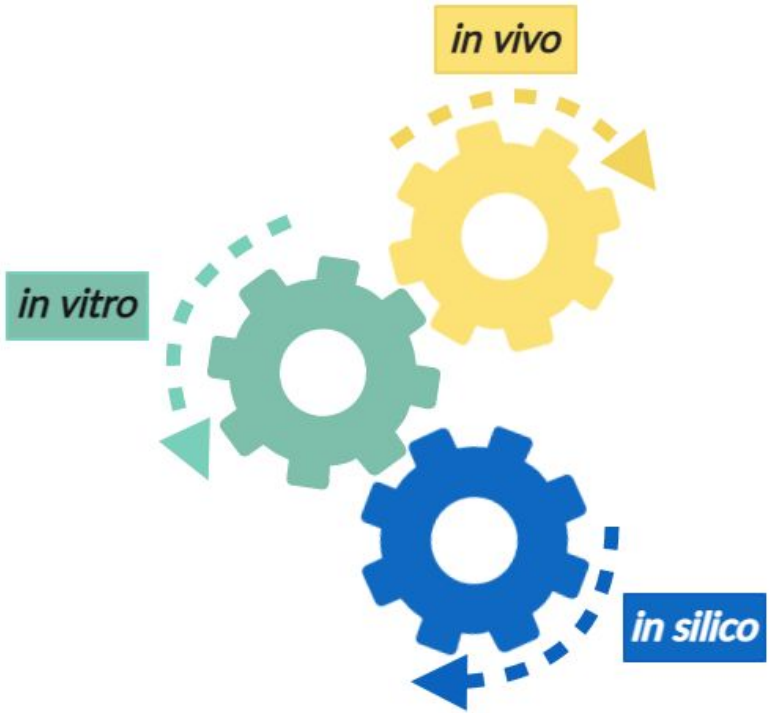
Simulating pathways and mutual exclusivity

Blanca Lacruz Pleguezuelos
Víctor Mateo Cáceres
Manuel Moradiellos Corpus

Cancer: Challenge and approaches

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

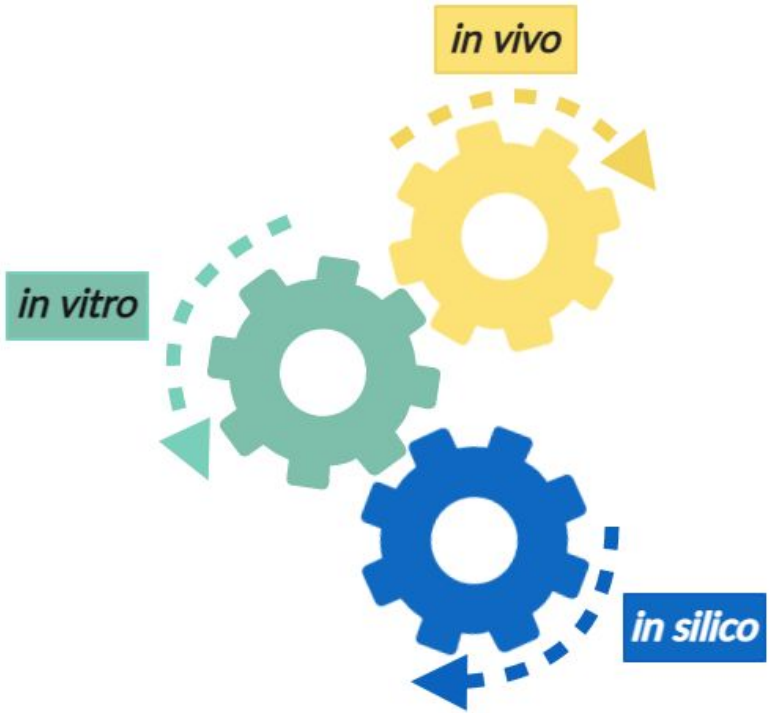
Cancer: Challenge and approaches



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: Challenge and approaches



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

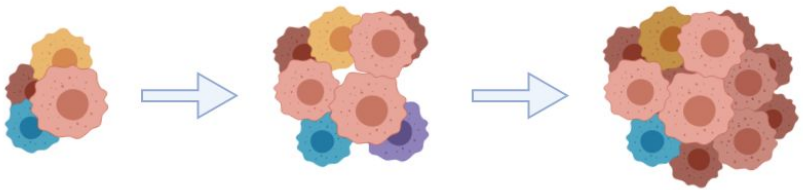
In silico **models** integrate data from experimental studies and analytical predictions

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

- Diagnosis
- Treatment
- Prognostic

Modeling cancer: CPMs and DAGs

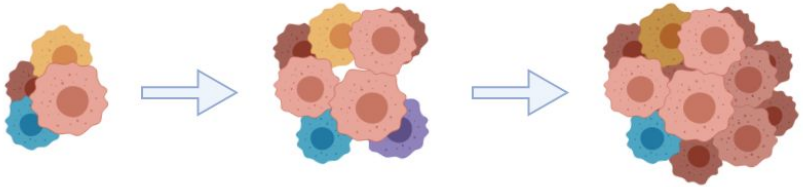
Cancer Progression Models (CPMs)



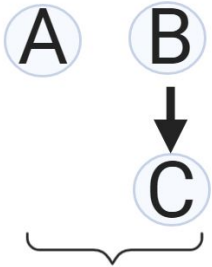
Genotype frequency data from
cross-sectional studies

Modeling cancer: CPMs and DAGs

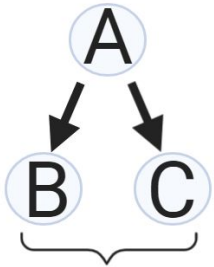
Cancer Progression Models (CPMs) → **Directed Acyclic Graphs** (DAGs)



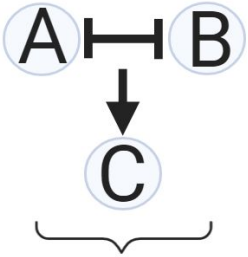
Genotype frequency data from cross-sectional studies



Some genes with no dependency



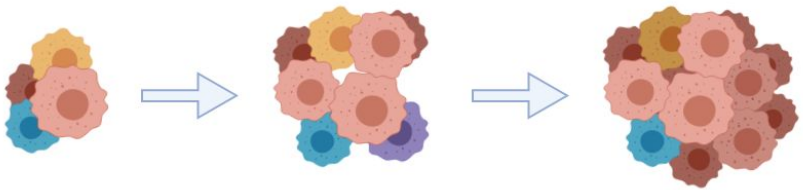
Temporal constraint: one mutation precedes



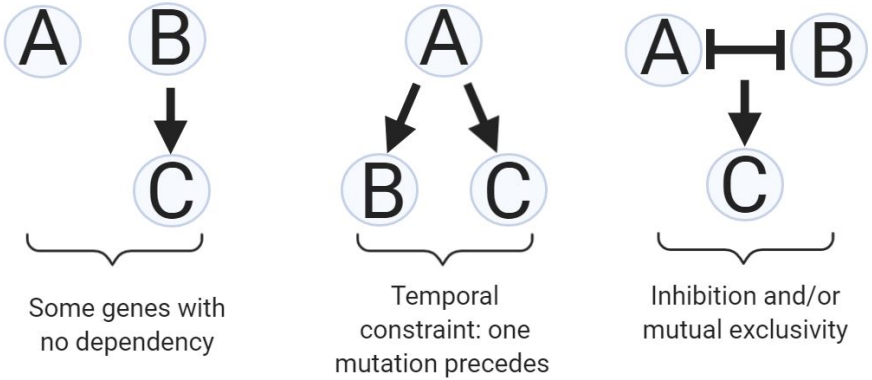
Inhibition and/or mutual exclusivity

Modeling cancer: CPMs and DAGs

Cancer Progression Models (CPMs) → **Directed Acyclic Graphs** (DAGs)

