X chromosome inactivation in female mammalian cells

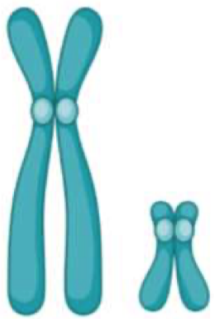
Alice Fohr

18/06/2026

PhD supervised by Frédérique Clément and Hélène Leman

Why does X chromosome inactivation (XCI) occur ?

Mammals :
Females **XX**, Males **XY** $\longrightarrow$ **Dosage compensation**



X and Y human chromosomes

X chromosome: > 1300 genes

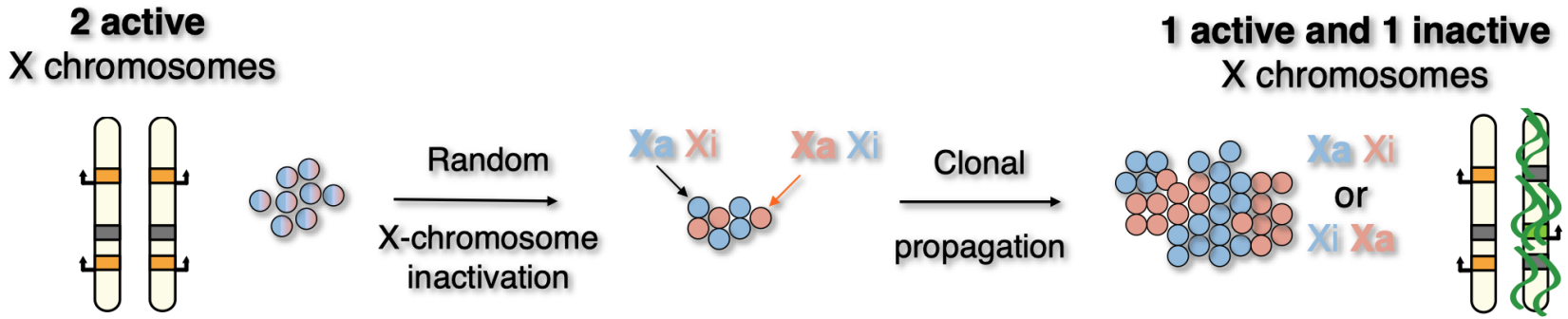
Y chromosome: approx. 200 genes

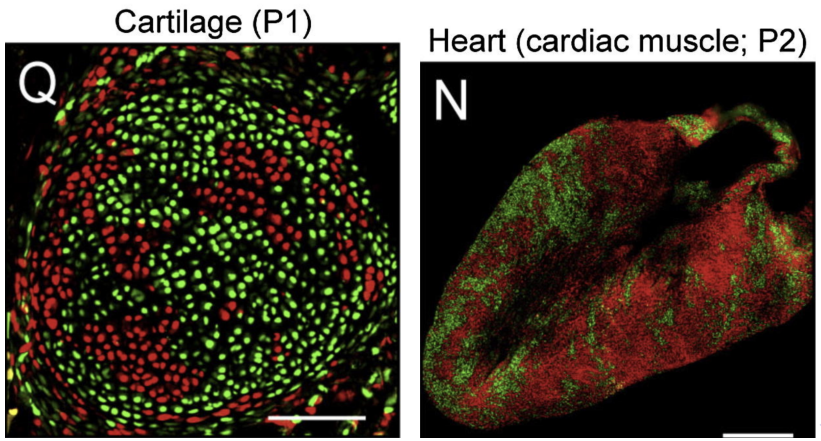
Transcriptional silencing of one of the 2 X chromosomes
=
Dosage compensation

The mechanism of inactivation

During inactivation, the paternal or maternal X chromosome is inactivated randomly, and this "choice" of the inactivated X chromosome is passed from the mother cell to the daughter cells during mitosis.

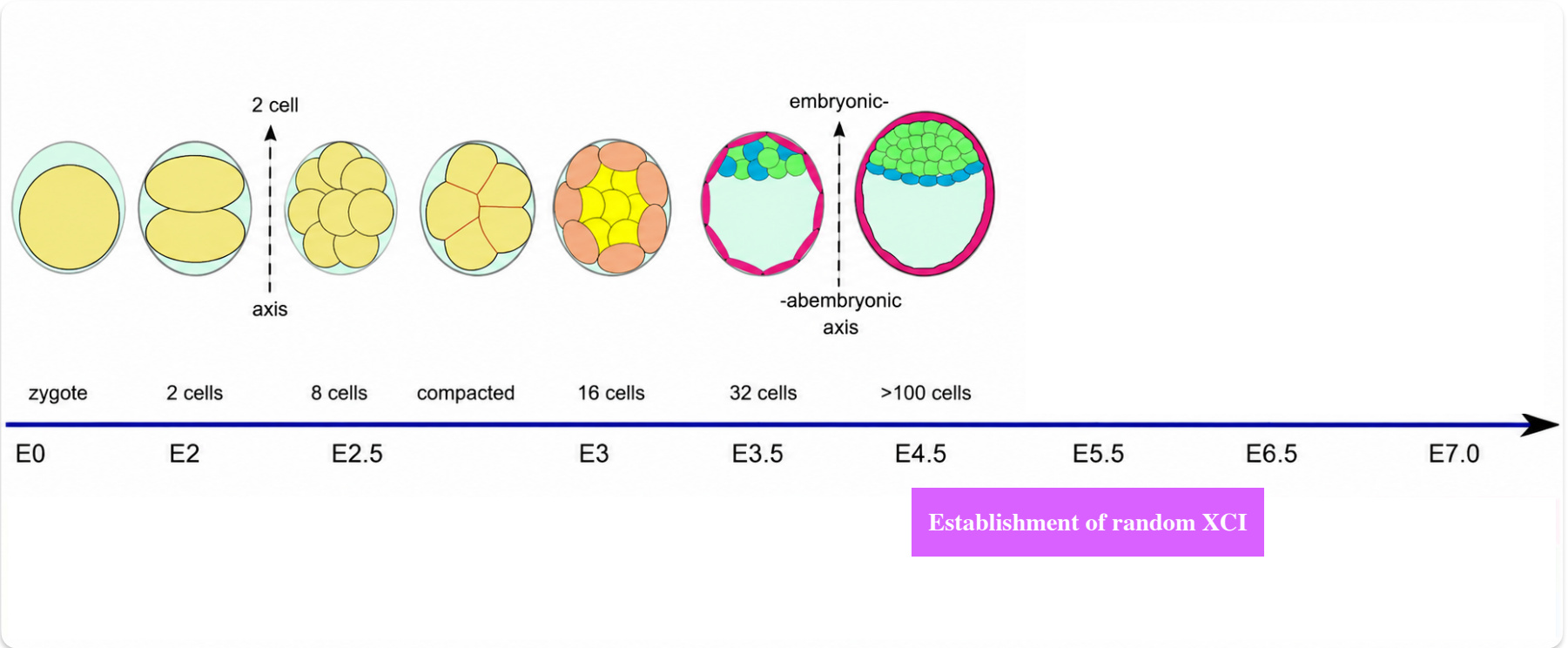
$\longrightarrow$ Appearance of a cellular mosaic.





Wu et al 2014

Developmental timeline of random XCI in the mouse embryo



Definition

XCI ratio

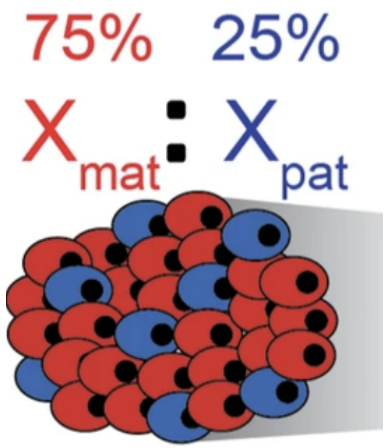
A priori :

$$P(X_m \text{ inactive}) = P(X_p \text{ inactive}) = \frac{1}{2}$$

$\Rightarrow$ expected 50 :50 distribution of XiXa(m) and XiXa(p) cells.

Observed deviations

Deviations from the expected 50 :50 ratio are frequently observed (**XCI imbalance**)



Importance of studying the XCI ratio

XCI ratio : powerful tool to **infer early developmental dynamics**.

Variability between individuals reflects the **number of precursor cells** present at XCI.

By the law of large numbers :

- Large founder pools $\rightarrow$ low variability
- Small founder pools $\rightarrow$ high variability

Existing models use XCI variability to estimate :

- Founder population size
- Lineage allocation dynamics

Existing models of XCI ratio variability

General approach

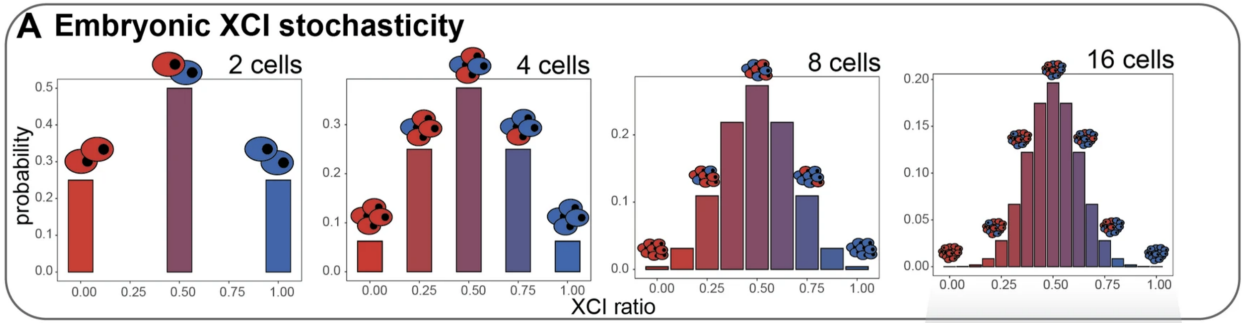
Use the observed variability of XCI ratios to infer developmental parameters, in particular the number of cells present at the time of XCI.

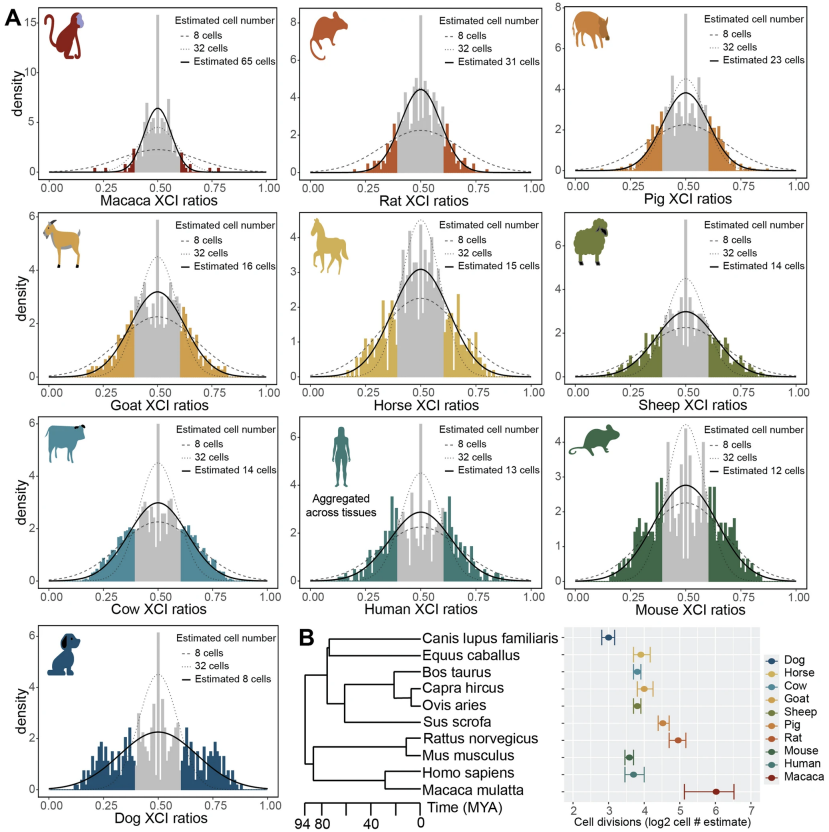
Almost all existing models (Werner 2024, Shvetsova 2019, Amos 2006, Kane 1979, Nesbit 1971) rely on **two key assumptions** :

1. **Instantaneous XCI**
   All N_0 precursor cells undergo XCI simultaneously.
2. **Deterministic growth**
   Subsequent proliferation preserves exactly the proportions established at XCI.

