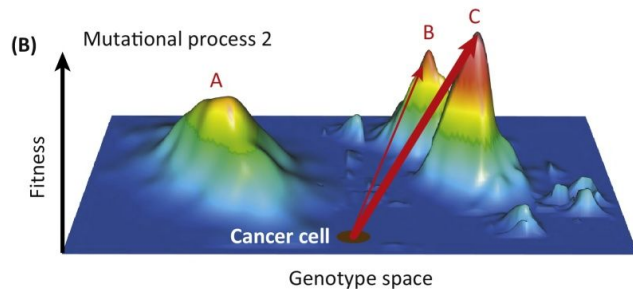


Lipinski et al. (2016)

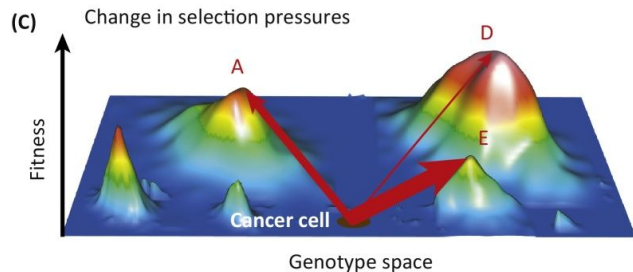
10 Modeling cancer: Fitness landscapes



Maps of genotypes and their fitness, delimit **multiple paths** for accumulation of mutations



Accessible genotypes: Mutational pathways along different genotypes where each one is separated by a single mutational step and fitness increases

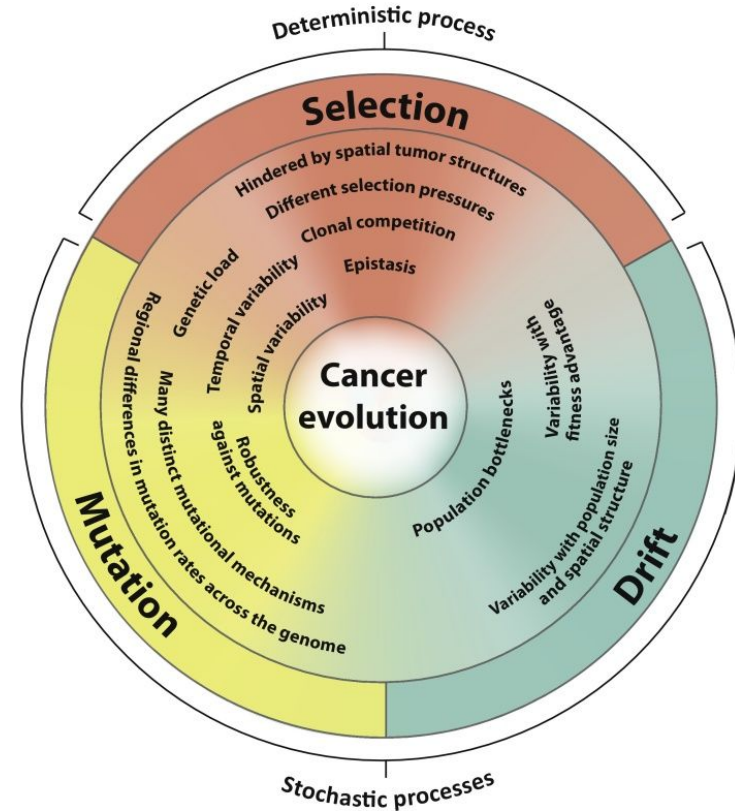


Tumor progression is an **evolutionary process**, fitness reinforces prediction of **most probable paths** (accessible genotypes)

11 Modeling cancer: Many possible paths

Multiple complex evolutionary paths and **high variance in fitness** due to biological interactions such as:

- Complementarity
- **Mutual exclusivity** (*synthetic lethality or no fitness gain in consecutive mutations in the same pathway*)

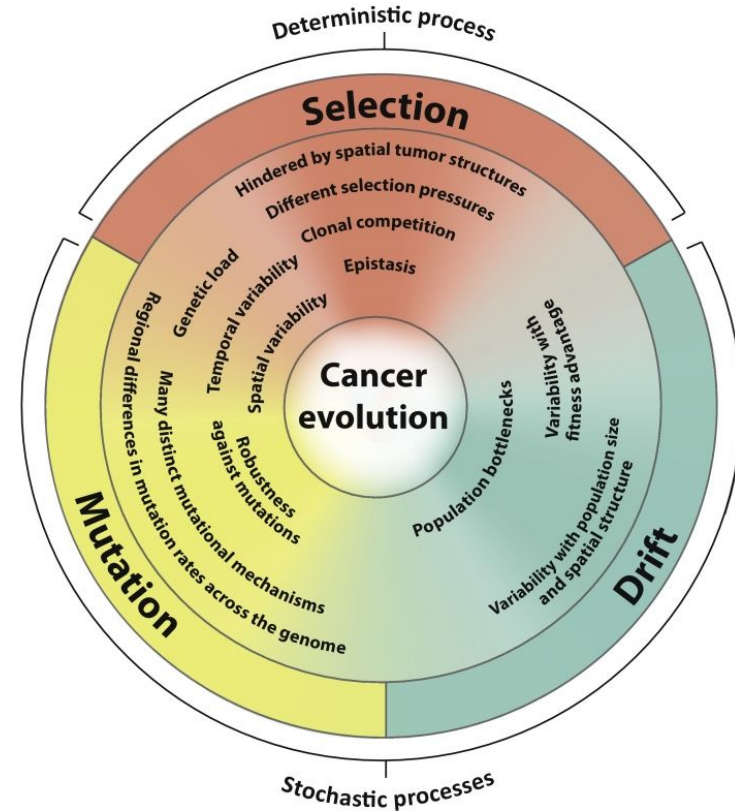


12 Modeling cancer: Many possible paths

Multiple complex evolutionary paths and **high variance in fitness** due to biological interactions such as:

- Complementarity
- **Mutual exclusivity** (*synthetic lethality or no fitness gain in consecutive mutations in the same pathway*)

CPMs cannot represent well this events (*only temporal order*) but **fitness landscape can**



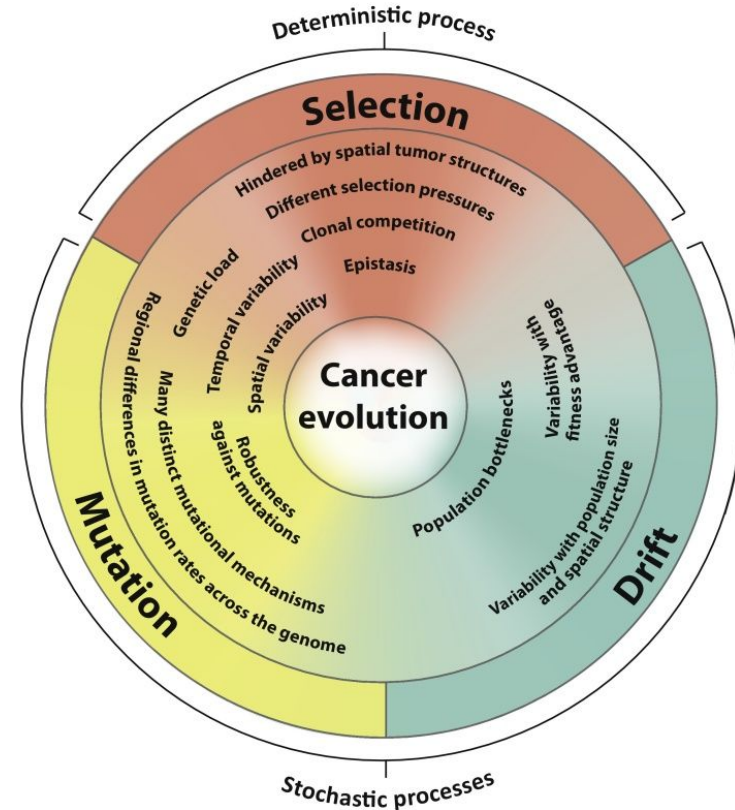
13 Modeling cancer: Many possible paths

Multiple complex evolutionary paths and **high variance in fitness** due to biological interactions such as:

- Complementarity
- **Mutual exclusivity** (*synthetic lethality or no fitness gain in consecutive mutations in the same pathway*)

CPMs cannot represent well this events (*only temporal order*) but **fitness landscape can**

Exhaustive **fitness landscapes are hard to produce** (evolution is complicated, duh), many models resort to using CPMs-DAGs and carry their limitations



Various algorithms (pathTiMEx, MEMO, etc.) incorporate **mutual exclusivity** in their models based on DAGs, but **they don't validate those with evolutionary information** → May not represent well enough accessible pathways

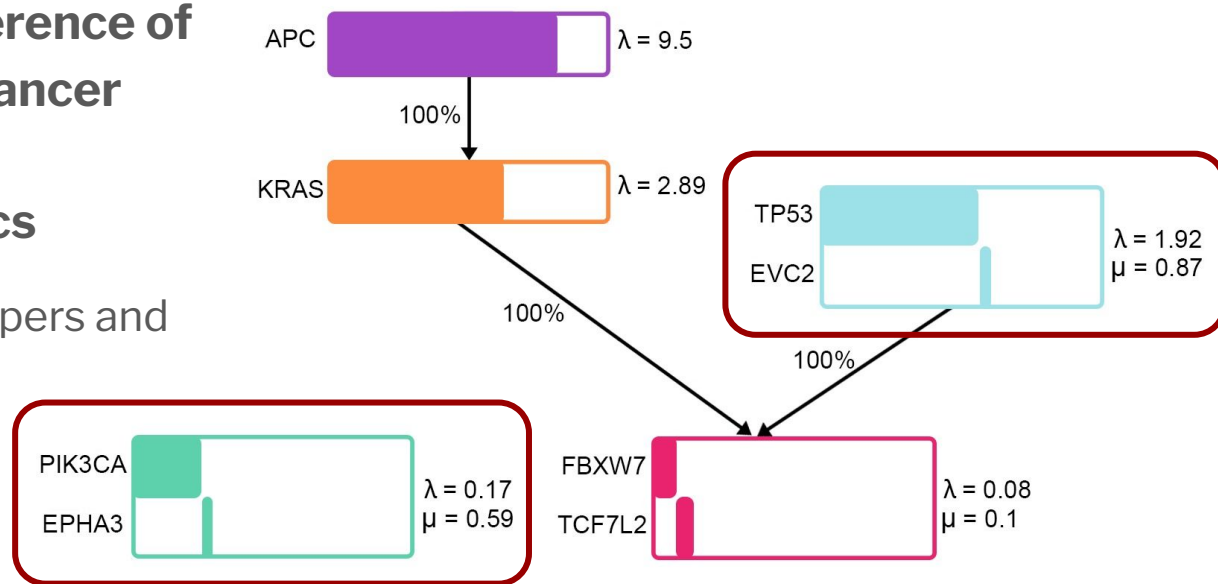
Various algorithms (pathTiMEx, MEMO, etc.) incorporate **mutual exclusivity** in their models based on DAGs, but **they don't validate those with evolutionary information** → May not represent well enough accessible pathways

Using OncoSimulR we worked on the data simulation from two papers/algorithms, critically assessing them by:

- Replicating some of their models and examples
- Extending on what they were modeling
- Identifying some of their limitations

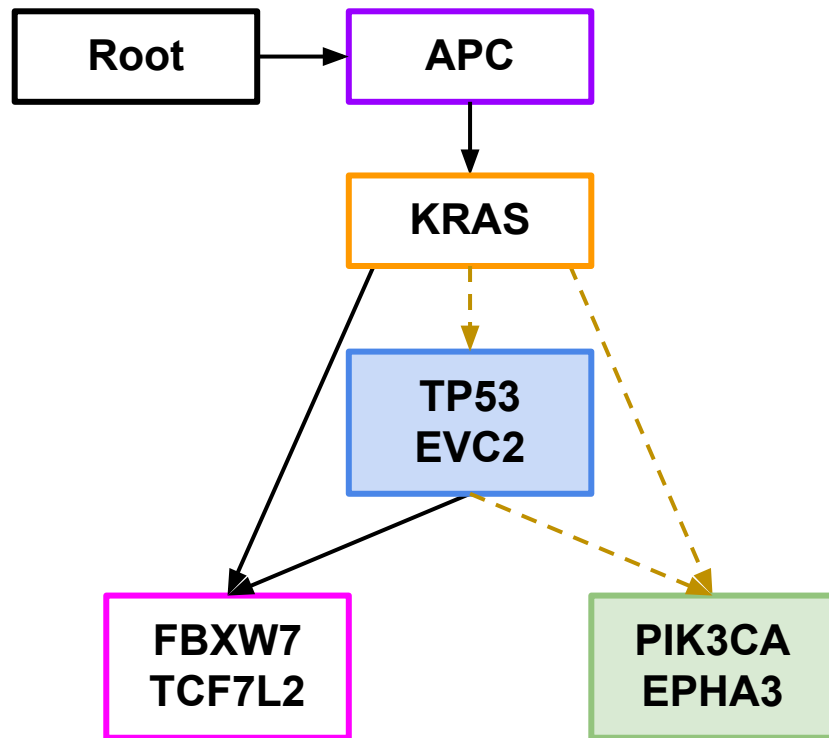
pathTiMEx: Joint Inference of Mutually Exclusive Cancer Pathways and Their Progression Dynamics

Simona Cristea, Jack Kuipers and Niko Beerenwinkel



Order effects:

- Modules TP53/EVC2 and PIK3CA/EPHA3 appear independently
- The positive effect on fitness is greater if order restrictions are respected

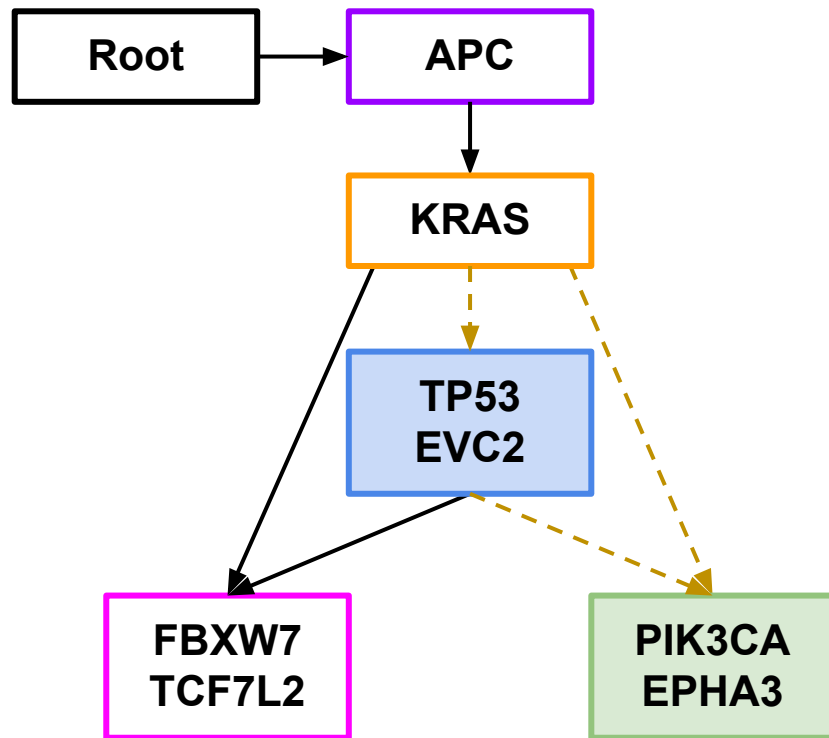


Order effects:

- Modules TP53/EVC2 and PIK3CA/EPHA3 appear independently
- The positive effect on fitness is greater if order restrictions are respected

```
eAG_wood_order[grep("^APC > TP53  
> KRAS$", eAG_wood_order[,1]),]
```

```
##           Genotype  Fitness  
##  104 APC > TP53 > KRAS      1.44
```



Order effects:

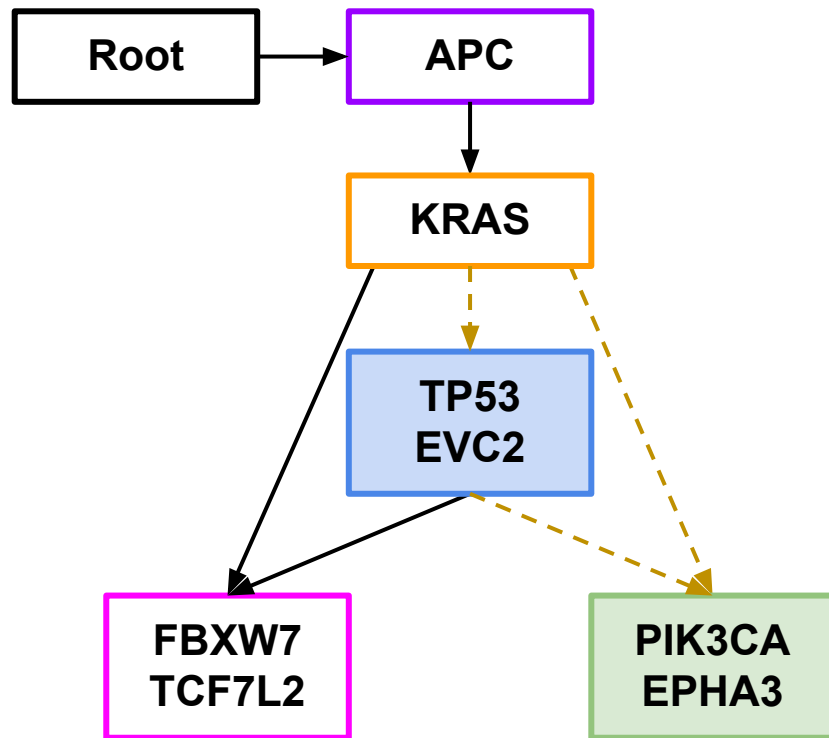
- Modules TP53/EVC2 and PIK3CA/EPHA3 appear independently
- The positive effect on fitness is greater if order restrictions are respected

```
eAG_wood_order[grep("^APC > TP53  
> KRAS$", eAG_wood_order[,1]),]
```

```
##           Genotype  Fitness
##  104 APC > TP53 > KRAS      1.44
```

```
"^APC > KRAS > TP53$"
```

```
##           Genotype  Fitness
##  104 APC > TP53 > KRAS    1.584
```

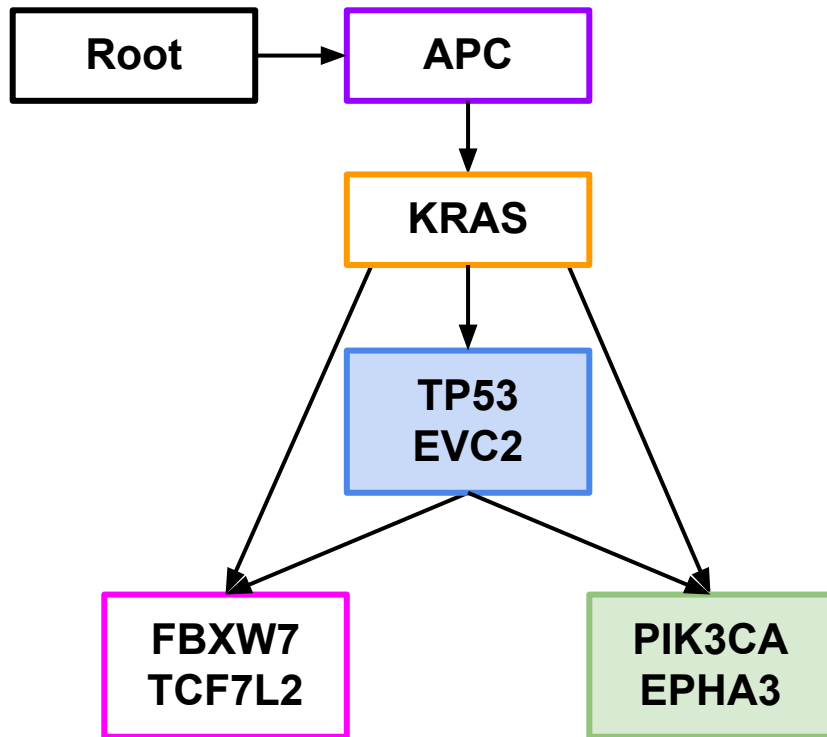


KRAS restrictions:

- If the restrictions are not fulfilled, a strong penalty is applied to the fitness
- **We are assuming a restriction that was not reported by the authors**

```
eAG_wood_kras[54, ]
```

```
##           Genotype  Fitness
##  54 APC, KRAS, TP53    1.728
```

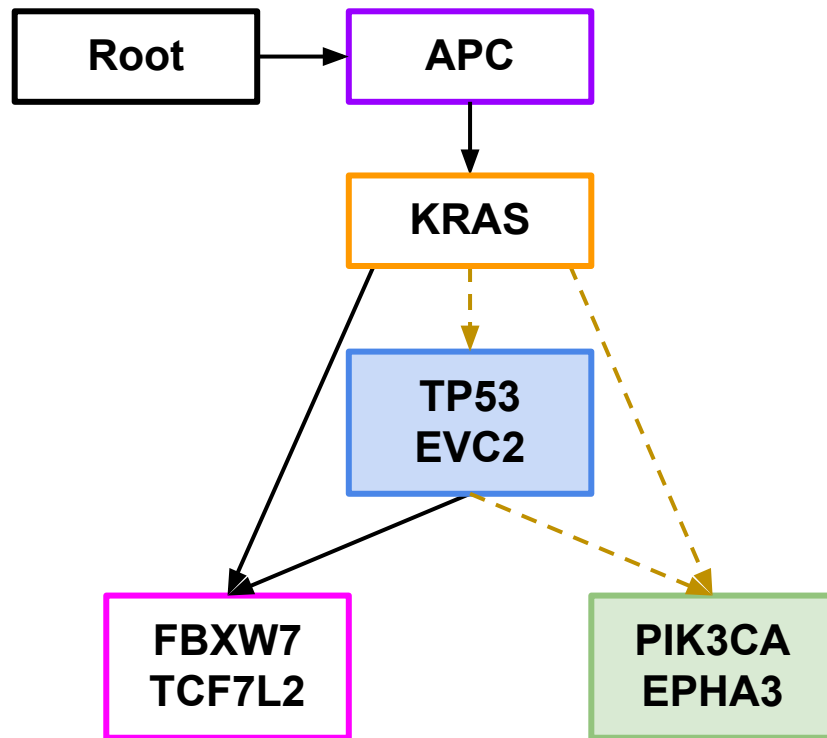


```
eAG_wood_order[50, ]
```

```
##           Genotype Fitness
##      50 PIK3CA > TP53      1
```

```
eAG_wood_order[63, ]
```

```
##           Genotype Fitness
##      63 TP53 > PIK3CA    1.05
```



```
eAG_wood_order[50, ]
```

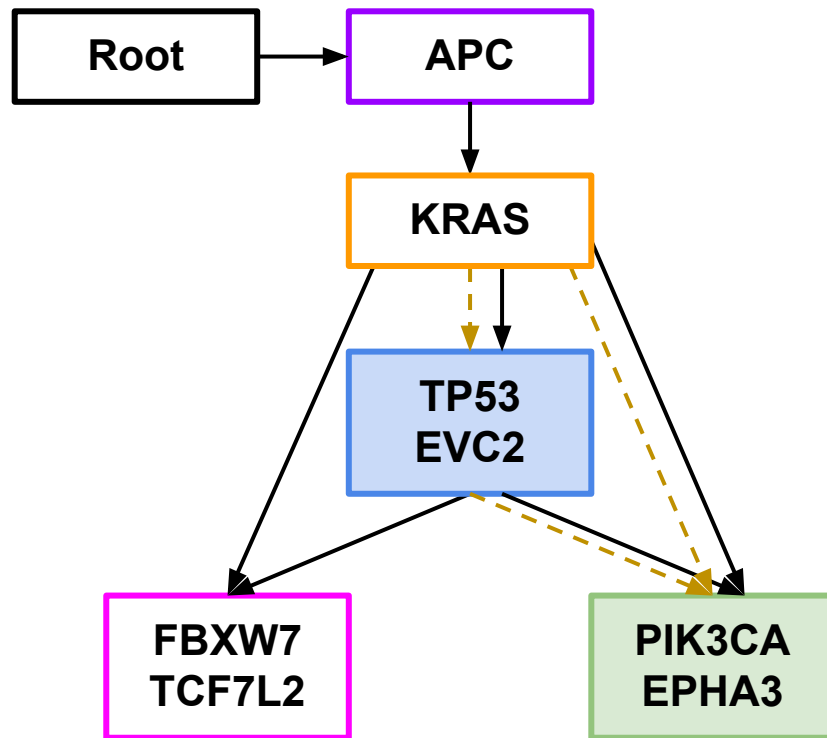
```
##           Genotype Fitness
##      50 PIK3CA > TP53      1
```

```
eAG_wood_order[63, ]
```

