Fast simulations of clonal expansion on networks

Cancer, copy number alterations, and age

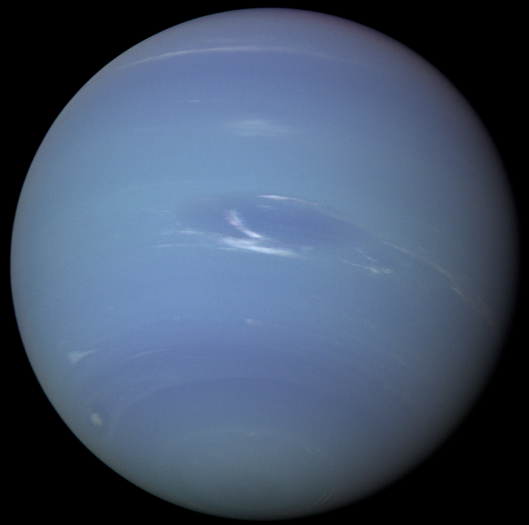
Chay Paterson

University of Manchester

18 October 2022

Introduction

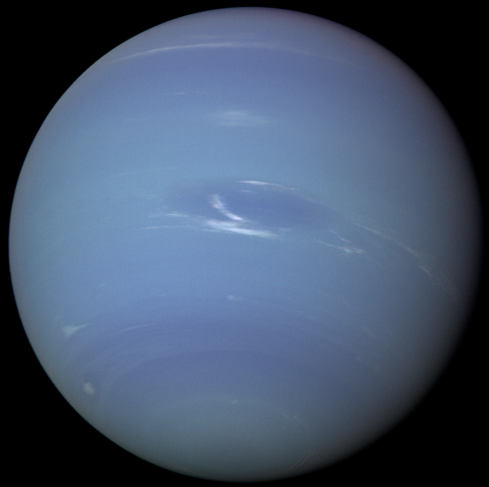
Introduction



The value of good models

The big picture

Mechanistic model ↓



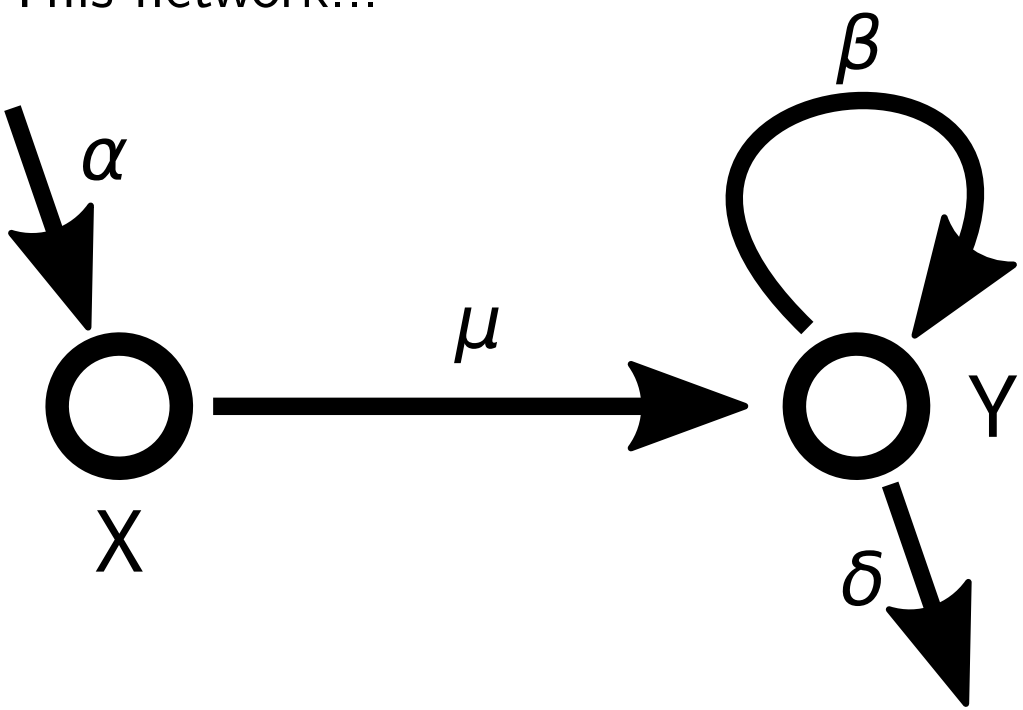
↑ within 1° of predicted location

[1] J.G. Galle, Monthly Notices of the Royal Astronomical Society 7 (1846): 153.

Warning

There will be very few equations in this talk!

This network...



corresponds to this stochastic process:...

$$\begin{aligned} \emptyset &\to X, && \text{rate } \alpha \\ X &\to X + Y, && \text{rate } \mu X \\ Y &\to Y + Y, && \text{rate } \beta Y \\ Y &\to \emptyset, && \text{rate } \delta Y \end{aligned}$$

and approximately this linear system...

$$\frac{d}{dt}\begin{pmatrix} E[X] \\ E[Y] \end{pmatrix} = \begin{bmatrix} 0 & 0 \\ \mu & \beta - \delta \end{bmatrix} \cdot \begin{pmatrix} E[X] \\ E[Y] \end{pmatrix} + \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$$

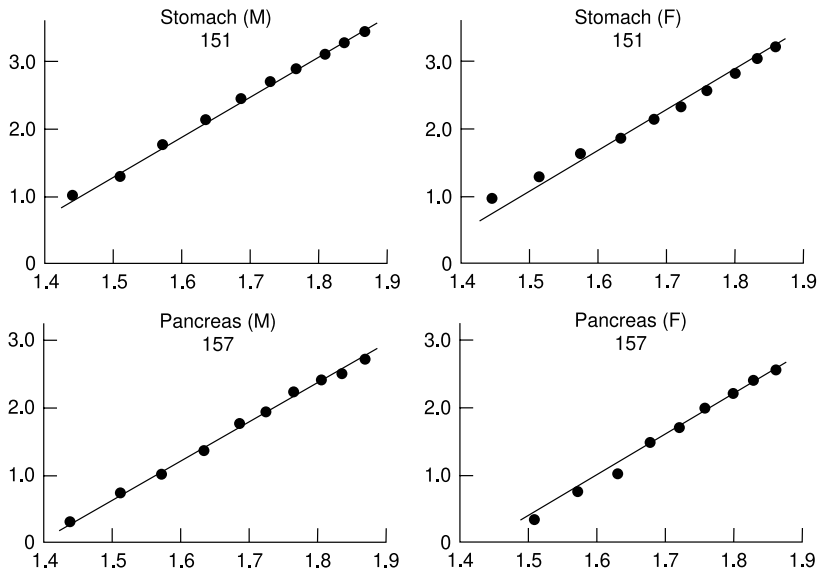
Most of our models are linear, high-dimensional and sparse[1]

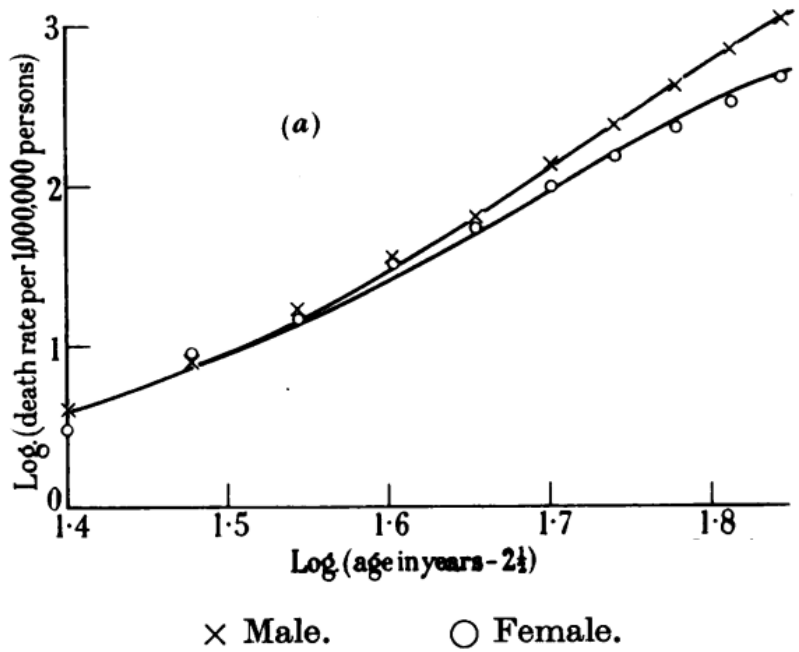
[1] C. Paterson, I. Bozic, H. Clevers, PNAS 2020; 117(34): 20681-20688

Age and cancer

P. Armitage and R. Doll[12]

Risk of cancer increases with age:





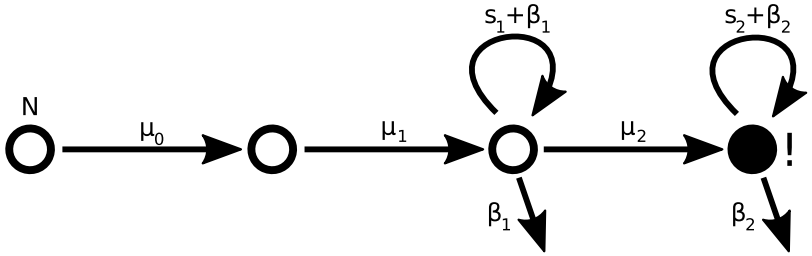
[1]P. Armitage and R. Doll, British Journal of Cancer 1954; 8: 1–12

