Network models of carcinogenesis and vestibular schwannoma

Chay Paterson

University of Manchester*

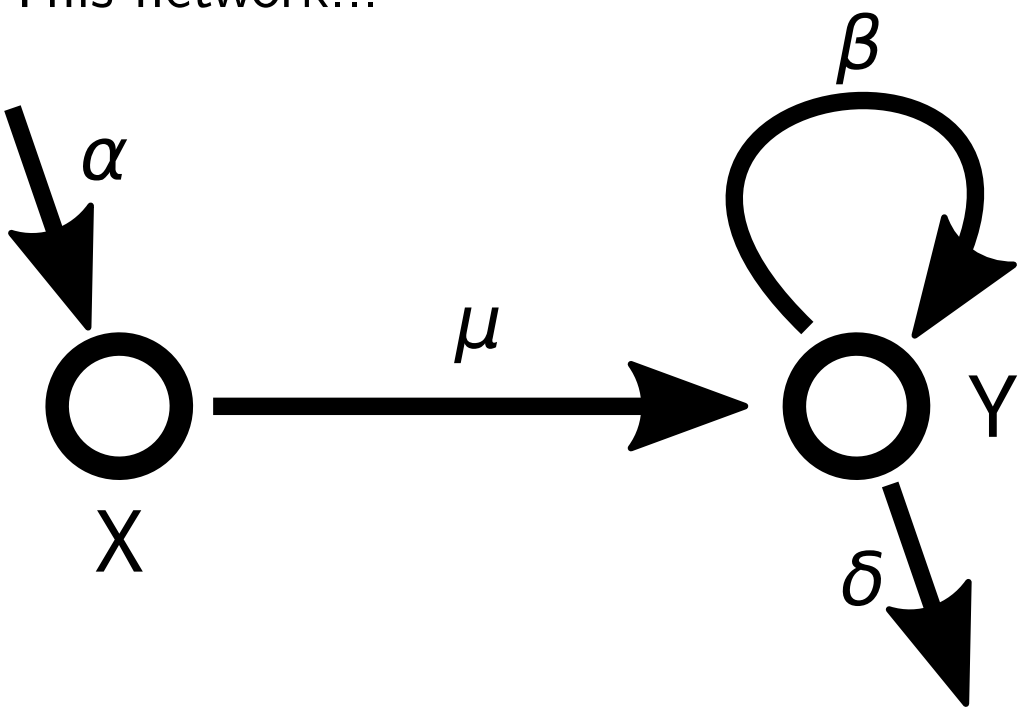
28 March 2022

*(sort of)

Warning

There will be very few equations in this talk!

This network...



corresponds to this stochastic process:...

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

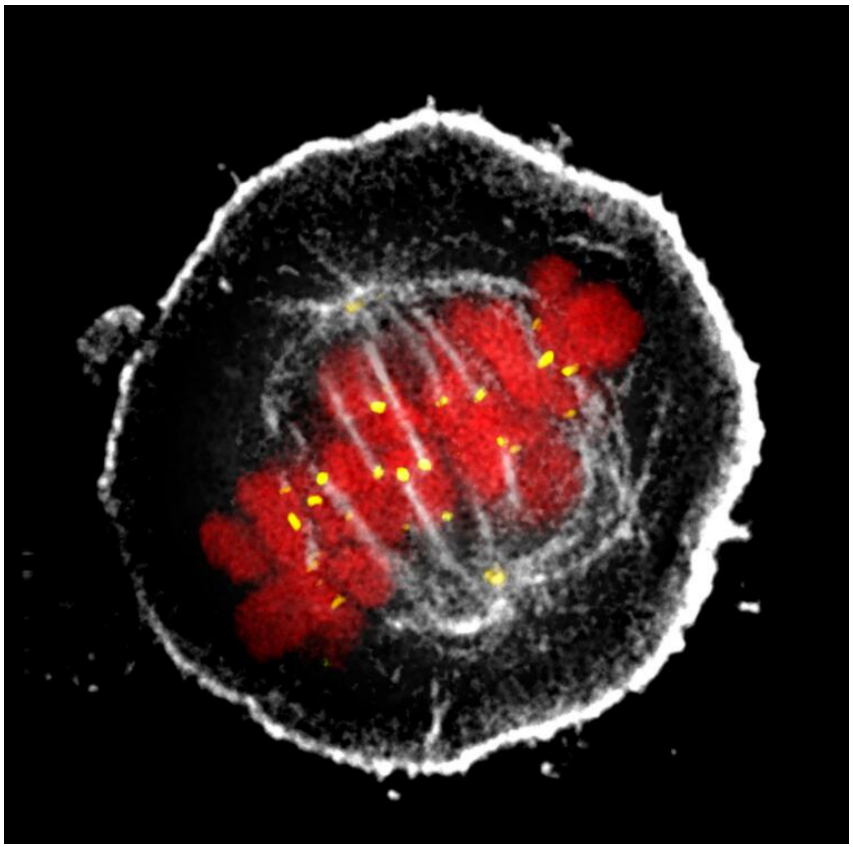
and approximately this linear system...

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

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

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

What is cancer?



- Various risk factors
- Mutations accumulate
- Loss of control of cell growth & death

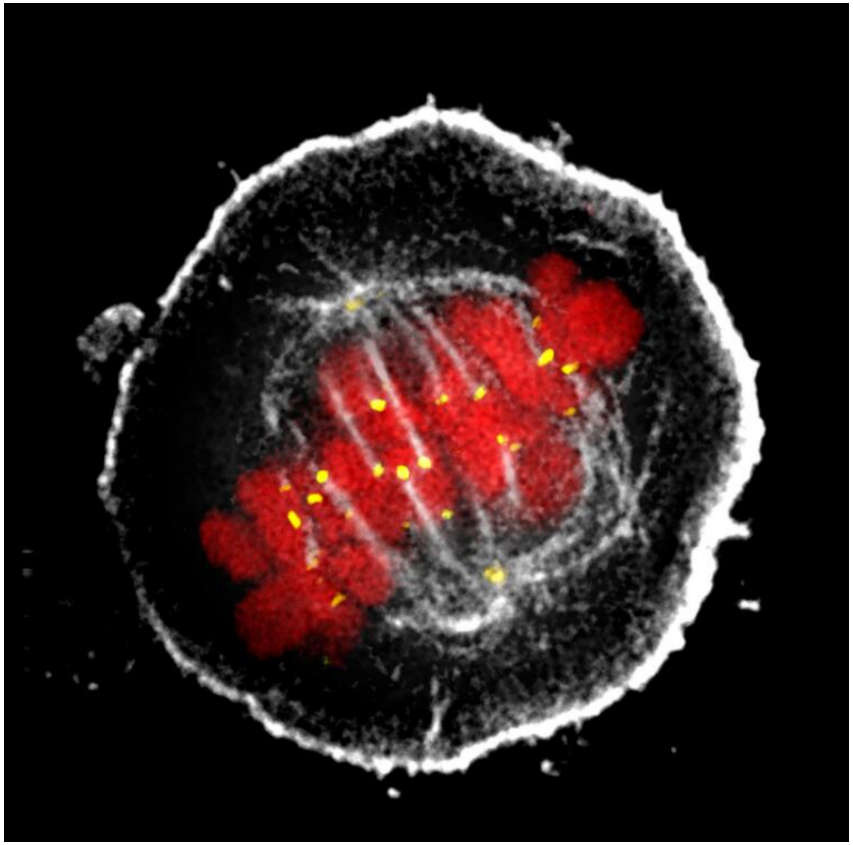
$\downarrow$ gets us a benign tumour/neoplasia...

$\downarrow$ something happens...

- Dedifferentiation, migration...
- Malignancy = cancer

[1] Figure: HeLa cell https://wellcomecollection.org/works/vpgx8zcd

What is cancer?



- Various risk factors
- Mutations accumulate
- Loss of control of cell growth & death

  ↓ gets us a benign tumour/neoplasia...

  ↓ something happens...

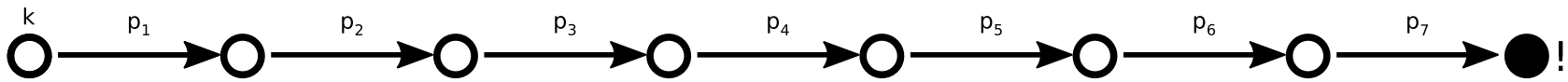
- Dedifferentiation, migration...
- Malignancy = cancer

Why do we believe this?

[1] Figure: HeLa cell https://wellcomecollection.org/works/vpgx8zcd

Multi-stage models

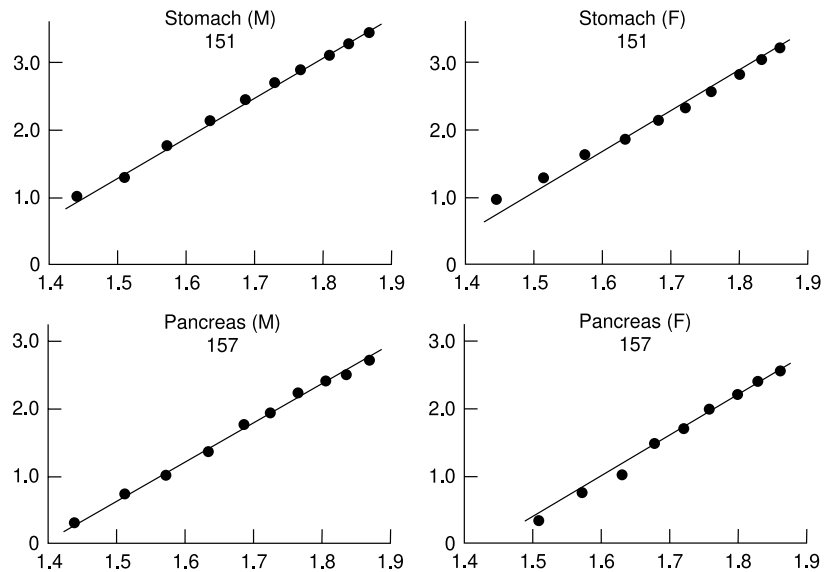
P. Armitage and R. Doll[2]