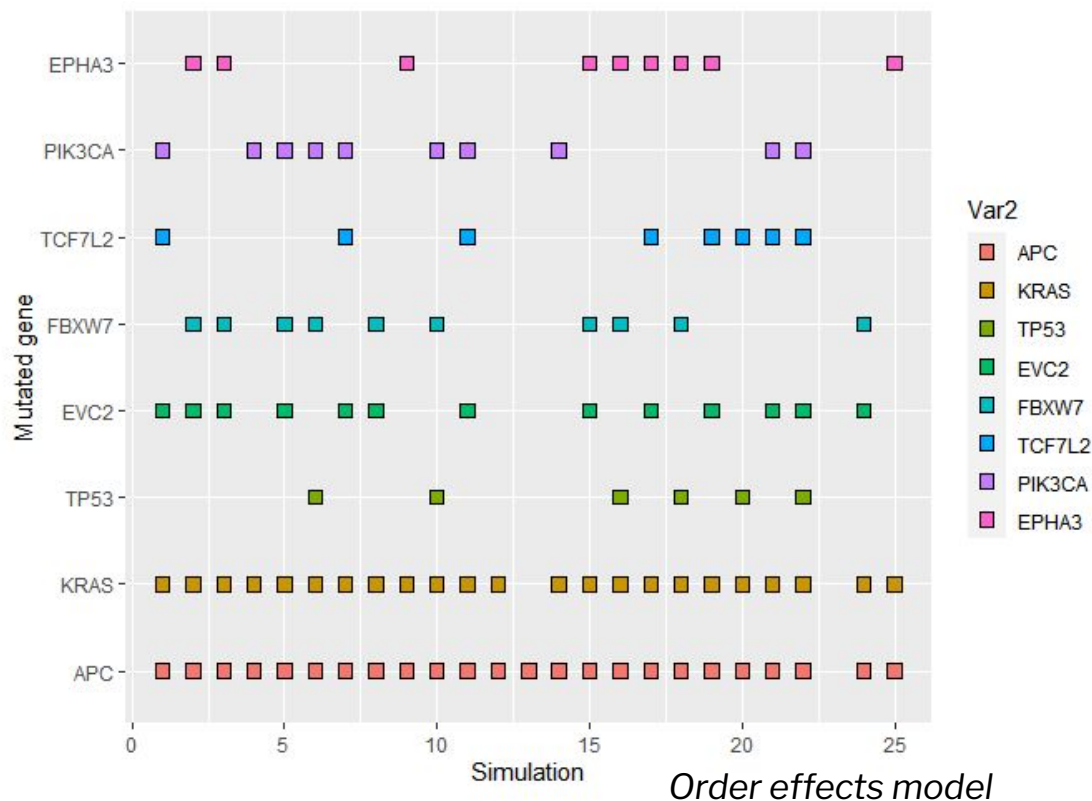
```
##           Genotype Fitness
##      63 TP53 > PIK3CA    1.05
```

```
eAG_wood_kras[35, ]
```

```
##           Genotype Fitness
##      35 PIK3CA, TP53    0.01
```



- Resulting genotypes after running the simulation 25 times with **oncoSimulPop**
- **Mutual exclusivity** is (mostly) maintained
- We cannot see **order effects**



- Models described in *Ciriello et al*
- Modules that share one or more genes
- “*The intersection of modules is the empty set*”

Mutual exclusivity analysis identifies oncogenic network modules

Giovanni Ciriello,^{1,3,4} Ethan Cerami,^{1,2,3} Chris Sander,¹ and Nikolaus Schultz¹

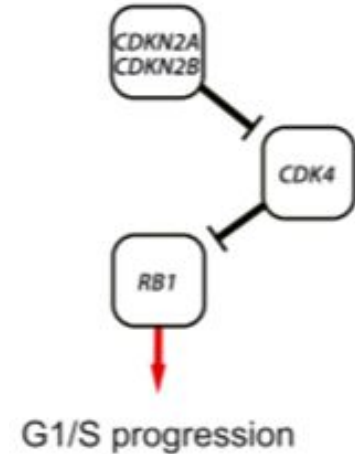
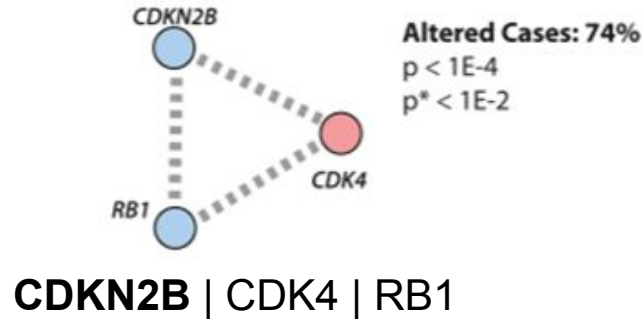
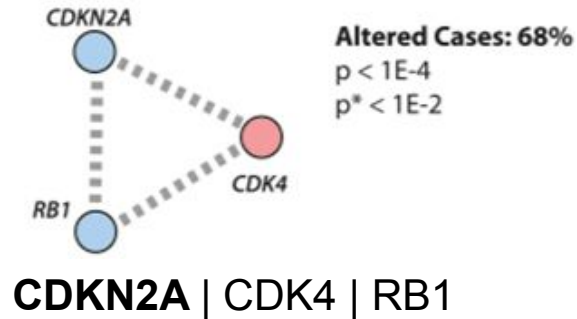


25 Glioblastoma: 1 pathway but 2 modules

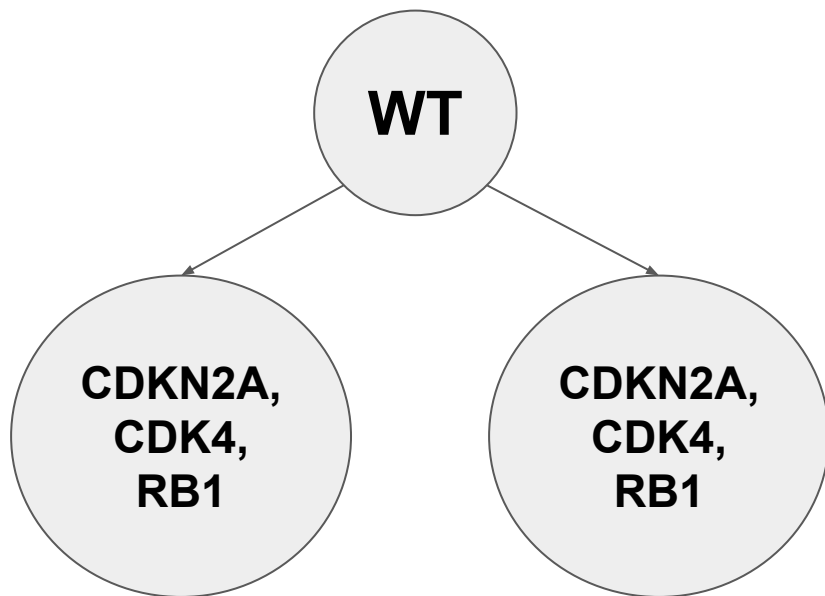
- Software applied to Glioblastoma dataset (TCGA)