[2]P. Armitage and R. Doll, British Journal of Cancer 1957; 11(2): 161-169

Age and cancer

P. Armitage and R. Doll[12]

Why? Accumulating mutations[123]:



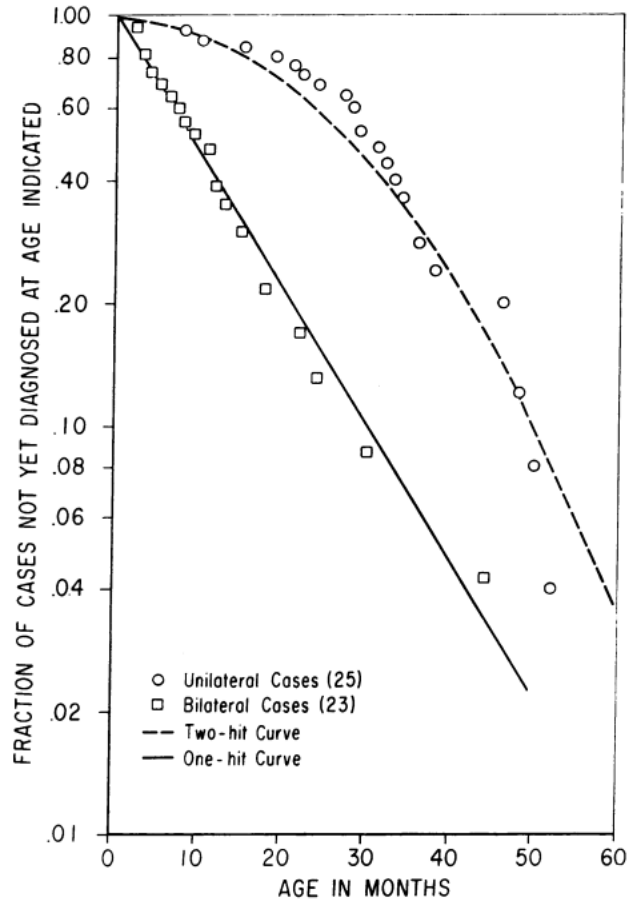


FIG. 1. Semilogarithmic plot of fraction of cases of retinoblastoma not yet diagnosed (S) vs. age in months (t). The one-hit curve was calculated from $\log S = -t/30$, the two-hit curve from $\log S = -4 \times 10^{-5} t^2$.

[1] P. Armitage and R. Doll, British Journal of Cancer 1954; 8: 1–12

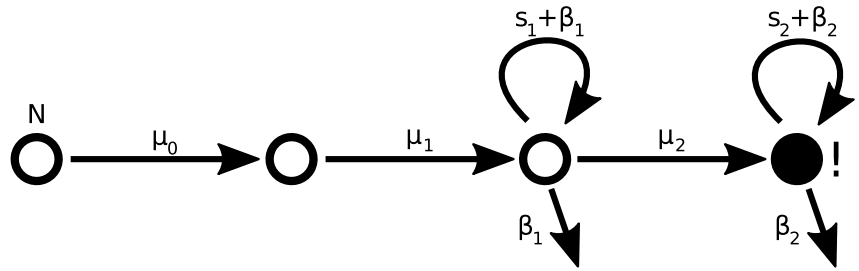
[2] P. Armitage and R. Doll, British Journal of Cancer 1957; 11(2): 161-169

[3] F. Michor, Y. Iwasa and MA. Nowak, Nature Reviews Cancer 2004; 4:

Multi-stage clonal expansion models

2-3 rate limiting steps[123]

Problem: how to compute $P(cancer, t)$ for a given model?



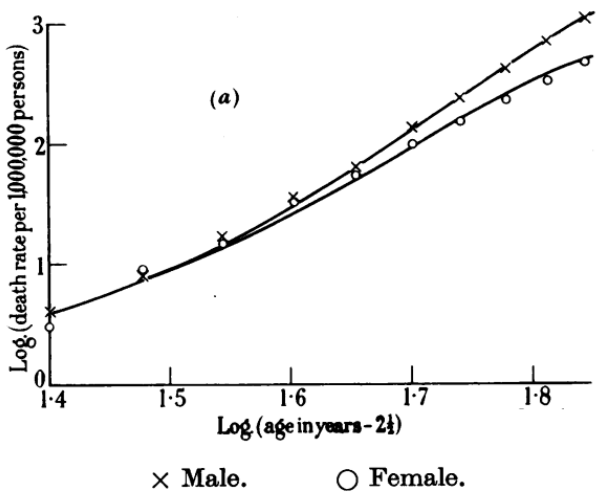
Different methods: Fast:

- Mean-field approximation[1]
- Numerical quadrature[2]

Slow:

- Gillespie algorithm + sampling [3]
- tau leaping + sampling [3]

↓?



[1] P. Armitage and R. Doll, British Journal of Cancer 1957; 11(2): 161-169

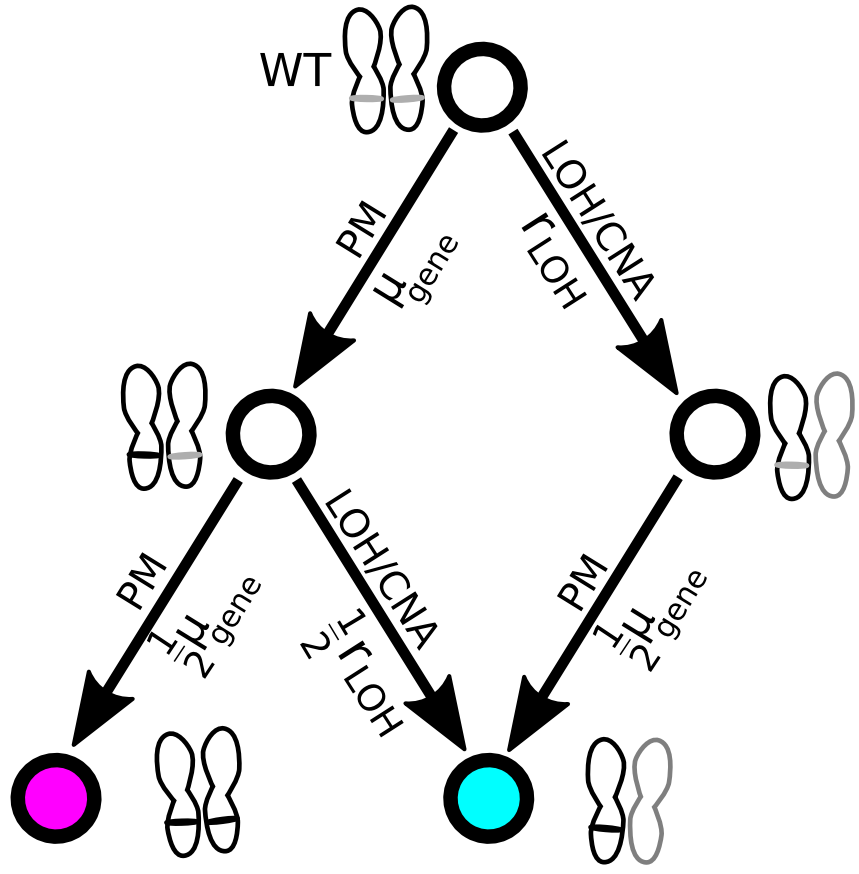
[2] S. Moolgavkar and G. Luebeck, JNCI 1992; 84(8): 610-618

[3] C. Paterson, I. Bozic, H. Clevers, PNAS 2020; 117(34): 20681-20688 (supp. material)

Network models

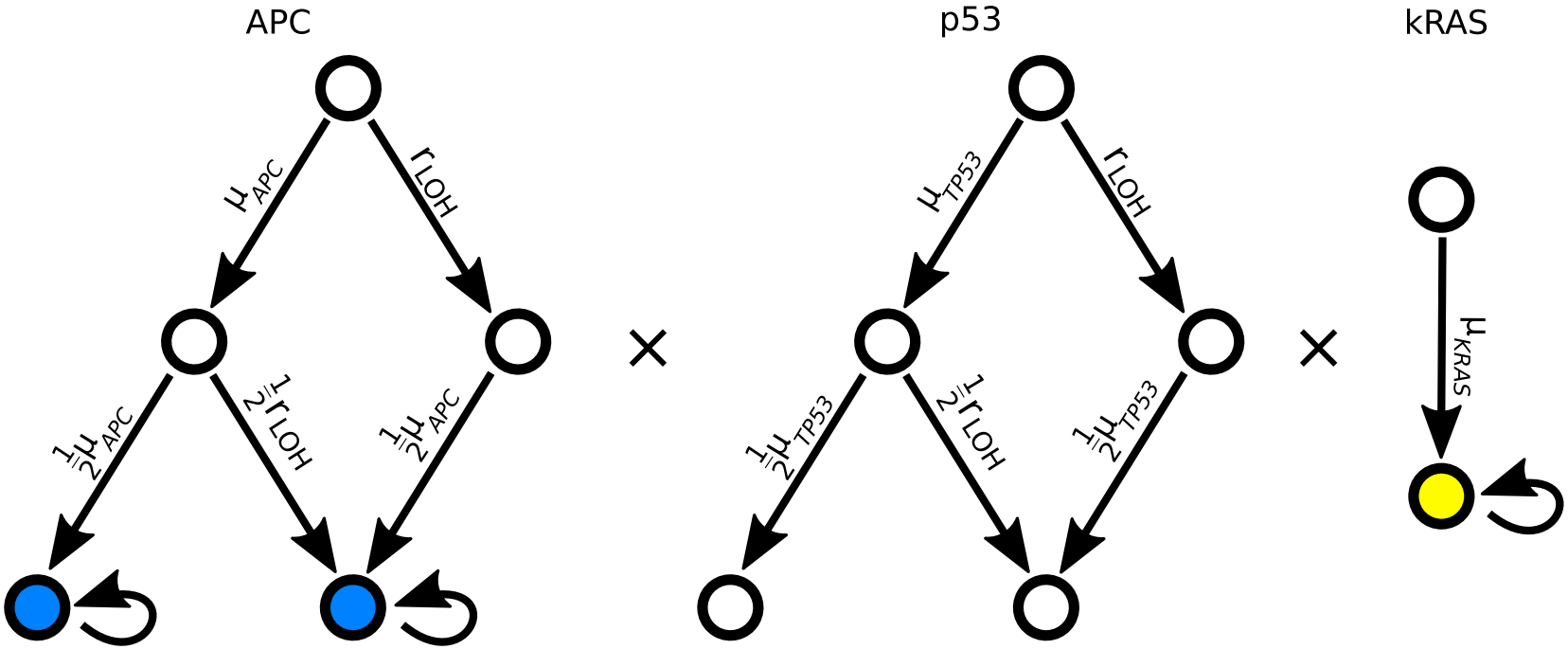
1. Study **specific genes** and mechanisms of interest (SNVs, LOH, CNA, etc.)
2. Fix parameters from sequences and experiments
3. Distinguish different orders of events