Existing literature (J. M. Werner, J. Hover, J. Gillis, 2024)

Population variability in X-chromosome inactivation across 10 mammalian species

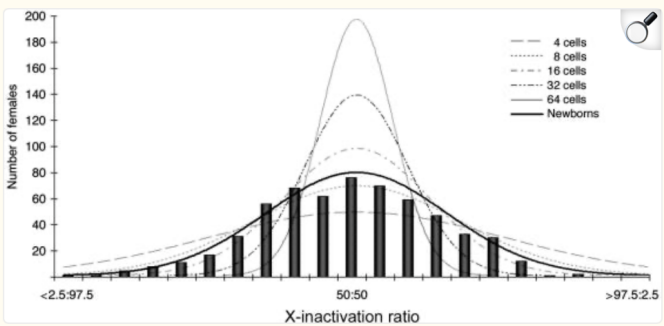




Existing literature (Shvetsova 2019, Amos 2006)

Previous studies determined the number of hematopoietic stem cells present at the time of X-inactivation choice by use of the variance of the X inactivation–pattern distribution.[26] To estimate the primordial stem-cell pool size (N), the formula for the variance of a binomial distribution was used ($variance=pq/N$) and setting the probability that either allele (p or q) is active to 0.5. Normal distributions were drawn by determining the variance (σ^2) for each of the values of N and drawing the curve with use of the equation for the normal distribution (fig. 3). To limit the potential effect that somatic mutations and environmental factors might impose on hematopoietic stem cells, we performed these calculations using only the 590 newborn samples. With the SD of 15.3%, the number of stem cells is predicted to be between 8 and 16, which is consistent with previous estimates[26,27,42] (fig. 3).

Figure 3. .



Open in a new tab

Comparison of the newborn distribution with predicted normal distributions of varying stem-cell pool size. The solid black curve follows a normal distribution with use of the actual SD of the newborn samples (15.4 [fig. 2]). The estimated number of precursor stem cells predicted by the newborn SD is between 10 and 12 cells.

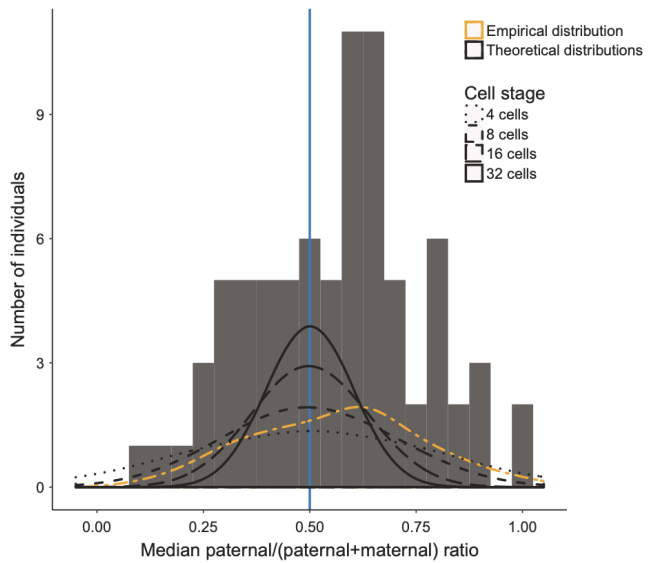
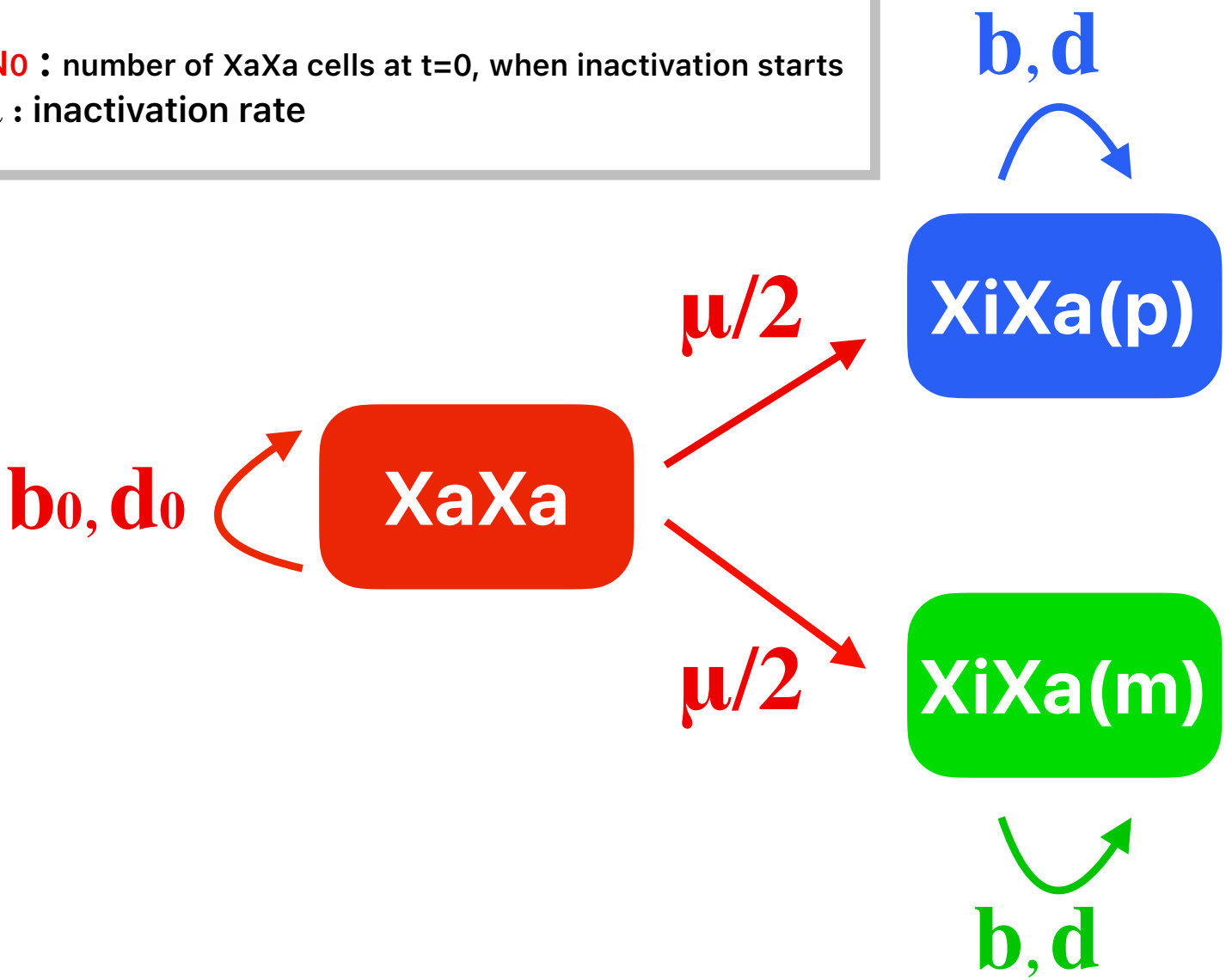


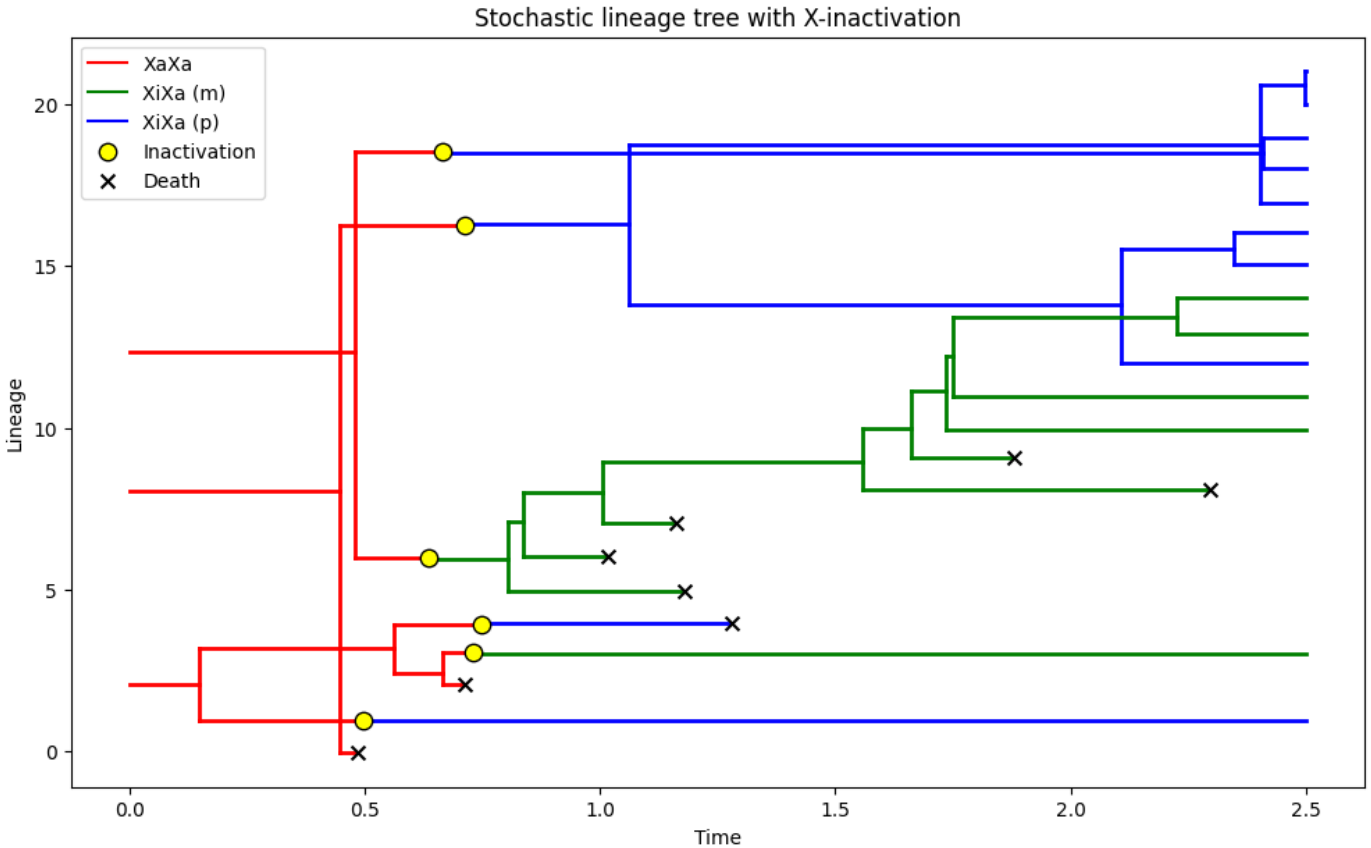
Fig. 3 Theoretical assessment of cell numbers at the time of X-inactivation. Comparison of the empirical median paternal ratio distribution for heterozygous SNVs with more than ten reads per individual (orange line) with theoretical distributions under the hypothesis that X-inactivation takes place at the 4 (dotted black line), 8 (dashed black line), 16 (long dashed black line), and 32 (solid black line) precursor stage. Theoretical distribution at eight initial lineage-restricted precursor cells is most comparable with empirical distribution (highest p value = 0.011, two-sample Kolmogorov–Smirnov test)

Model presentation

N0 : number of XaXa cells at t=0, when inactivation starts
μ : inactivation rate



Model presentation



Sources of variability

- Number of times Xp or Xm is chosen for inactivation : on average it's 50/50, but there is variance.
  $\longrightarrow$ The number of clones arising from XiXa(p) or XiXa(m) cells can vary.