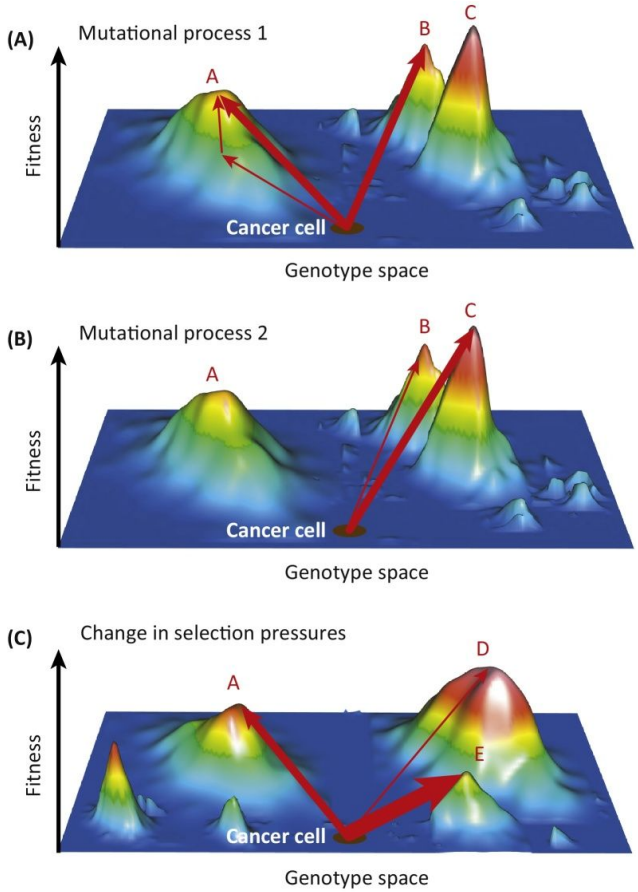


Genotype frequency data from cross-sectional studies



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

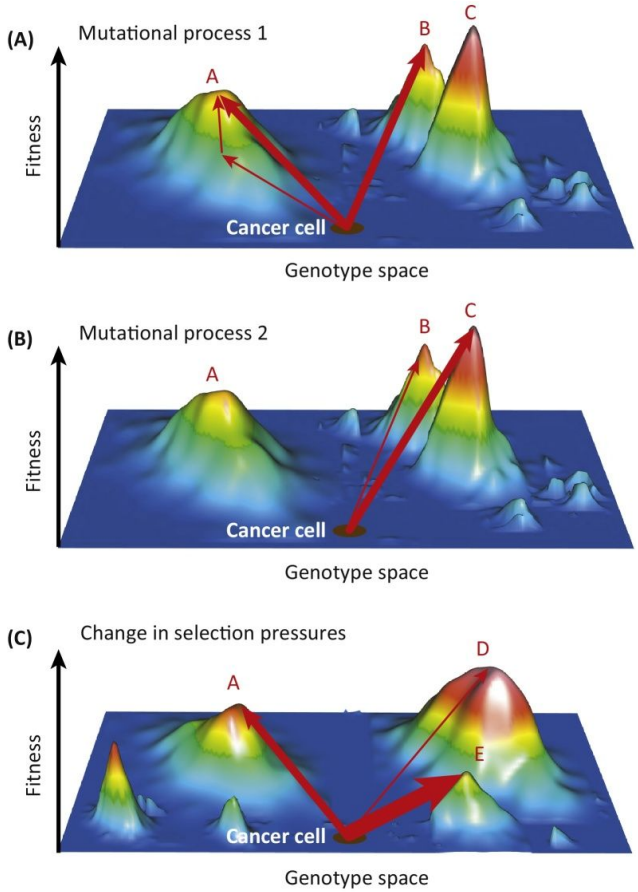
8 **Modeling cancer:** Fitness landscapes



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

Lipinski et al. (2016)

Modeling cancer: Fitness landscapes

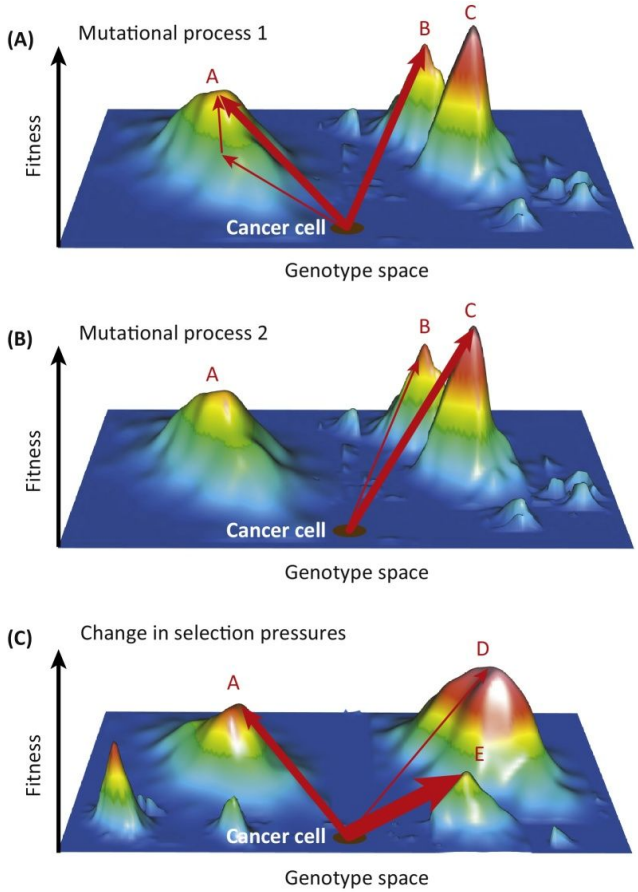


Maps of genotypes and their fitness, delimit **multiple paths** for accumulation of mutations

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

Lipinski et al. (2016)

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

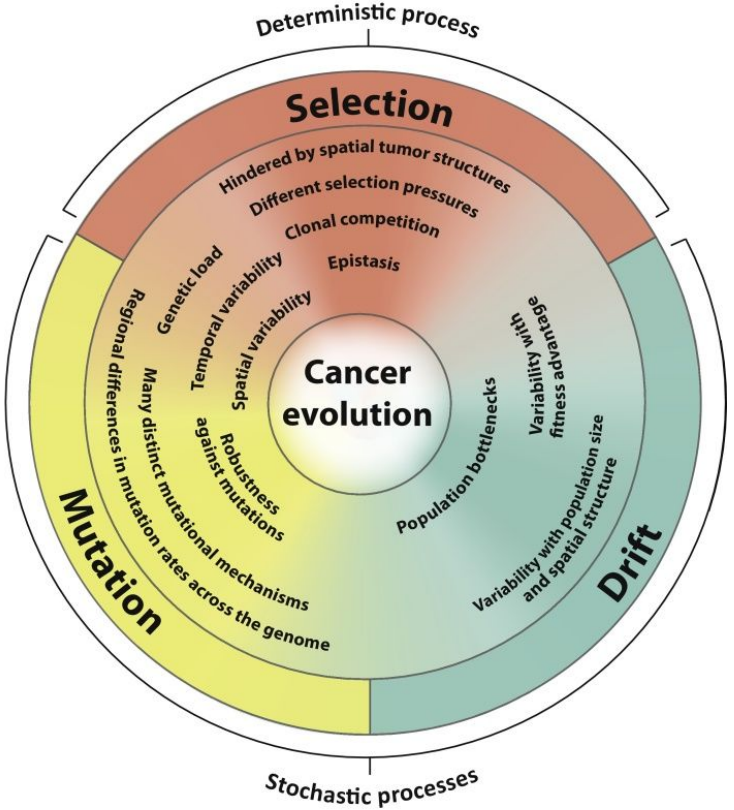
Tumor progression is an **evolutionary process**, fitness reinforces prediction of **most probable paths** (accessible genotypes)

Lipinski et al. (2016)

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*)



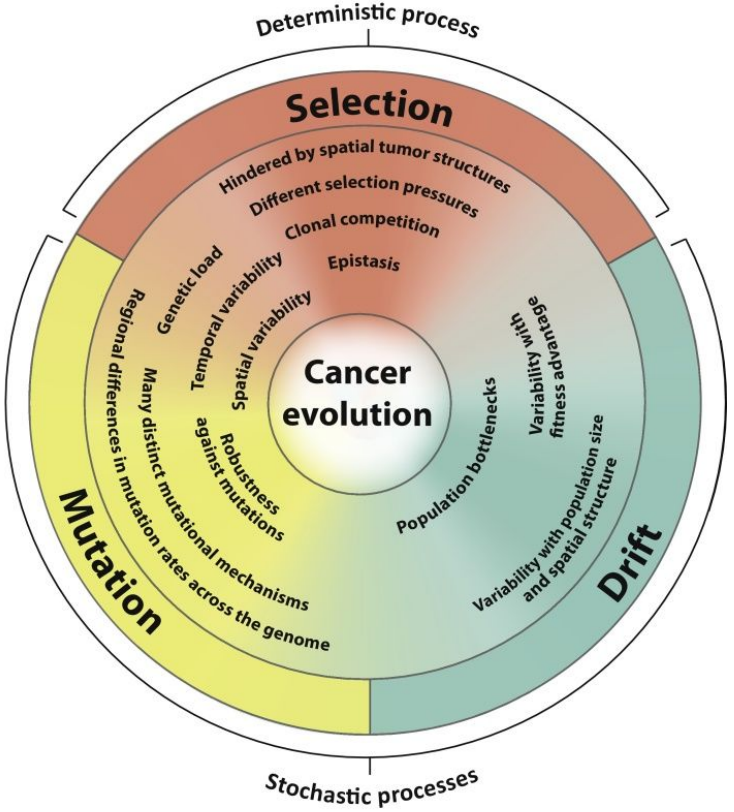
Lipinski et al. (2016)

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**



Lipinski et al. (2016)