- 2 mutually exclusive modules. Rb1 pathway



- ¿2 modules DAG? Not possible in OncoSimul.
 - Probably it doesn't make sense



```
m1 <- allFitnessEffects(data.frame(
  parent = c("Root", "Root"),
  child = c("A", "B"),
  s = 1,
  sh = -1,
  typeDep = "OR"),
  geneToModule = c("Root" = "Root",
    "A" = "CDKN2A, CDK4, RB1",
    "B" = "CDKN2B, CDK4, RB1"))
```

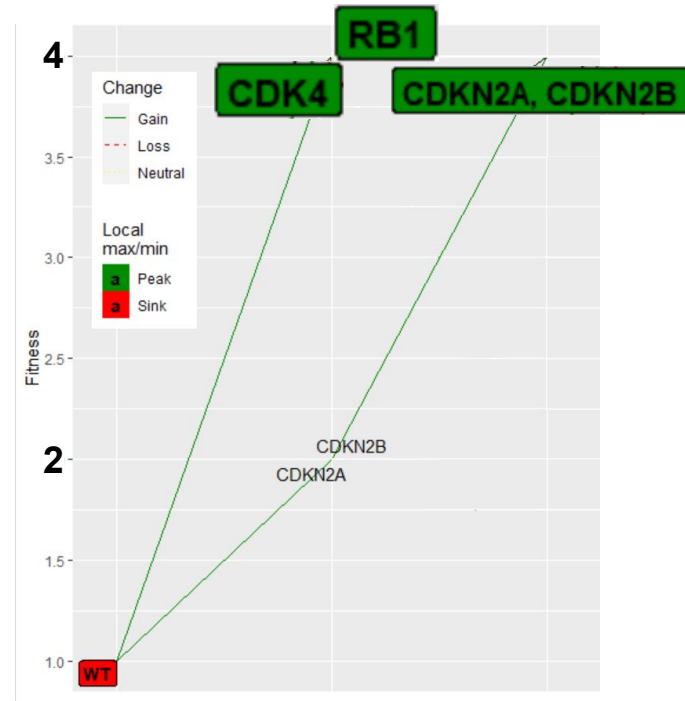
Error in gm.to.geneModuleL(geneToModule, one.to.one = gMOneToOne) :
Are there identical gene names in different modules?

- Define effects of each possible genotype manually
 - CDKN2A+CDKN2B, CDK4 and Rb1 should have the same fitness



