Genes, environment, and “bad luck”

Martin A. Nowak and Bartlomiej Waclaw

Mobolaji Williams — Shakhnovich Journal Club— April 25, 2017

*(*All figures are from the paper unless otherwise cited)*

Is cancer caused by bad luck?

"It is a human trait to search for explanation for catastrophic events and rule out mere "chance" or "bad luck".

- Nowak and Wallow

Authors' Question: Is cancer mostly caused by bad luck (i.e., random mutations) and thus cannot be deliberately prevented?

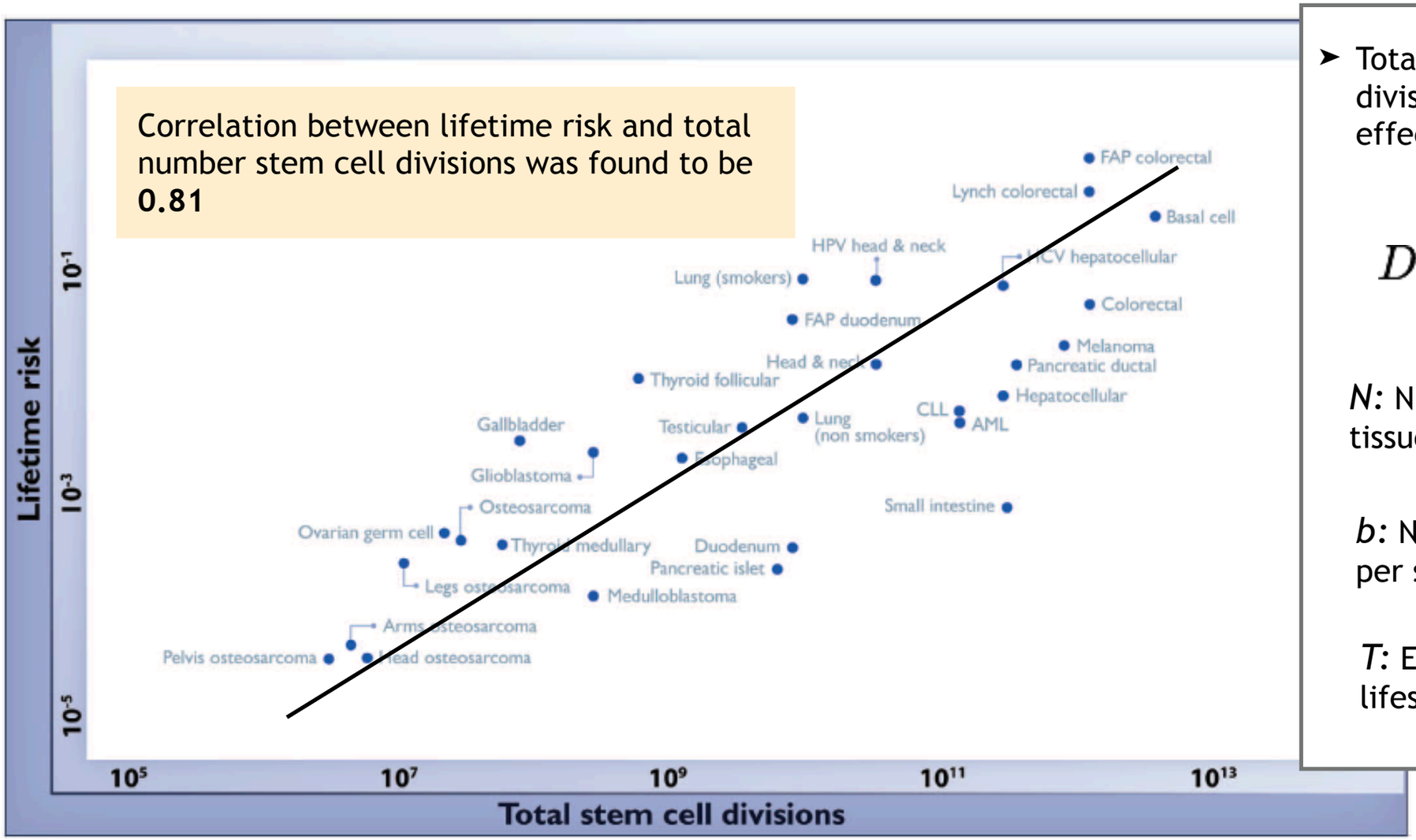
Authors' Answer: This question cannot be answered from data. Mathematical models of cancer are required to supplement the existing data analysis.