$$\downarrow$$

$$P(\text{cancer}) \approx k p_1 p_2 p_3 p_4 p_5 p_6 p_7 \frac{t^7}{7!}$$

$$\Longrightarrow \text{ incidence} \propto \frac{t^6}{6!}$$

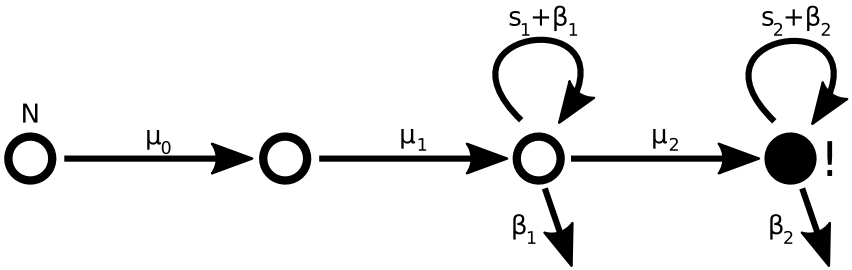


[1] P. Armitage and R. Doll, British Journal of Cancer 1954; 8: 1–12

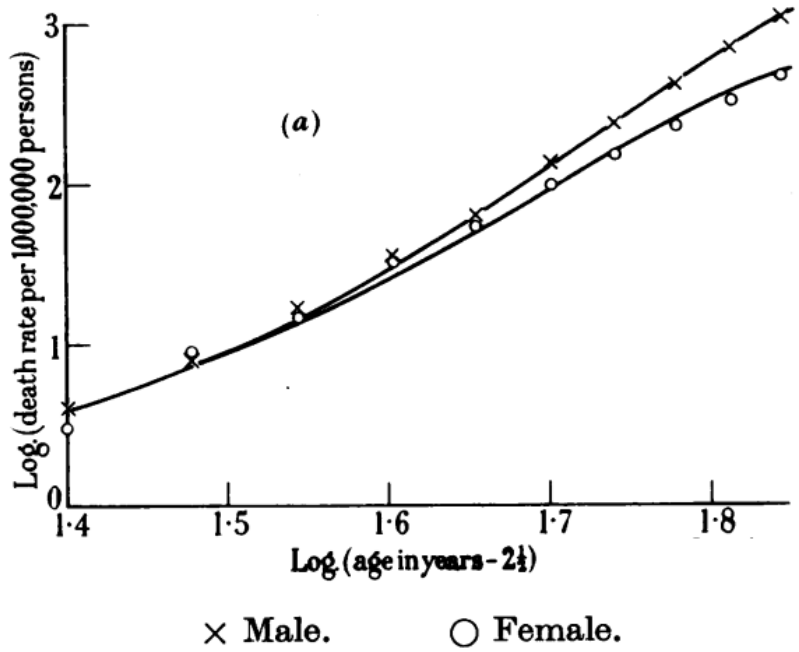
[2] note that $P(t) = 1 - S(t)$, other authors (e.g. Knudson)

Multi-stage models

2-3 rate limiting steps[1][2][3]



- Add selection s_k
- Better fit, more plausible μ_k



[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

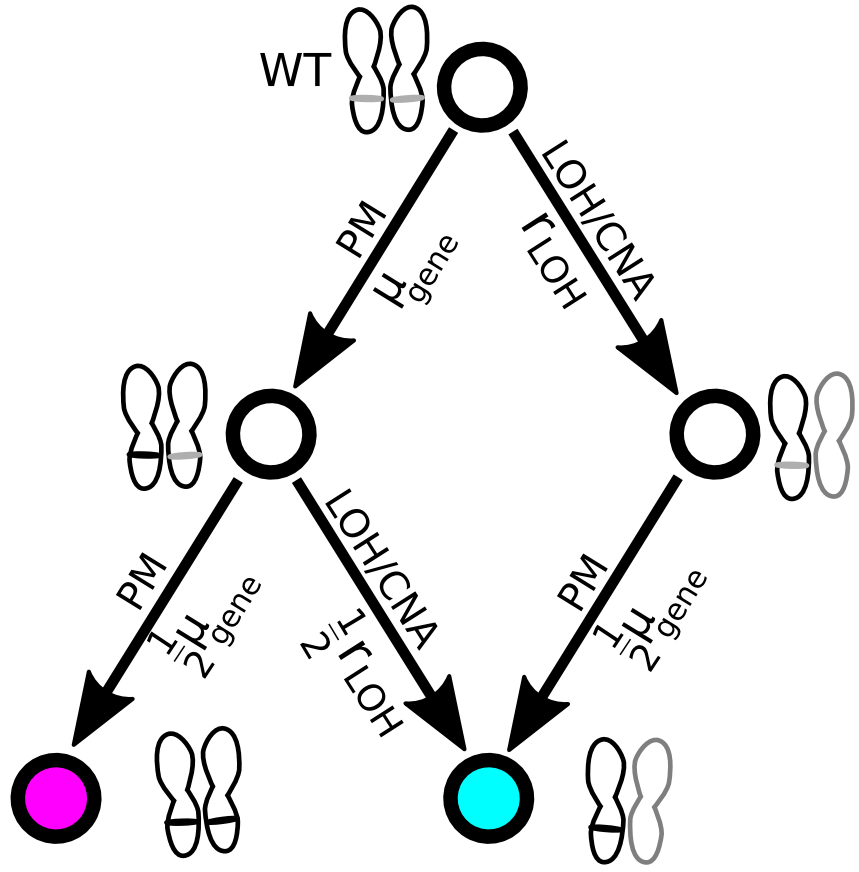
[3] F. Michor, Y. Iwasa and MA. Nowak, Nature Reviews Cancer 2004; 4: 197-205 doi:10.1038/nrc1295

Network models

1. Study **specific genes** and mechanisms of interest
2. Fix parameters from sequences and experiments
3. Distinguish different orders of events

Predict copy number alterations (etc.)

example model

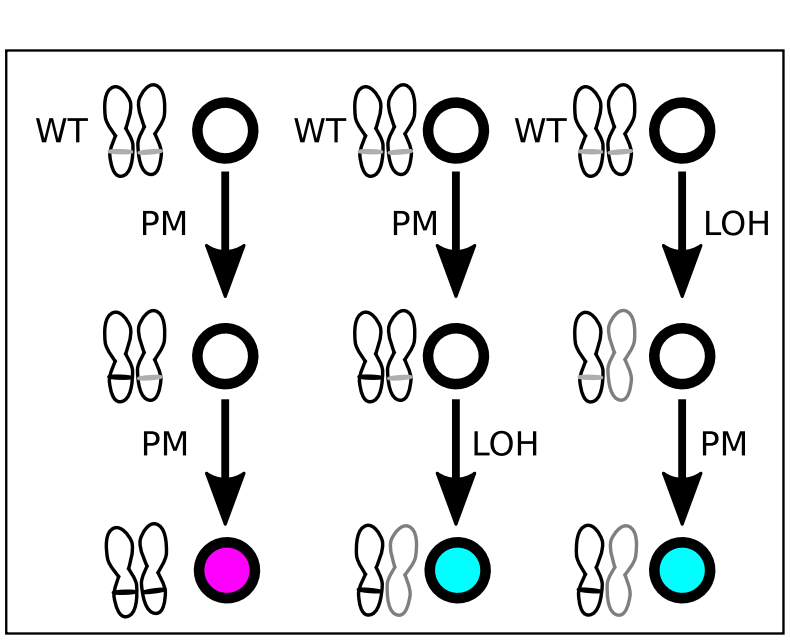


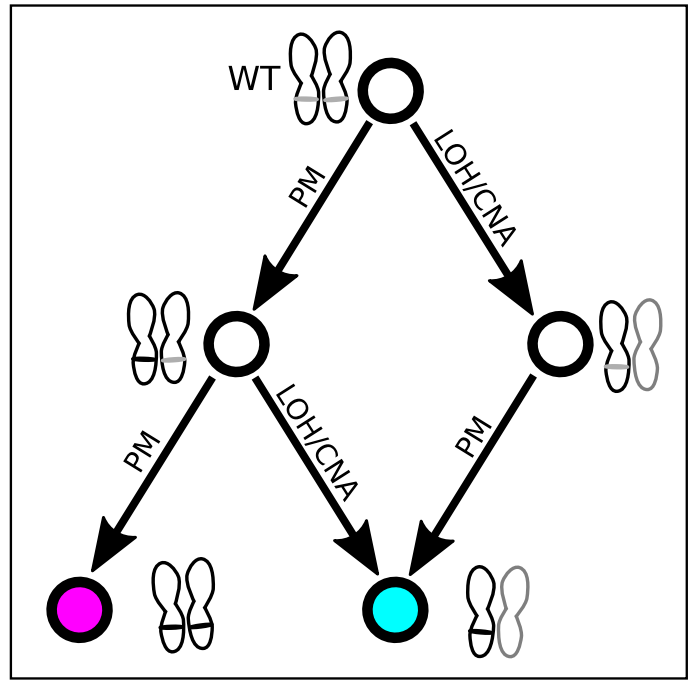
This gets us the incidence of *specific karyotypes*

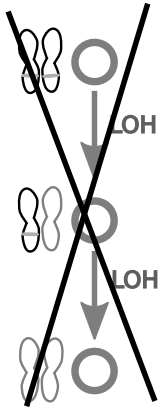
Network models

Worked example: tumour suppressor loss-of-function

Two events of interest, 3 possible orders:



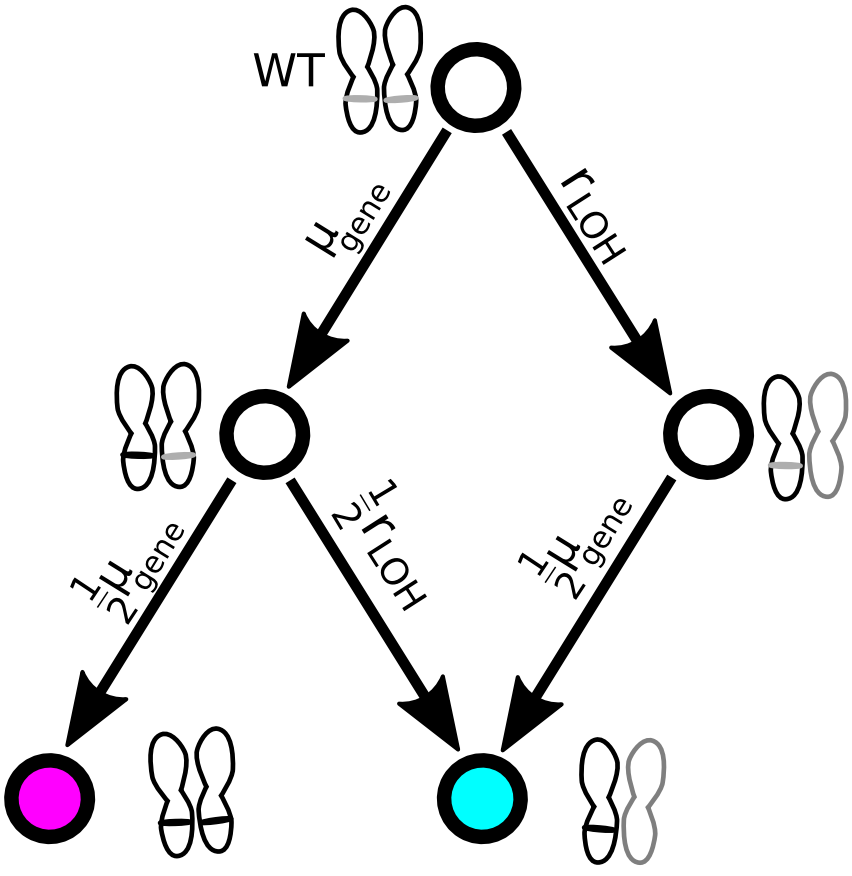




← this order of events is impossible
(resulting cells are non-viable)