- Timing of inactivation for each cell : inactivation is not simultaneous, the timing is random. Cells inactivated earlier tend to produce larger clones on average.

- Clonal growth : Even for clones originating from cells inactivated at the same time, clone sizes are random (even if the average growth rate is the same).

Notation

A_t : number of XaXa cells at time t
I_{m_t} : number of XiXa(m) cells at time t
I_{p_t} : number of XiXa(p) cells at time t

$$A_t = N_0 + \int_0^t \int_{\theta \in R^+} \mathbb{1}_{\theta \leq b_0 A_{s^-}} N_1(ds, d\theta) - \int_0^t \int_{R^+} \mathbb{1}_{\theta \leq d_0 A_{s^-}} N_2(ds, d\theta) - \int_0^t \int_{R^+} \mathbb{1}_{\theta \leq \frac{\mu}{2} A_{s^-}} N_3(ds, d\theta) - \int_0^t \int_{R^+} \mathbb{1}_{\theta \leq \frac{\mu}{2} A_{s^-}} N_4(ds, d\theta)$$

$$I_{p_t} = \int_0^t \int_{R^+} \mathbb{1}_{\theta \leq b_1 I_{p_{s^-}}} N_5(ds, d\theta) - \int_0^t \int_{R^+} \mathbb{1}_{\theta \leq d_1 I_{p_{s^-}}} N_6(ds, d\theta) + \int_0^t \int_{R^+} \mathbb{1}_{\theta \leq \mu_1 A_{s^-}} N_3(ds, d\theta)$$

$$I_{m_t} = \int_0^t \int_{R^+} \mathbb{1}_{\theta \leq b_2 I_{m_{s^-}}} N_7(ds, d\theta) - \int_0^t \int_{R^+} \mathbb{1}_{\theta \leq d_2 I_{m_{s^-}}} N_8(ds, d\theta) + \int_0^t \int_{R^+} \mathbb{1}_{\theta \leq \mu_2 A_{s^-}} N_4(ds, d\theta)$$

A new quantity : N_{clones}

Distinguishing two quantities

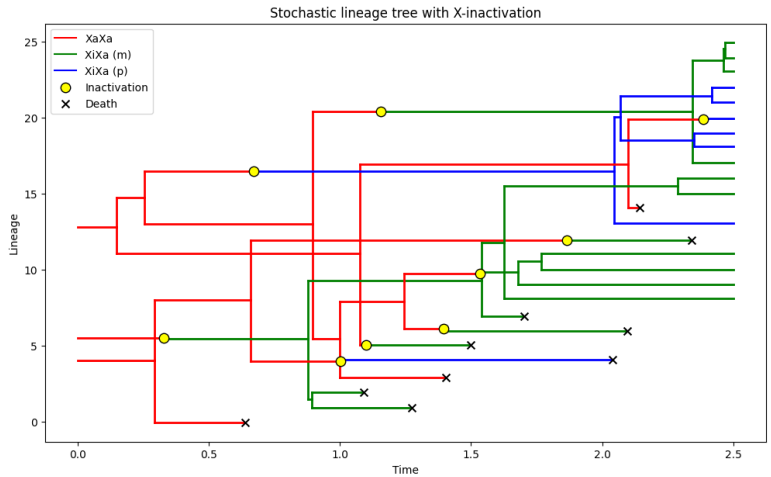
- N_0 : number of cells present at the onset of XCI
- N_{clones} : number of independent XCI choices : $X_aX_a \rightarrow X_iX_a$

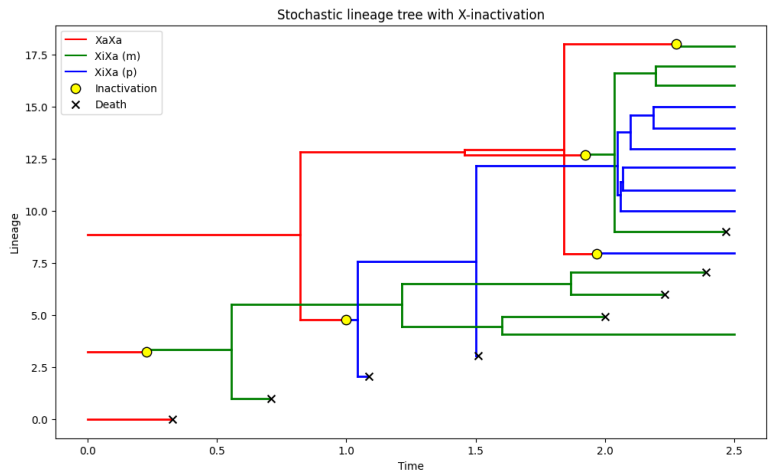
Each XCI event initiates a clone whose cells share the same inactive X.

Key point

The amount of stochasticity generated by XCI is determined by N_{clones}, not directly by N_0.

$$\text{Instantaneous XCI} \implies N_{\text{clones}} = N_0$$
$$\text{Asynchronous XCI} \implies E(N_{\text{clones}}) > N_0$$





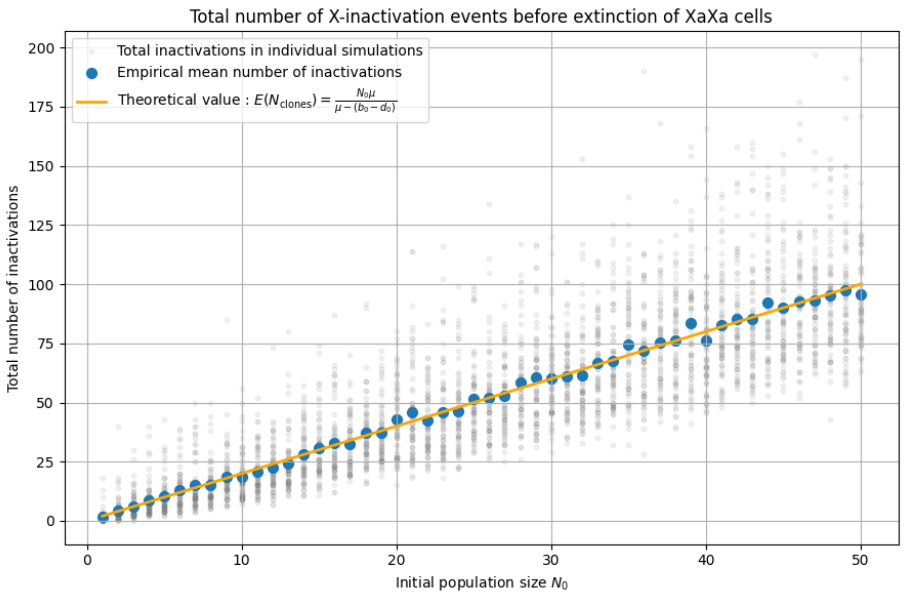
Expected number of XCI events

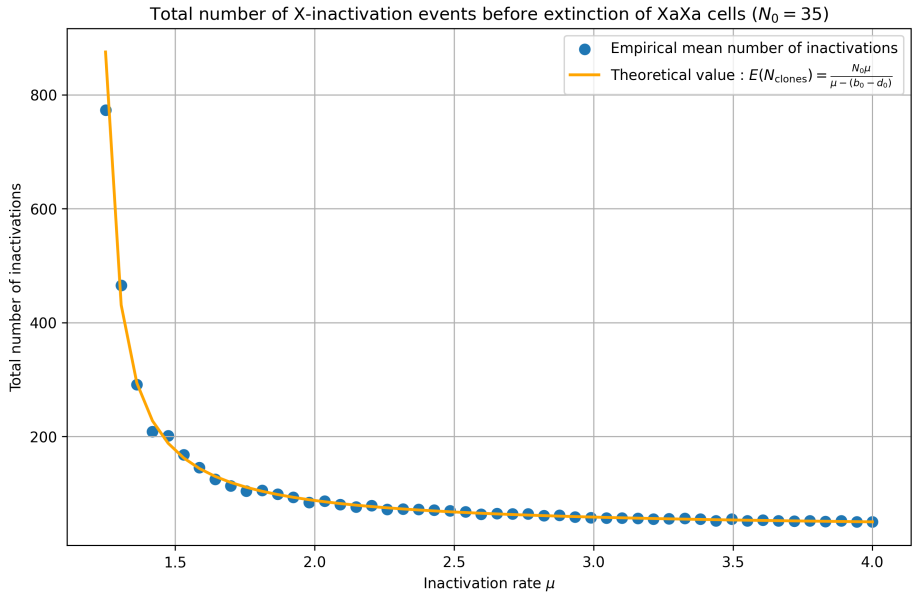
Number of initiated clones

$$C_t = \int_0^t \int_{\mathbb{R}_+} 1_{\{\theta \leq \mu A_{s-}\}} N_1(ds, d\theta), \text{ and } N_{\text{clones}} = \lim_{t \to \infty} C_t$$

Expected number of independent XCI choices

$$E(N_{\text{clones}}) = \frac{N_0 \mu}{\mu - (b_0 - d_0)} > N_0 \qquad \text{if } b_0 - d_0 > 0$$





Variance of absolute population sizes

An explicit expression for $Var(I_{m_t} - I_{p_t})$ can be derived by solving a set of ordinary differential equations (ODEs).

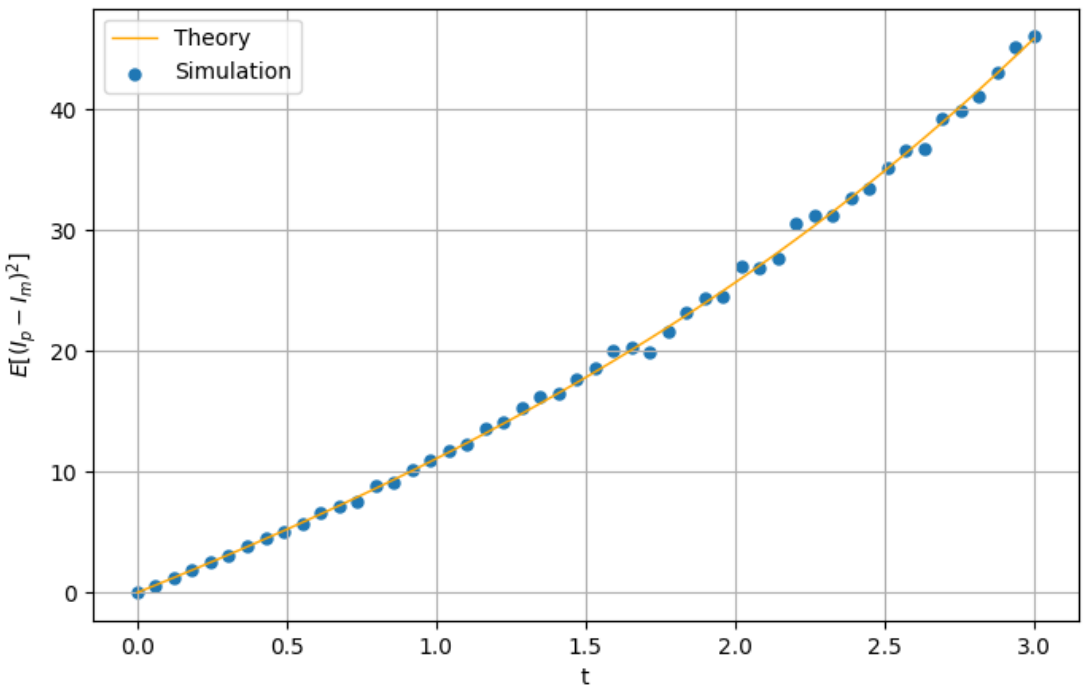
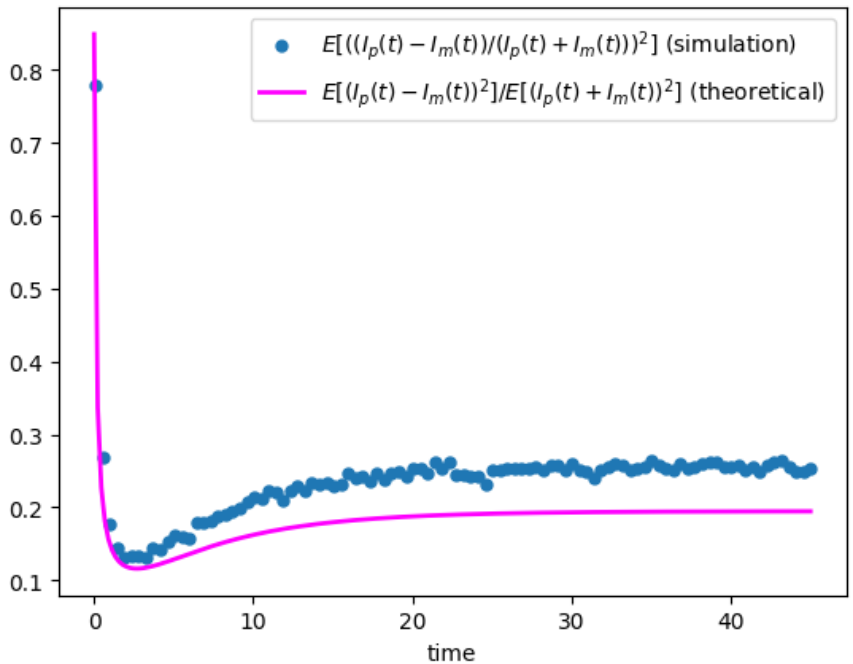