Risk and stem cells

In a 2015 study, Tomasetti and Vogelstein concluded that 65% of the variation in the risk of certain cancers was due to random stem cell divisions.

- Lifetime risk was taken from the Surveillance, Epidemiology, and End Results (SEER) database.

Correlation between lifetime risk and total number stem cell divisions was found to be **0.81**



FAP = Familial Adenomatous Polyposis ◆ HCV = Hepatitis C virus ◆ HPV = Human papillomavirus ◆ CLL = Chronic lymphocytic leukemia ◆ AML = Acute myeloid leukemia

(From Tomasetti *et al* (2015))

- Total number of stem cell divisions was calculated with an effective formula:

$$D = N \times b \times T$$

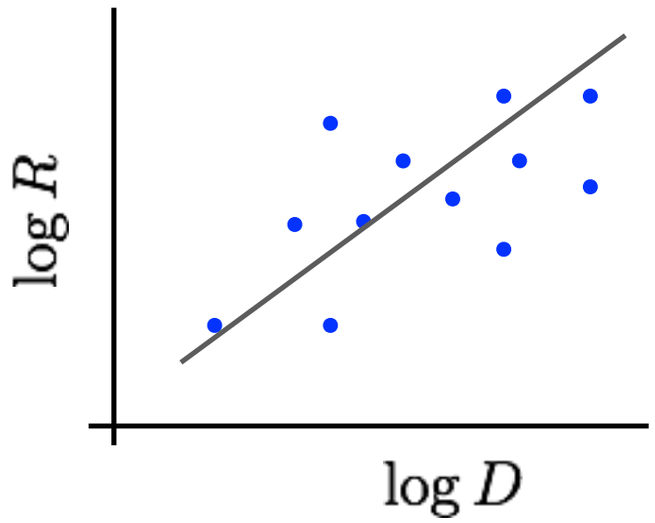
N: Number of stem cells in tissue where the cancer

b: Number of divisions per stem cell

T: Estimated human lifespan

Understanding the results

Correlation between lifetime risk and total number stem cell divisions was found to be **0.81**



Tomasetti and Vogelstein's result presents a **statistical account** of the relationship between lifetime risk and # of stem cell divisions, but no **biological account.**

How can we find an analytic relationship between $\log R$ and $\log D$?

Nowak and Waclaw's question: How do environment, heredity, and mutation rates enter into a calculation of the lifetime risk of cancer?

Deconstructing lifetime risk

Q: How, approximately, does "risk of death" depend on "initiation" and "progression"?

probability density for cancer cell arising when person is at age *t*

Lifetime probability of death from cancer

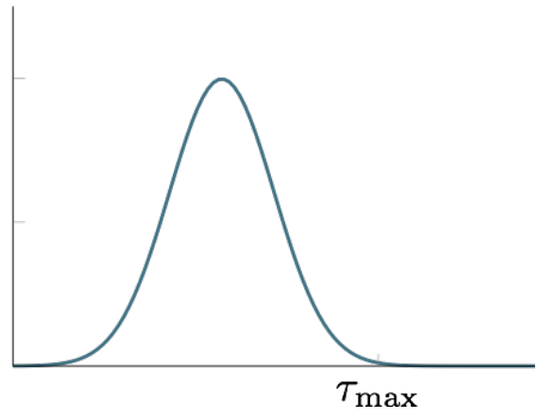
$$R = \int_0^T dt\, f(t) \int_0^{T-t} d\tau\, g(\tau)$$

probability density for full progression of disease at some time further time τ.