Network models

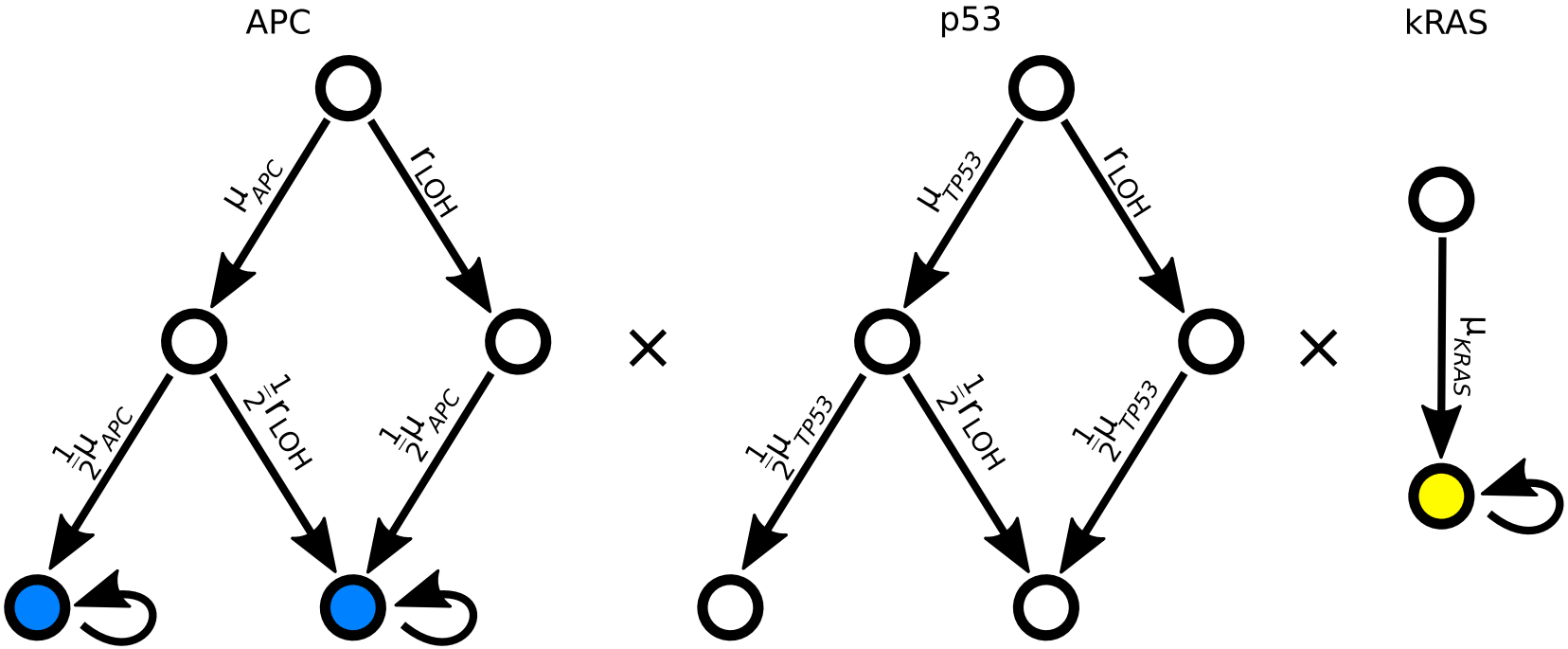
Worked example: tumour suppressor loss-of-function



Assign rates to events:

- Get μ_{gene} from $n_{gene} \times u \times b$
- To get n_{gene}, count possible nonsense codons in a known sequence
- Fix r_{LOH} by looking at relative frequency of LOH

Colorectal adenocarcinoma model



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

---

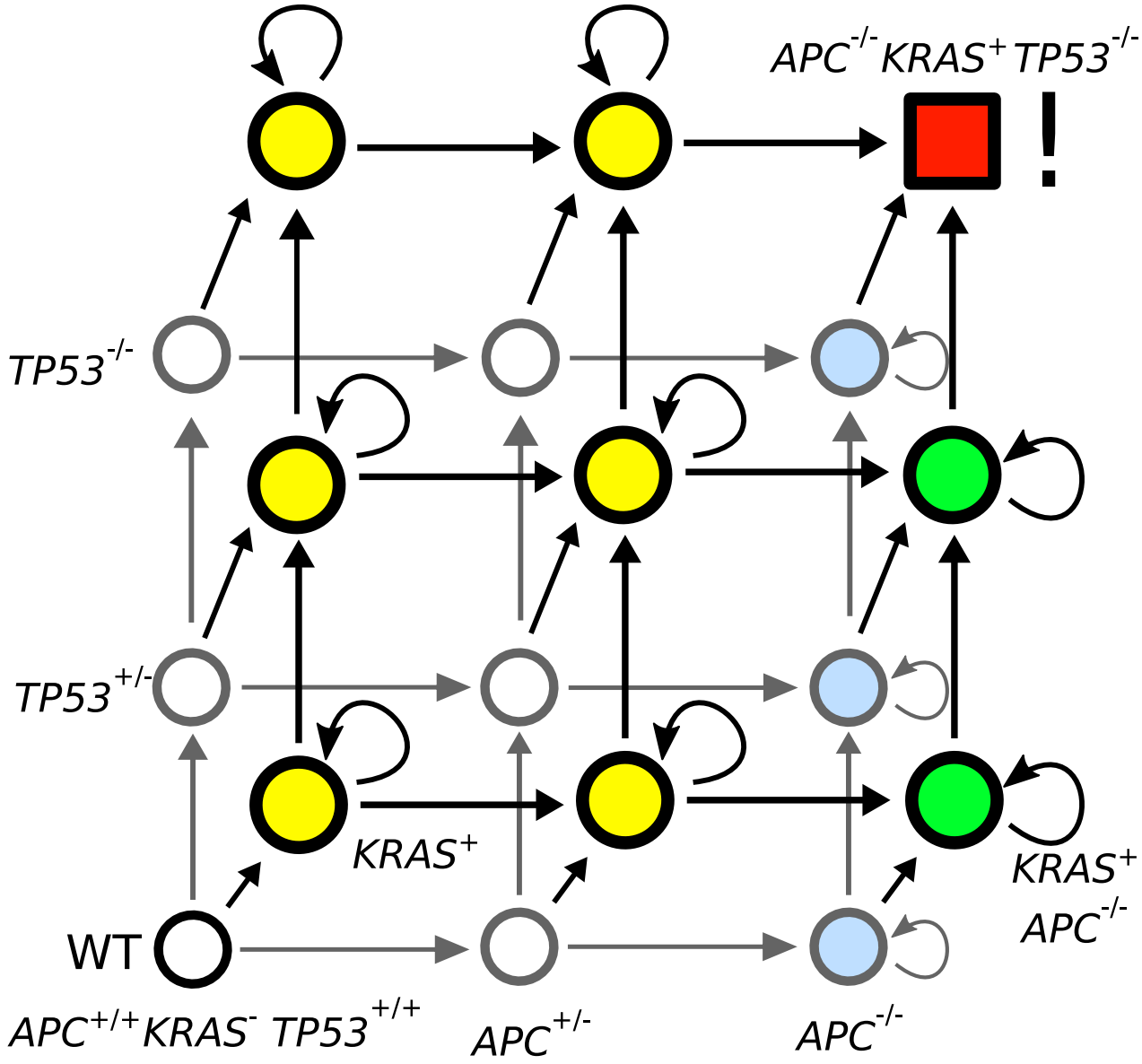
[1] Fearon et al. TODO

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

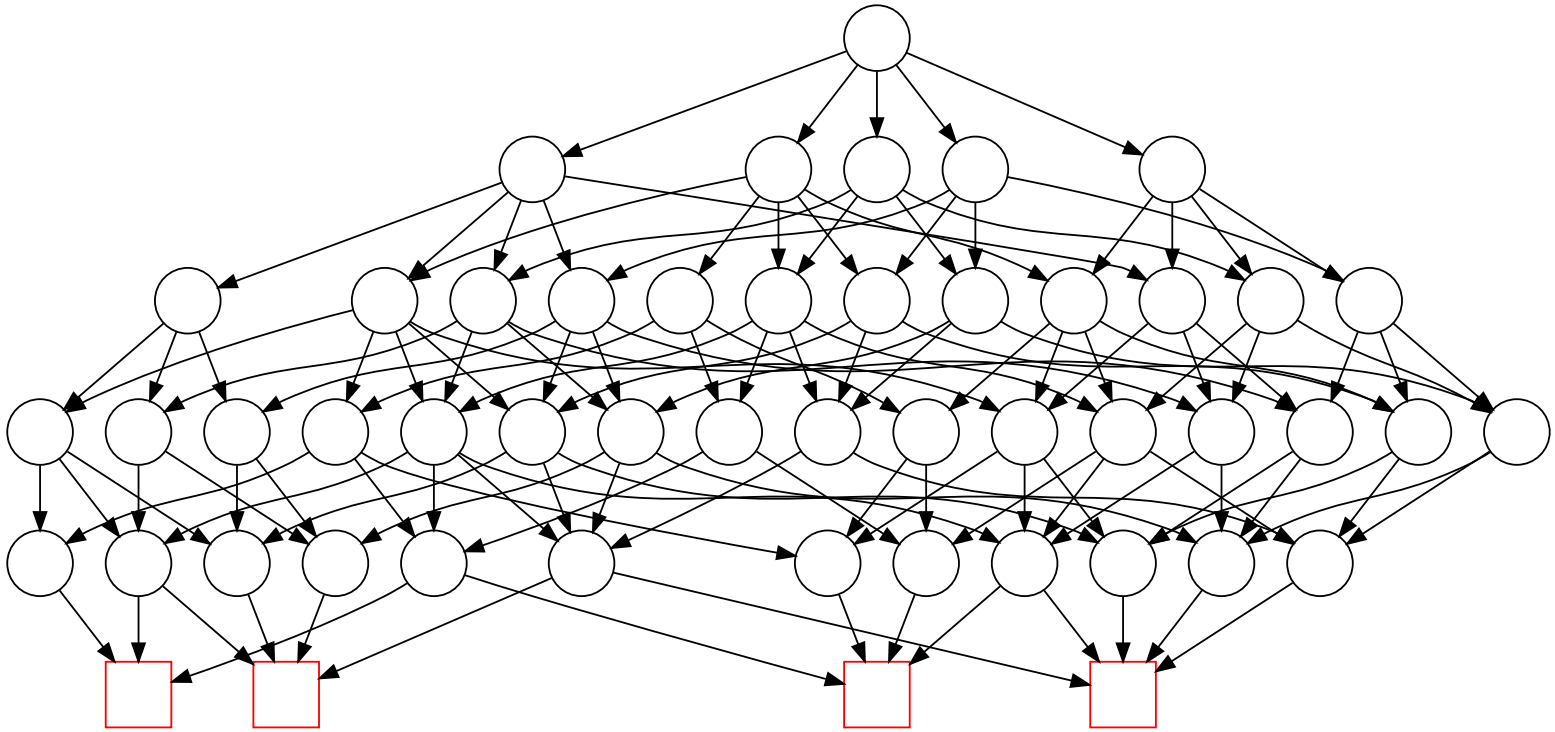
[2] Office for National Statistics, England 2019