example model



This gets us the incidence of *specific karyotypes*

Colorectal adenocarcinoma model



- *APC-p53-kRAS* combo accounts for about 15% of incidence
- 5-year survival about 60% (any stage)

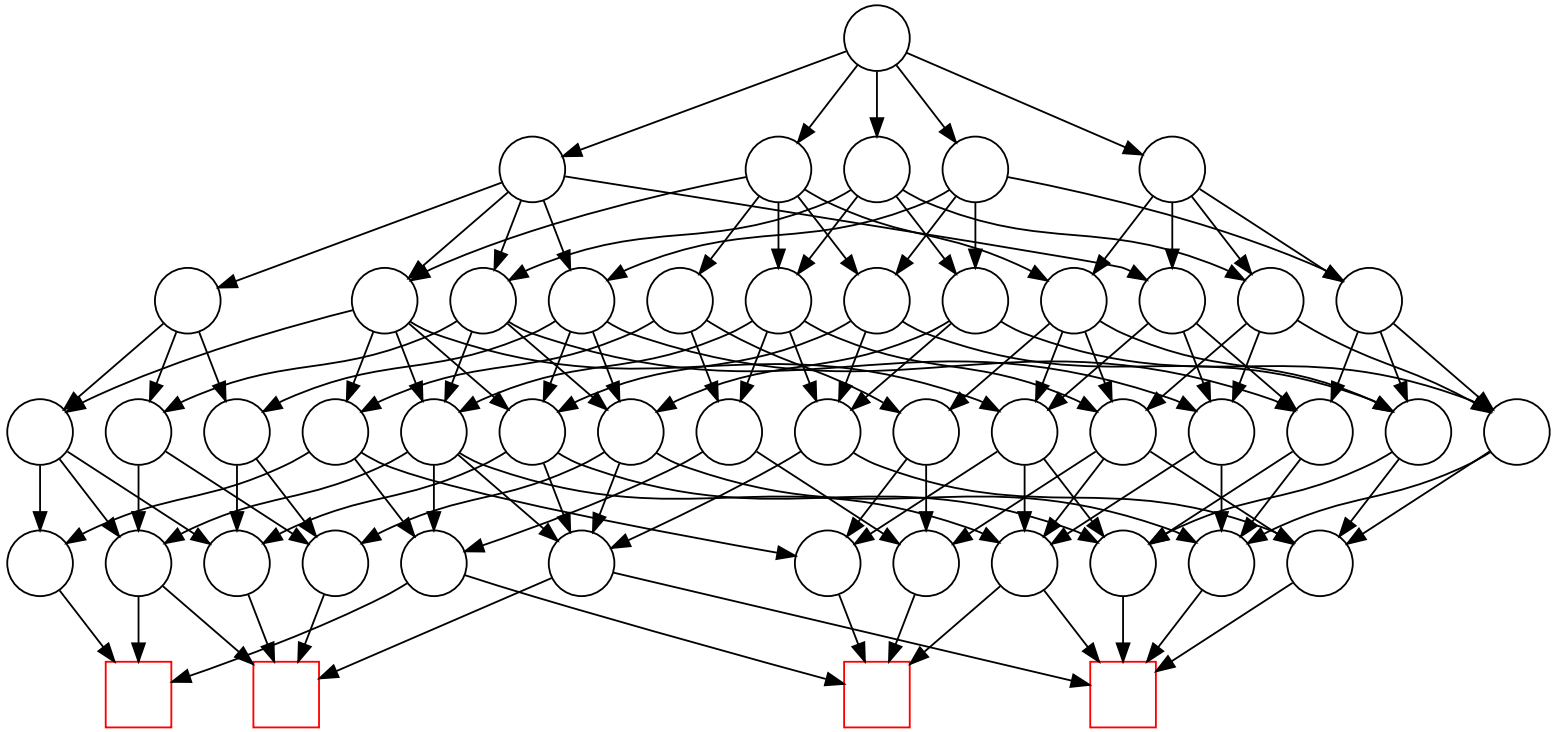
---

[1] Fearon et al. TODO

[1] M. S. Lawrence et al., Nature 2014; 505: 495–501

[2] Office for National Statistics, England 2019

Colorectal adenocarcinoma model

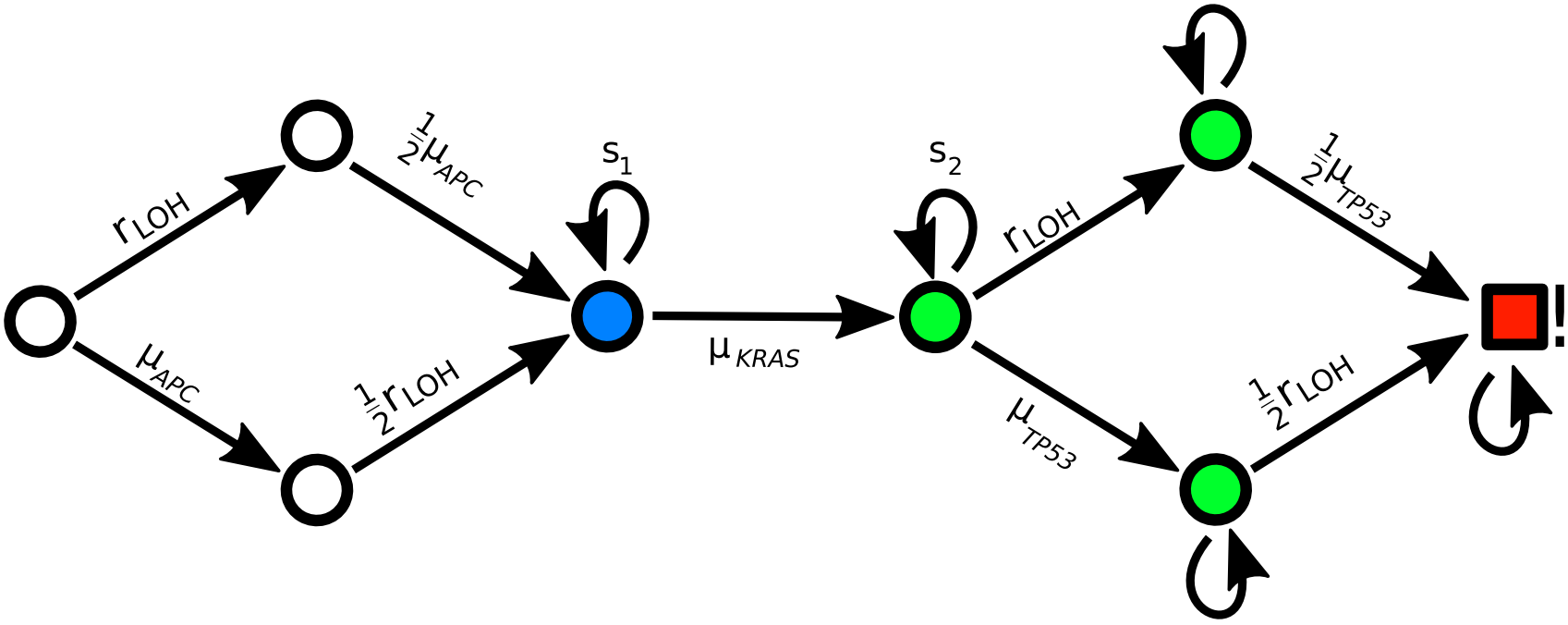


each end node is a different copy number profile

e.g. $(-17p, -5q)$, etc

[1] C. Paterson, I. Bozic, H. Clevers, PNAS 2020; 117(34): 20681-20688 (supp. material)