Simplifying Assumption: We take g(τ) to be peaked at some τ = 5-20 years and to fall off by some τ_max

$$0 < \tau < \tau_{\max} \ll T$$

$$\int_0^{T-t} d\tau\, g(\tau) \simeq \int_0^{\tau_{\max}} d\tau\, g(\tau)$$



$$R \simeq \int_0^T dt\, f(t) \times \int_0^{\tau_{\max}} d\tau\, g(\tau)$$
$$= P \times Q$$

Lifetime probability of cancer initiation

Lifetime probability of cancer progression

Lifetime probability of cancer initiation

Lifetime probability of cancer initiation

Lifetime probability of death from cancer

$$R = P \times Q$$

Lifetime probability of cancer progression

Q: How does the lifetime probability of cancer initiation depend on # of stem cell divisions?

Main Formula:

$P \simeq$ (Number of stem cells)
$\times$ (Probability of ocogenic mutation at time t)
$\times$ (Probability of fixation)

N: Number of stem cells present in tissue where cancer originates
b: Rate of division of a stem cell
μ: Probability of activating an oncogene
ρ: Average fixation probability of oncogene mutation

$$D = Nbt$$

(# of divisions of stem cell at time t)

"One-Hit" Model

(Single mutation causes initiation)

Probability of ocogenic mutation at time t: $b\mu t$

$$P \simeq N \times b\mu t \times \rho$$

$$P \simeq \rho\mu D$$

"Two-Hit" Model

(Two mutations are required for initiation)

Probability of ocogenic mutation at time t: $b\mu t \times b\mu t/2$

$$P \simeq N \times \frac{b\mu t \times b\mu t}{2} \times \rho$$

$$P \simeq \frac{\rho\mu^2 D^2}{2N}$$

Data and model comparisons

Lifetime probability of cancer initiation

Lifetime probability of death from cancer

$$R = P \times Q$$

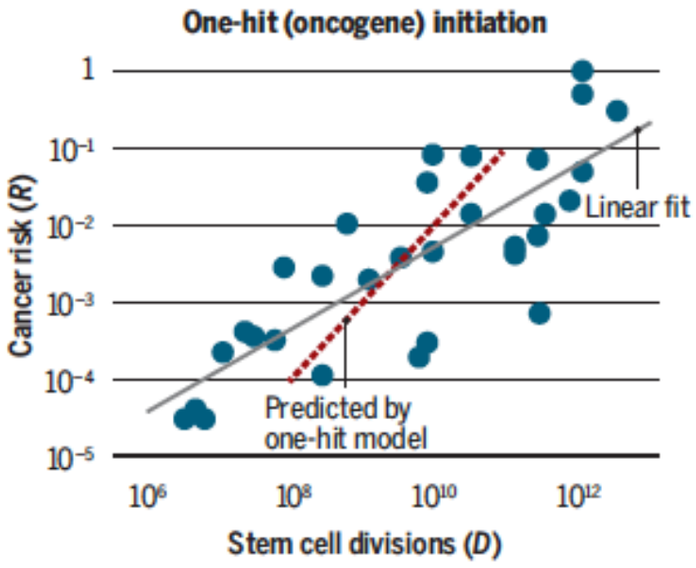
Lifetime probability of cancer progression