Colorectal adenocarcinoma model

Simplified



Colorectal adenocarcinoma model

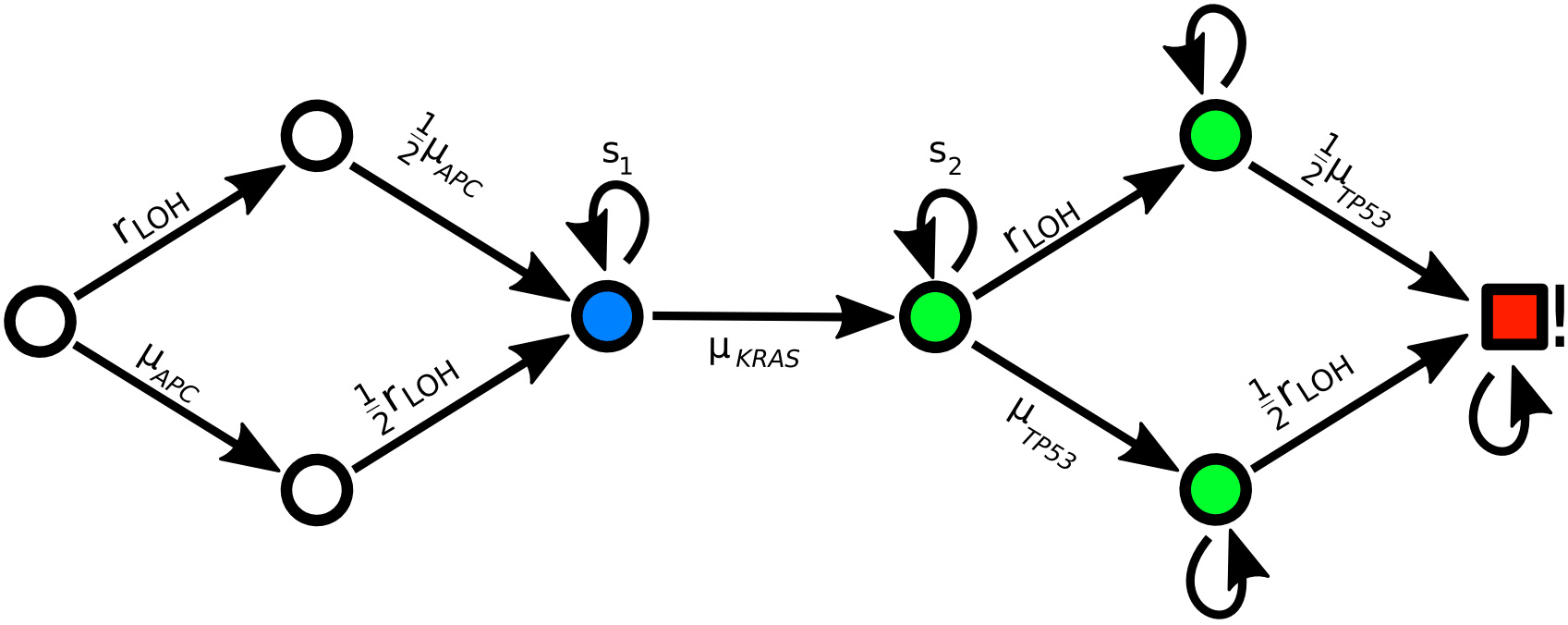


each end node is a different copy number profile

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

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

Colorectal adenocarcinoma model



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

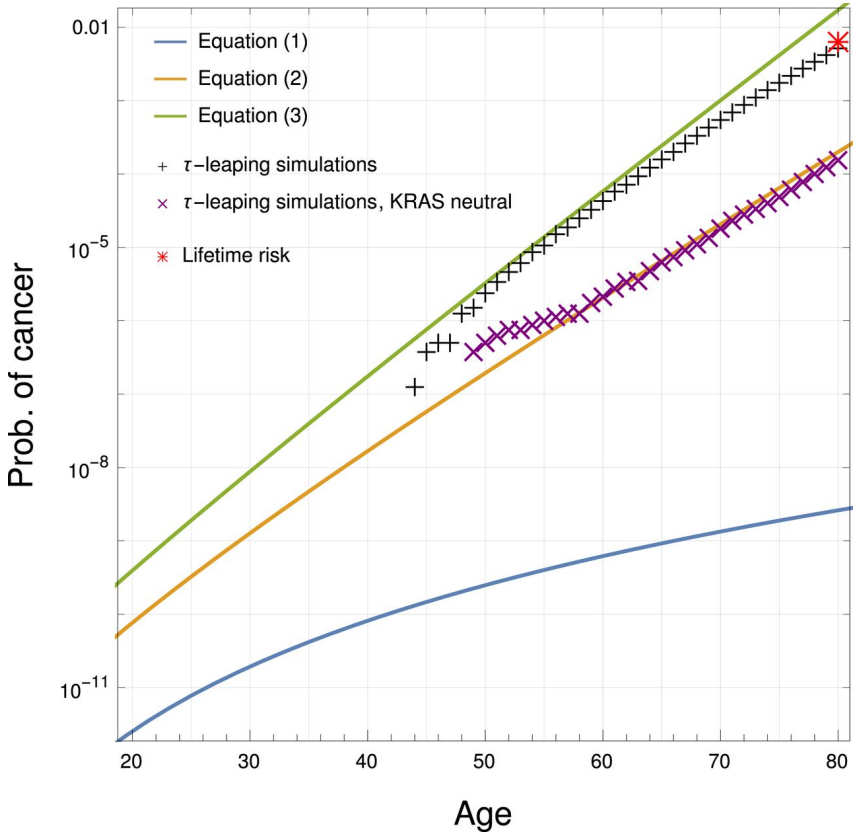
- The 4 most likely paths account for 50% of the risk
- Consistent with classic model[2]

[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



- Gets proportion of incidence right – *with no fitting!*
- Can constrain *APC/KRAS* epistasis ($s_2 < 0.31/\text{yr}$)
- Timing of *p53* inactivation: must be *late*[2]
- *Compatible with* 3-hit models, similar curve: *p53 not rate limiting*
- Conditional path probabilities $P(X_i)$ encode fitnesses

---

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

[2] (relative to the others)

Colorectal adenocarcinoma model

Science is very hard

1. model only accounts for 15% of lifetime risk[1,2]
2. often malignant on discovery: difficult to constrain timing[3] or effect of drivers
3. 5 hits, 50 nodes, 120 edges, 270 paths, strong selection: complicated
4. approximations inaccurate when probabilities large

---

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

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

[3] (relative to malignant transformation)

[1] https://commons.wikimedia.org/wiki/File:Sun_(Earth_POV).jpg

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**

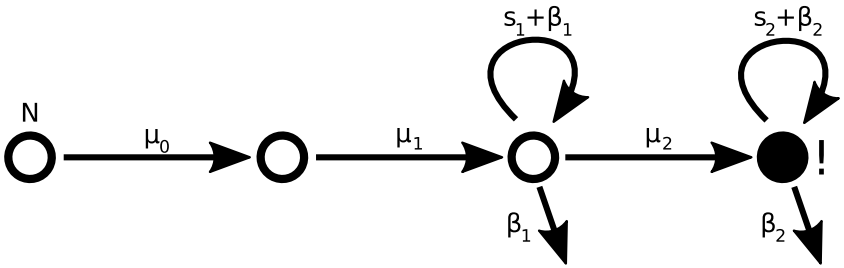
[1] ML Carlson et al., Otology & Neurotology: 2018;39(9):860 – 871

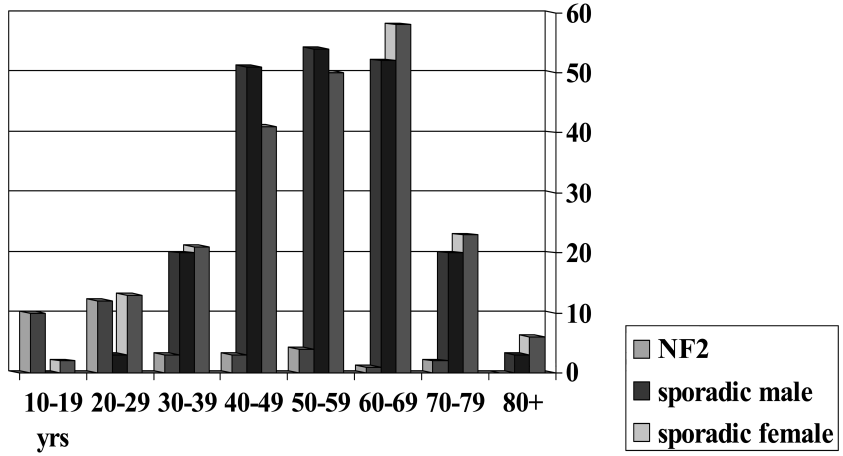
[2] AL Håvik et al., Journal of Neurosurgery JNS. 2017;128(3):911 – 922.