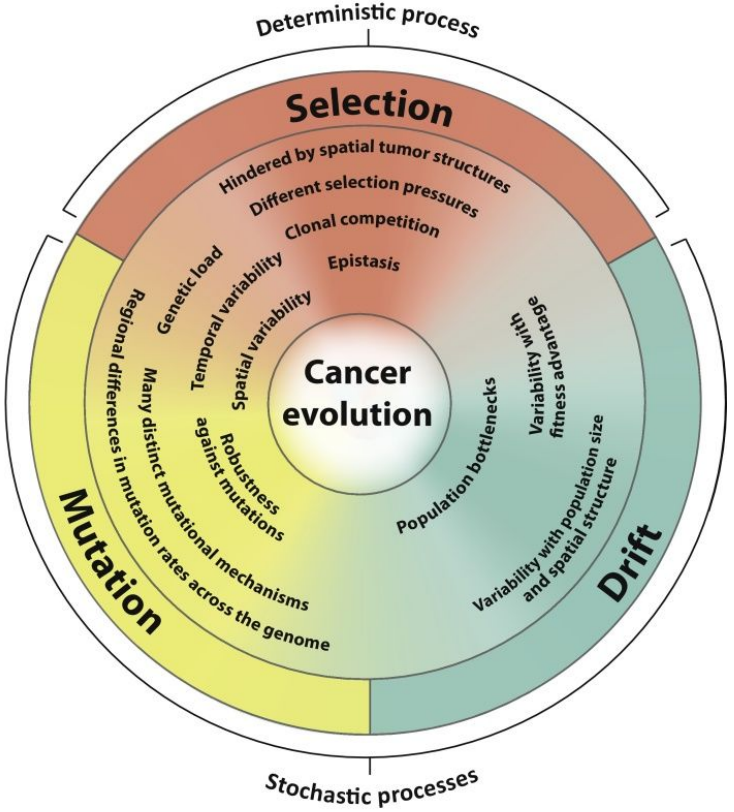
Modeling cancer: Many possible paths

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



Lipinski et al. (2016)

Objectives of our work

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

Objectives of our work

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

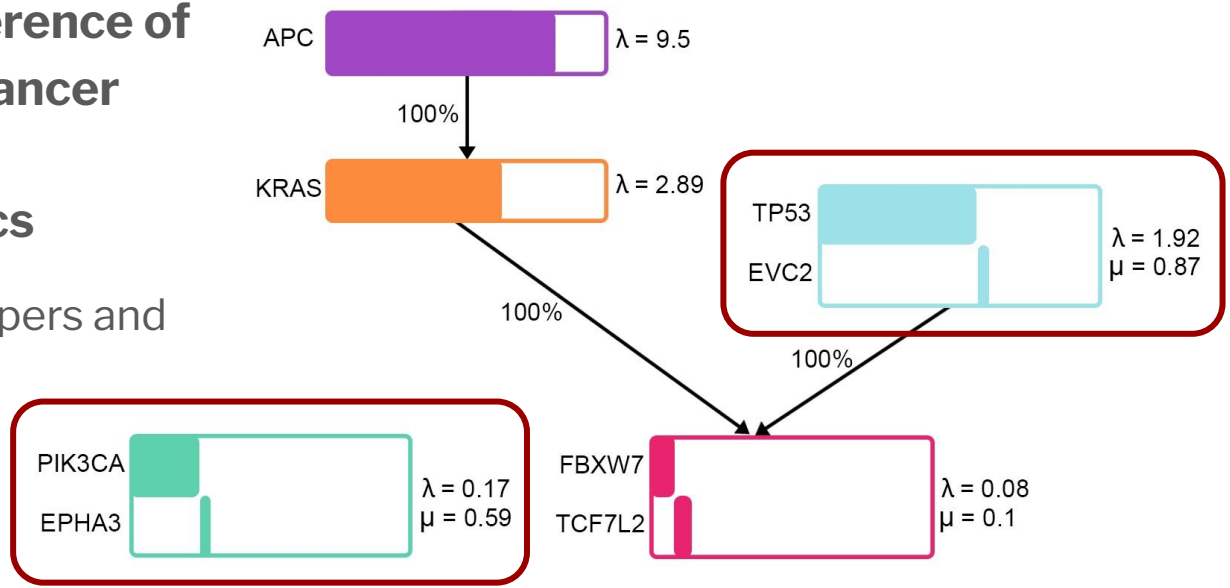
Using OncoSimulR we worked on the data simulation from two papers/algorithms, critically assessing them by:

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

Colorectal cancer: Model found by pathTiMEx

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

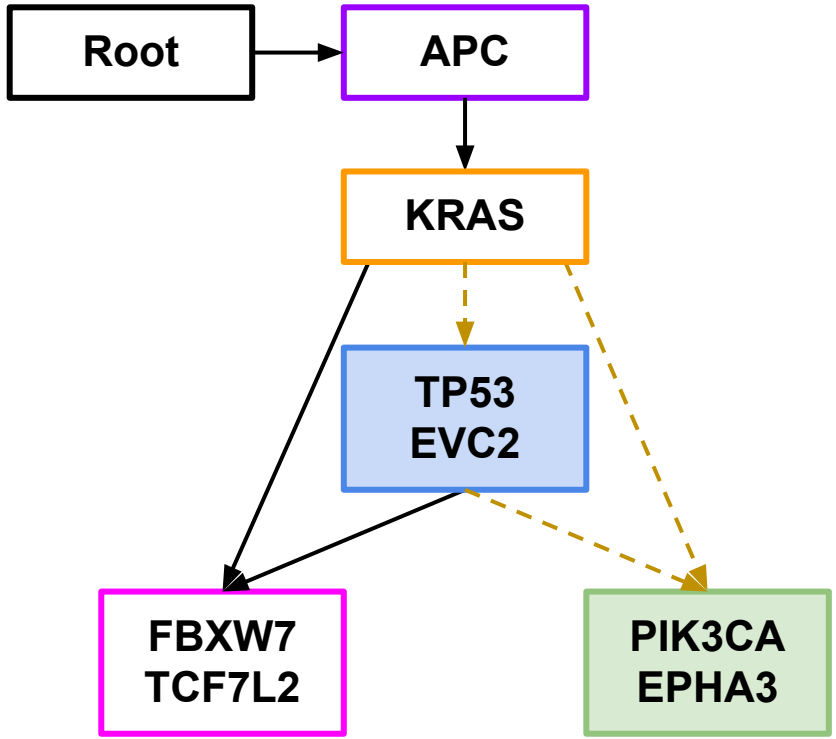
Simona Cristea, Jack Kuipers and Niko Beerenwinkel



Colorectal cancer: Model with order effects

Order effects:

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



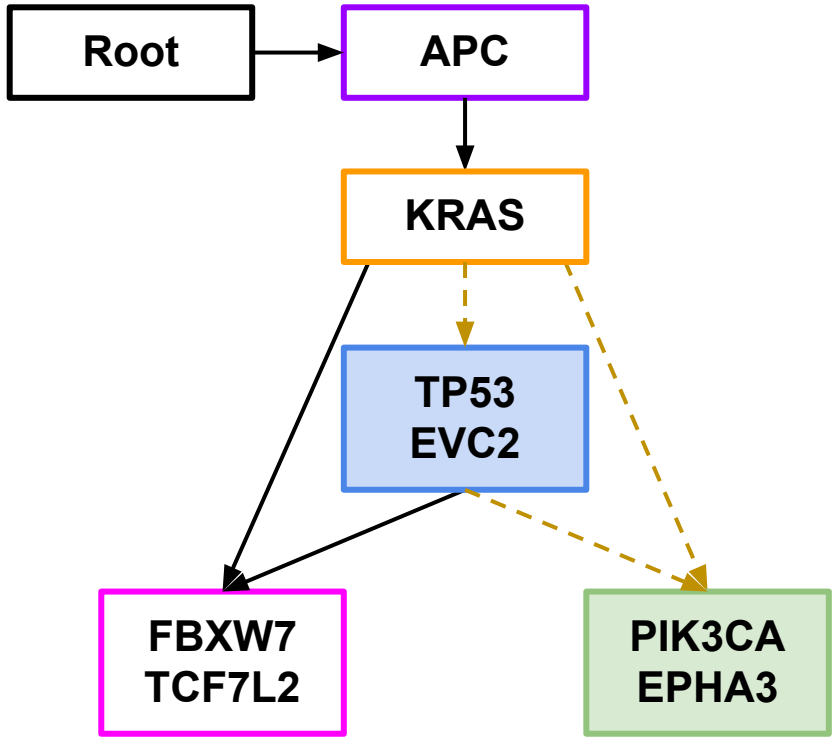
Colorectal cancer: Model with order effects