Colorectal adenocarcinoma model



$$P(t) \sim te^{s_2 t}$$

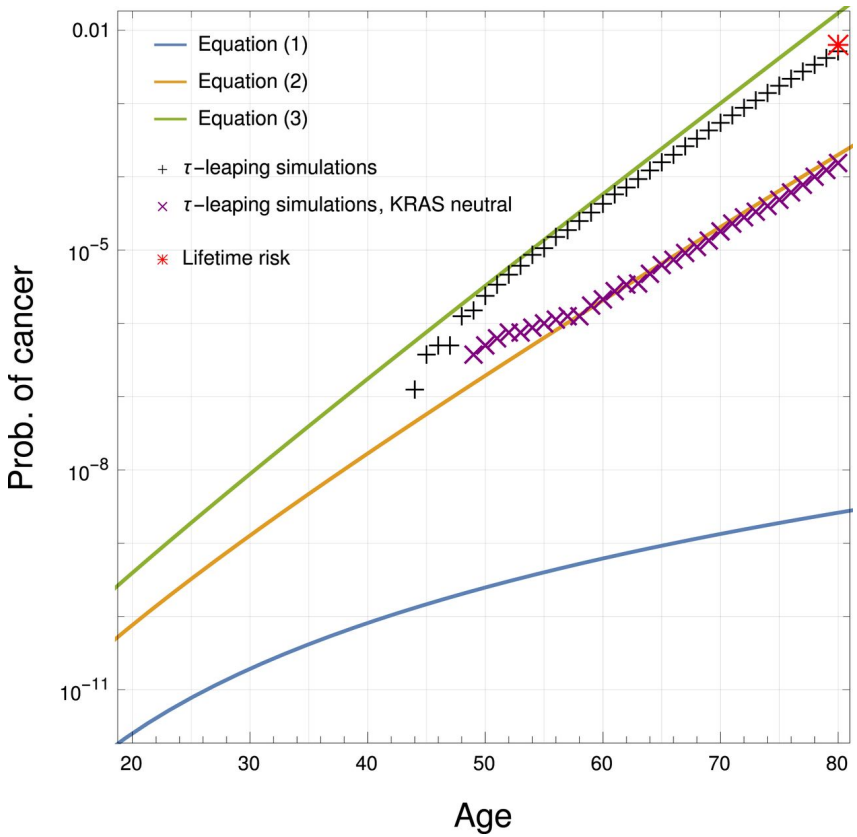
- The 4 most likely paths account for 50% of the risk
- Consistent with classic model[2]

[1] C. Paterson, I. Bozic, H. Clevers, PNAS 2020; 117(34): 20681-20688

[2] Kinzler and Vogelstein 1990

Colorectal adenocarcinoma model

Successful *ab initio* model



- Can constrain *APC/KRAS* epistasis ($s_2 < 0.31$/yr)
- Timing of *p53* inactivation: must be *late*
- *Compatible with* 3-hit models, similar curve: *p53 not rate limiting*

but:

- Mean-field breaks down at old ages / large probabilities
- Stochastic simulations are *extremely slow*

---

[1] C. Paterson, I. Bozic, H. Clevers, PNAS 2020; 117(34): 20681-20688

[2] (relative to the others)

Why study vestibular schwannoma?

Sometimes rare events make more interesting science possible

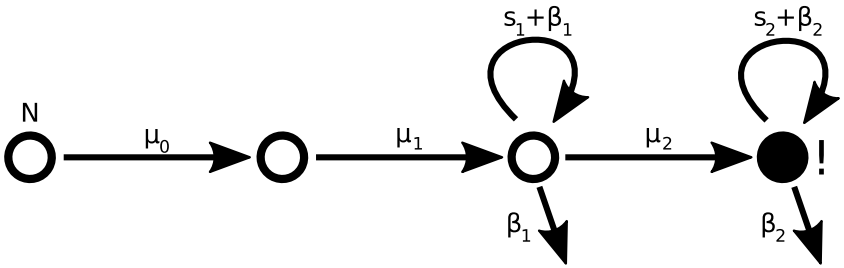
1. genomic subtypes better characterised: *NF2/Merlin* altered in $85 - 100\%$ of cases[1,2], *TP53* in $\approx 0\%$
2. usually benign = **easier to study timing!** (of drivers, malignant transformation...)
3. only 3 hits, weak selection (almost neutral)
4. probabilities low: approximations v. accurate **because it's rare**

[1] ML Carlson et al., Otology & Neurotology: 2018;39(9):860 – 871

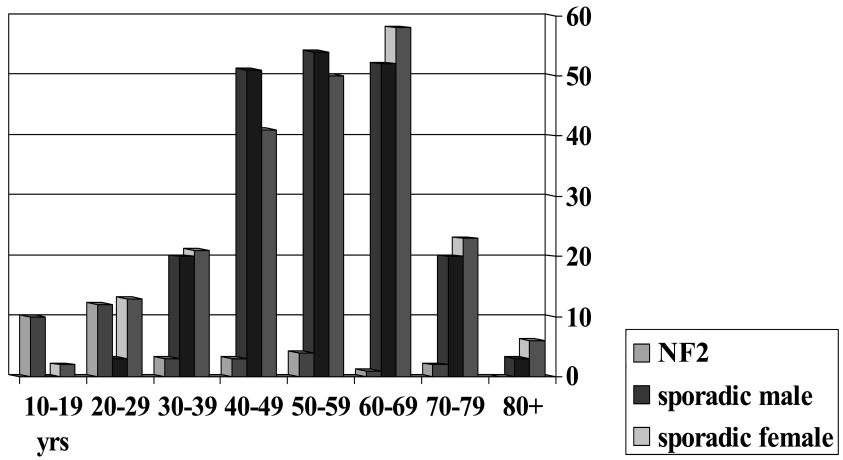
[2] AL Håvik et al., Journal of Neurosurgery JNS. 2017;128(3):911 – 922.

Vestibular schwannoma

3-*event* model



- Fitness suspiciously low, $s \approx 0.005/\text{yr}$ [1]
- Suggests nearly-neutral 3-hit model [3]



Gareth Evans 2005[2]

---

[1] R. Woods *et al.* Genetic Epidemiology (2003)24: 265–272

[2] DGR. Evans *et al.* Otology & Neurotology (2005)26:93–97

[3] C Paterson, I Bozic, MJ Smith, X Hoad, DGR Evans, https://doi.org/10.1101/2021.10.03.457528

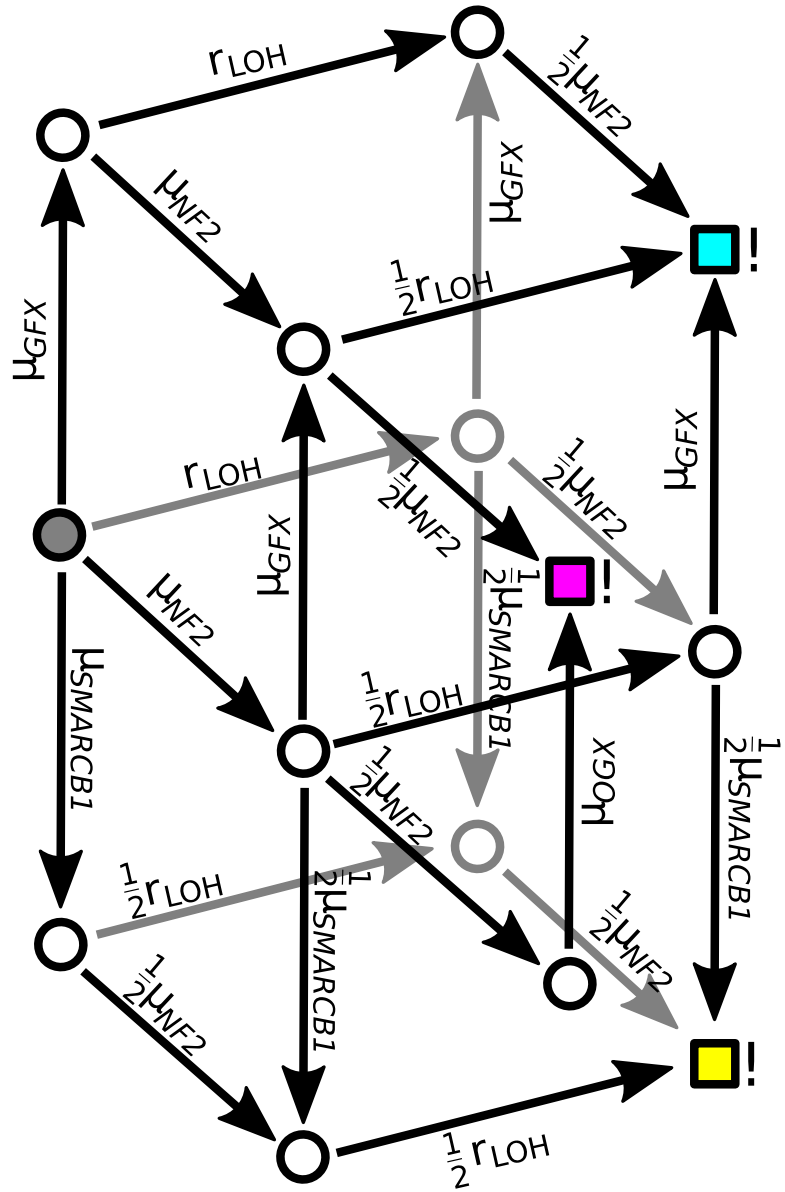
Vestibular schwannoma incidence

Our model for sporadic VS

- Include *NF2*, *SMARCB1* and (simplified) linkage
- Add hypothetical oncogene *GFX*

Risk of each subtype looks like

$$P(\blacksquare) \propto \frac{t^3}{3!}$$



[1] C Paterson, I Bozic, MJ Smith, X Hoad, DGR Evans, https://doi.org/10.1101/2021.10.03.457528