[1] https://apod.nasa.gov/apod/ap211209.html

Vestibular schwannoma

3-*event* model





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

---

[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

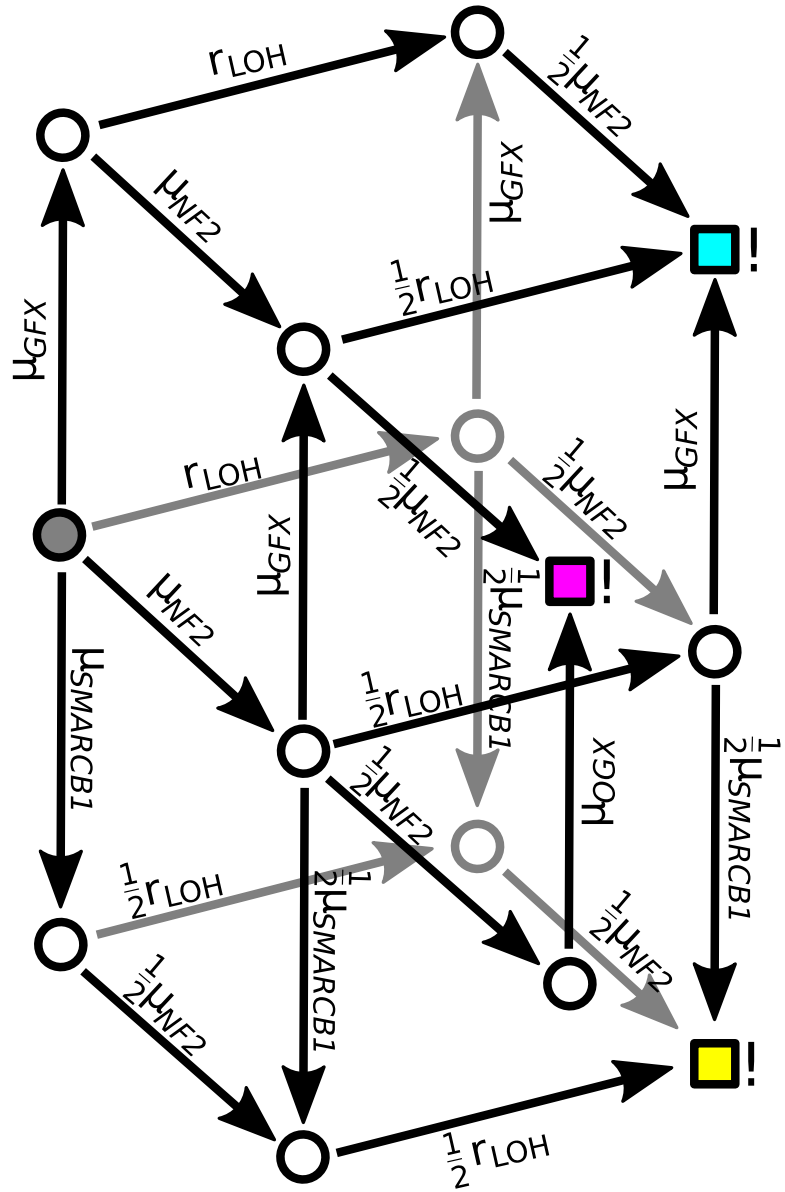
Vestibular schwannoma incidence

Our model for sporadic VS

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

Risk of each subtype looks like

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

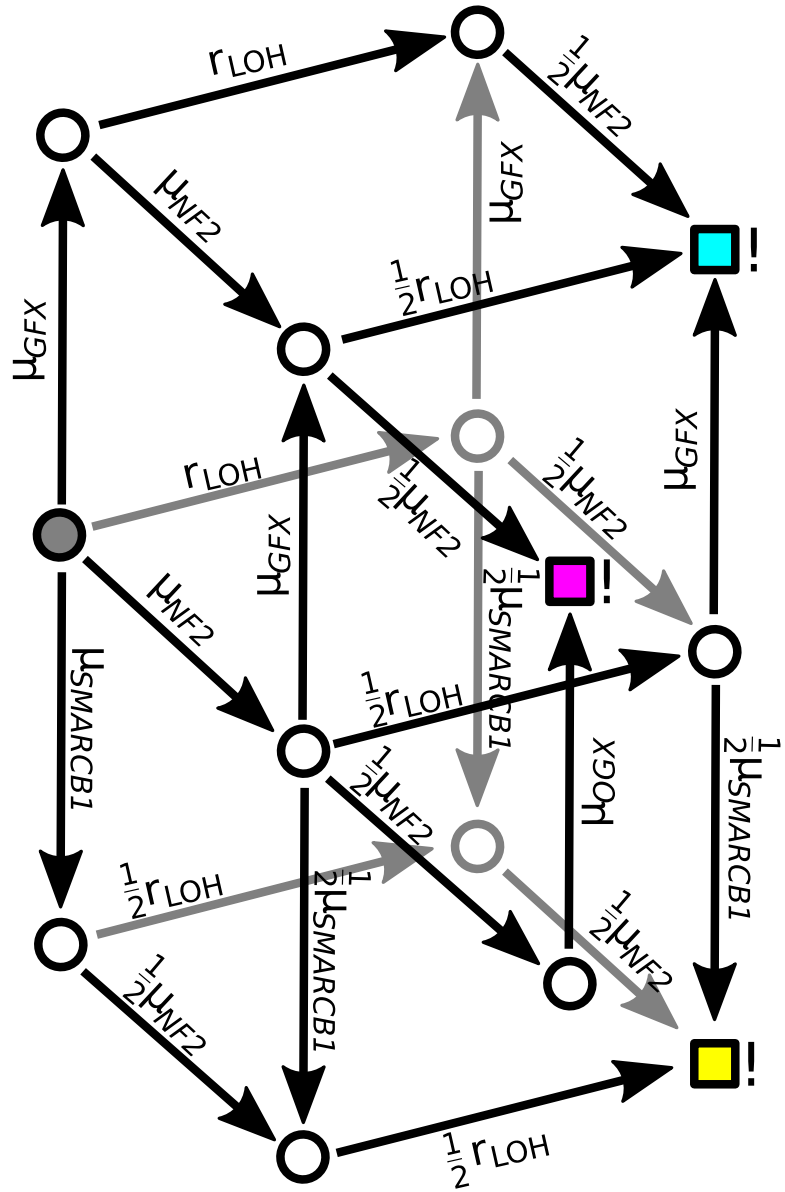


[1] C Paterson, I Bozic, MJ Smith, X Hoad, DGR Evans, https://doi.org/10.1101/2021.10.03.457528

Vestibular schwannoma incidence

Our model for sporadic VS

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



Risk of each subtype looks like

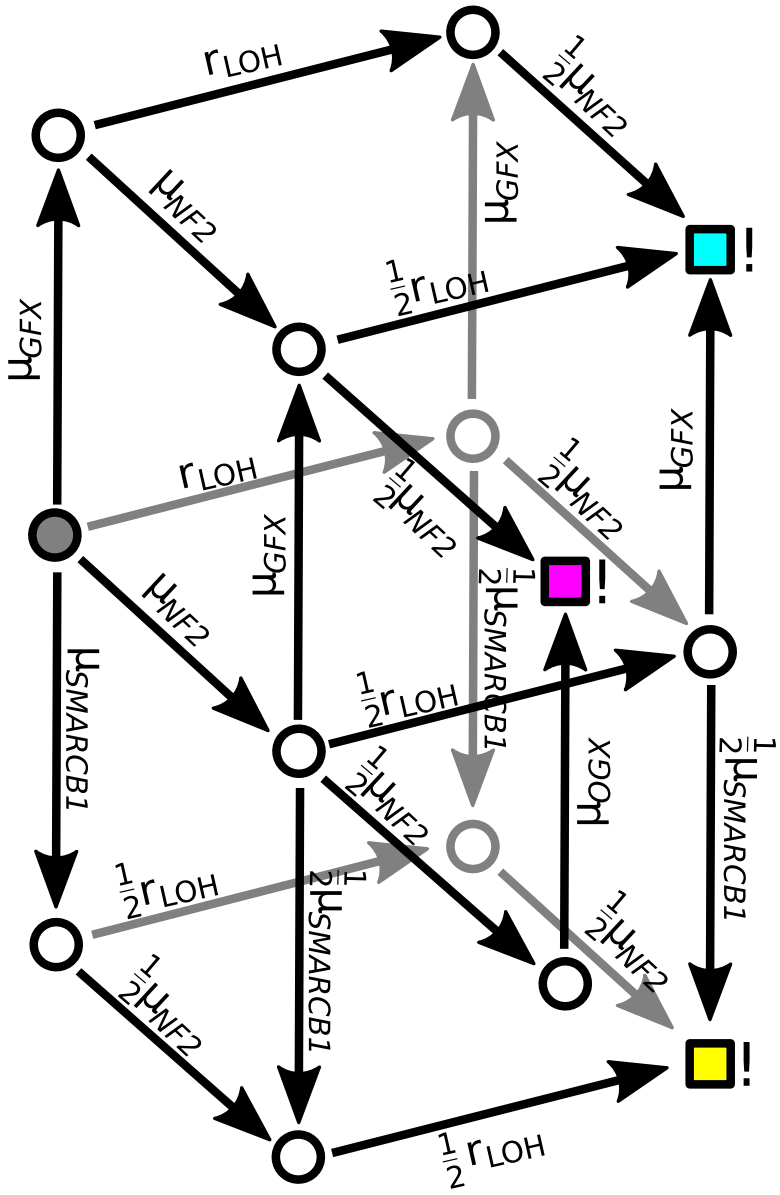
$$P(\blacksquare) \propto \frac{t^3}{3!}$$

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

[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



- ■ + ■ = have LOH on 22q
- frequency of LOH = $f_{LOH} = (■+■)/(■+■+■)$
- ■ = *SMARCB1*$^{-/-}$
- frequency of *SMARCB1*$^{-/-}$ = $f_{SMARCB1}$

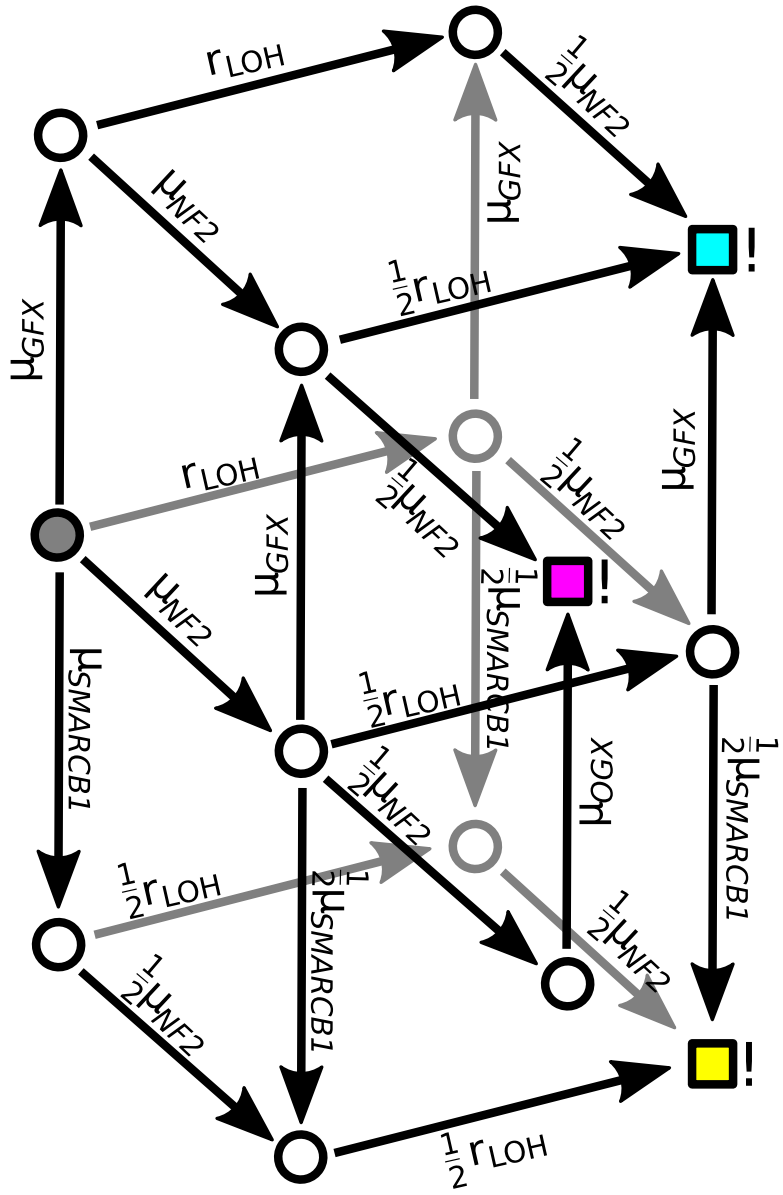
  = ■/(■ + ■ + ■)

Can use these to fix parameters!

