

Cancer, copy number alterations, and age:

Fast simulations of clonal expansion on networks

Chay Paterson

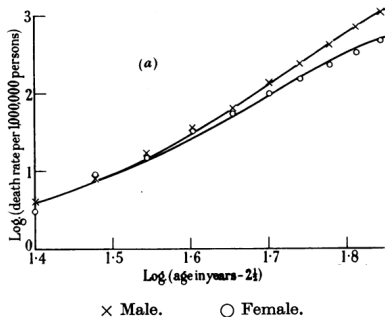
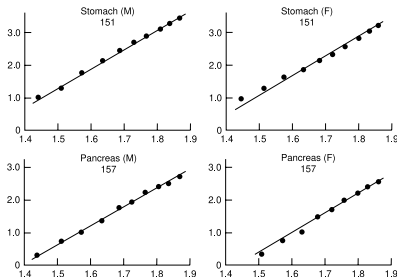
University of Manchester

18 October 2022

Age and cancer

P. Armitage and R. Doll¹²

Risk of cancer increases with age:



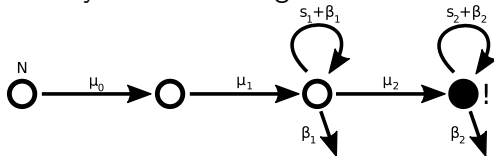
¹P. Armitage and R. Doll, British Journal of Cancer 1954; 8: 1-12

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

Age and cancer

P. Armitage and R. Doll¹²

Why? Accumulating mutations¹²³:



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

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

³F. Michor, Y. Iwasa and M.A. Nowak, Nature Reviews Cancer 2004; 4: 197-205 doi:10.1038/nrc1295

Multi-stage clonal expansion models

2-3 rate limiting steps¹²³

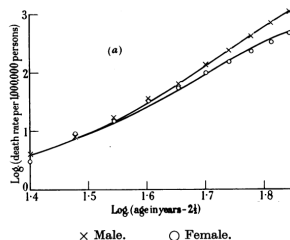
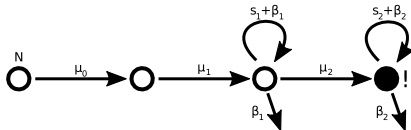
Problem: how to compute $P(\text{cancer}, t)$ for a given model?

Different methods: Fast:

- ▶ Mean-field approximation¹
- ▶ Numerical quadrature²

Slow:

- ▶ Gillespie algorithm + sampling³
- ▶ tau leaping + sampling³



↓?

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

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

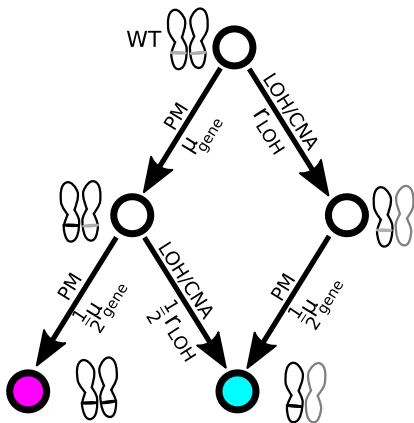
³C. Paterson, I. Bozic, H. Clevers, PNAS 2020; 117(34): 20681-20688

(supp. material)