Figure – $Var(I_{m_t} - I_{p_t})$ over time, theoretically computed (orange curve) and estimated by simulations (blue points).

Limitation of analyzing $Var(I_{m_t} - I_{p_t})$ (variance of absolute cell counts) :

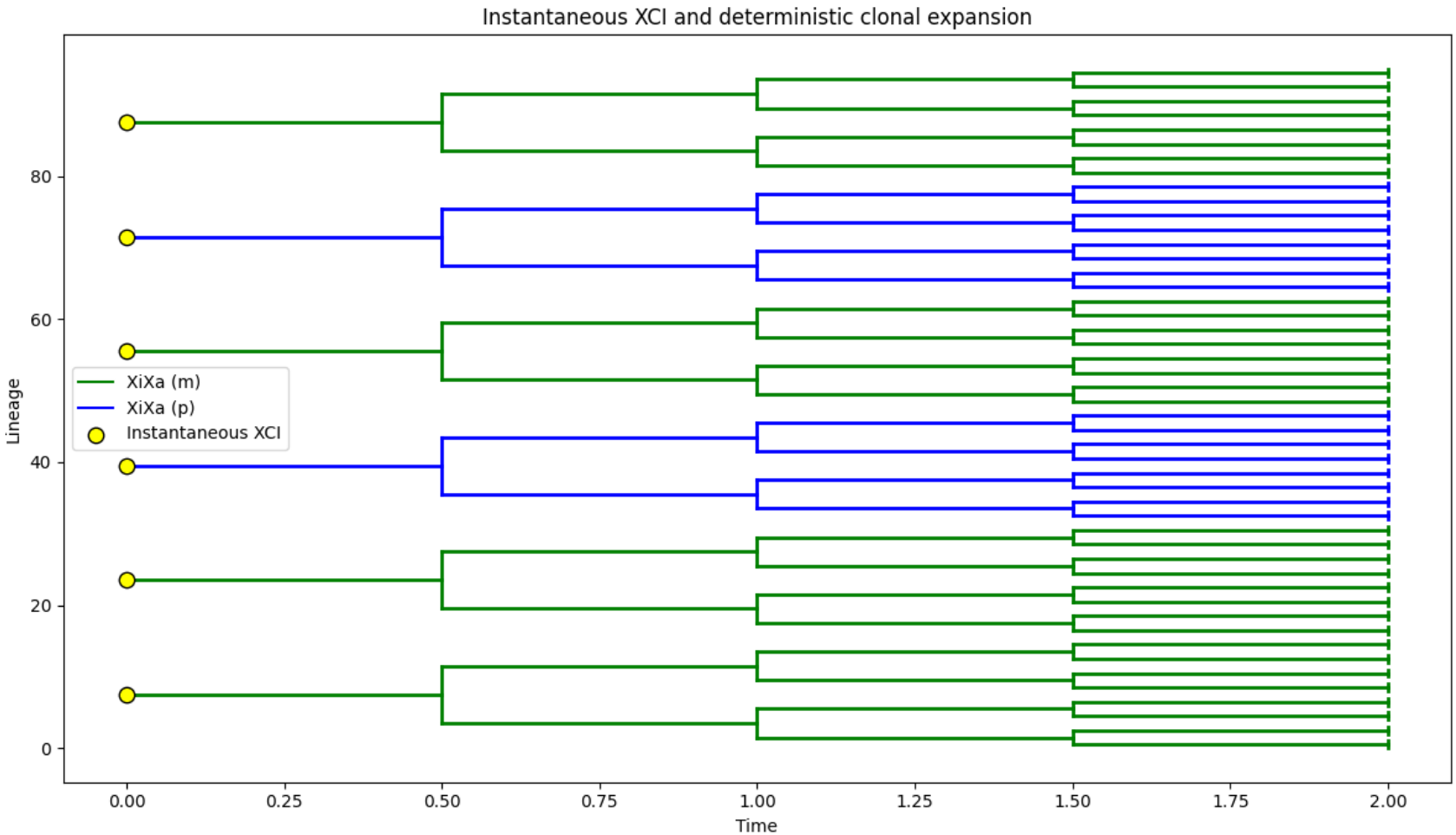
At later stages, it is practically impossible to count the total number of cells in a whole organism.
$\Rightarrow$ Variance of total counts is not well suited for comparison with real biological data.

Biological data correspond to the distribution of $\frac{I_{m_t} - I_{p_t}}{I_{m_t} + I_{p_t}}$, and $\frac{Var(I_{m_t} - I_{p_t})}{E((I_{m_t} + I_{p_t})^2)} \neq Var(\frac{I_{m_t} - I_{p_t}}{I_{m_t} + I_{p_t}})$.



1. Biological context
   - X chromosome inactivation
2. XCI ratio
   - Definition
   - Importance of studying the XCI ratio
3. Existing models
4. A multi-type stochastic birth-death process with asynchronous XCI
5. Computation of $\lim_{t\to\infty} Var\left(\frac{I_{m_t}-I_{p_t}}{I_{m_t}+I_{p_t}} \mid \text{survival}\right)$ in intermediate models

Instantaneous inactivation and derministic growth



Instantaneous inactivation and derministic growth

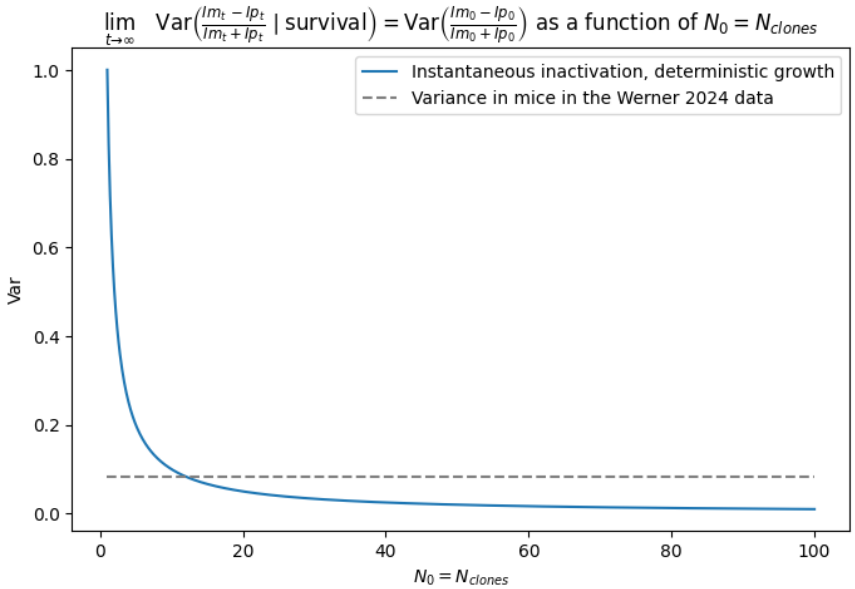
$N := N_0 = N_{\text{clones}}$

Binomial distribution : $Var\left(I_{p_0}\right) = Npq = \frac{N}{4}$

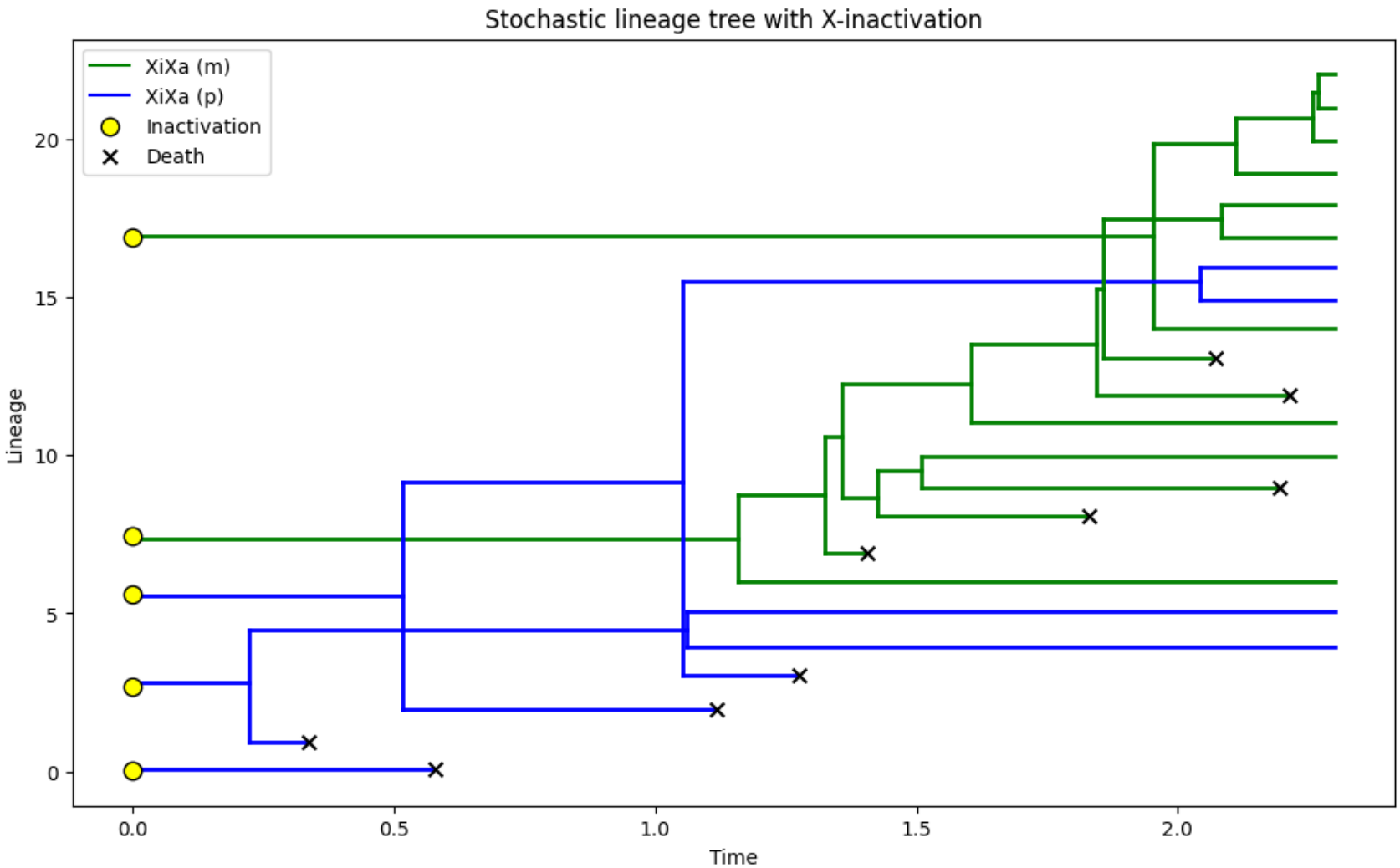
$$Var\left(\frac{I_{m_0} - I_{p_0}}{I_{m_0} + I_{p_0}}\right) = Var\left(1 - \frac{2I_{p_0}}{N}\right) = \frac{4}{N^2}\frac{N}{4} = \frac{1}{N}$$