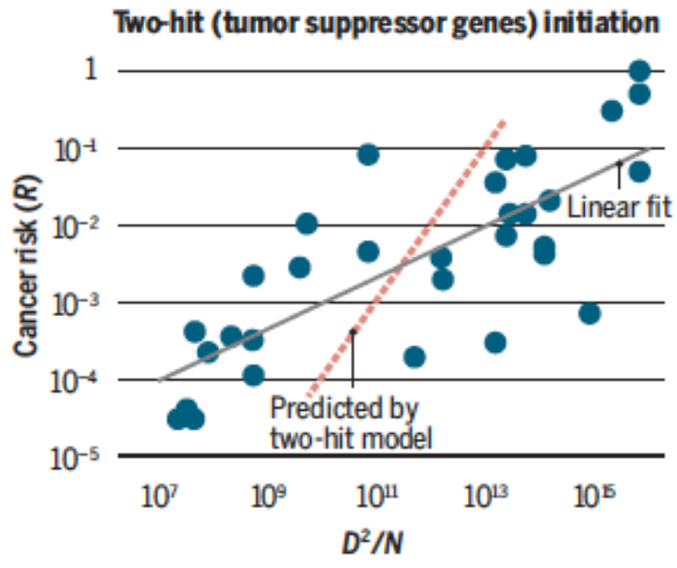
Q: How does the lifetime probability of cancer initiation depend on # of stem cell divisions?

"One-Hit" Model

(Single mutation causes initiation)

$$P \simeq \rho \mu D$$





"Two-Hit" Model

(Two mutations are required for initiation)

$$P \simeq \frac{\rho \mu^2 D^2}{2N}$$

$$\log R = \log D + \log(\rho \mu Q)$$

Both models predict a slope which is larger than that exhibited by the data.

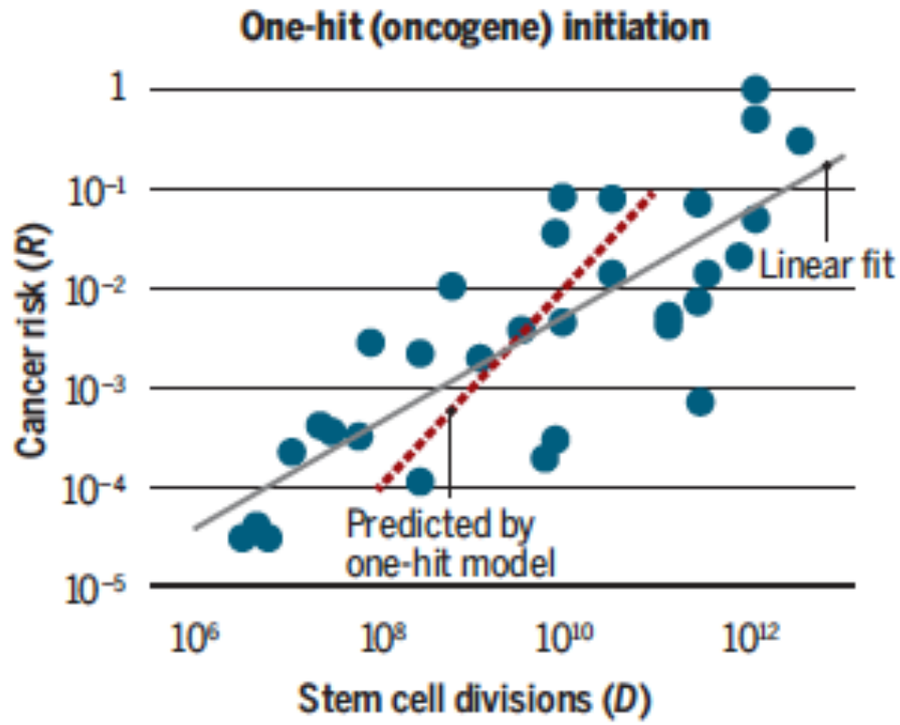
$$\log R = \log(D^2/N) + \log(\rho \mu^2 Q/2)$$

Modifying model for better fit

Q: How can we modify model to encompass the results in the data?

A: We need to decrease the predicted slope!

*We'll focus on the "one-hit" model because it provides a good proof of concept.



<u>"One-Hit" Model</u>

(Single mutation causes initiation)

$$\log R = \log D + \log(\rho\mu Q)$$

How can we decrease the predicted slope? (And what does this ability mean?)