Network models

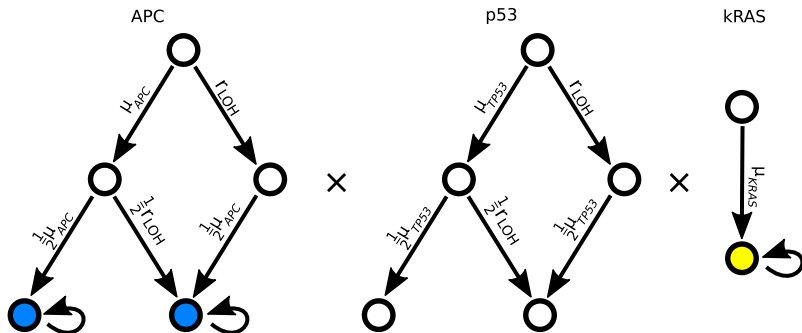
example model

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



This gets us the incidence of *specific karyotypes*

Colorectal adenocarcinoma model



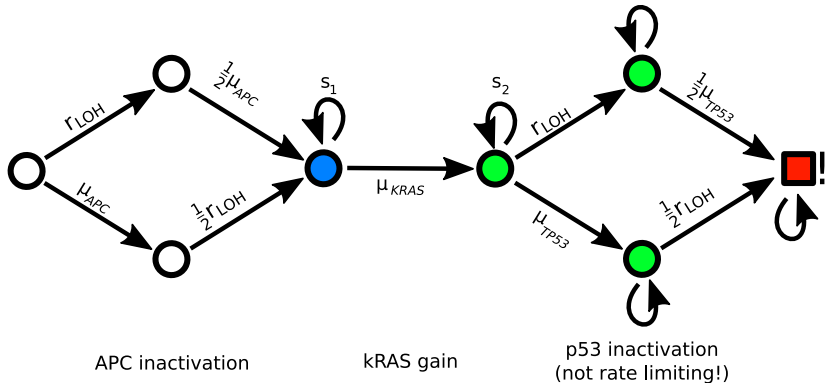
- ▶ *APC-p53-kRAS* combo accounts for about 15% of incidence
- ▶ 5-year survival about 60% (any stage)

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



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

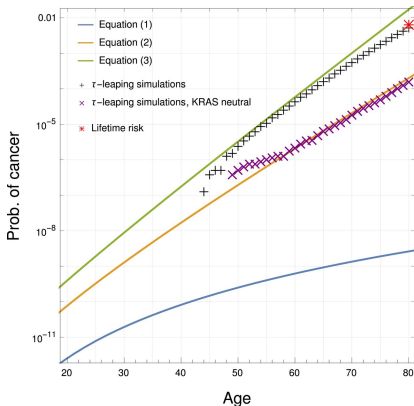
- ▶ The 4 most likely paths account for 50% of the risk
- ▶ Consistent with classic model²

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

²Kinzler and Vogelstein 1990

Colorectal adenocarcinoma model

Successful *ab initio* model



- ▶ Can constrain *APC/KRAS* epistasis ($s_2 < 0.31/\text{yr}$)
- ▶ 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*

¹C. 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

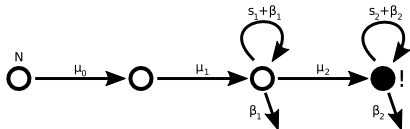
1. genomic subtypes better characterised: *NF2/Merlin* altered in 85 – 100% of cases^{1,2}, *TP53* in $\approx 0\%$
2. usually benign = **easier to study timing!** (of drivers, malignant transformation...)
3. only 3 hits, weak selection (almost neutral)
4. probabilities low: approximations v. accurate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

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

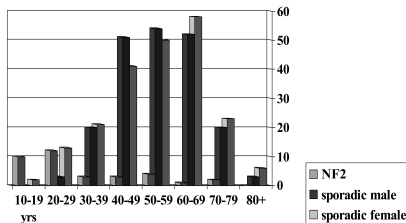
²AL Håvik et al., *Journal of Neurosurgery JNS*. 2017;128(3):911 – 922.

Vestibular schwannoma

3-event model



- ▶ Fitness suspiciously low, $s \approx 0.005/\text{yr}$ ¹
- ▶ Suggests nearly-neutral 3-hit model³