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
```



Colorectal cancer: Model with order effects

Order effects:

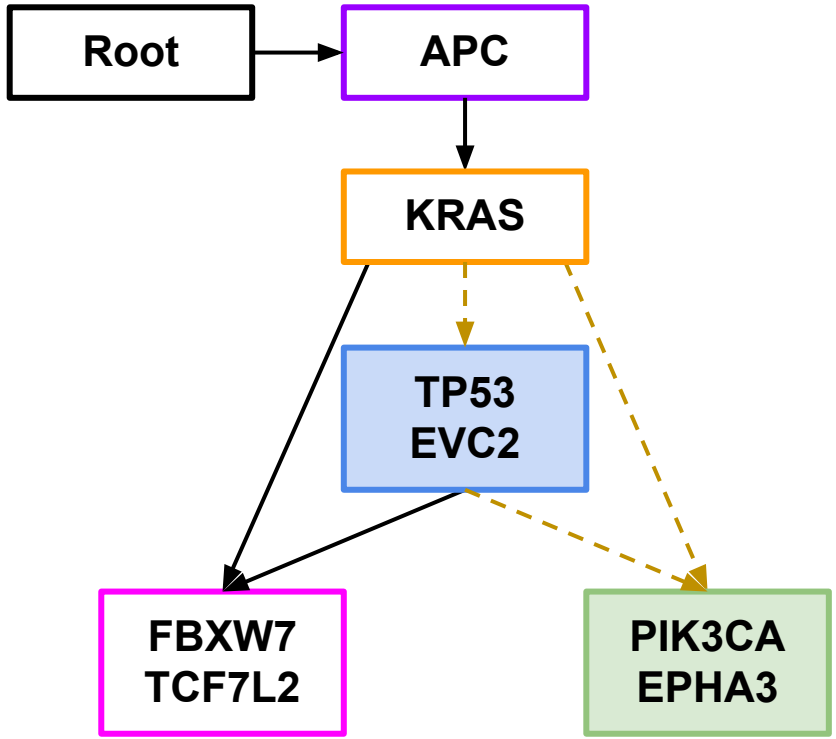
- Modules TP53/EVC2 and PIK3CA/EPHA3 appear independently
- The positive effect on fitness is greater if order restrictions are respected

```
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
```



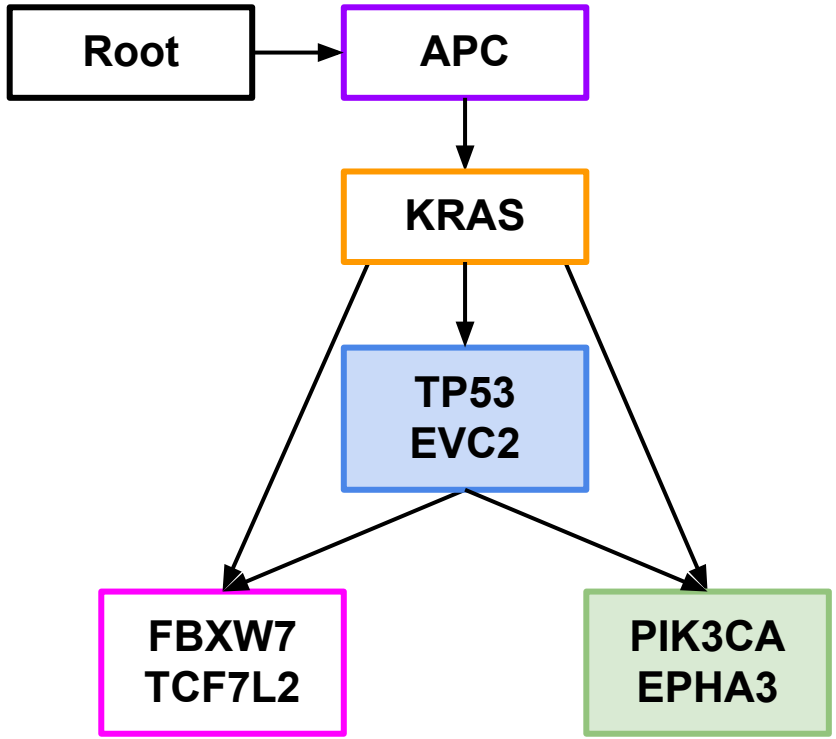
Colorectal cancer: What if there was a restriction?

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
```



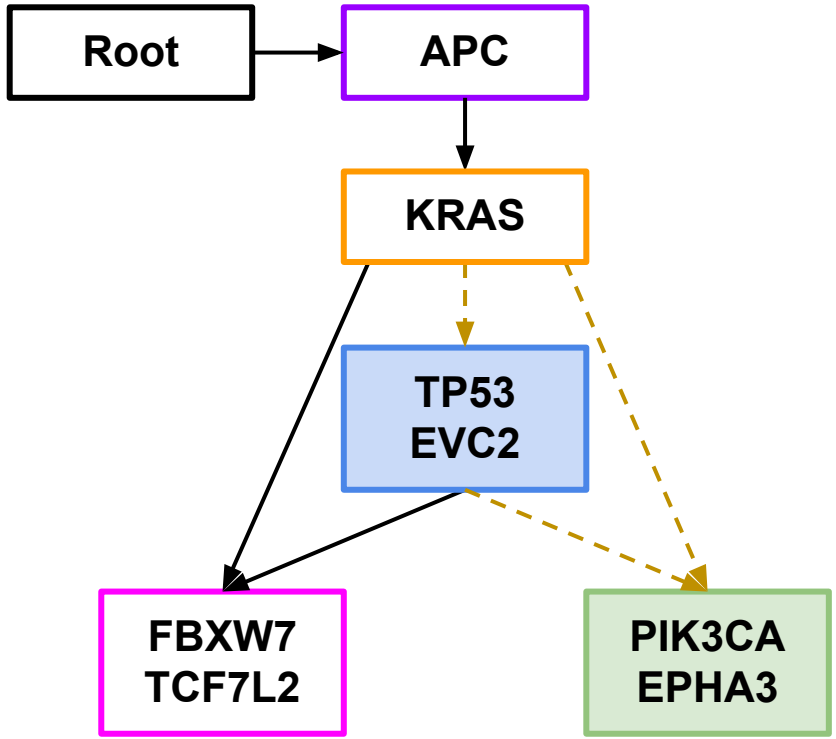
Colorectal cancer: Comparing both models

```
eAG_wood_order[50, ]
```

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

```
eAG_wood_order[63, ]
```

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



Colorectal cancer: Comparing both models

```
eAG_wood_order[50, ]
```

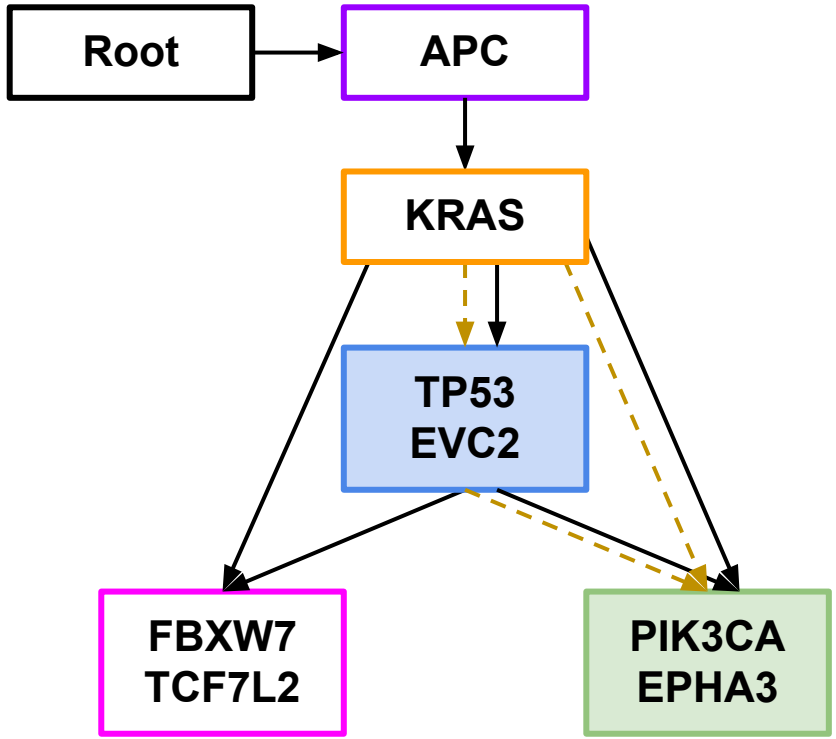
```
##                Genotype Fitness
##       50 PIK3CA > TP53       1
```

```
eAG_wood_order[63, ]
```