Vestibular schwannoma incidence

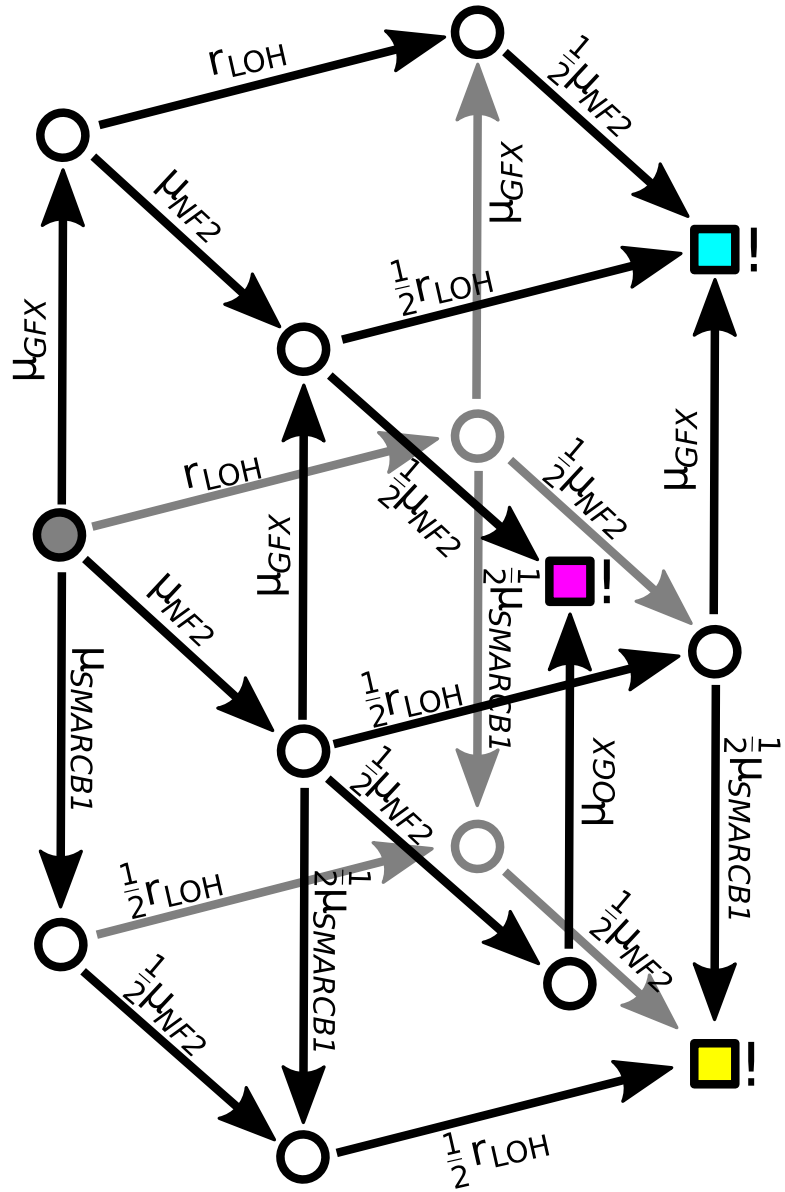
Our model for sporadic VS

- Include *NF2*, *SMARCB1* and (simplified) linkage
- Add hypothetical oncogene *GFX*

Risk of each subtype looks like

$$P(\blacksquare) \propto \frac{t^3}{3!}$$

$$P(\blacksquare) \approx N_{WT}\mu_{GFX} r_{LOH} \frac{1}{2}\mu_{NF2} \frac{t^3}{3!} \times 6$$

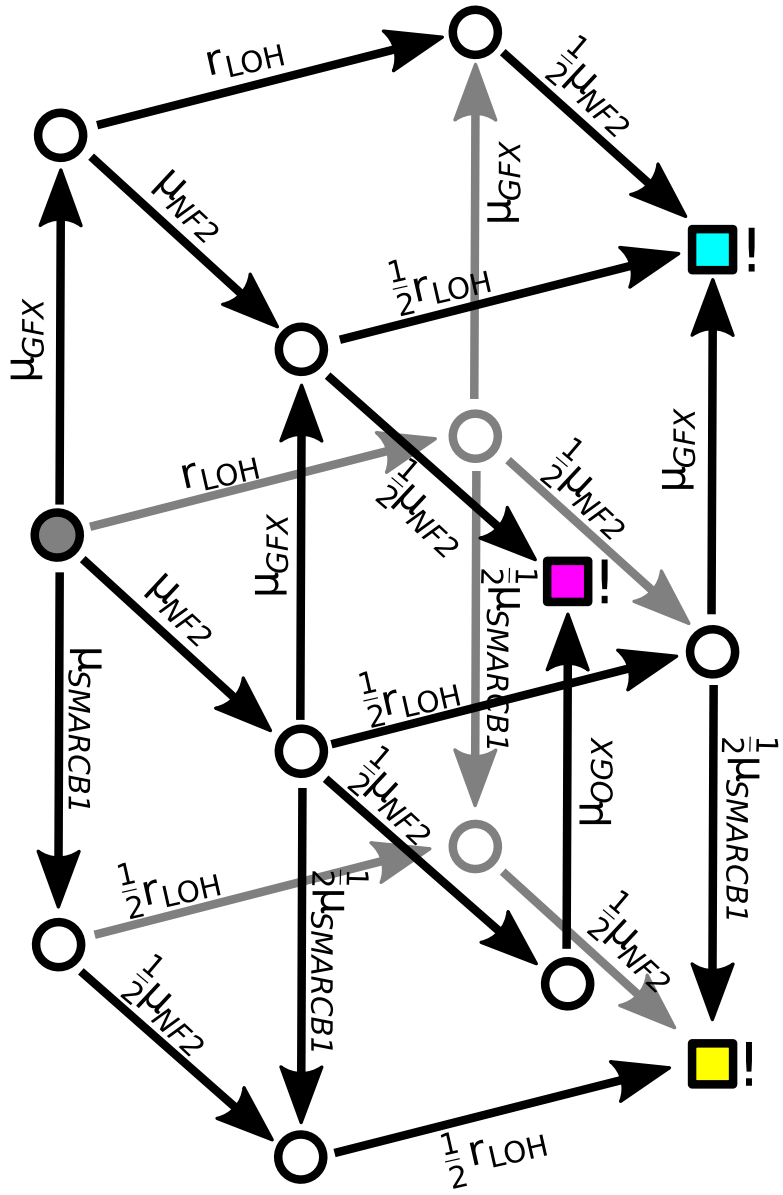


[1] C Paterson, I Bozic, MJ Smith, X Hoad, DGR Evans, https://doi.org/10.1101/2021.10.03.457528