Gareth Evans 2005²

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

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

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

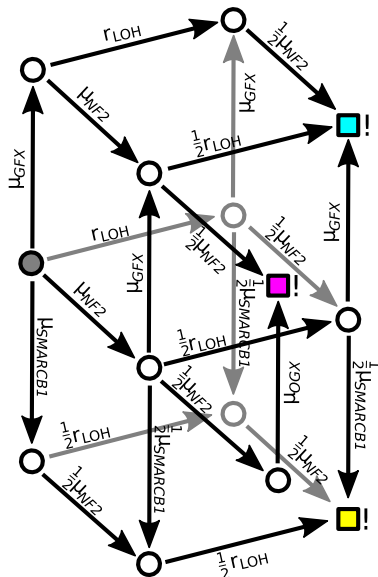
Vestibular schwannoma incidence

Our model for sporadic VS

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

Risk of each subtype looks like

$$P(\text{cyan square}) \propto \frac{t^3}{3!}$$



Vestibular schwannoma incidence

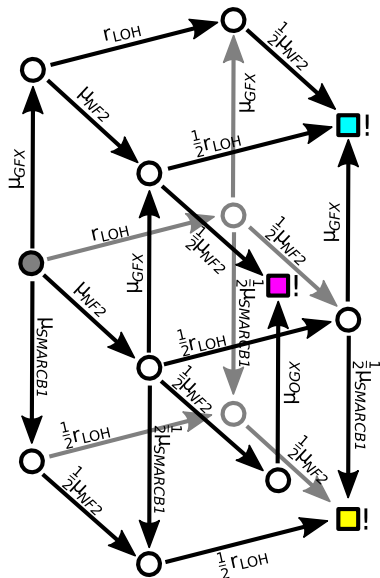
Our model for sporadic VS

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

Risk of each subtype looks like

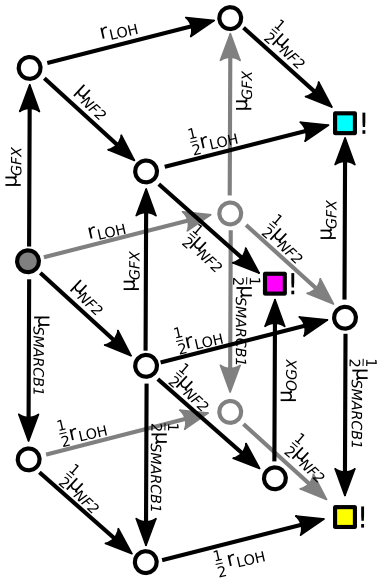
$$P(\text{cyan}) \propto \frac{t^3}{3!}$$

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

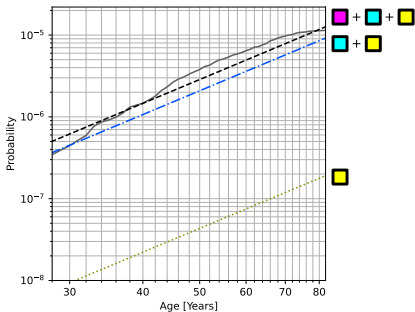


Vestibular schwannoma incidence

Our model for sporadic VS

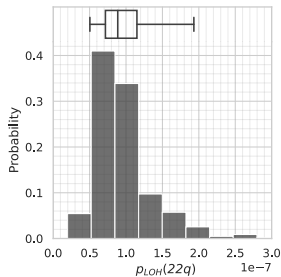
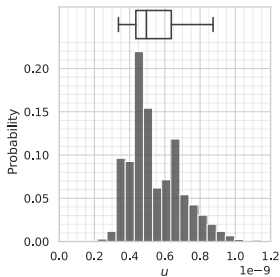
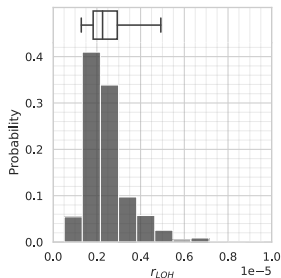
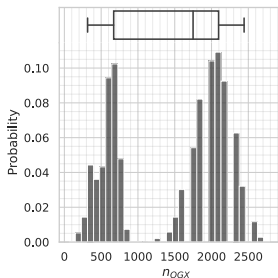


Cumulative incidence of vestibular schwannoma



Vestibular schwannoma incidence

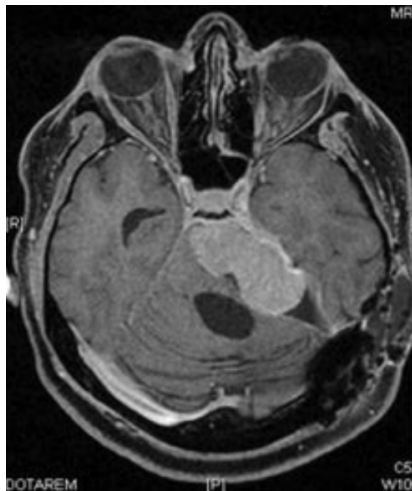
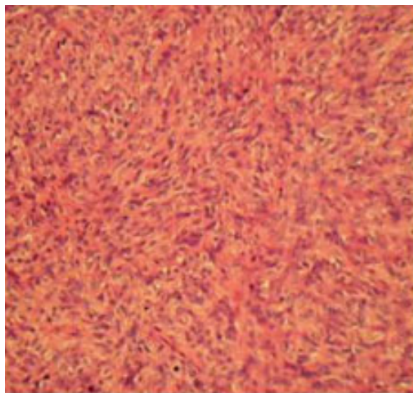
New parameter estimates