Deterministic growth : $Var\left(\frac{I_{m_t} - I_{p_t}}{I_{m_t} + I_{p_t}}\right) = \frac{1}{N}$ for all t (and $\mathbb{P}(\text{survival}) = 1$).

$$\lim_{t\to\infty} Var\left(\frac{I_{m_t} - I_{p_t}}{I_{m_t} + I_{p_t}} \mid \text{survival}\right) = \frac{1}{N}$$



Instantaneous inactivation and stochastic growth

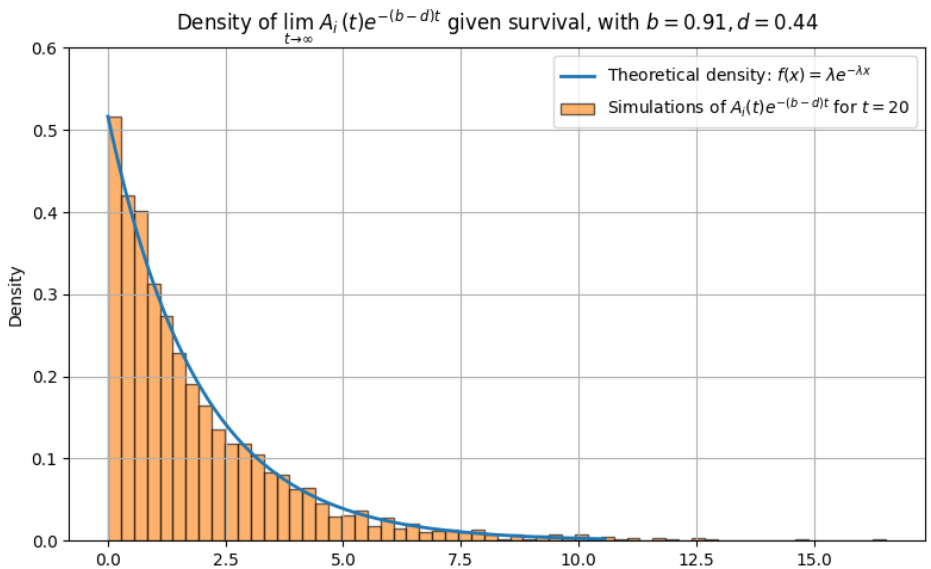


Instantaneous inactivation and stochastic growth

$N := N_0 = N_{\text{clones}}$

$A_i(t) =$ size of the clone originating from the i-th primordial XiXa cell, at time t

$\lim_{t\to\infty} A_i(t)e^{-(b-d)t} = w_i$ a.s. with $w_i = \chi_i\psi_i$, $\chi_i \sim \mathcal{B}(\lambda)$, $\psi_i \sim \mathcal{E}xp(\lambda)$, $\lambda = \frac{b-d}{b}$, and all the variables χ_i, ψ_j mutually independent.



Instantaneous inactivation and stochastic growth

$N := N_0 = N_{\text{clones}}$

$\lambda = \frac{b-d}{b}$

$$\lim_{t\to\infty} Var\left(\frac{I_{m_t}-I_{p_t}}{I_{m_t}+I_{p_t}} \mid \text{survival}\right)$$

$$= \frac{1}{1-(1-\lambda)^N} \sum_{n=1}^{N} \binom{N}{n} \lambda^n (1-\lambda)^{N-n} \frac{1}{2^n} \left(\lambda^n \sum_{k=1}^{n-1} \frac{\binom{n}{k}}{(k-1)!(n-k-1)!} \int_0^{+\infty} \int_0^{+\infty} \left(\frac{y-x}{x+y}\right)^2 x^{k-1} y^{n-k-1} e^{-\lambda(x+y)} dxdy + 2 \right)$$

Instantaneous inactivation and stochastic growth

$N := N_0 = N_{\text{clones}}$

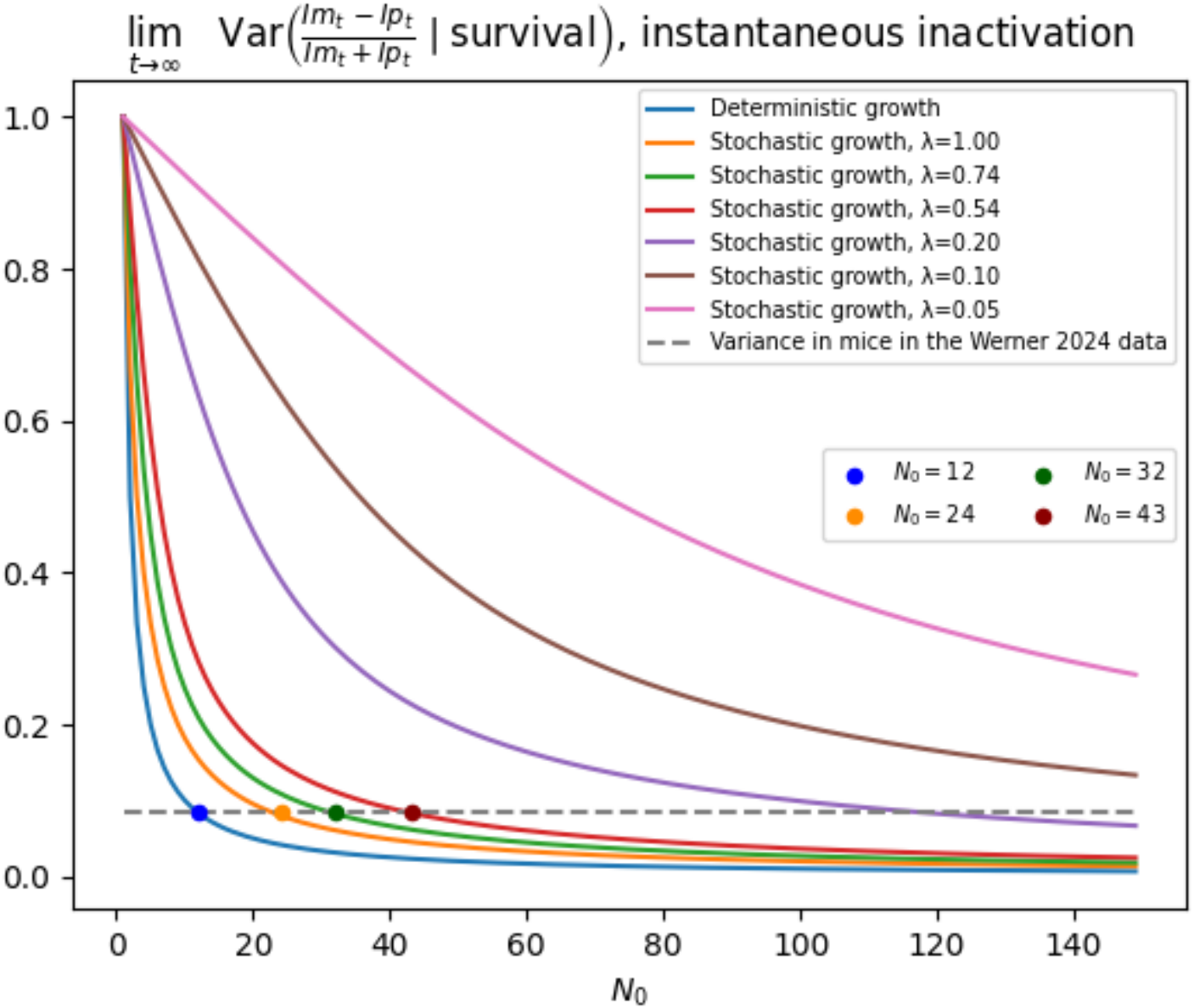
$\lambda = \frac{b-d}{b}$

$$\lim_{t\to\infty} Var\left(\frac{I_{m_t}-I_{p_t}}{I_{m_t}+I_{p_t}} \mid \text{survival}\right)$$

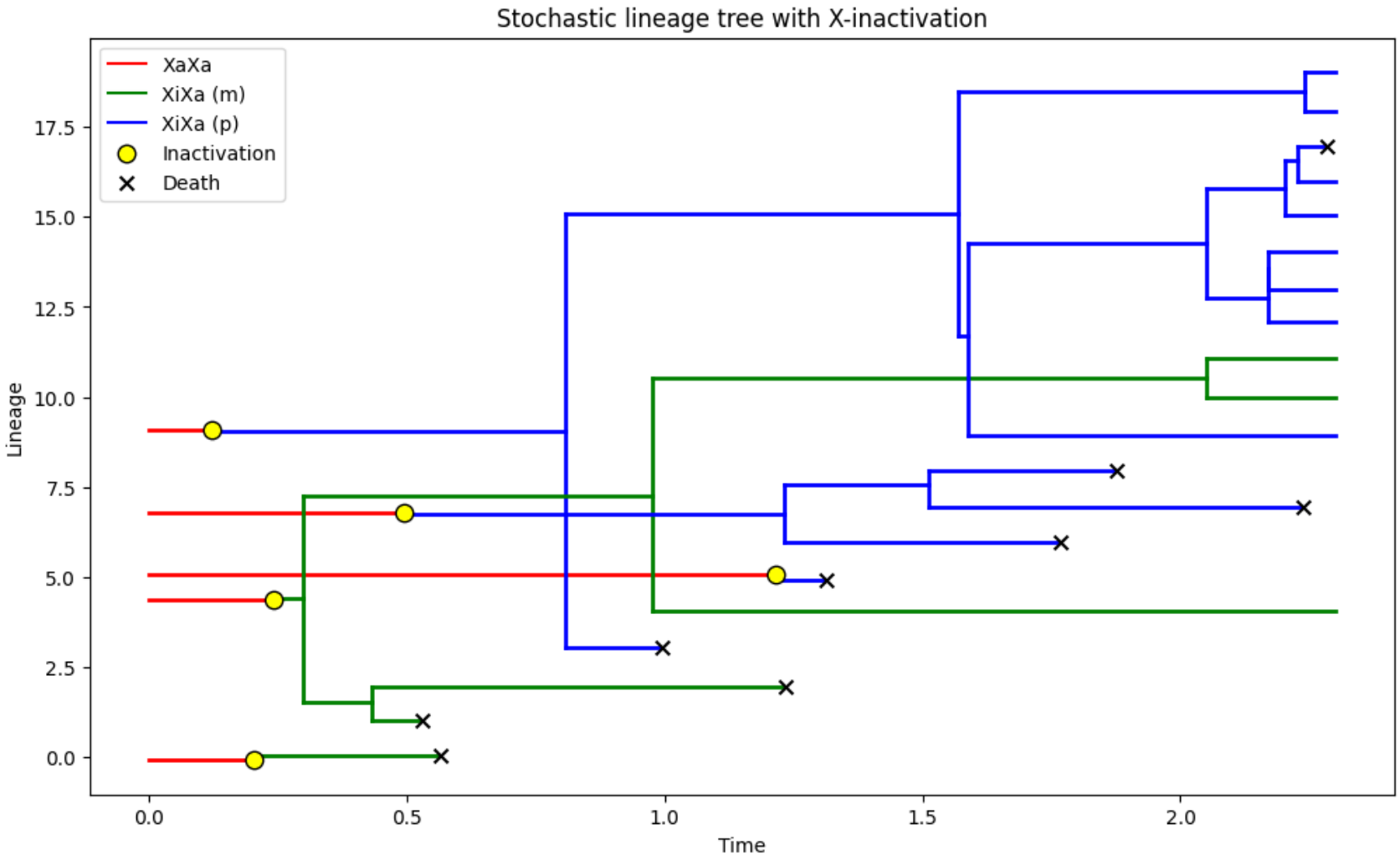
$$= \frac{1}{1-(1-\lambda)^N}\sum_{n=1}^{N}\binom{N}{n}\lambda^n(1-\lambda)^{N-n}\frac{1}{2^n}\left(\lambda^n\sum_{k=1}^{n-1}\frac{\binom{n}{k}}{(k-1)!(n-k-1)!}\int_0^{+\infty}\int_0^{+\infty}\left(\frac{y-x}{x+y}\right)^2 x^{k-1}y^{n-k-1}e^{-\lambda(x+y)}dxdy + 2\right)$$

$$= \frac{1}{1-(1-\lambda)^N}\sum_{n=1}^{N}\binom{N}{n}\lambda^n(1-\lambda)^{N-n}\frac{1}{2^n}\left(2+\sum_{k=1}^{n-1}\binom{n}{k}\left(1-\frac{4k(n-k)}{n(n+1)}\right)\right)$$