Vestibular schwannoma incidence

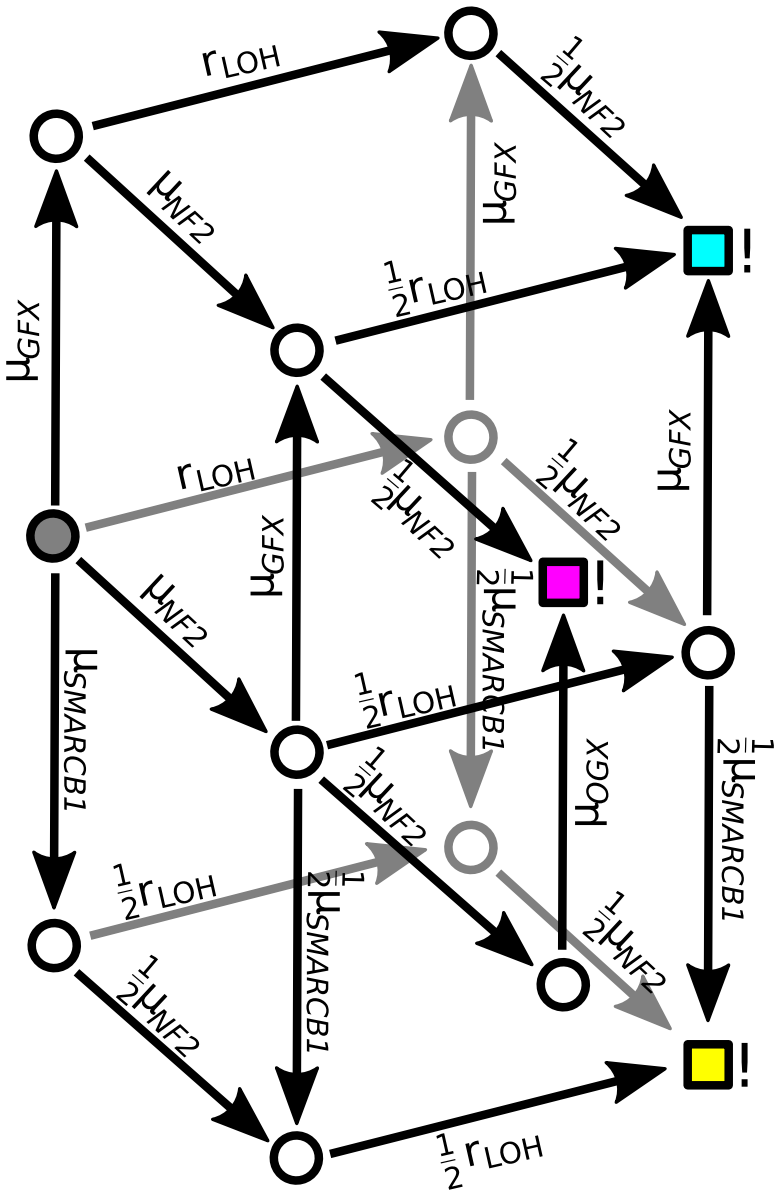
Our model for sporadic VS

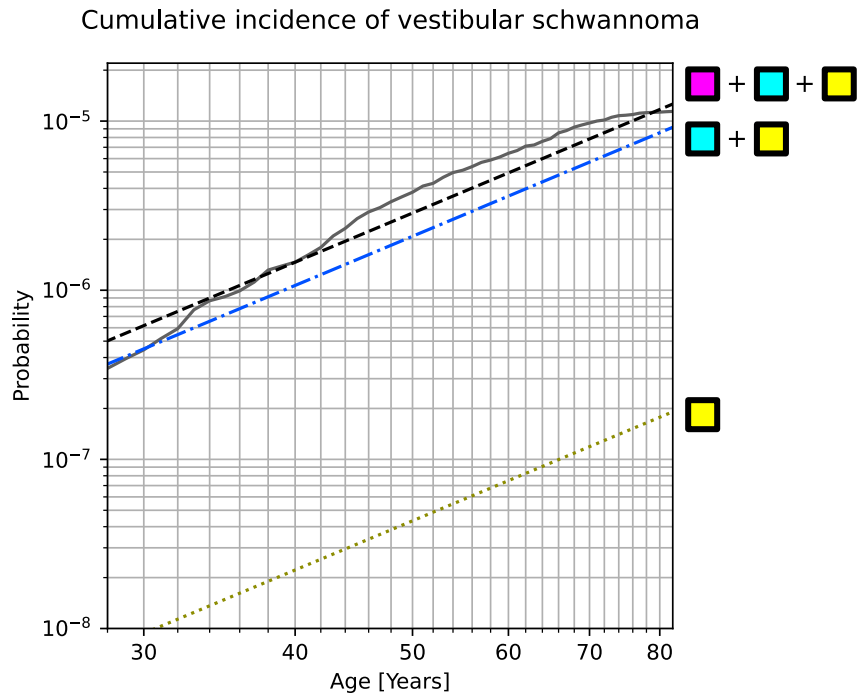
- *Don't* use *ab initio* point estimates for u, r_{LOH}, n_{GFX} this time...
- Instead use $P(t) = \square + \square + \square$, $f_{LOH} = \frac{\square + \square}{\square + \square + \square}$, and $f_{SMARCB1} = \frac{\square}{\square + \square + \square}$ to fix the parameters!



Vestibular schwannoma incidence

Our model for sporadic VS

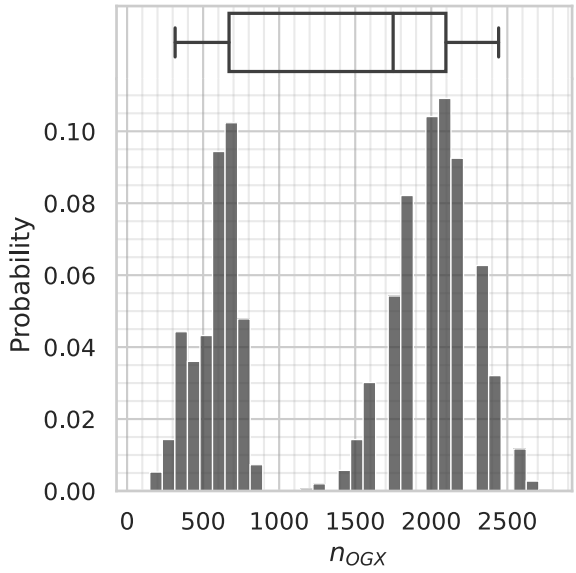


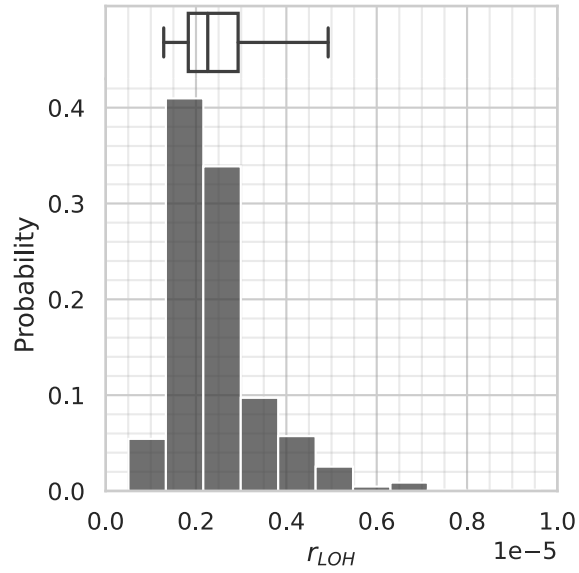


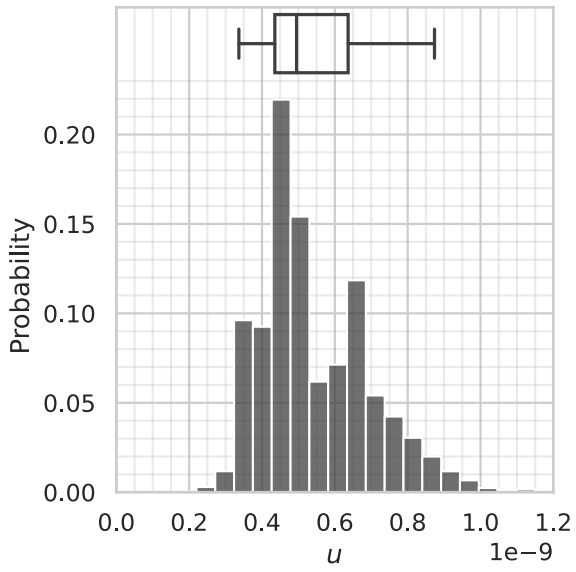
[1] C Paterson, I Bozic, MJ Smith, X Hoad, DGR Evans, https://doi.org/10.1101/2021.10.03.457528