Vestibular schwannoma incidence

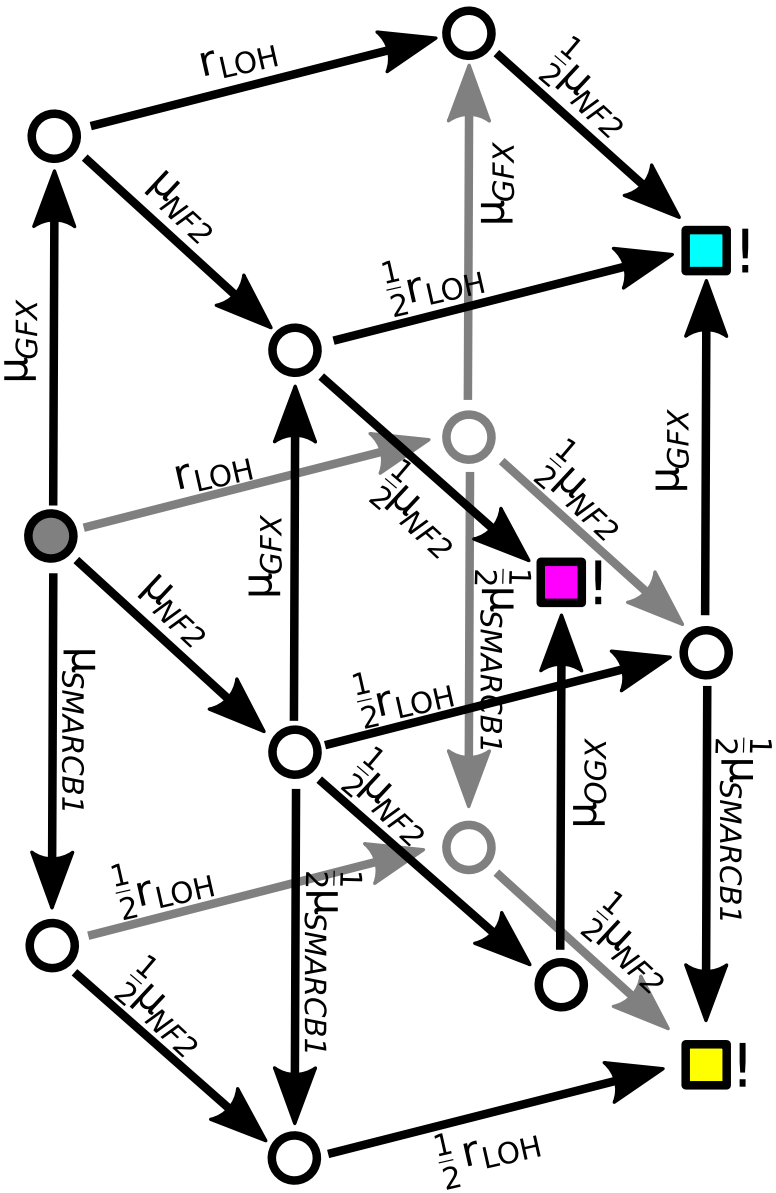
Our model for sporadic VS

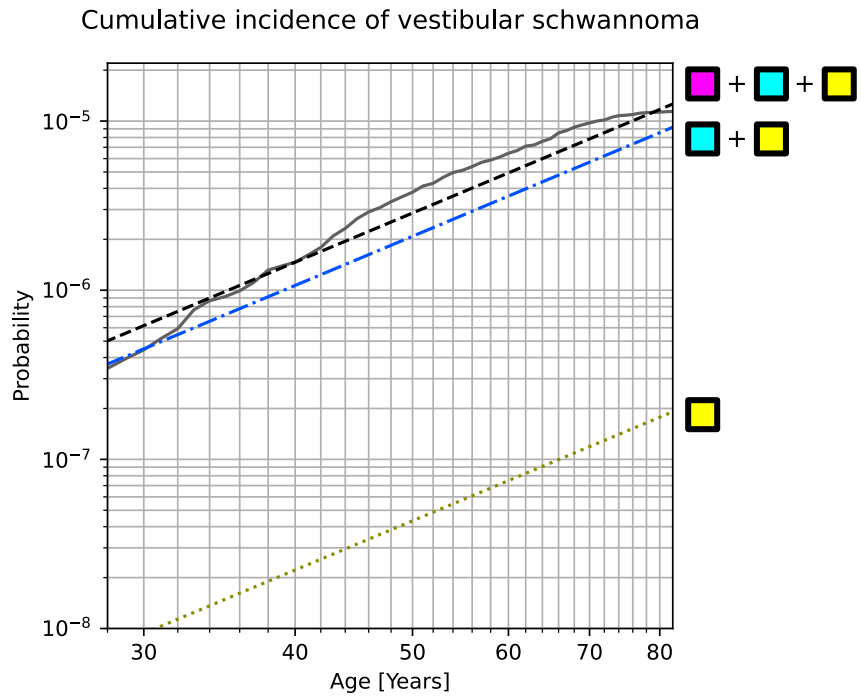
- *Don't* use *ab initio* point estimates for u, r_{LOH}, n_{GFX} this time...
- Instead use $P(t) = \square + \square + \square$, $f_{LOH} = \frac{\square + \square}{\square + \square + \square}$, and $f_{SMARCB1} = \frac{\square}{\square + \square + \square}$ to fix the parameters!



Vestibular schwannoma incidence

Our model for sporadic VS

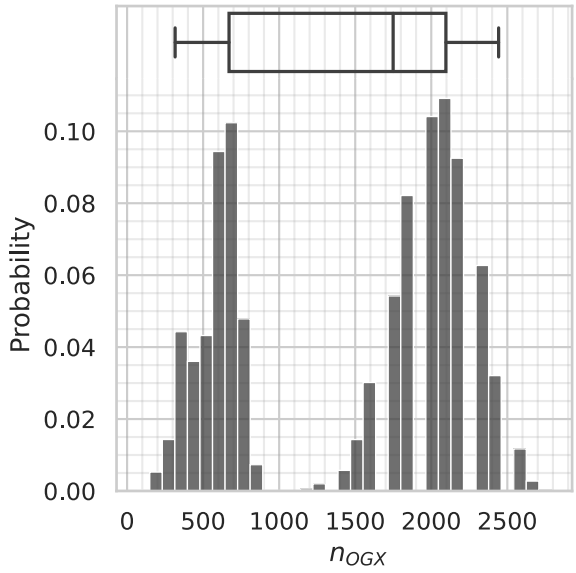


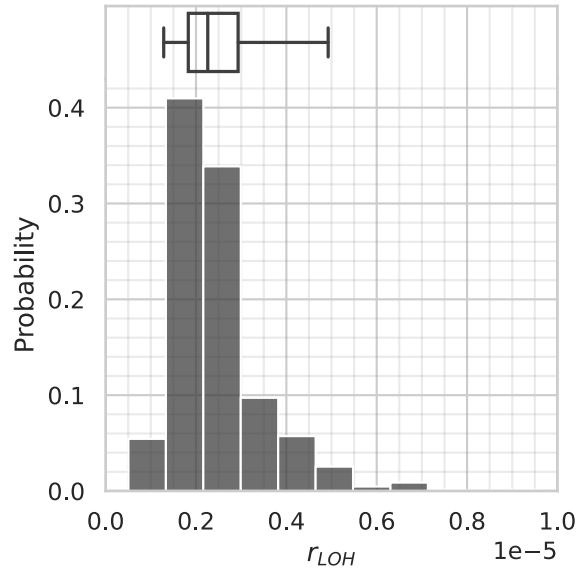


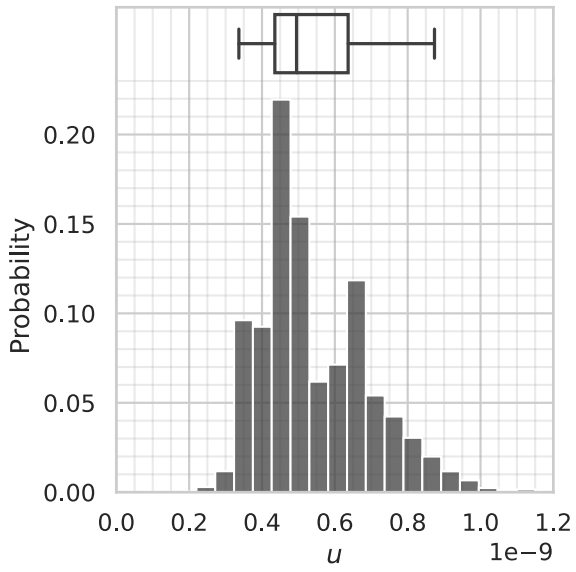
[1] C Paterson, I Bozic, MJ Smith, X Hoad, DGR Evans, https://doi.org/10.1101/2021.10.03.457528