2. CDKN2A and CDKN2B should have lower fitness

Genotype	Fitness
CDK4	4
RB1	4
CDKN2A, CDKN2B	4
CDKN2A	2
CDKN2B	2

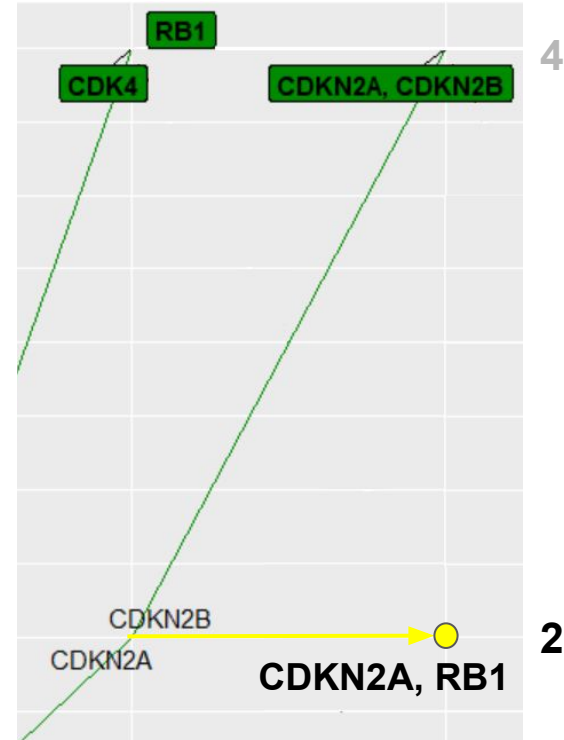


3. Fitness of more complex clones?

a) Mutating in RB1 to a CDKN2A clone



Fitness doesn't change



3. Fitness of more complex clones?

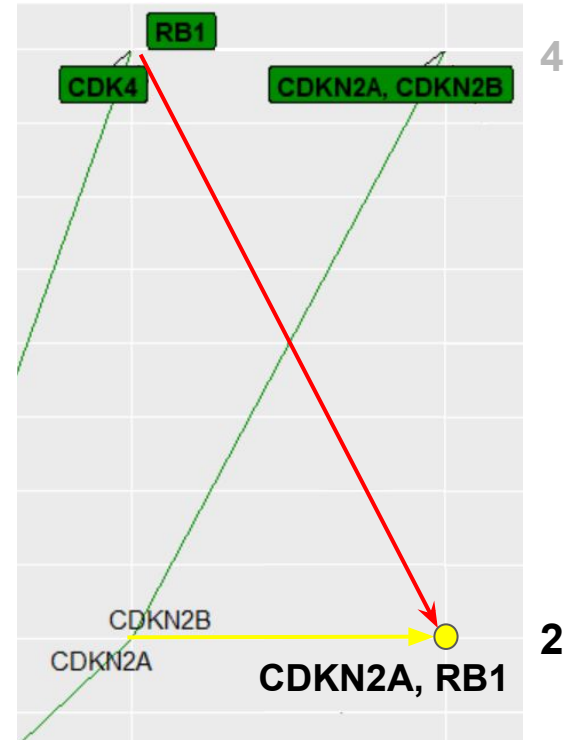
a) Mutating in RB1 to a CDKN2A clone



Fitness doesn't change



Path from RB1 involves a decrease



3. Fitness of more complex clones?

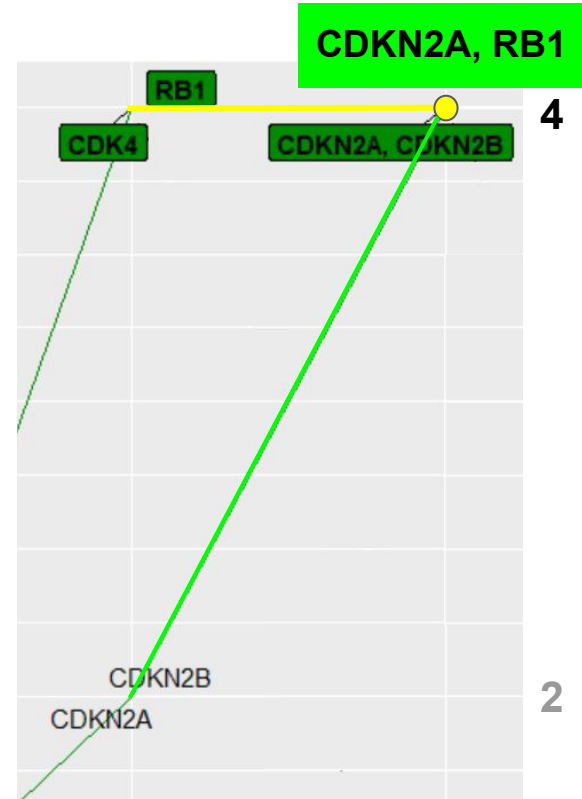
b) Mutating CDKN2 in a RB1 clone



Fitness doesn't change



Path from CDKN2 breaks exclusivity



3. Fitness of more complex clones?

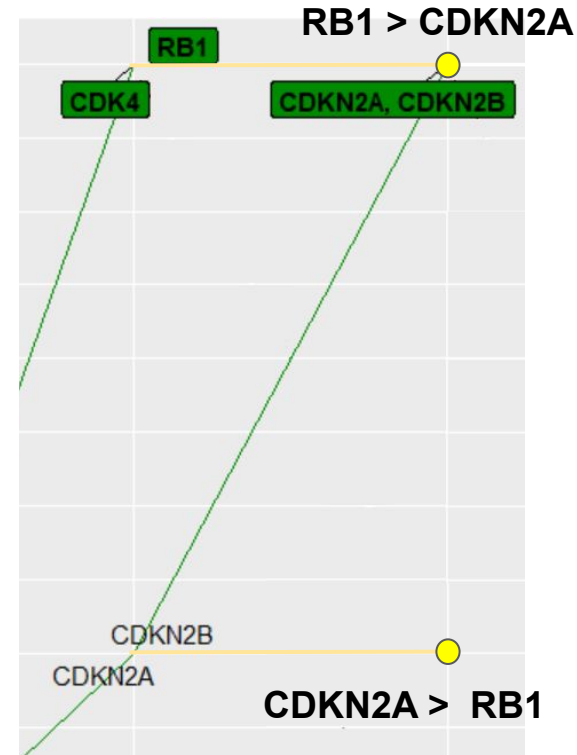
c) Specify order effects



Fitness doesn't change

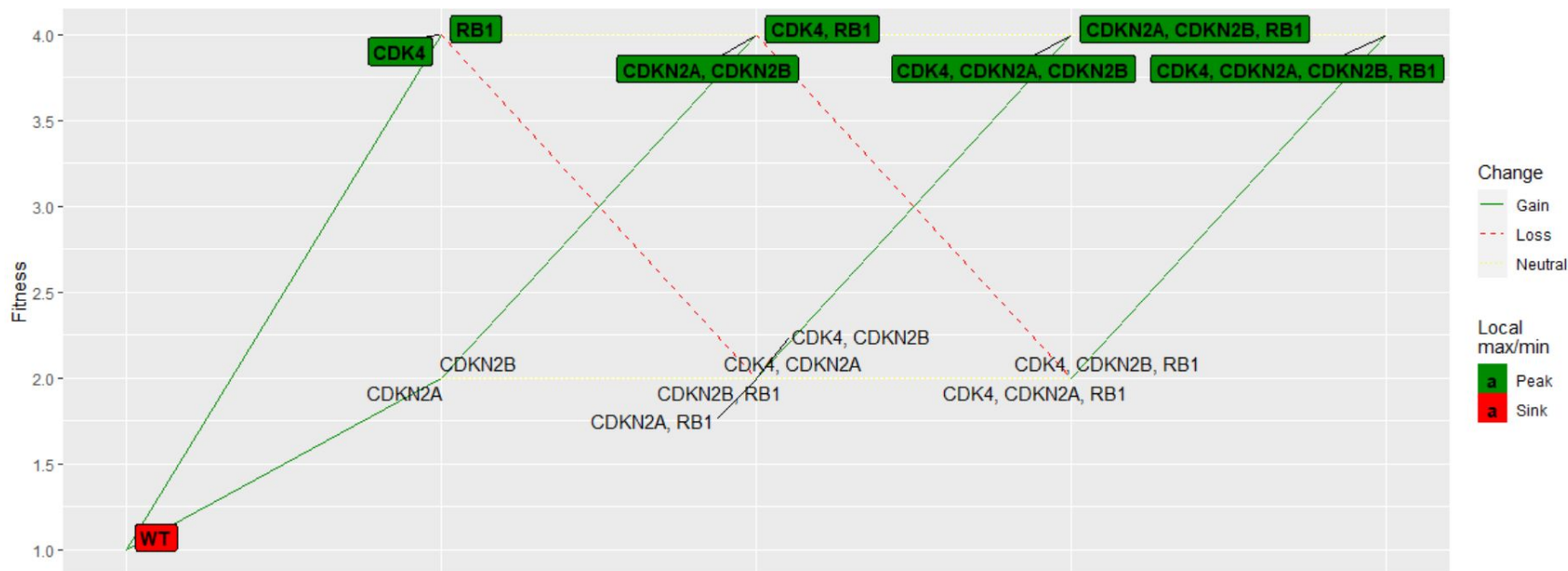


Biological justification?

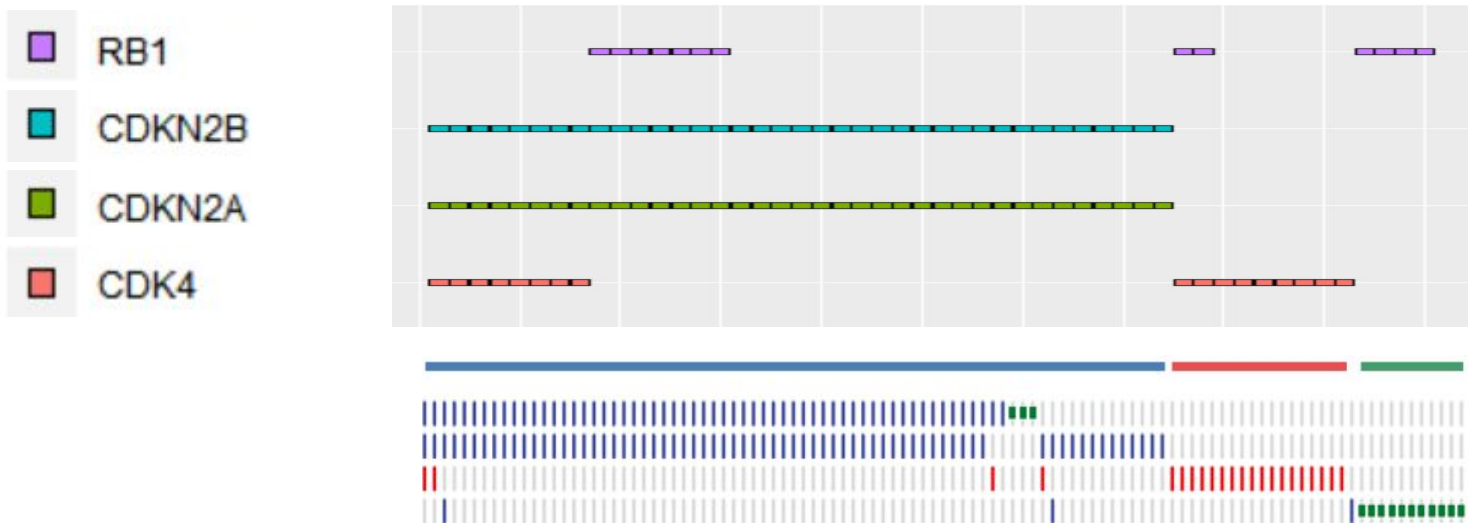


33 Glioblastoma: 1 pathway but 2 modules

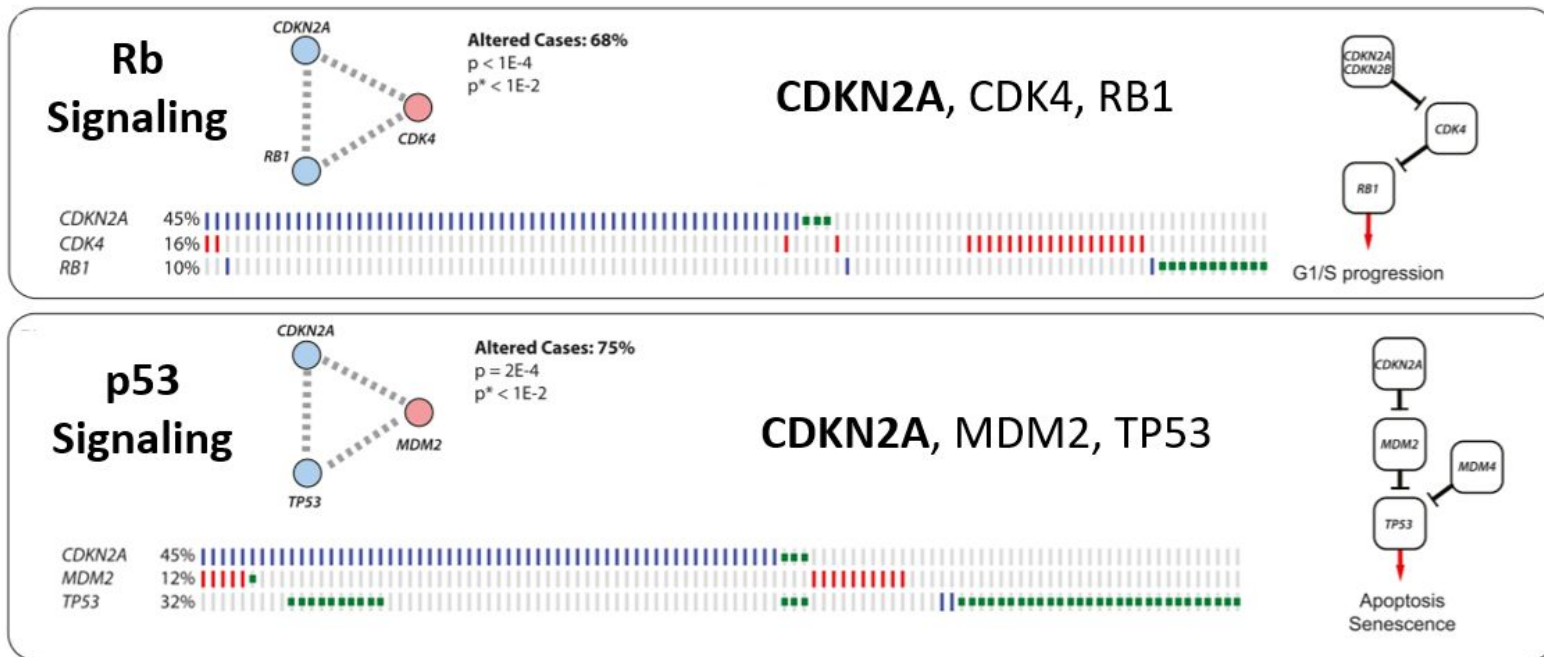
- No solution that satisfies all restrictions
- But following the first approach we reach this landscape:



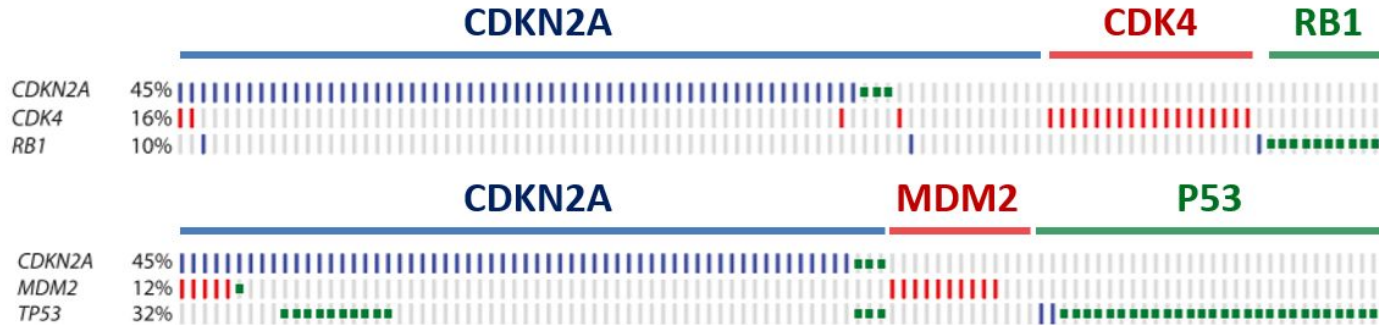
- Does our model replicate the data properly?
- Multiple simulations (**OncoSimulPop**):



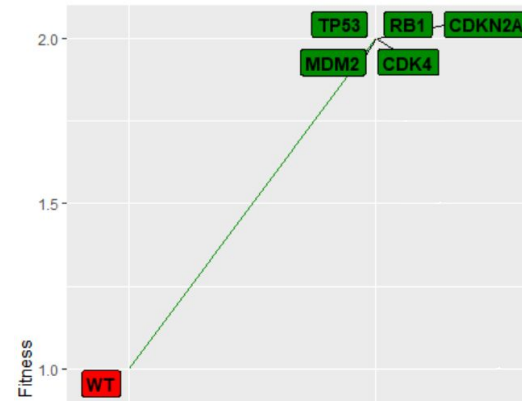
- Model with 2 different pathways which share one gene:



Glioblastoma: 1 pathway but 2 modules



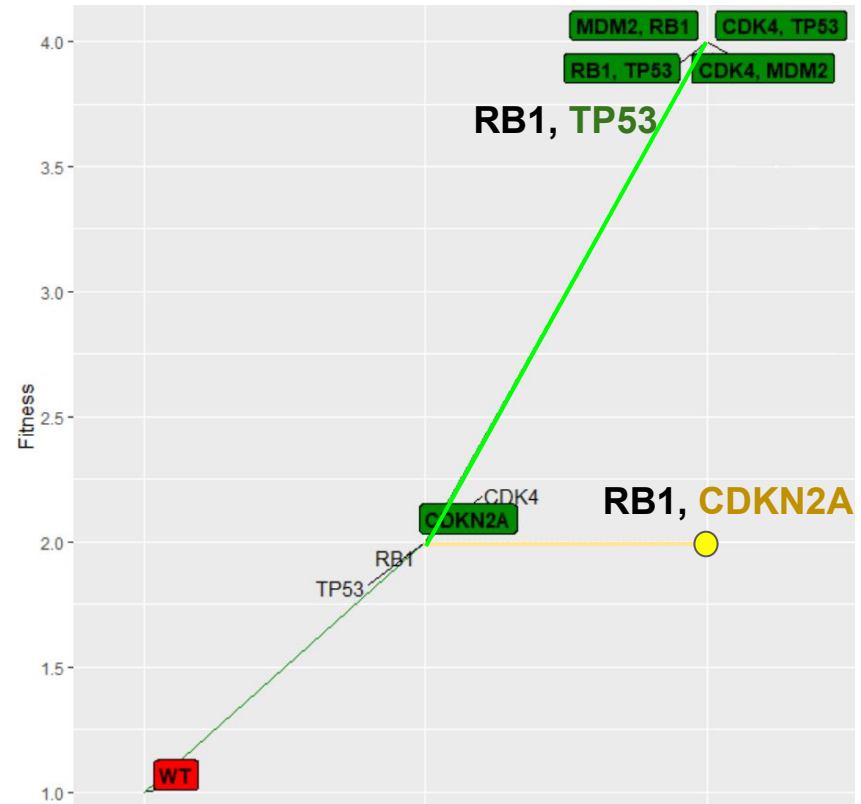
- All genotypes with one mutation should have the same fitness
 - Otherwise, we'd see just 1 genotype



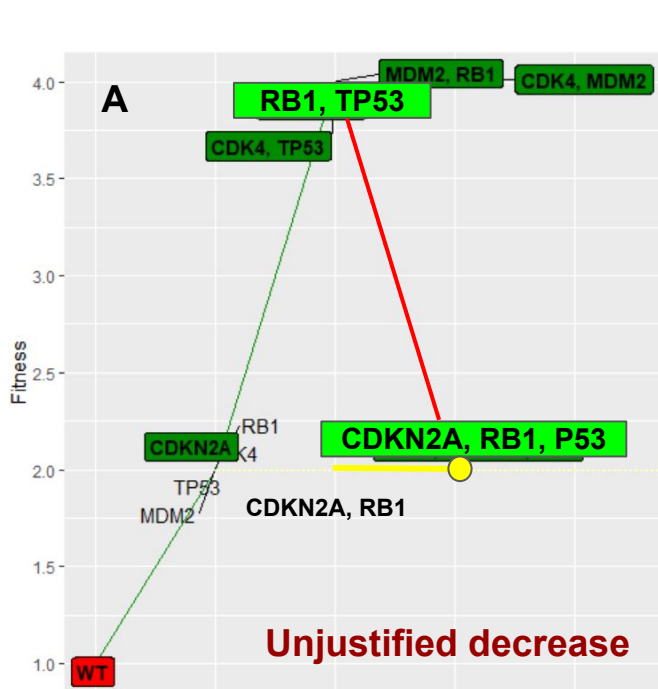
- Pair combinations:
 - Non-exclusive? **Greater fitness**