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

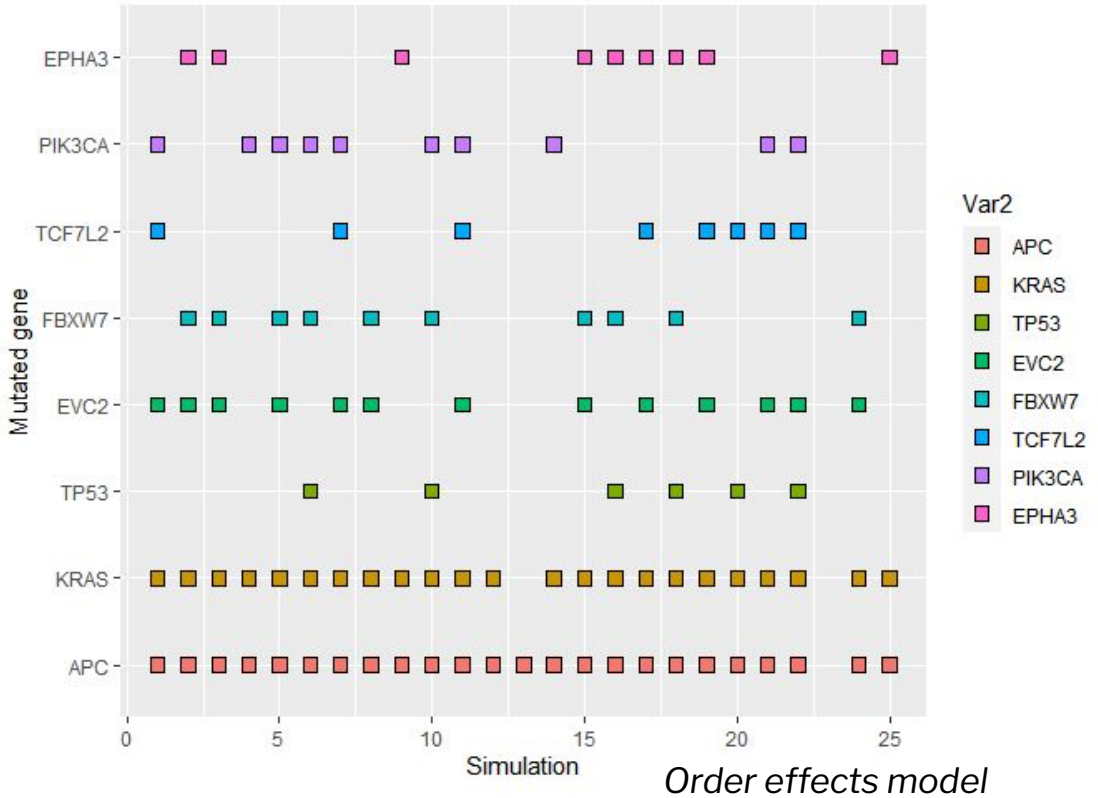
```
eAG_wood_kras[35, ]
```

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



Colorectal cancer: Simulating tumor progression

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



Order effects model

Glioblastoma: Defying the concept of "*Module*"

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

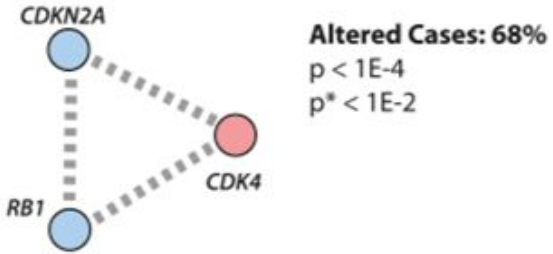
Mutual exclusivity analysis identifies oncogenic network modules

Giovanni Ciriello,[1,3,4] Ethan Cerami,[1,2,3] Chris Sander,[1] and Nikolaus Schultz[1]

Glioblastoma: 1 pathway but 2 modules

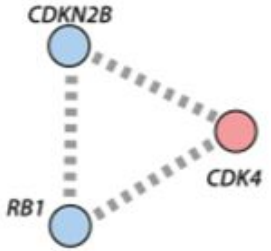
- Software applied to Glioblastoma dataset (TCGA)