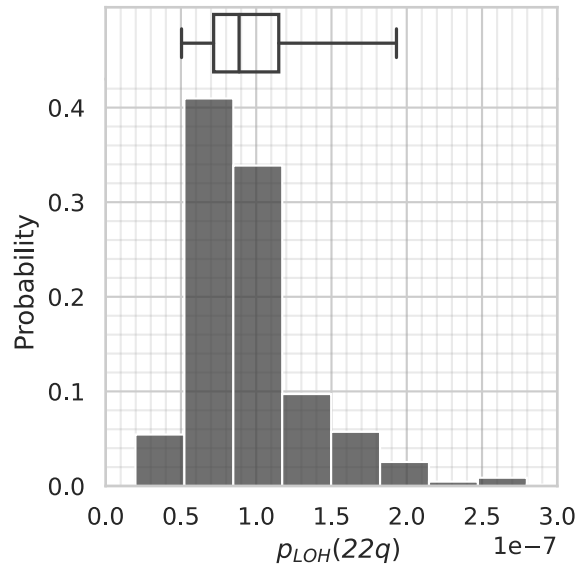
Vestibular schwannoma incidence

New parameter estimates





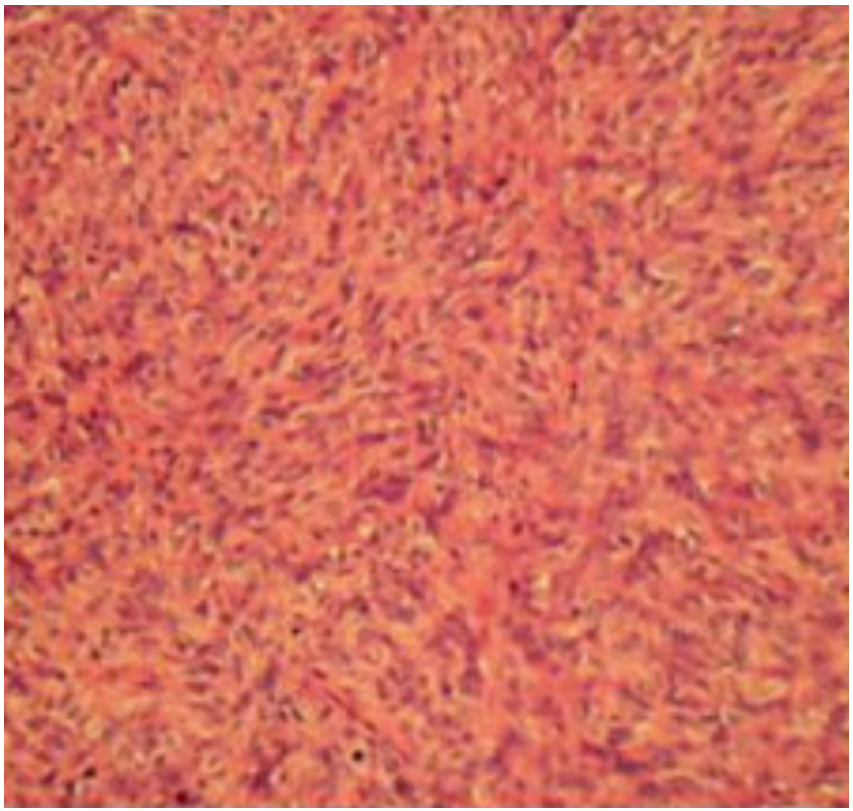




Malignant transformation in vestibular schwannoma

Very rare, very bad

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

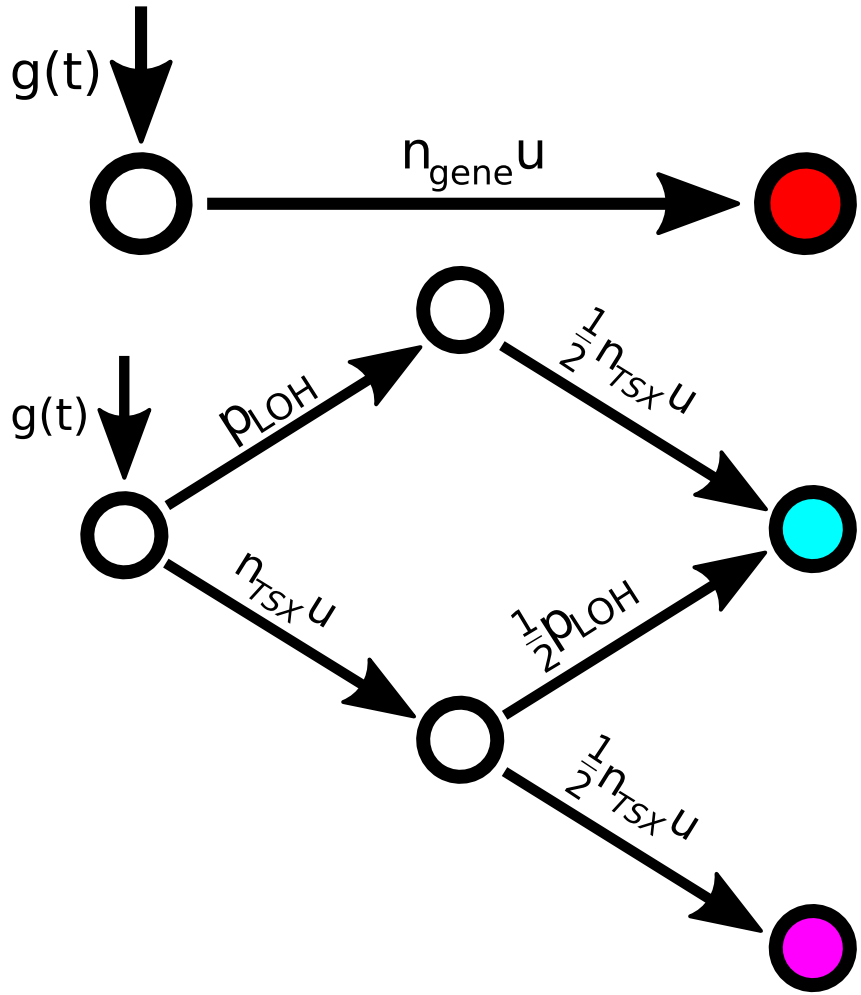


[2] AK Demetriades et al. Skull Base (2010)20:381–387.

Malignant schwannoma: two models

Timing and identity of drivers

Oncogene activation:



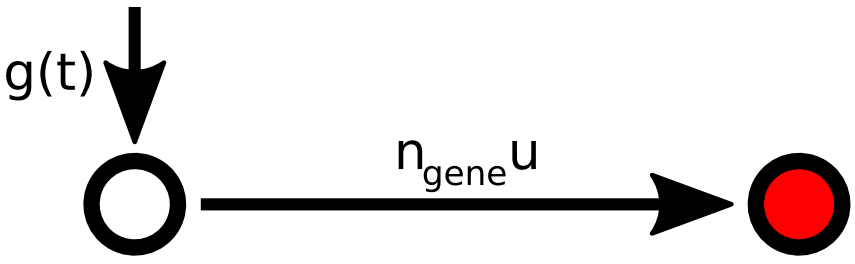
TSX inactivation:

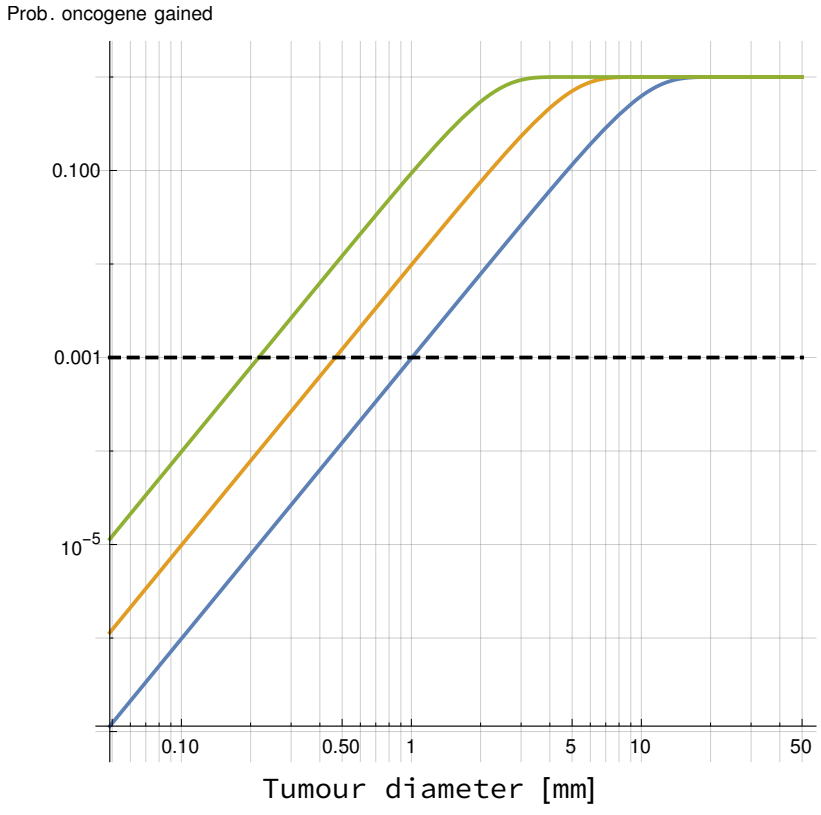
Malignant VS is *extremely* rare!

[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

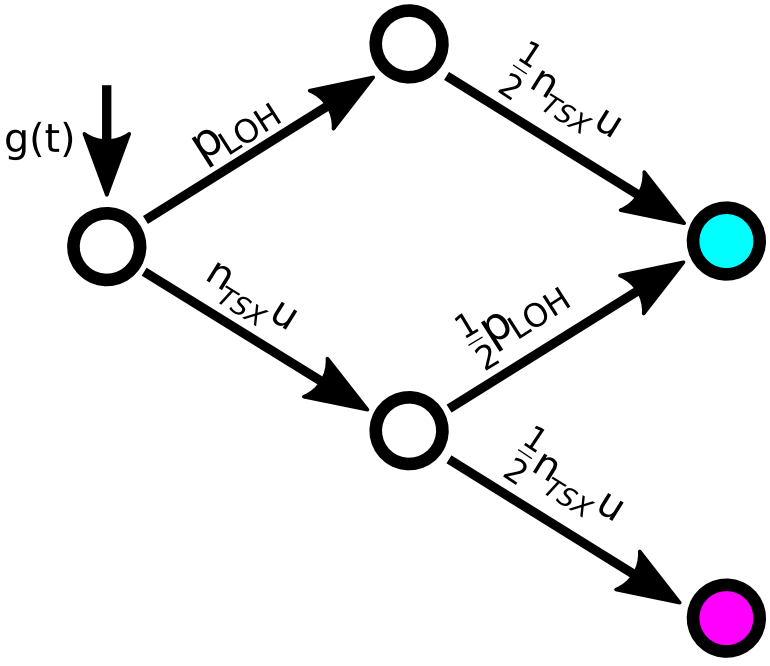


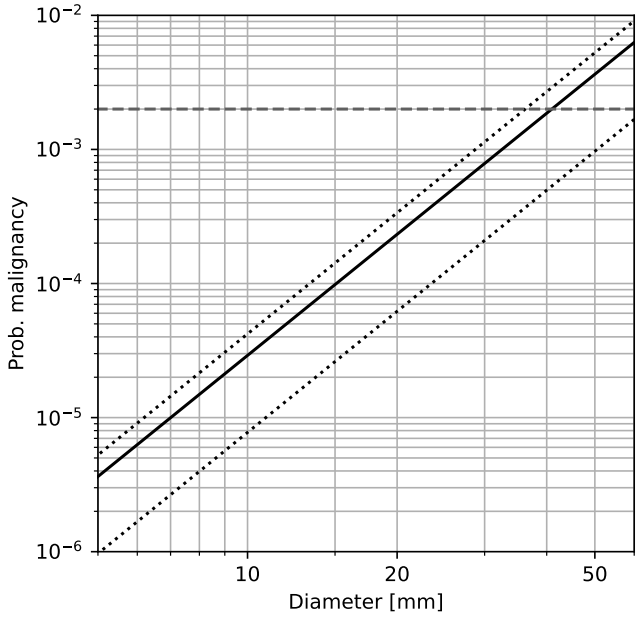


- Oncogene activation $\Longrightarrow$ high risk
- But it's a rare outcome
- So it's probably not caused by oncogene activation

Malignant schwannoma: second model

Tumour suppressor *TSX* inactivation

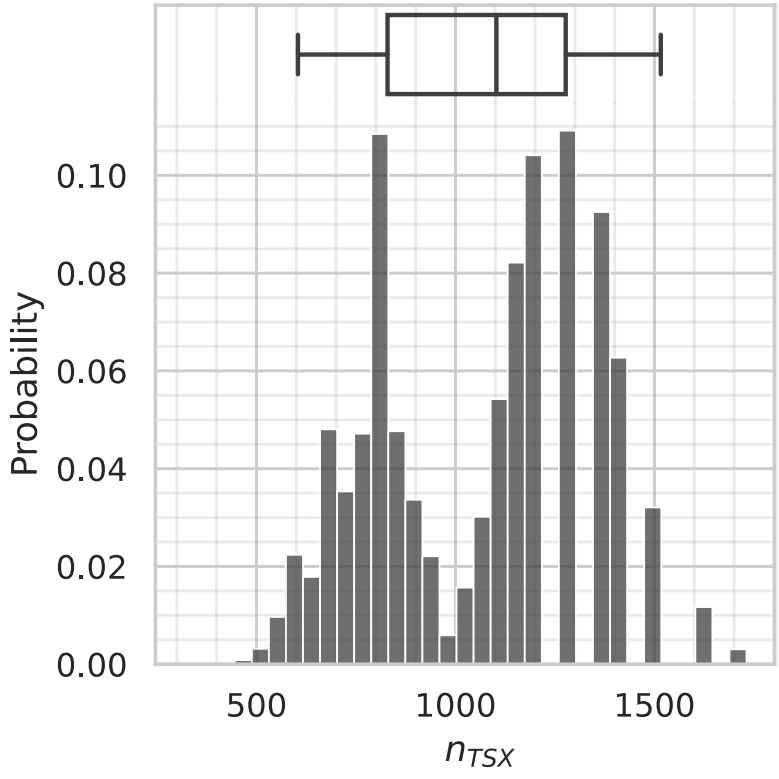




- *TSX* inactivation $\Longrightarrow$ low risk
- Can also estimate n_{TSX} that's consistent with incidence

Who is *TSX*?