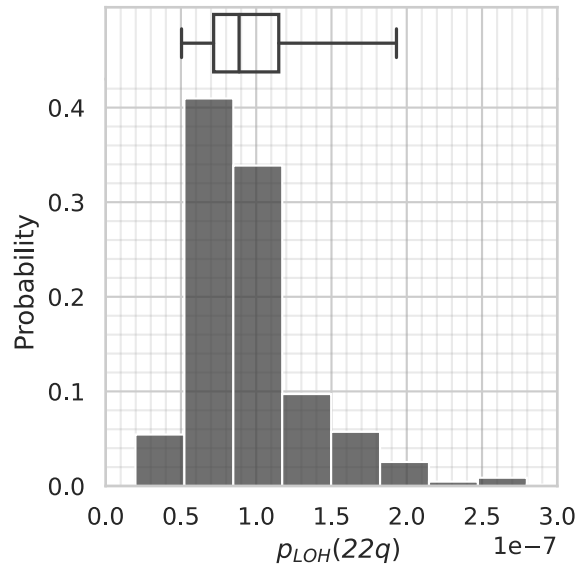
Vestibular schwannoma incidence

New parameter estimates





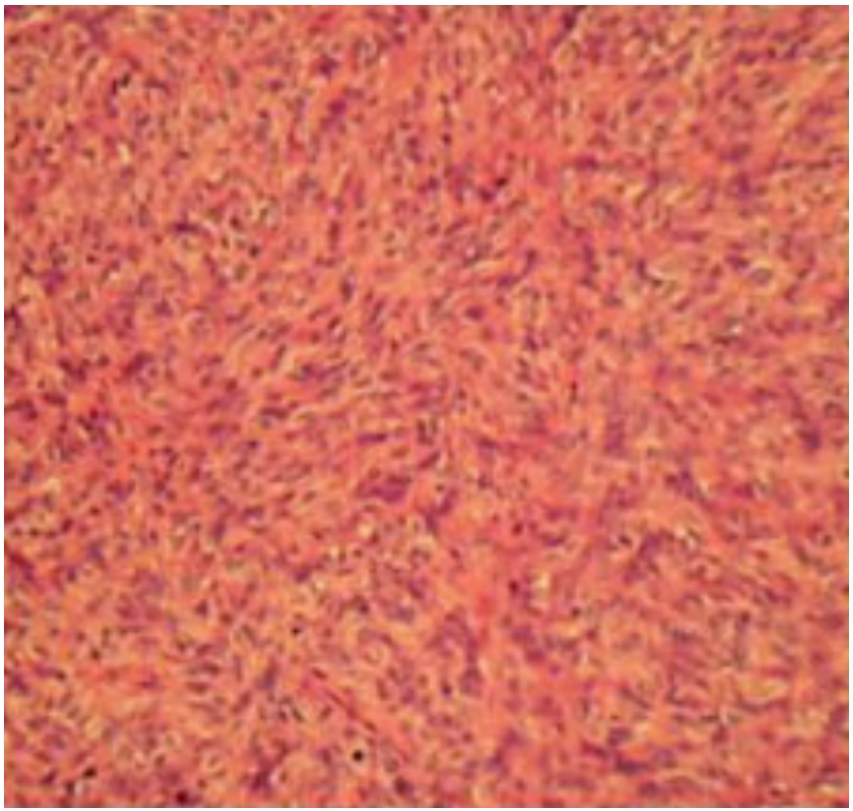




Malignant transformation in vestibular schwannoma

Very rare, very bad

- Risk $\approx 0.1\%$ of VS cases
- 5-year survival $\approx 12 - 20\%$

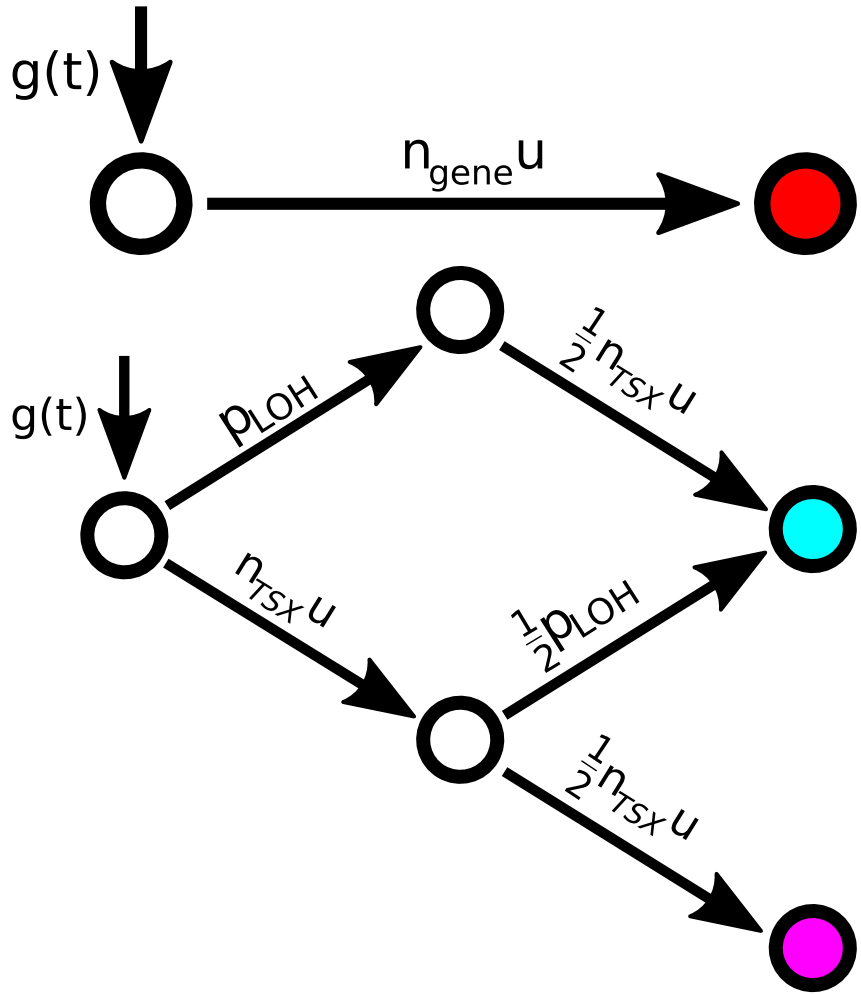


[1] AK Demetriades et al. Skull Base (2010)20:381–387.

Malignant schwannoma: two models

Timing and identity of drivers

Oncogene activation:



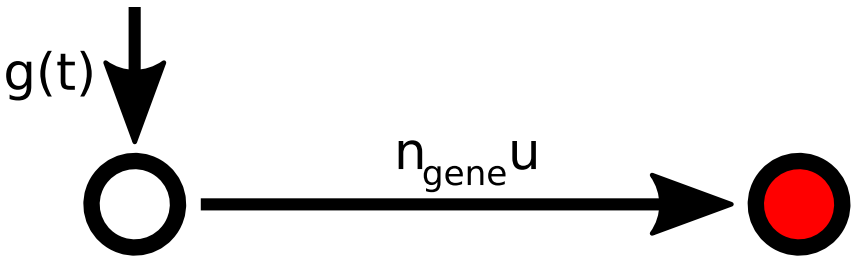
TSX inactivation:

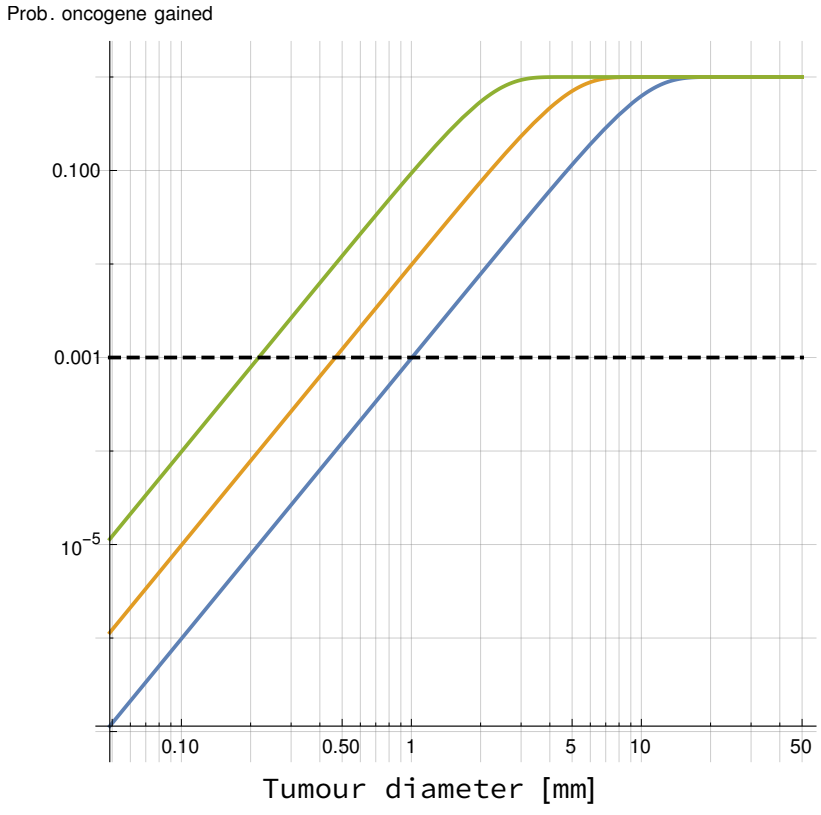
Malignant VS is *extremely* rare!

[1] C Paterson, I Bozic, MJ Smith, X Hoad, DGR Evans, https://doi.org/10.1101/2021.10.03.457528

Malignant schwannoma: first model

Oncogene activation

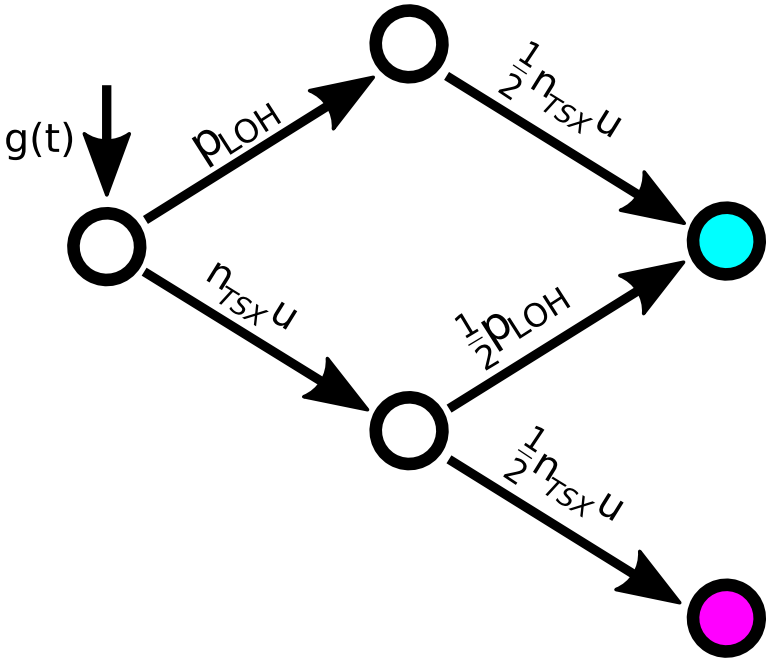


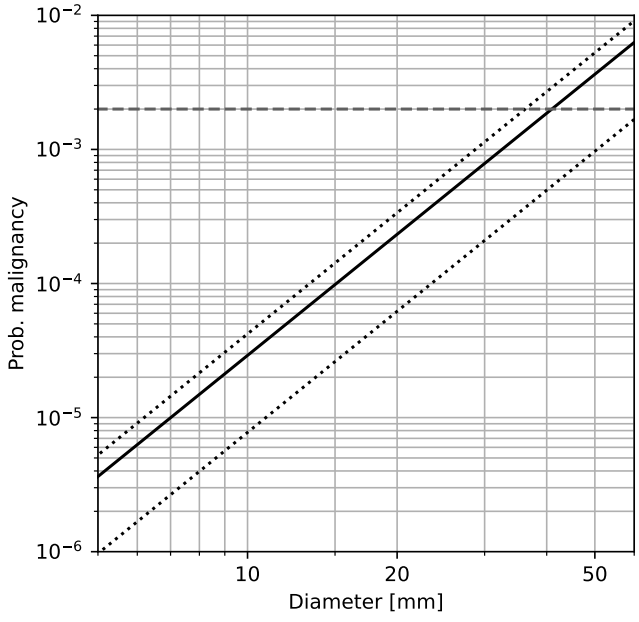


- Oncogene activation $\Longrightarrow$ high risk
- But it's a rare outcome
- So it's probably not caused by oncogene activation