Instantaneous inactivation and stochastic growth



Asynchronous inactivation, without XaXa birth and death during inactivation



Asynchronous inactivation, without XaXa birth and death during inactivation

$N := N_0 = N_{\text{clones}}$

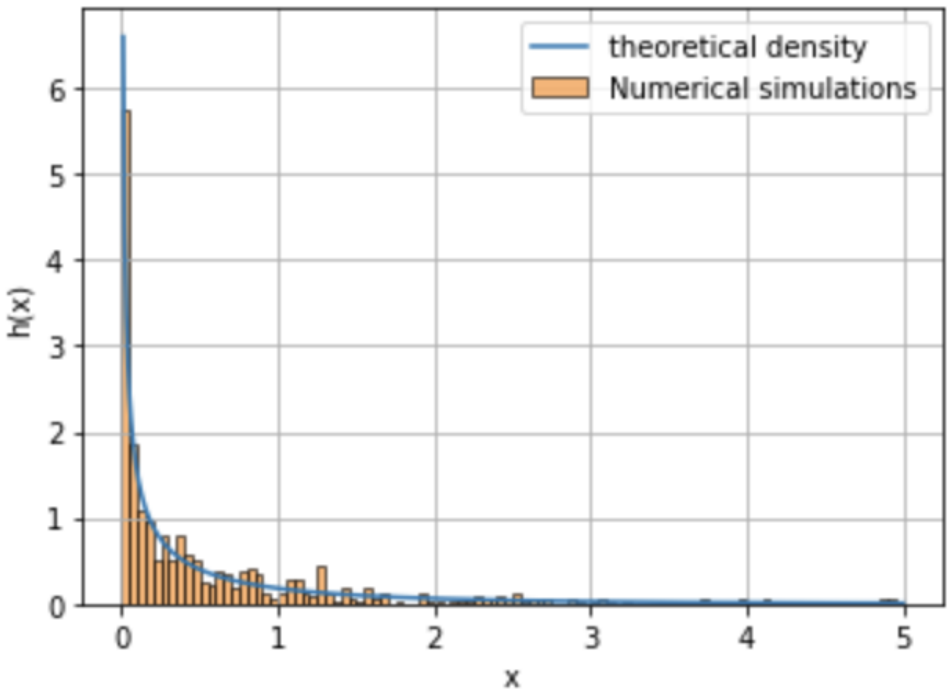
$A_i(t)e^{-(b-d)t} \underset{t\to+\infty}{\longrightarrow} \chi_i\psi_i e^{-(b-d)T_i}$ a.s., with $T_i \sim \mathcal{E}xp(\mu)$ i.i.d, $\chi_i \sim \mathcal{B}(\lambda)$, $\psi_i \sim \mathcal{E}xp(\lambda)$, and all the variables χ_i, ψ_j, T_k mutually independent.

$X_i = \psi_i e^{-(b-d)T_i}$ **(limiting rescaled clone size conditional on survival)**

X_i density function :
$h(x) = \beta\lambda^\beta x^{\beta-1}\int_{u=\lambda x}^{\infty} e^{-u}u^{-\beta}du = \beta\lambda^\beta x^{\beta-1}\Gamma(1-\beta,\lambda x)$

with $\lambda = \frac{b-d}{b}$ and $\beta = \frac{\mu}{b-d}$

Asynchronous inactivation, without XaXa birth and death during inactivation



Density of $X = \psi_i e^{-(b-d)T_i} = \lim_{t\to\infty} A_i(t)e^{-(b-d)t}$

Asynchronous inactivation, without XaXa birth and death during inactivation

Laplace transform :

$$E(e^{-sX_i}) = \int_0^\infty e^{-sx} h(x) dx = \beta\lambda^\beta \int_{x=0}^\infty e^{-sx} x^{\beta-1} \left(\int_{u=\lambda x}^\infty e^{-u} u^{-\beta}\, du \right) dx$$

$$= \frac{\lambda}{s} \ln\left(1 + \frac{s}{\lambda}\right) \text{ if } \beta = \frac{\mu}{b-d} = 1$$

$$= \frac{2\lambda}{s} \left(1 - \frac{\lambda}{s} \ln\left(1 + \frac{s}{\lambda}\right)\right) \text{ if } \beta = \frac{\mu}{b-d} = 2$$

$$= -\beta! \sum_{i=0}^{\beta-1} \frac{1}{(\beta-i-1)!} \left(-\frac{\lambda}{s}\right)^{i+1} \left(\frac{\left(1+\frac{s}{\lambda}\right)^i}{i!} \left(\ln\left(1 + \frac{s}{\lambda}\right) - H_i \right) + \sum_{j=1}^{i} \left(\frac{s}{\lambda}\right)^{i-j} \frac{H_j}{j!(i-j)!} \right)$$

if β integer ≥ 1, where $H_j = \sum_{k=1}^{j} \frac{1}{k}$ is the j-th harmonic number.

If $S_n = X_1 + ... + X_n$ with $X_1, ..., X_n$ independent :

$$E(e^{-sS_n}) = (E(e^{-sX_i}))^n$$

Asynchronous inactivation, without XaXa birth and death during inactivation

$$\mathrm{Var}\left(\frac{I_{m_t}-I_{p_t}}{I_{m_t}+I_{p_t}}\,\middle|\,\text{survival}\right) \underset{t\to+\infty}{\longrightarrow} \frac{1}{1-(1-\lambda)^N}\sum_{n=1}^{N}\frac{1}{2^n}\binom{N}{n}\lambda^n(1-\lambda)^{N-n}\left(2+\sum_{k=1}^{n-1}\binom{n}{k}\int_0^{+\infty}\int_0^{+\infty}\left(\frac{x-y}{x+y}\right)^2 f_k(x)f_{n-k}(y)\,dx\,dy\right)$$

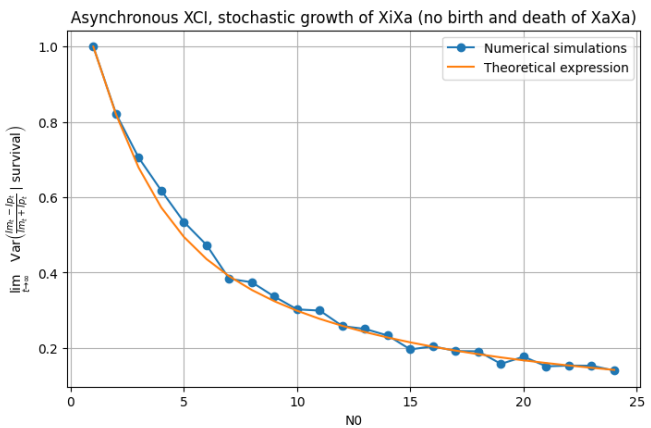
with f_n the inverse Laplace transform of $s \mapsto \frac{\lambda^n}{s^n}\left(\ln\left(1+\frac{s}{\lambda}\right)\right)^n$ if $\beta = \frac{\mu}{b-d} = 1$

with f_n the inverse Laplace transform of $s \mapsto \frac{2^n\lambda^n}{s^n}\left(1-\frac{\lambda}{s}\ln\left(1+\frac{s}{\lambda}\right)\right)^n$ if $\beta = \frac{\mu}{b-d} = 2$

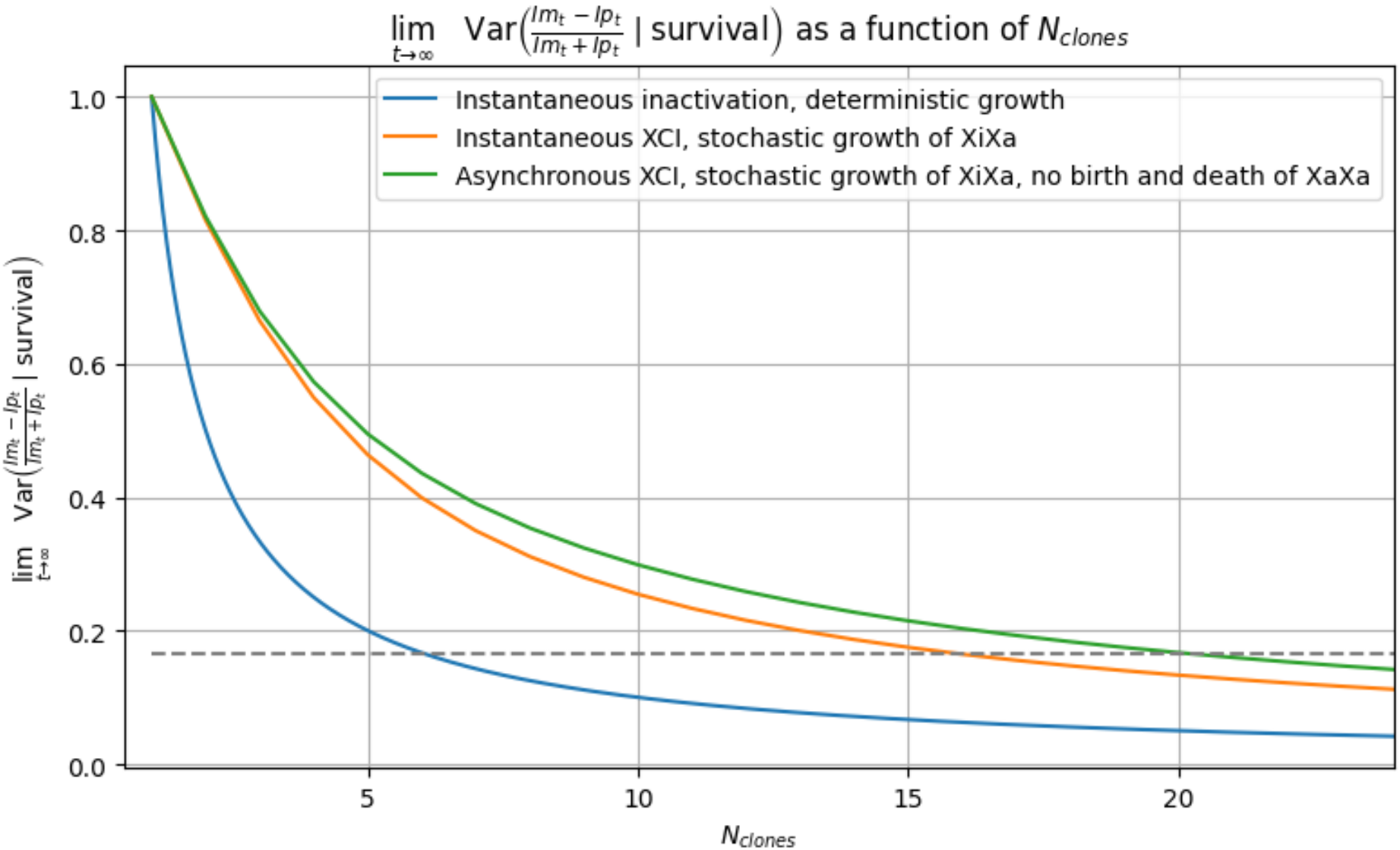
with f_n the inverse Laplace transform of

$$s \mapsto \left(-\beta!\sum_{i=0}^{\beta-1}\frac{1}{(\beta-i-1)!}\left(-\frac{\lambda}{s}\right)^{i+1}\left(\frac{\left(1+\frac{s}{\lambda}\right)^i}{i!}\left(\ln\left(1+\frac{s}{\lambda}\right)-H_i\right)+\sum_{j=1}^{i}\left(\frac{s}{\lambda}\right)^{i-j}\frac{H_j}{j!(i-j)!}\right)\right)^n \text{ if } \beta \text{ integer} \geq 1,$$