RB1 + P53 -> fitness = 4
 - Mutual exclusivity? **Same fitness**

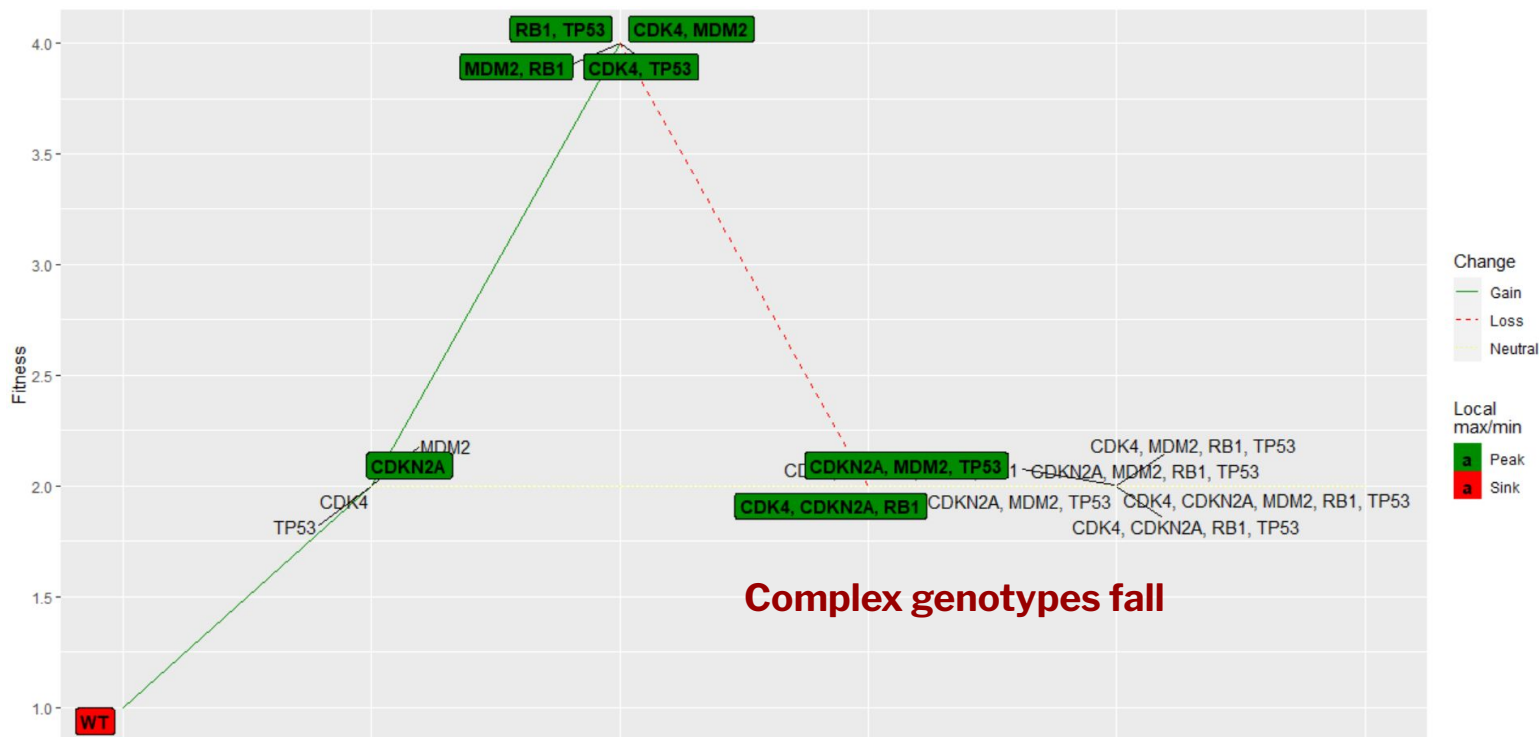
RB1 + CDKN2A -> fitness = 2



- Can define fitness of clones with 3 mutations again...

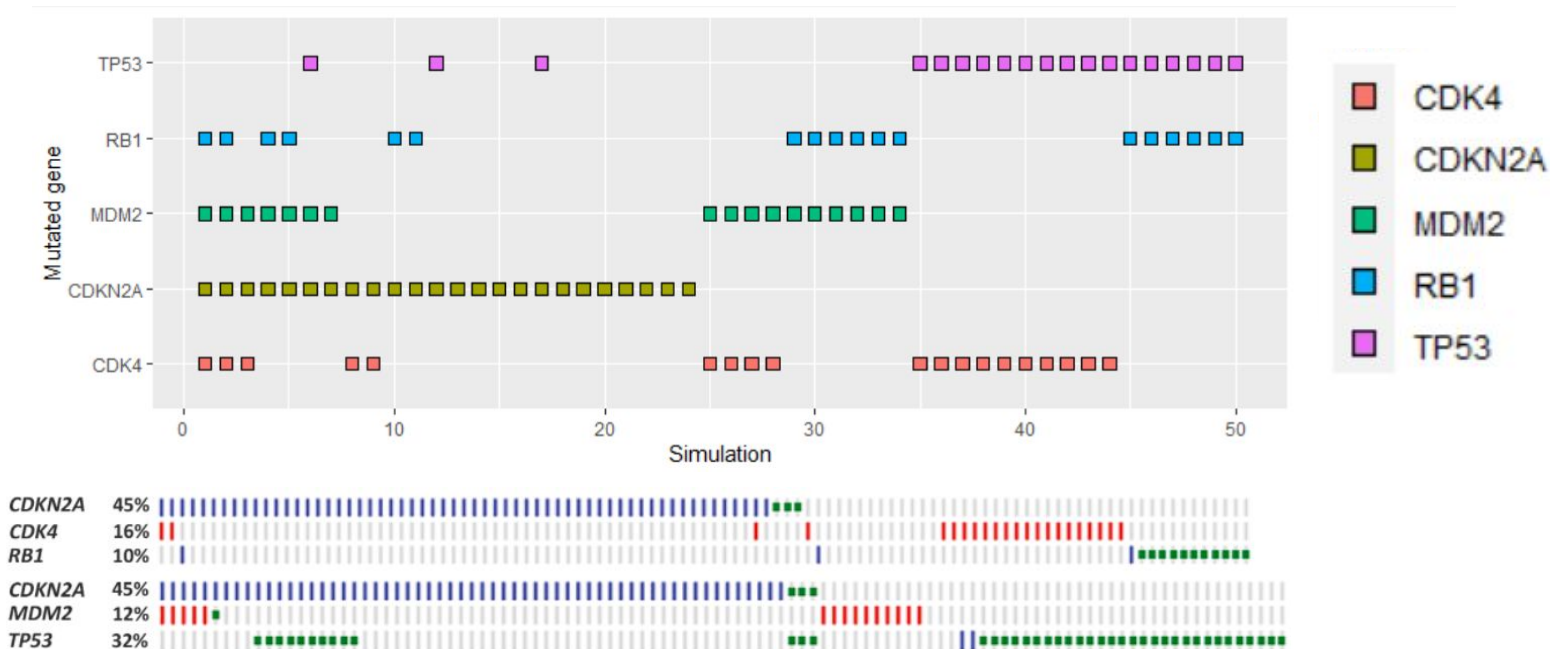


- Apply first option (as in previous case):



Glioblastoma: 1 pathway but 2 modules

- Multiple simulations:



(Can't really compare data)

- We were **able to show simple models** with a small number of genes and pathways involved

- We were **able to show simple models** with a small number of genes and pathways involved, but **different interpretations** of these have a **great impact on fitness** values

- We were **able to show simple models** with a small number of genes and pathways involved, but **different interpretations** of these have a **great impact on fitness** values
- More **complex models are challenging to translate into code**, more intricate paths and gene interactions are involved

- We were **able to show simple models** with a small number of genes and pathways involved, but **different interpretations** of these have a **great impact on fitness** values
- More **complex models are challenging to translate into code**, more intricate paths and gene interactions are involved → May break DAGs' assumptions
(shown in our examples of glioblastoma)

- We were **able to show simple models** with a small number of genes and pathways involved, but **different interpretations** of these have a **great impact on fitness** values
- More **complex models are challenging to translate into code**, more intricate paths and gene interactions are involved → May break DAGs' assumptions
(shown in our examples of glioblastoma)
- One **possible improvement** would be to **relax some DAGs constraints** (*such as allowing genes to be present in more than one model*)

- We were **able to show simple models** with a small number of genes and pathways involved, but **different interpretations** of these have a **great impact on fitness** values
- More **complex models are challenging to translate into code**, more intricate paths and gene interactions are involved → May break DAGs' assumptions
(shown in our examples of glioblastoma)
- One **possible improvement** would be to **relax some DAGs constraints** (*such as allowing genes to be present in more than one model*)

Tumor progression characterization is an ongoing challenge, but the improvements to the synergistic tools used are a great approach

Thank you
for your
attention!