- 2 mutually exclusive modules. Rb1 pathway



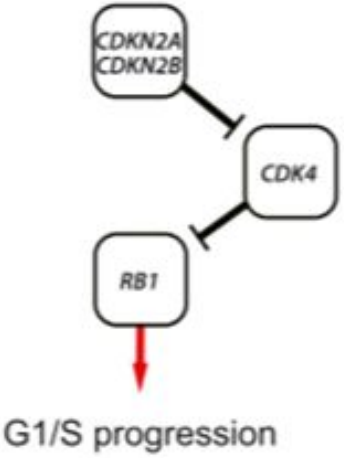
Altered Cases: 68%
p < 1E-4
p* < 1E-2

CDKN2A | CDK4 | RB1



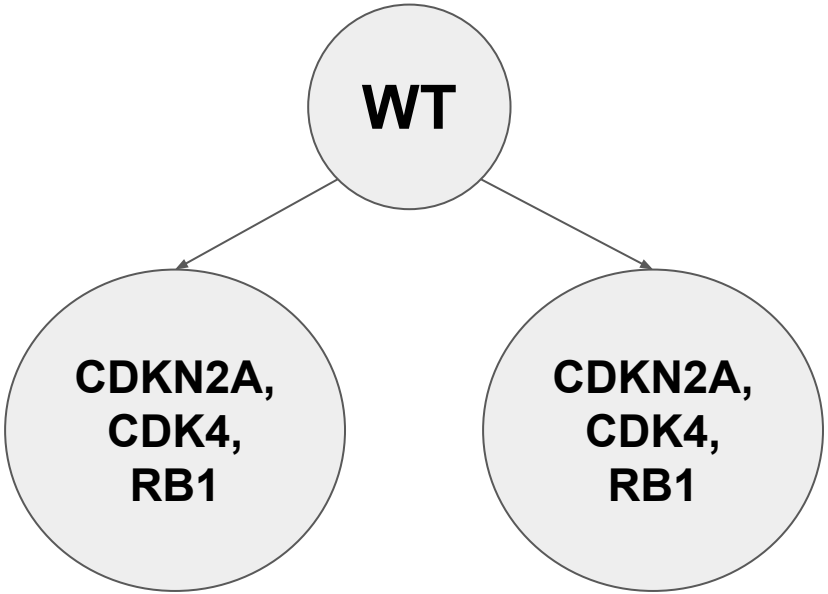
Altered Cases: 74%
p < 1E-4
p* < 1E-2

CDKN2B | CDK4 | RB1



Glioblastoma: 1 pathway but 2 modules

- ¿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, o
ne.to.one = gMOneToOne) :
  Are there identical gene names in differ
ent modules?
```

Glioblastoma: 1 pathway but 2 modules

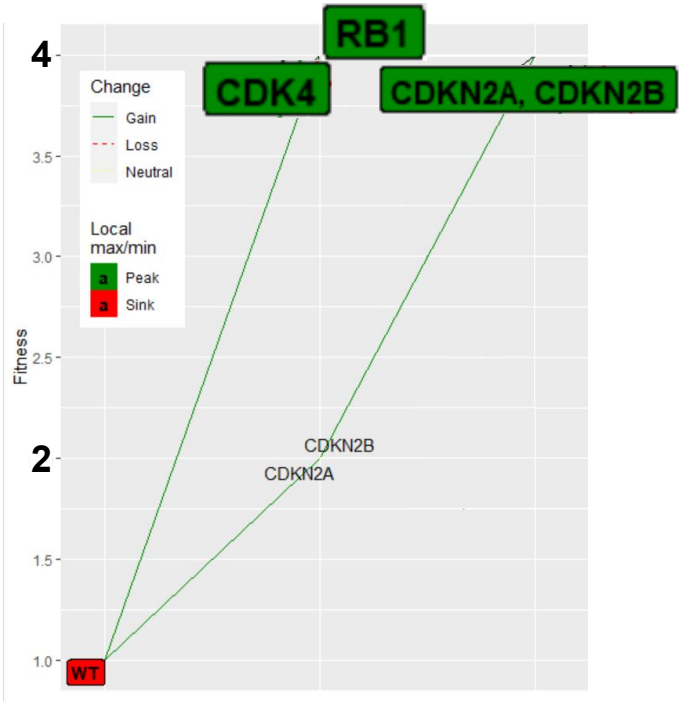
- Define effects of each possible genotype manually

1. CDKN2A+CDKN2B, CDK4 and Rb1 should have the same fitness

Glioblastoma: 1 pathway but 2 modules

2. CDKN2A and CDKN2B should have lower fitness

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



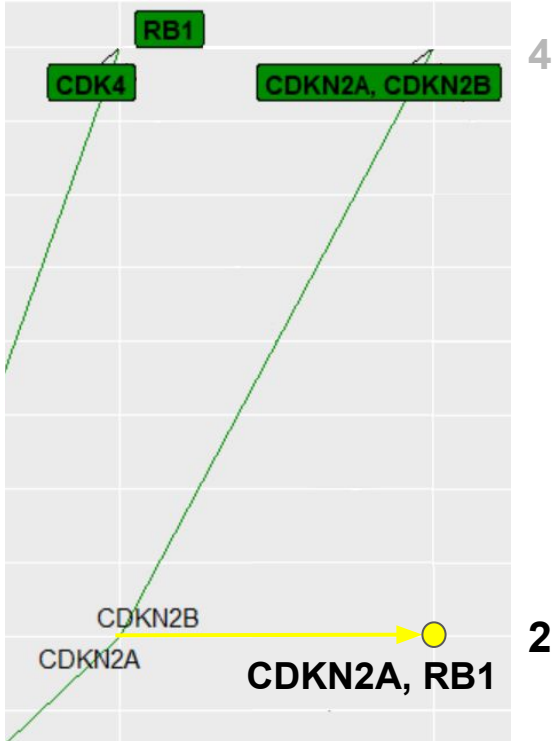
Glioblastoma: 1 pathway but 2 modules

3. Fitness of more complex clones?

   a) Mutating in RB1 to a CDKN2A clone

↓

Fitness doesn't change



Glioblastoma: 1 pathway but 2 modules

3. Fitness of more complex clones?