1. Progression probability *Q*, decreases with number of stem cell divisions *D*

$$Q(D) \propto D^{-k} \quad \text{for } 0 < k < 1$$

yields

$$\log R = (1-k)\log D + \cdots$$

2. There is an effective number of stem cells D_eff which is the more relevant parameter

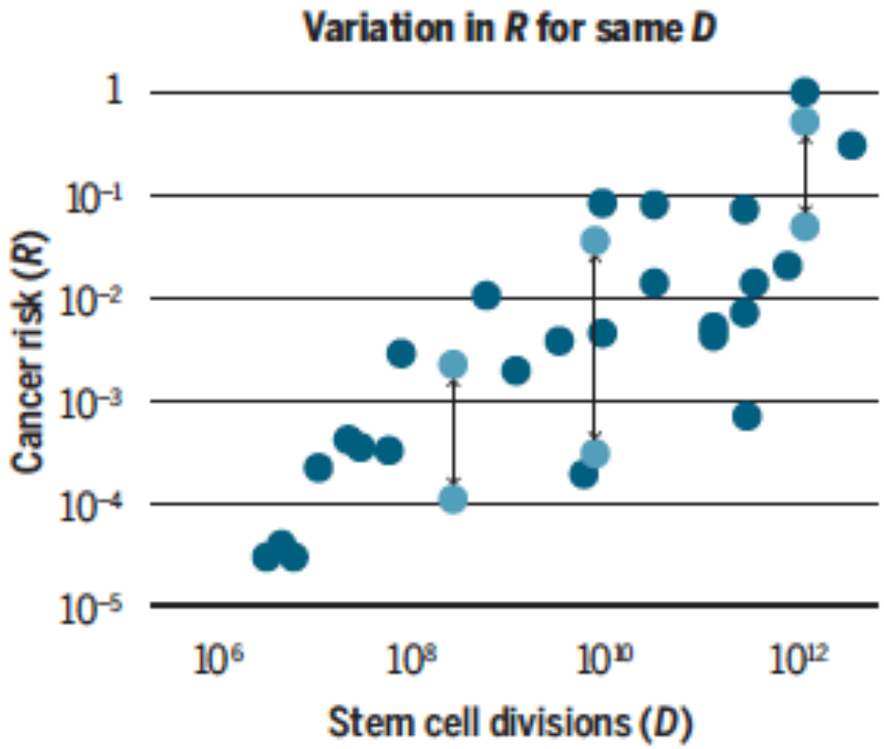
$$D_{\text{eff}} \propto D^{m} \quad \text{for } 0 < m < 1$$

yields

$$\log R = m \log D + \cdots$$

Author's point: Model provides a more precise accounting of how # of stem cell divisions can relate to cancer risk

Variation in risk for constant # of divisions



Q: How can the model explain variations in risk (*R*) for the same number of divisions (*D*)

Probability of progression of cancer

$$\log R = \log D + \log(\rho\mu Q)$$

Probability of fixation

Probability of activation of oncogene

More Specifically: There can be variations in

i) # of target genes leading to cancer initiation

ii) # of additional hits needed for progression

iii) Various rates of cell division and death

iv) exposure to environmental agents that change mutation rate

Generally:
Any of the parameters, in the second term can vary yielding different *R* for the same *D*.

Is cancer mostly not preventable?

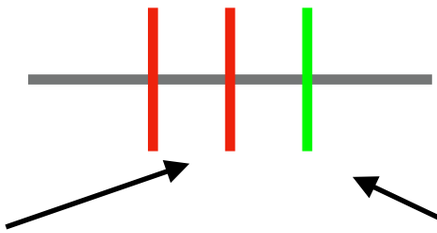
Tomasetti and Vogelstein find that most (i.e., 66%) of the mutations leading to cancer are due to random replication errors.

Q: Does this mean that cancer mostly cannot be prevented?

Author's deeper point: There needs to be a precise mathematical understanding of cancer in order to better interpret Tomasetti's and Vogelstein's results.

Author's claim: No!

e.g.



oncogenes that are activated randomly

oncogene that is activated due to environmental factors

If progression requires all three, cancer is still preventable

End