Malignant schwannoma: second model

Tumour suppressor *TSX* inactivation

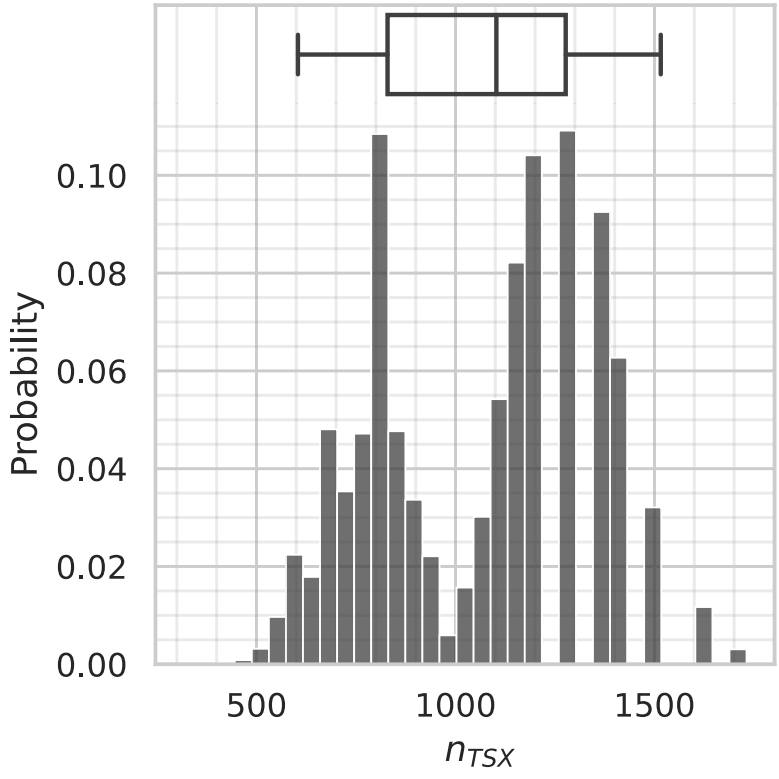




- *TSX* inactivation $\Longrightarrow$ low risk
- Can also estimate n_{TSX} that's consistent with incidence

Who is *TSX*?

Parameter estimates for n_{TSX}



Probably multiple (10?) distinct tumour suppressors

i.e. not (just) *TP53*: $n_{TP53} = 73$

Main outputs

1. Better estimates of event rates in Schwann cells
2. Can constrain timing of "*TSX*" (resp. for malignancy)
3. Can constrain size of *GFX* and *TSX*

Main outputs

but...

1. Uncertainties still large
2. Identity of *TSX* unknown
3. Constraints weak: *GFX* and *TSX* probably multiple genes

Multiple genes

Recent progress: new algorithm!

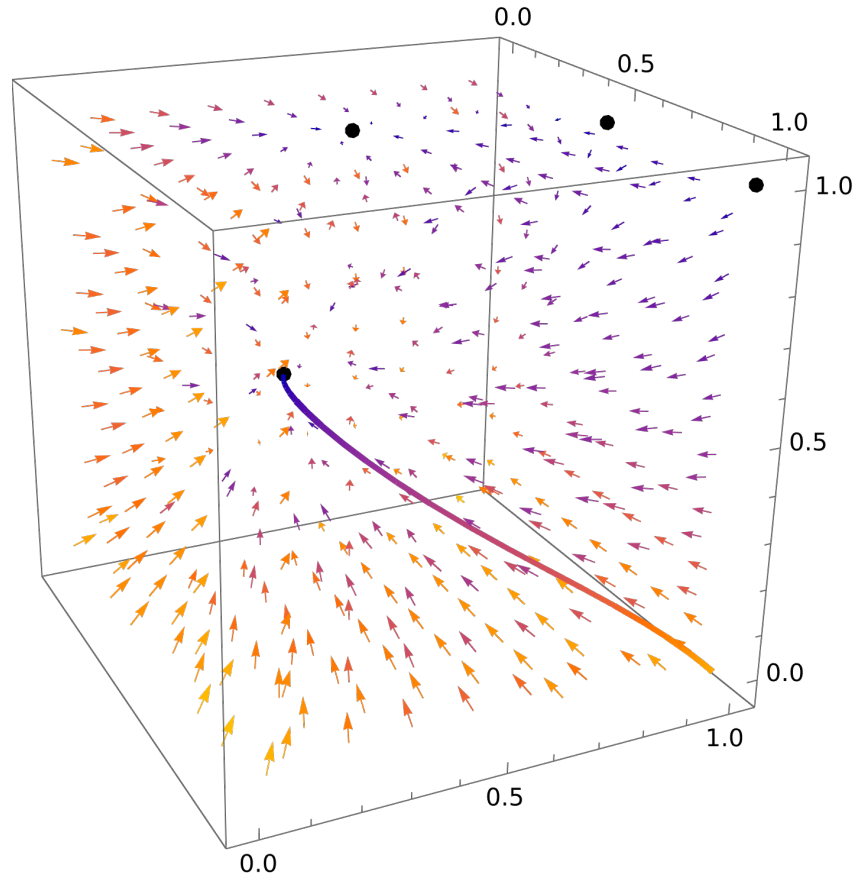
Laplace transform turns master equation in $\vec{N}$ into a wave equation in $\vec{q}$:

$$\frac{dP(t,\vec{N})}{dt} = \sum_{j,k} \mu_{jk} N_j P(t,\vec{N}') + \cdots$$

$$\downarrow$$

$$\frac{\partial \Psi}{\partial t} = \vec{X} \cdot \nabla \Psi$$

Much easier to solve: just numerical integration



New algorithm

Compare to sampling... Stochastic algorithms take

$$N \sim \mathcal{O}(\epsilon^{-2}) \tag{1}$$

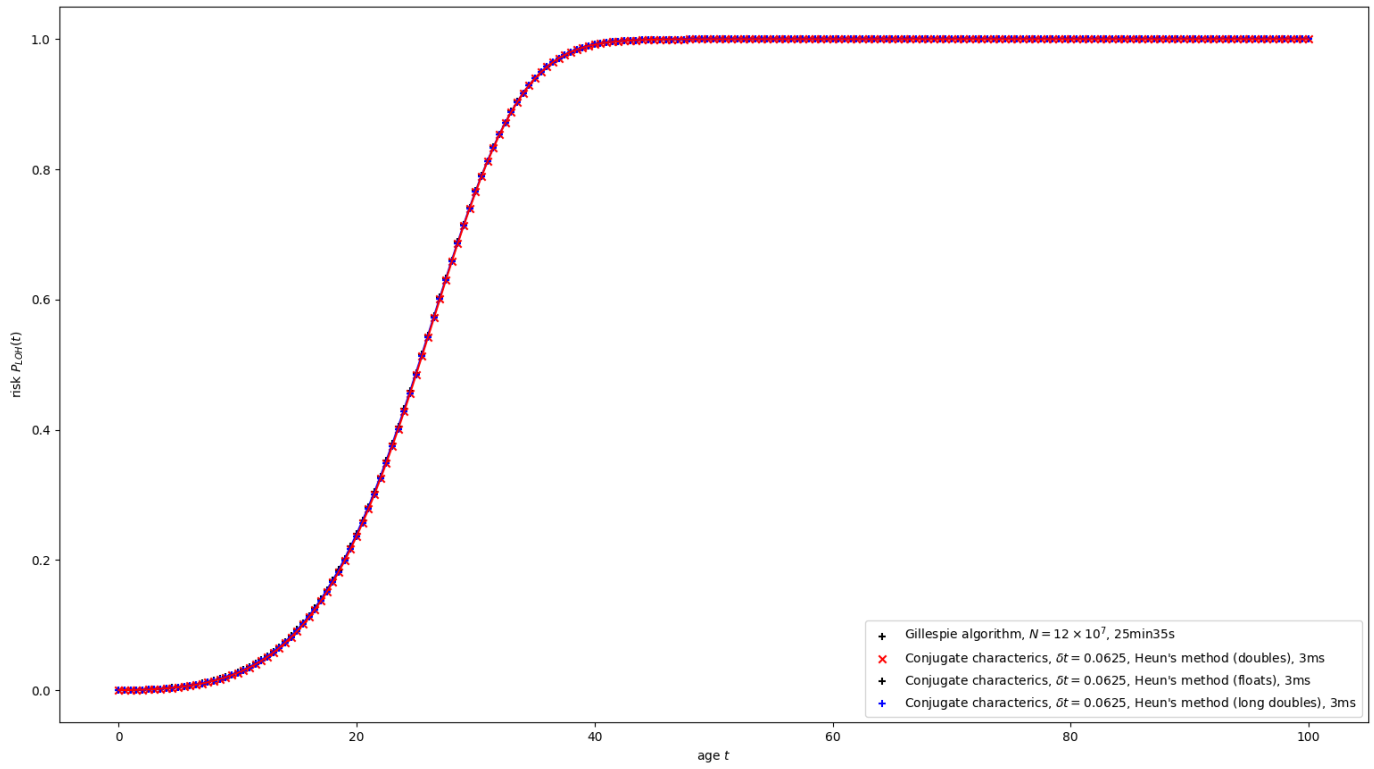
runs, and runtime $T \propto N$.

New algorithm runs in

$$T \sim \mathcal{O}(\epsilon^{-1/3}) \tag{2}$$

so new runtime $\sim \mathcal{O}\left(\text{ old runtime }^{1/6}\right)$.

New algorithm



To do list

Experiments:

- Sporadic VS to constrain $f_{SMARCB1}$: $n > 300$
- Empirical CNA/NGS in MPNST (**rare!**)?: $n > 30$

Theory:

- Convert n_{GFX} and n_{TSX} to estimates for multiple genes using *integer partitions*
- Implement new efficient algorithm in parameter inference with max. likelihood

Acknowledgements

The University of Washington

In order of appearance...

- Ivana Božić
- Hans Clevers
- Gareth Evans
- Xanthe Hoad
- Miriam Smith
- David Wedge