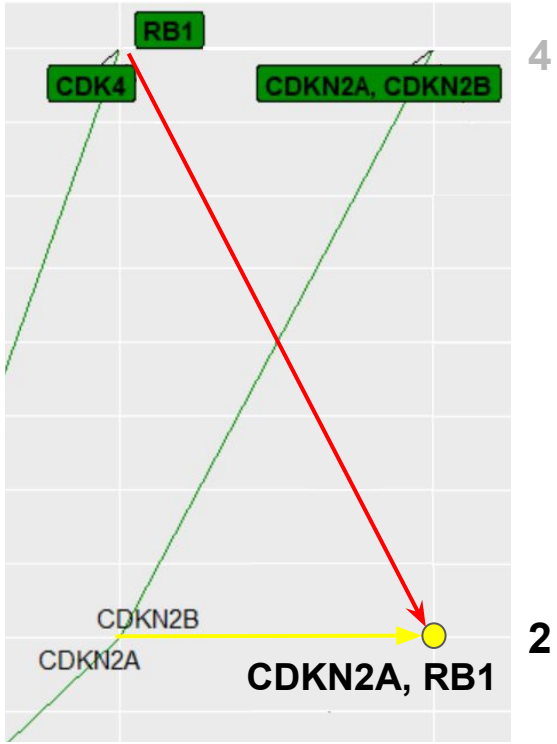
   a) Mutating in RB1 to a CDKN2A clone

↓

Fitness doesn't change

↓

Path from RB1 involves a decrease



Glioblastoma: 1 pathway but 2 modules

3. Fitness of more complex clones?

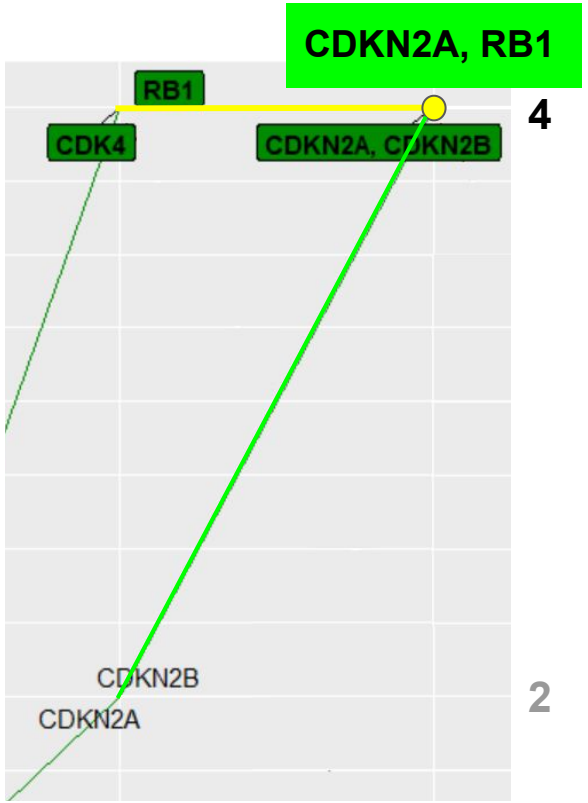
   b) Mutating CDKN2 in a RB1 clone

↓

Fitness doesn't change

↓

Path from CDKN2 breaks exclusivity



Glioblastoma: 1 pathway but 2 modules

3. Fitness of more complex clones?

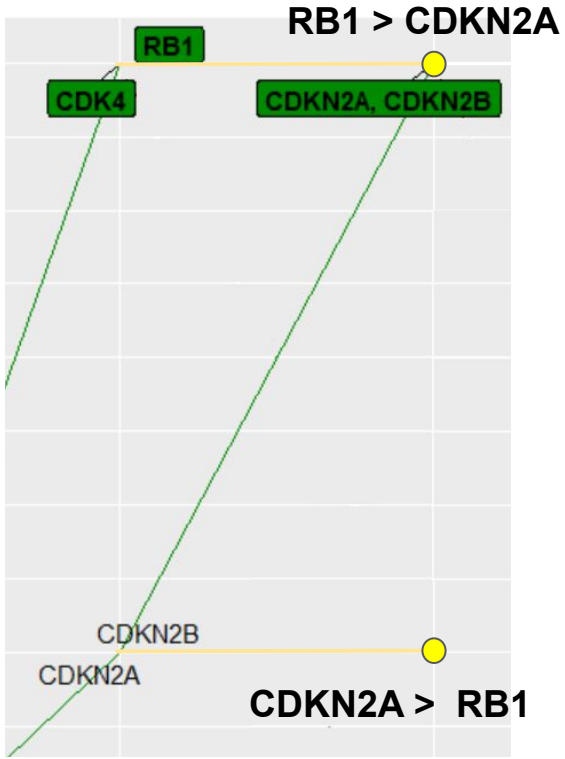
c) Specify order effects

↓

Fitness doesn't change

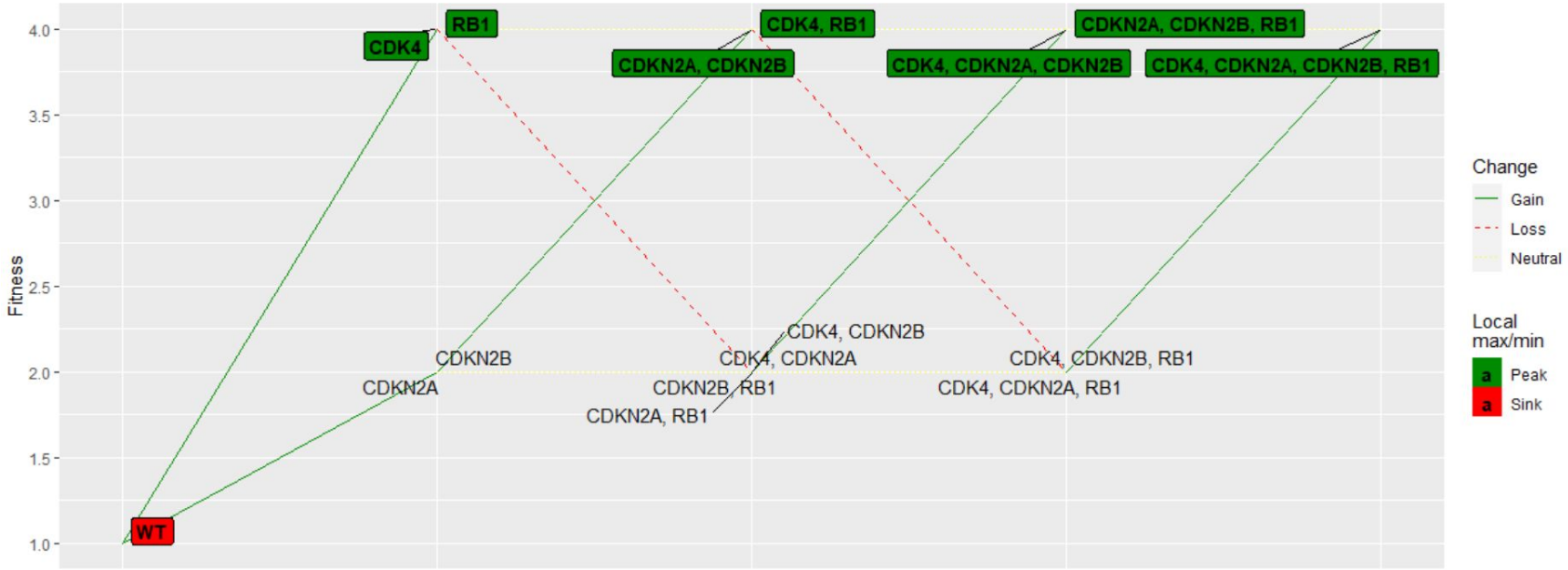
↓

Biological justification?



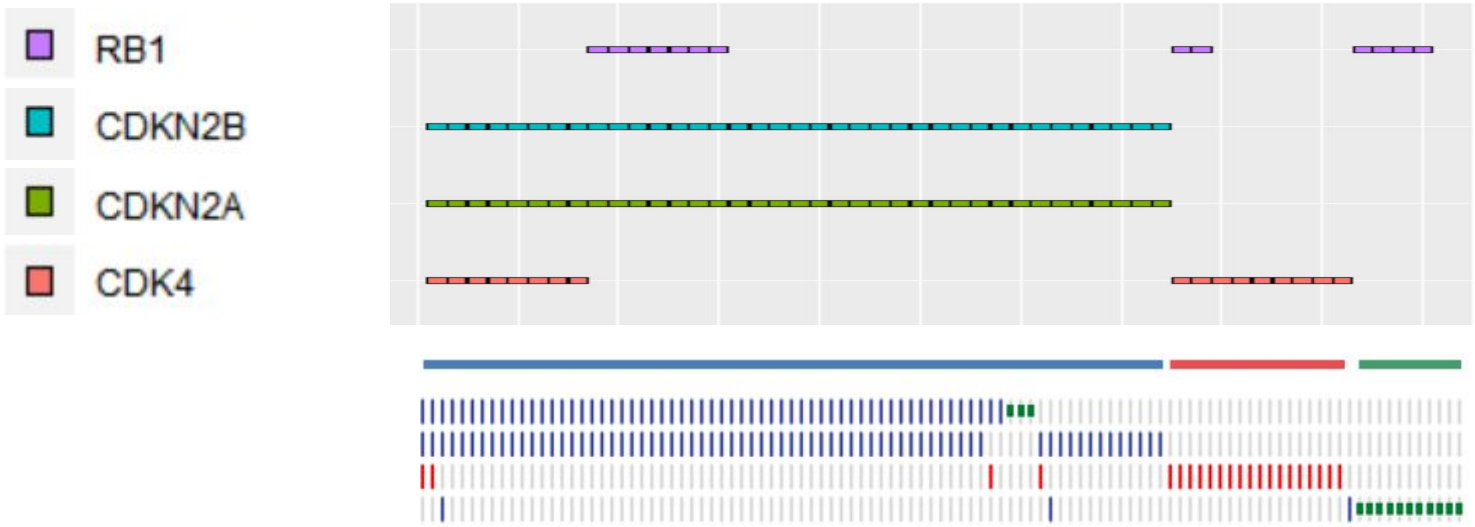
Glioblastoma: 1 pathway but 2 modules

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



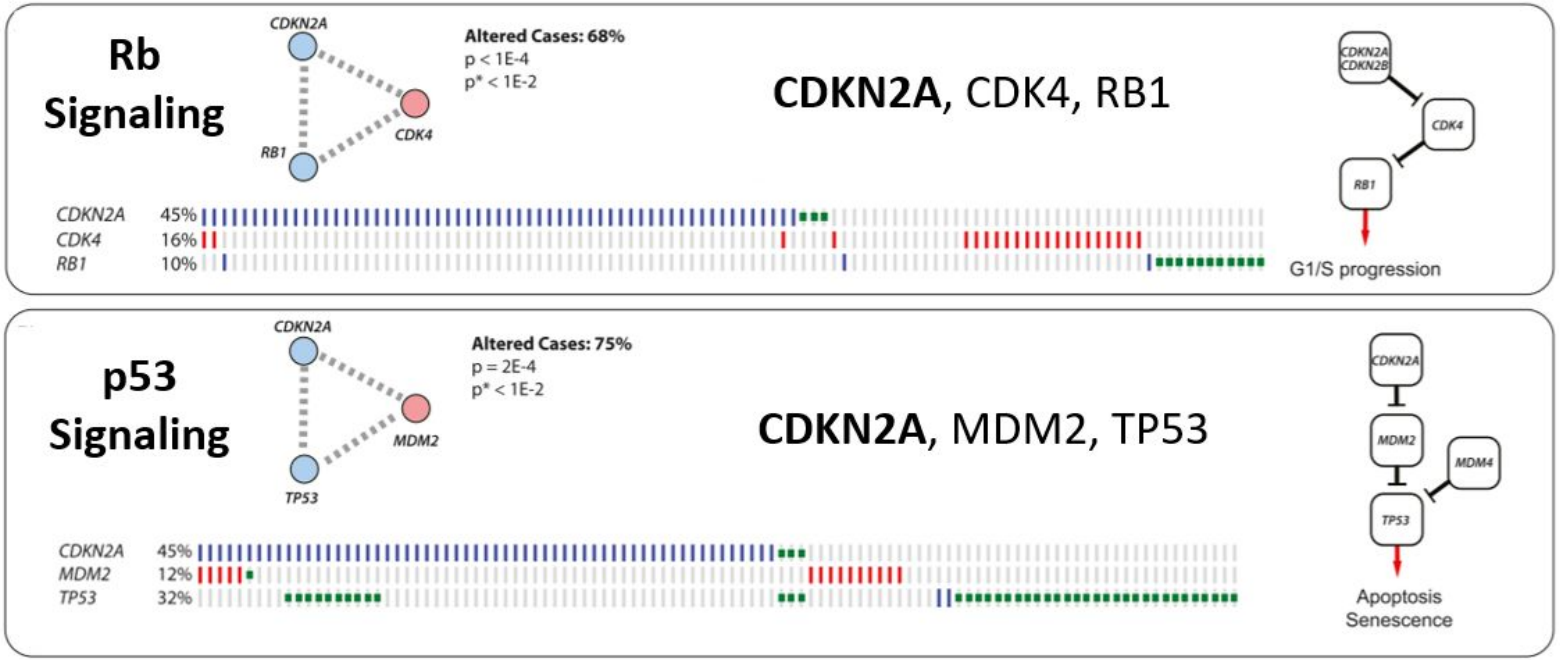
Glioblastoma: 1 pathway but 2 modules

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



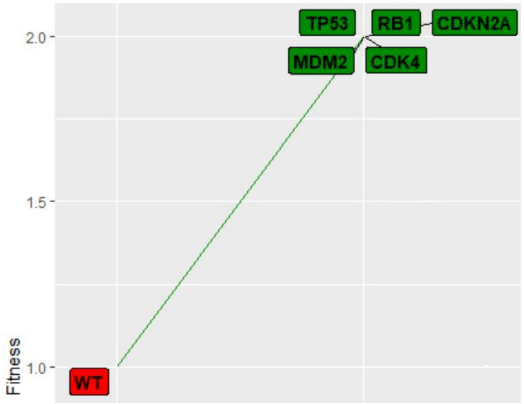
Glioblastoma: 2 pathways and 1 shared gene

- 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



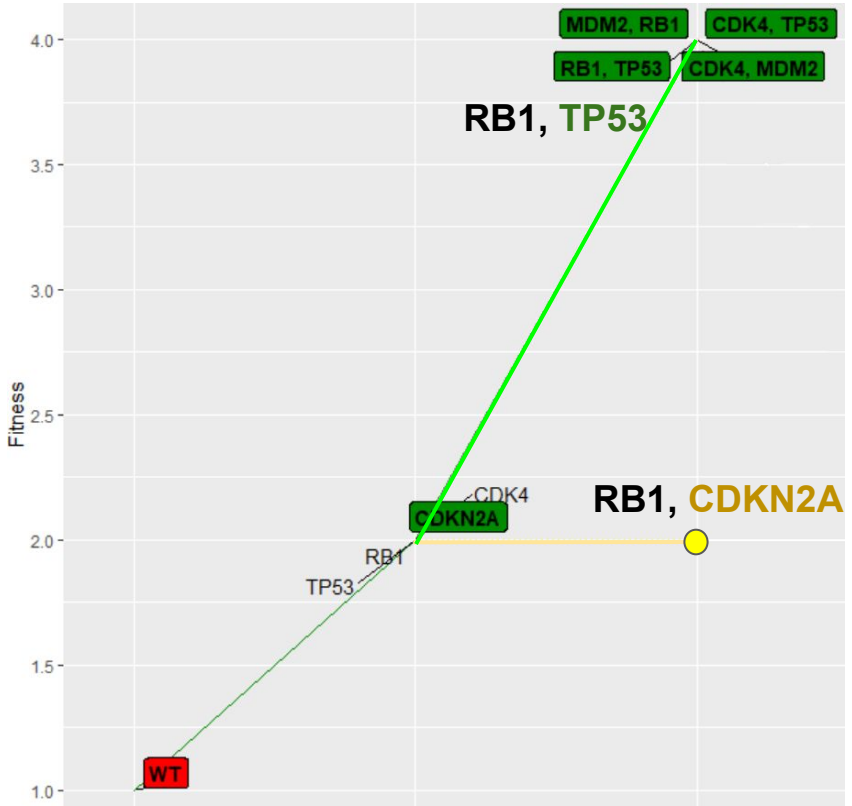