Malignant transformation in vestibular schwannoma

Very rare, very bad

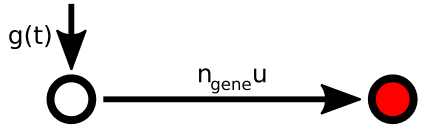
- ▶ Risk $\approx 0.1\%$ of VS cases
- ▶ 5-year survival $\approx 12 - 20\%$



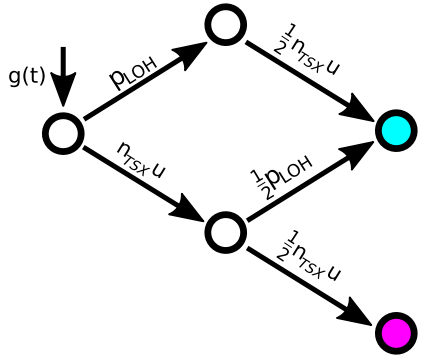
Malignant schwannoma: two models

Timing and identity of drivers

Oncogene activation:



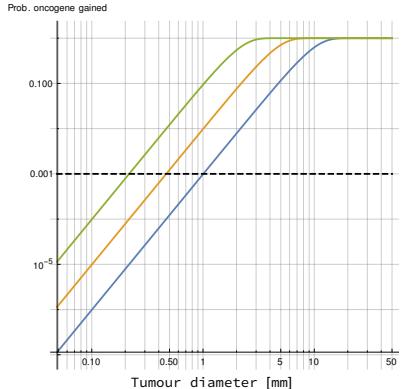
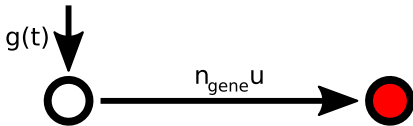
TSX inactivation:



Malignant VS is *extremely* rare!

Malignant schwannoma: first model

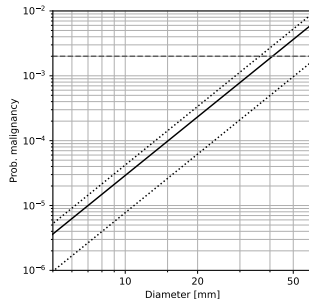
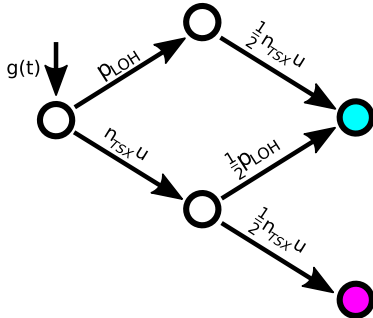
Oncogene activation



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

Malignant schwannoma: second model

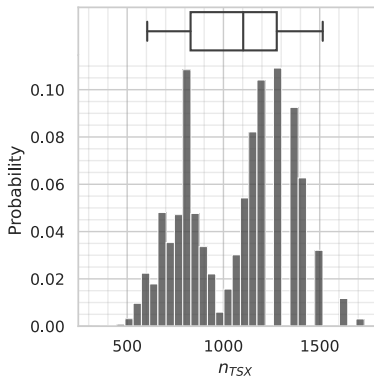
Tumour suppressor *TSX* inactivation



- ▶ *TSX* inactivation \Rightarrow 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!

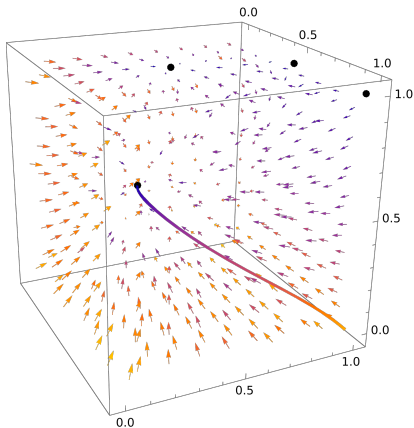
Laplace transform turns master equation in \vec{N} into a wave equation in \vec{q} :

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

↓

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

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

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