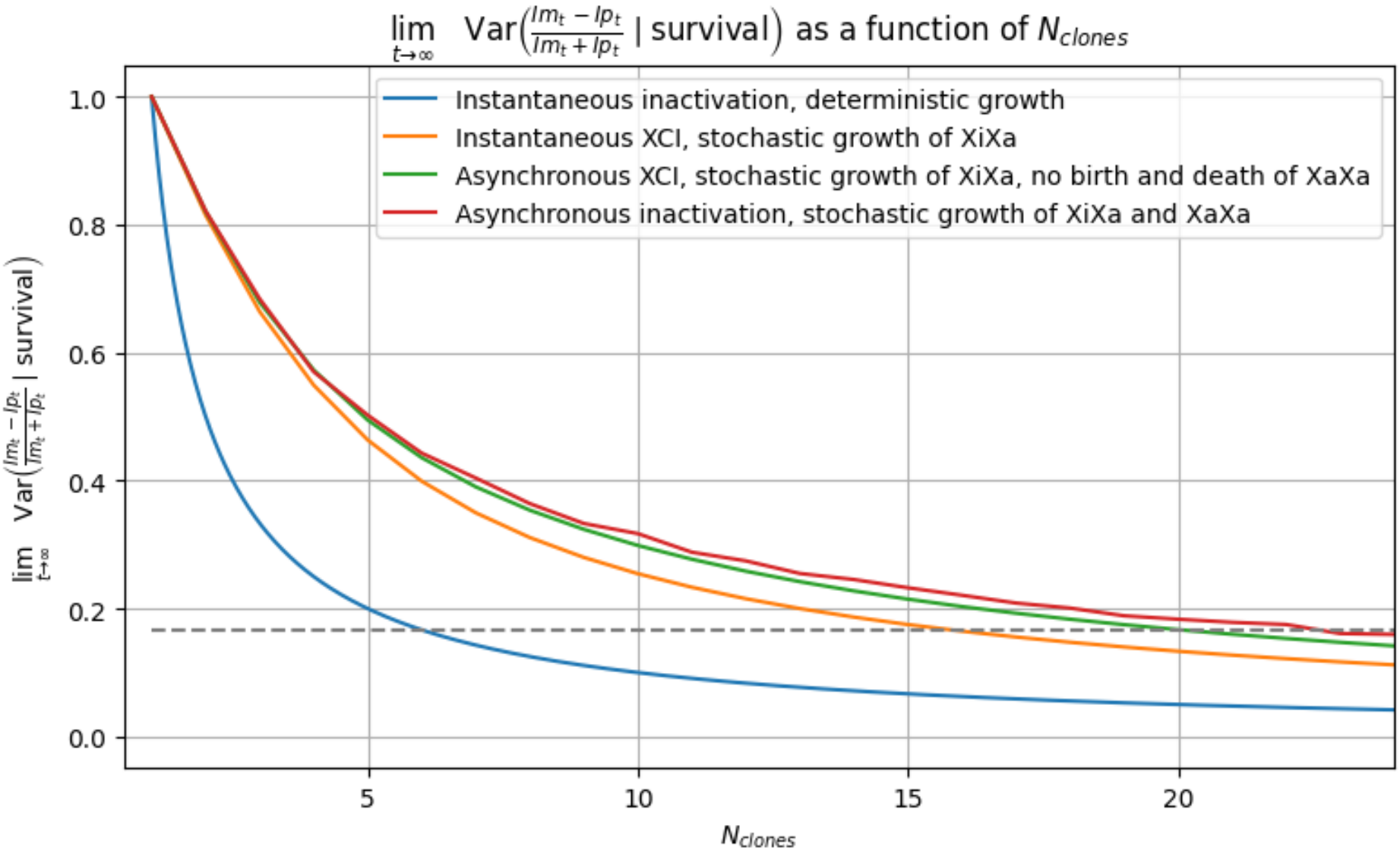
where $H_j = \sum_{k=1}^{j}\frac{1}{k}$ is the j-th harmonic number.



Asynchronous inactivation, without XaXa birth and death during inactivation



Asynchronous inactivation, with growth of the XaXa cell population during inactivation



Conclusion and Perspectives

Future extension to this work

- Gamma distributed lifetimes
- Incorporating successive samplings

Other ongoing related work

- Estimation of model parameters from experimental mouse data :
  - initial population size N_0,
  - cell proliferation rates (b_i, d_i),
  - XCI rate (μ_i).

Published related work

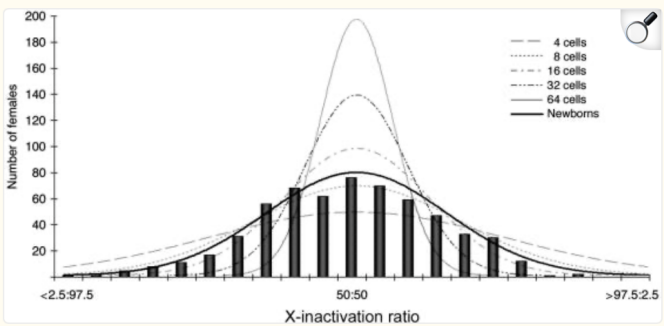
- Molecular mechanisms of XCI (gene regulatory network, establishment and maintenance of XiXa) : deterministic ODE-based modeling.

  Published in SIAM Journal on Applied Dynamical Systems (2026) : *Bifurcation Analysis for Gene Regulatory Networks Embedding a Toggle-Switch Model : Application to X Chromosome Inactivation*.

Existing literature (Shvetsova 2019, Amos 2006)

Previous studies determined the number of hematopoietic stem cells present at the time of X-inactivation choice by use of the variance of the X inactivation–pattern distribution.[26] To estimate the primordial stem-cell pool size (N), the formula for the variance of a binomial distribution was used ($variance=pq/N$) and setting the probability that either allele (p or q) is active to 0.5. Normal distributions were drawn by determining the variance (σ^2) for each of the values of N and drawing the curve with use of the equation for the normal distribution (fig. 3). To limit the potential effect that somatic mutations and environmental factors might impose on hematopoietic stem cells, we performed these calculations using only the 590 newborn samples. With the SD of 15.3%, the number of stem cells is predicted to be between 8 and 16, which is consistent with previous estimates[26,27,42] (fig. 3).

Figure 3. .



Open in a new tab

Comparison of the newborn distribution with predicted normal distributions of varying stem-cell pool size. The solid black curve follows a normal distribution with use of the actual SD of the newborn samples (15.4 [fig. 2]). The estimated number of precursor stem cells predicted by the newborn SD is between 10 and 12 cells.

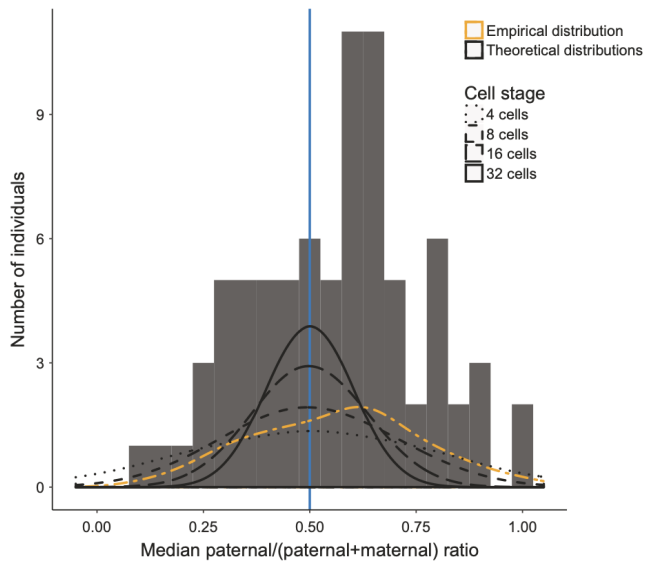


Fig. 3 Theoretical assessment of cell numbers at the time of X-inactivation. Comparison of the empirical median paternal ratio distribution for heterozygous SNVs with more than ten reads per individual (orange line) with theoretical distributions under the hypothesis that X-inactivation takes place at the 4 (dotted black line), 8 (dashed black line), 16 (long dashed black line), and 32 (solid black line) precursor stage. Theoretical distribution at eight initial lineage-restricted precursor cells is most comparable with empirical distribution (highest p value = 0.011, two-sample Kolmogorov–Smirnov test)

Existing literature with sampling (Kane 1979, Nesbit 1971)

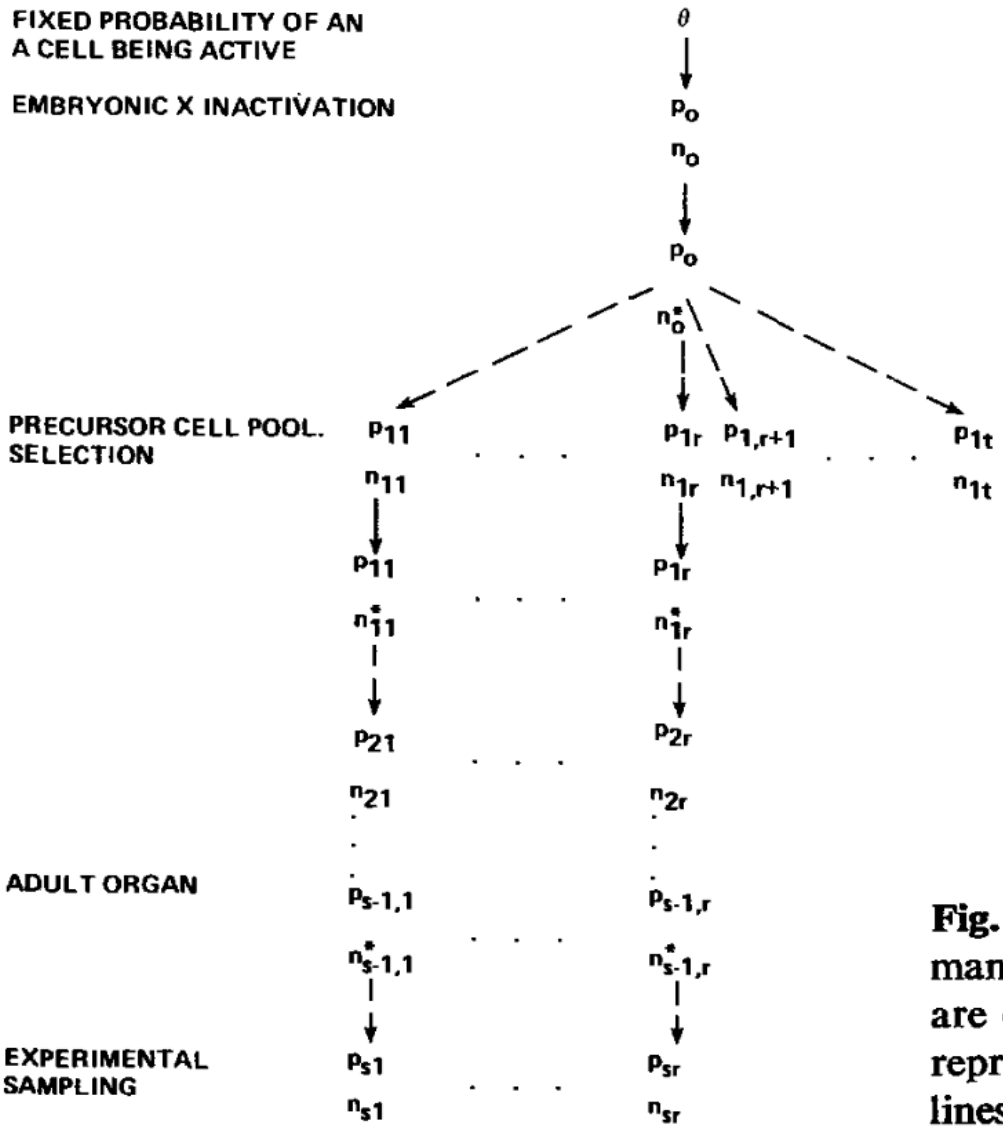


Fig. 1. Illustration of an s stage model for mammalian development where r organs are experimentally sampled. Dashed lines represent a sampling event and continuous lines represent growth