Parameter estimates for n_{TSX}



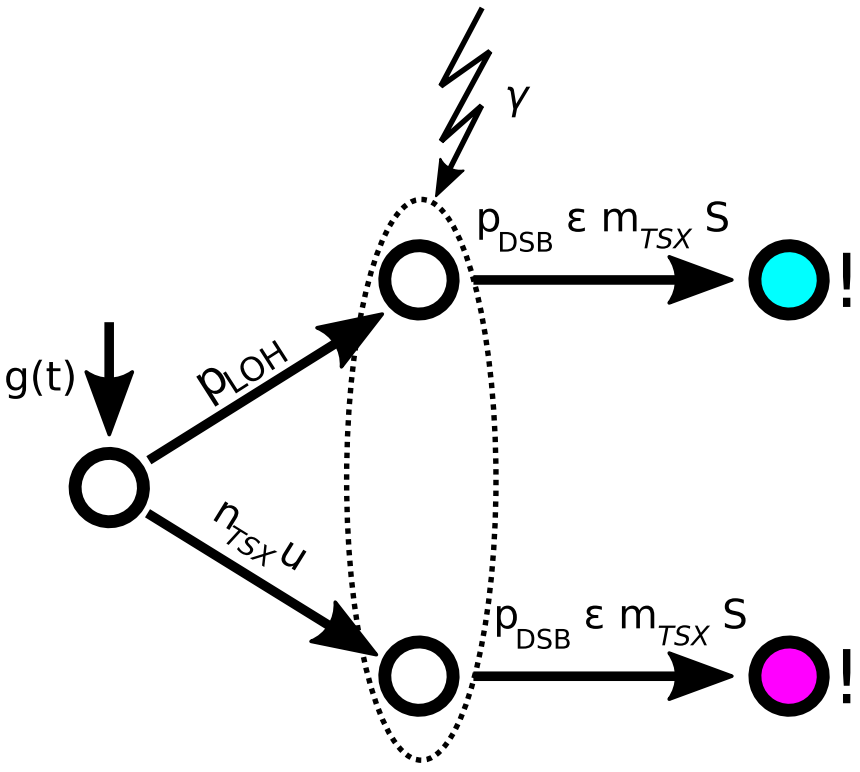
Probably multiple (10?) distinct tumour suppressors

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

Malignant schwannoma

Radiation

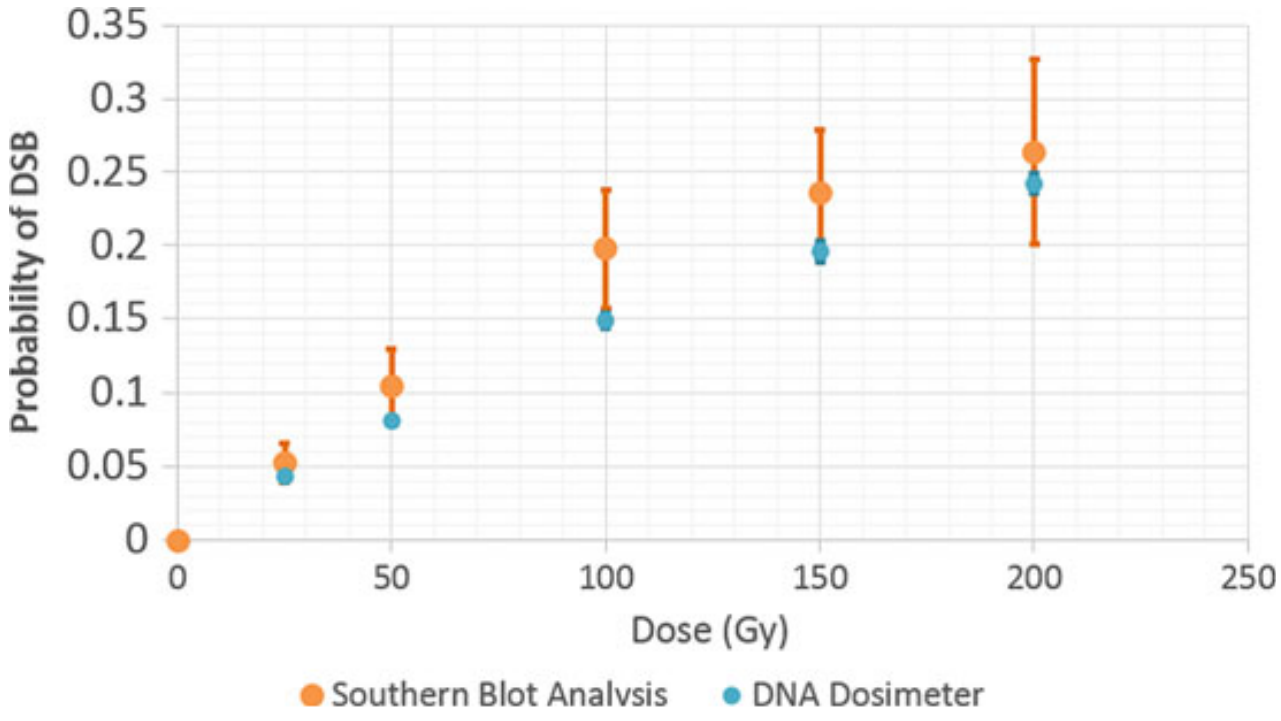
Why do we care about *TSX* anyway?



3 dose-dependent effects:

- DSB induction $p_{DSB}(D)$
- DSB misrepair $\varepsilon(D)$
- cell survival $S(D)$

DSB induction model



DSB misrepair

- Should be dose dependent $\varepsilon(D)$...
- Maybe also dose-*rate* dependent $\varepsilon(D, \dot{D})$

$$\varepsilon \approx \tag{1}$$

Cell survival

Why do we care about *TSX* anyway?

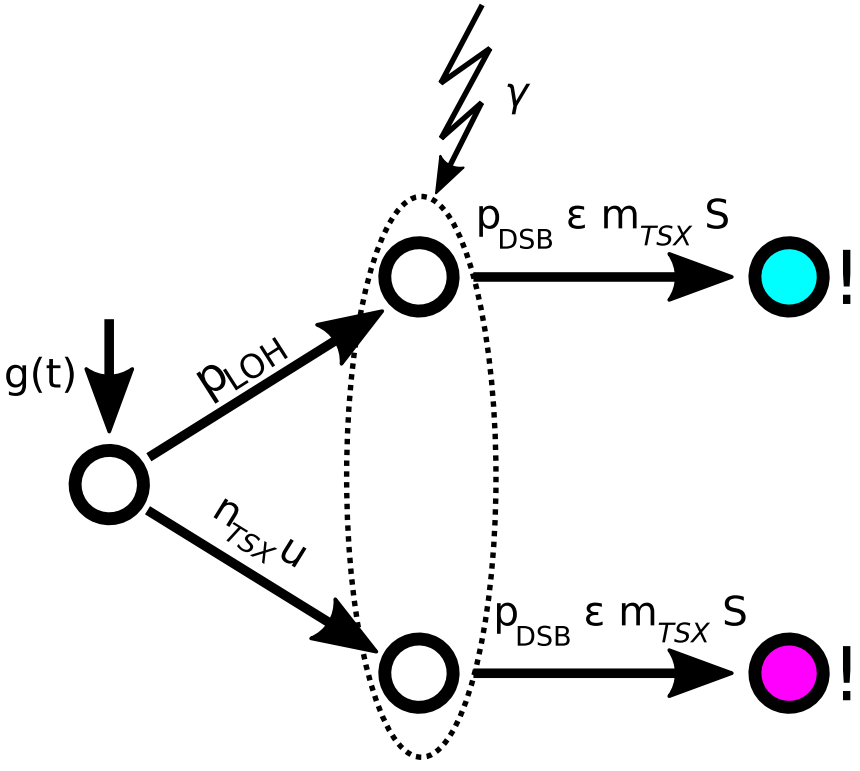
Need to know how many indel sites m_{TSX} result in loss of function:

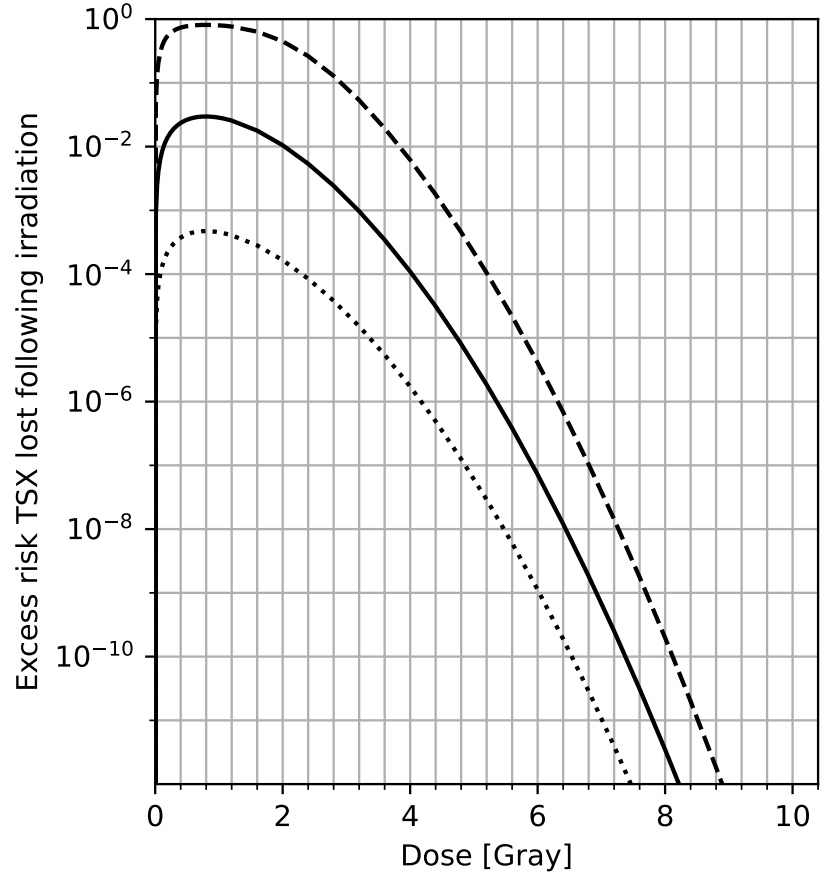
$$m_{TSX} \approx 0.42 n_{TSX} \tag{2}$$

Malignant schwannoma

Radiation

Why do we care about *TSX* anyway?





Malignant schwannoma

Radiation

Uncontroversial conclusions, but...

- Model is missing mechanisms (what is ε? multiple genes?)
- Multifocality and familial NF2?
- Margin clearance & recurrence?

What experiments could constrain these?

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*
4. Radiotherapy probably OK (w. caveats + huge error bars)

Main outputs

but...

1. Uncertainties still large
2. Identity of *TSX* unknown
3. Constraints weak: *GFX* and *TSX* probably multiple genes
4. Don't know about NF2

To do list

Next gen sequencing...

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

but also...

- Better models!!! CNA/VAF models, machine learning, optimal clustering, $\mu_{C>T}$ etc.
- SEER data too?
- Multiple genes *GFX* and *TSX*?
- Haploinsufficiency, selection?

Lots to do!

I need collaborators, send me data

Acknowledgements + collaborators

for their *in kind* support

The University of Washington

In order of appearance...

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


Thank you for listening!