Glioblastoma: 1 pathway but 2 modules

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

    RB1 + P53 -> fitness = 4

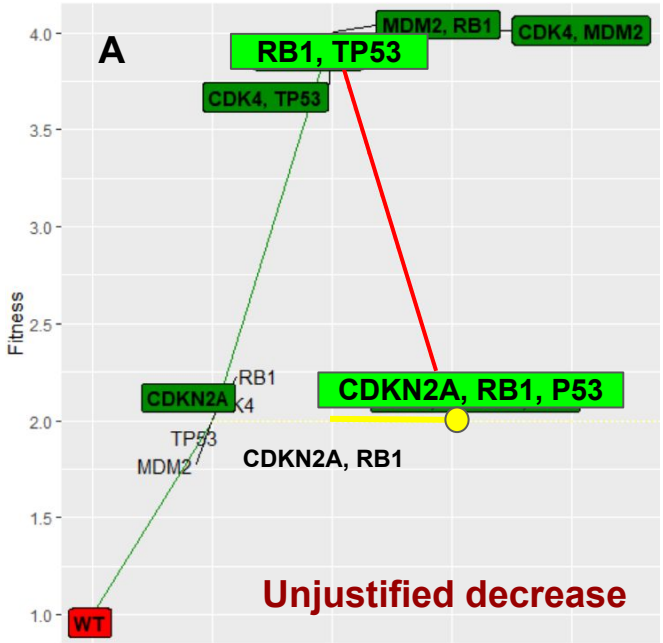
  - Mutual exclusivity? **Same fitness**

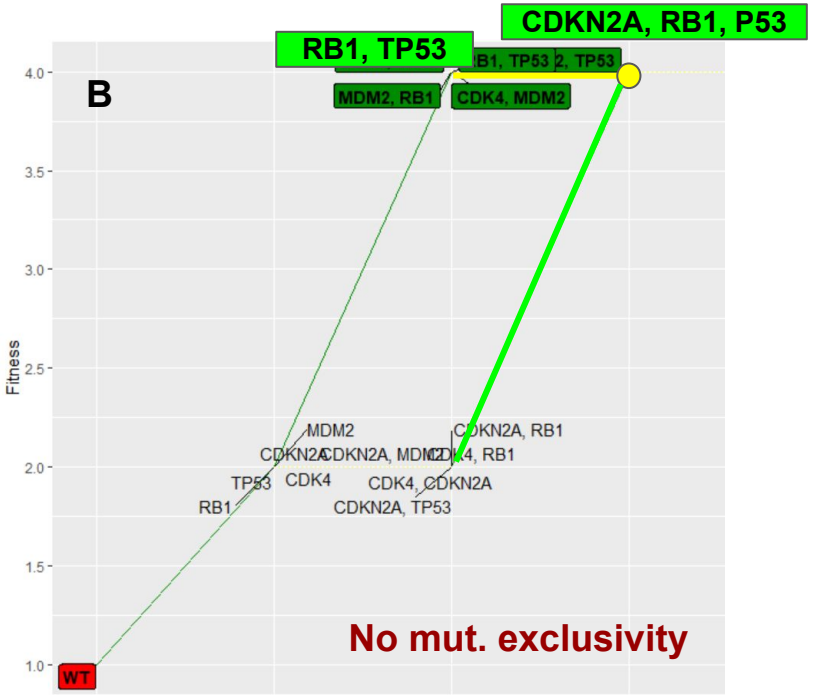
    RB1 + CDKN2A -> fitness = 2



Glioblastoma: 1 pathway but 2 modules

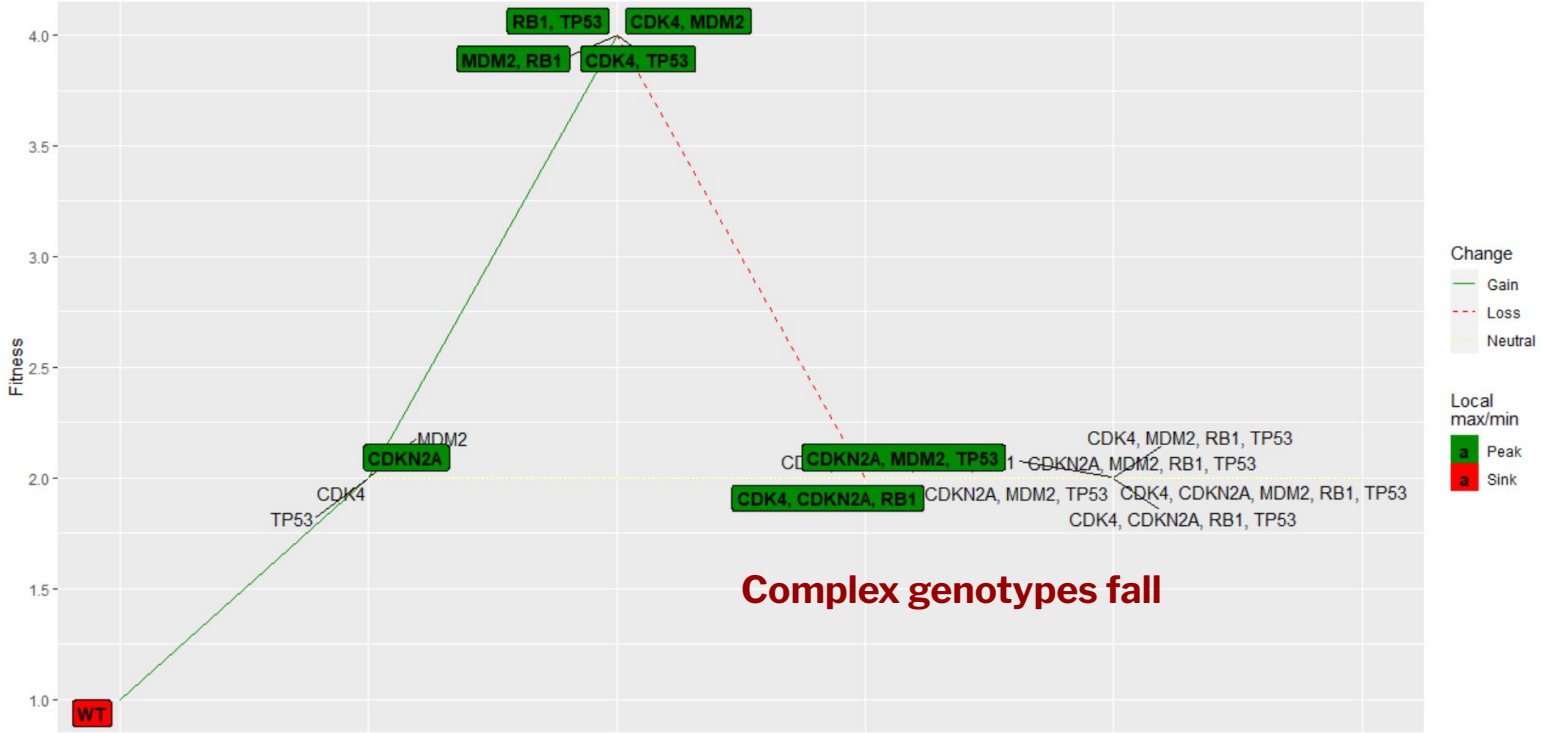
- Can define fitness of clones with 3 mutations again...





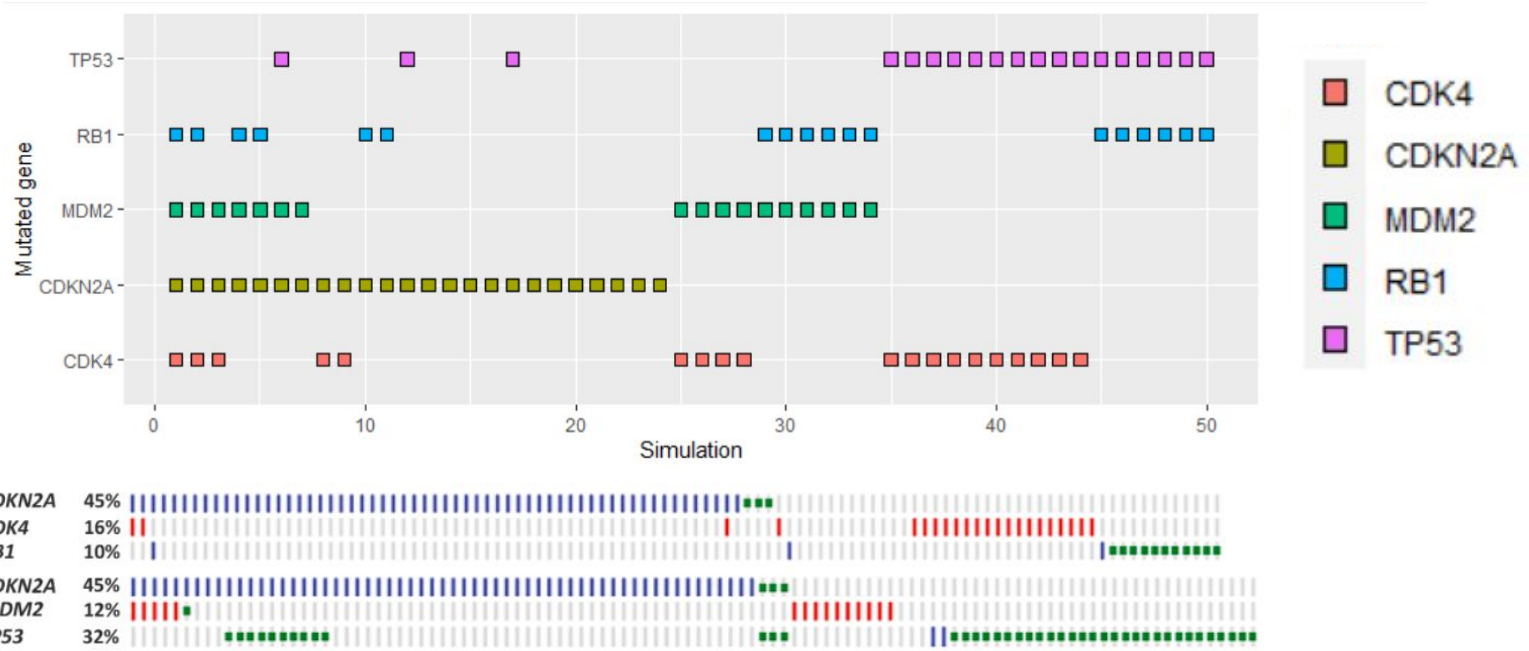
Glioblastoma: 1 pathway but 2 modules

- Apply first option (as in previous case):



Glioblastoma: 1 pathway but 2 modules

- Multiple simulations:



(Can’t really compare data)

Summary

Summary

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

Summary

- 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

Summary

- 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

Summary

- 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)

Summary

- 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*)

Summary

- 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