An alternative model based on stochastic growth (Ropers 1977)

- Initial condition :
  - deterministic XCI
  - perfectly balanced population

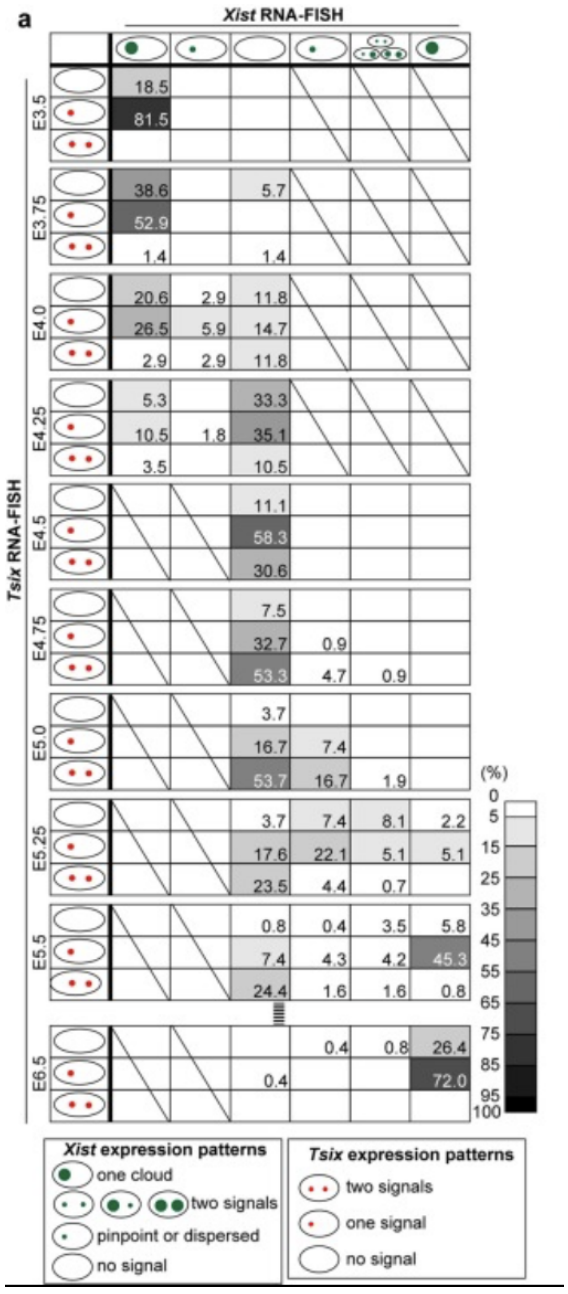
$$50\%\ XiXa(m) \qquad 50\%\ XiXa(p)$$

- Source of variability : stochastic cell proliferation (variability generated *after* XCI)
- Mathematical formulation :
  Pólya urn process : division $\leftrightarrow$ addition of a ball of the same color

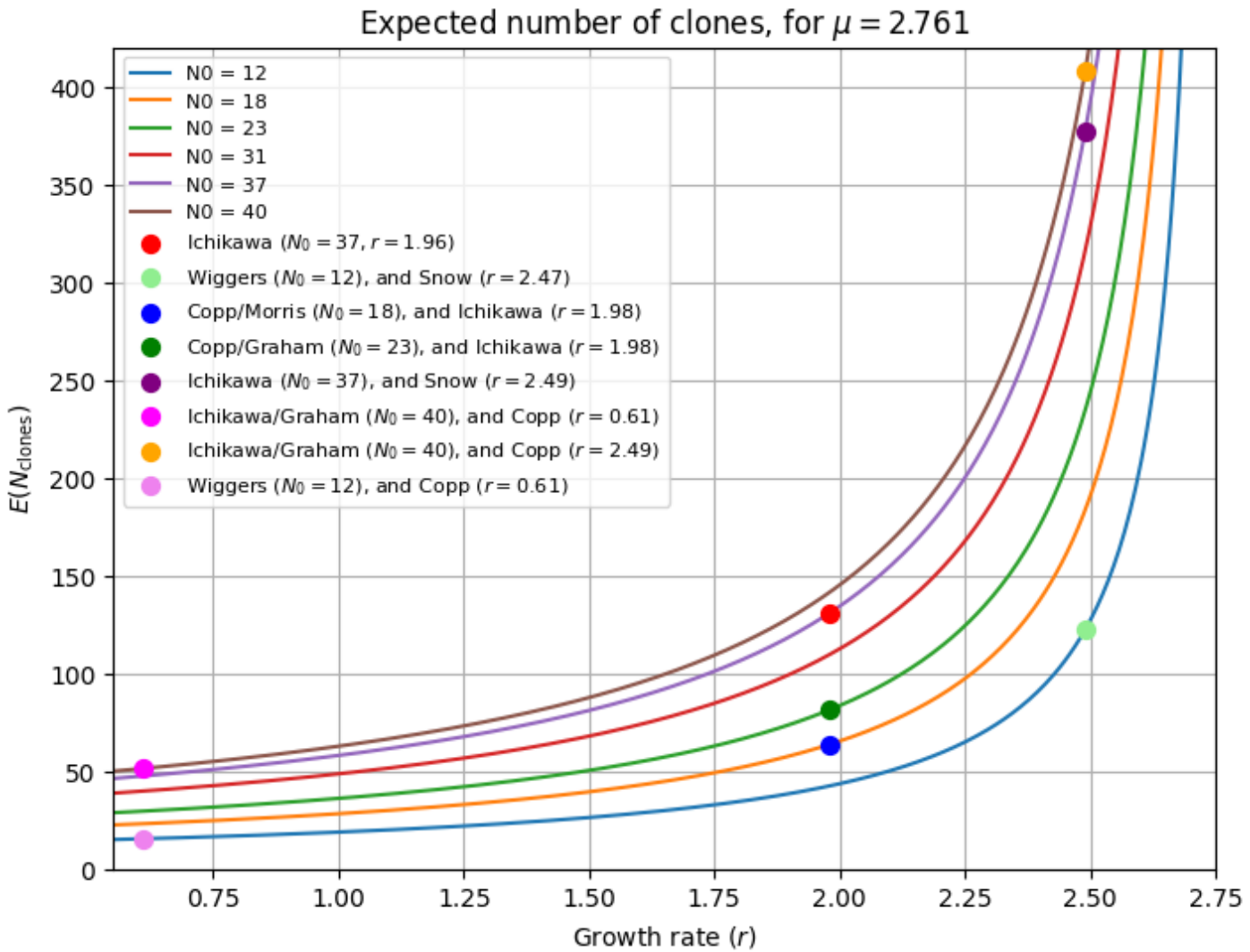
Remark

Stochastic growth model with cell proliferation only (no cell death).

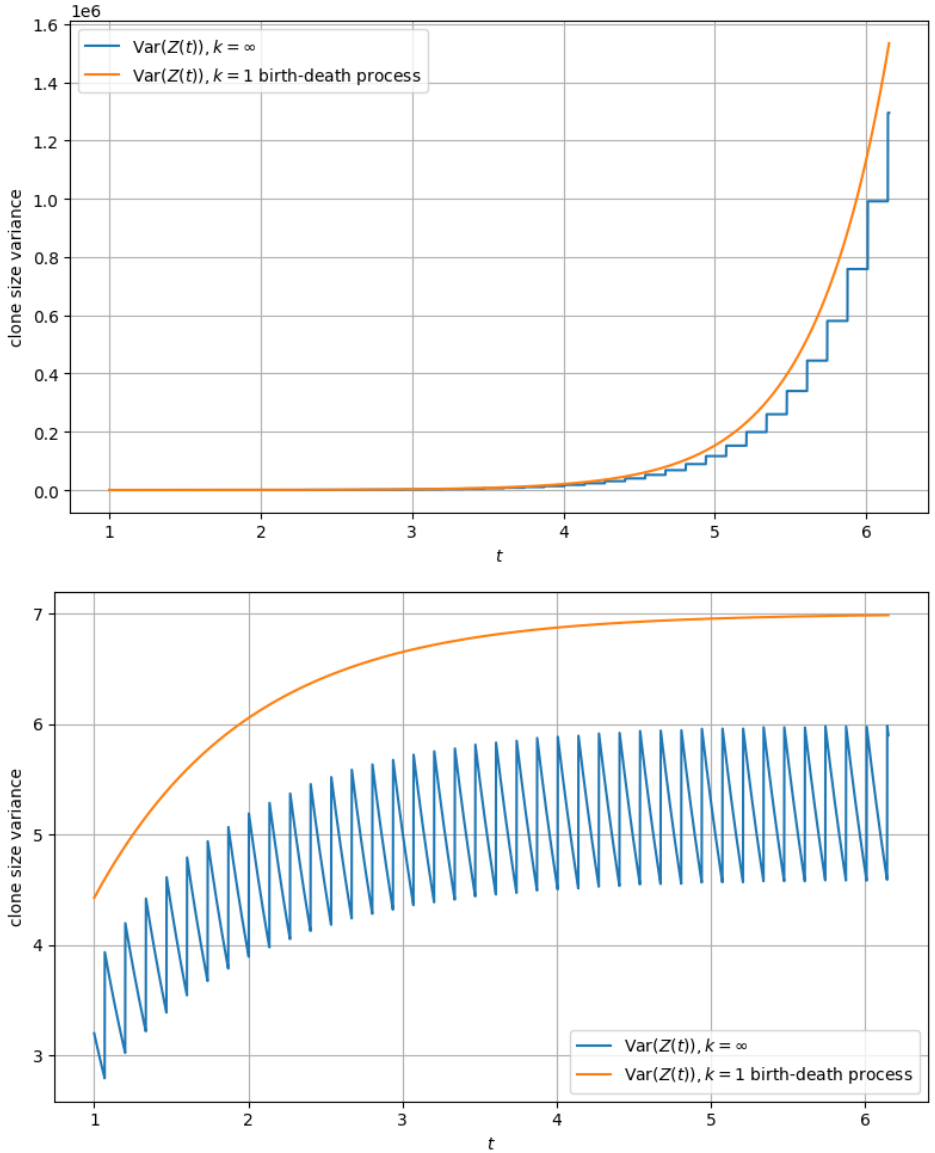
Data


a
Xist RNA-FISH
Tsix RNA-FISH
E3.5
E3.75
E4.0
E4.25
E4.5
E4.75
E5.0
E5.25
E5.5
E6.5
(%)
0
5
15
25
35
45
55
65
75
85
95
100
Xist expression patterns
one cloud
two signals
pinpoint or dispersed
no signal
Tsix expression patterns
two signals
one signal
no signal

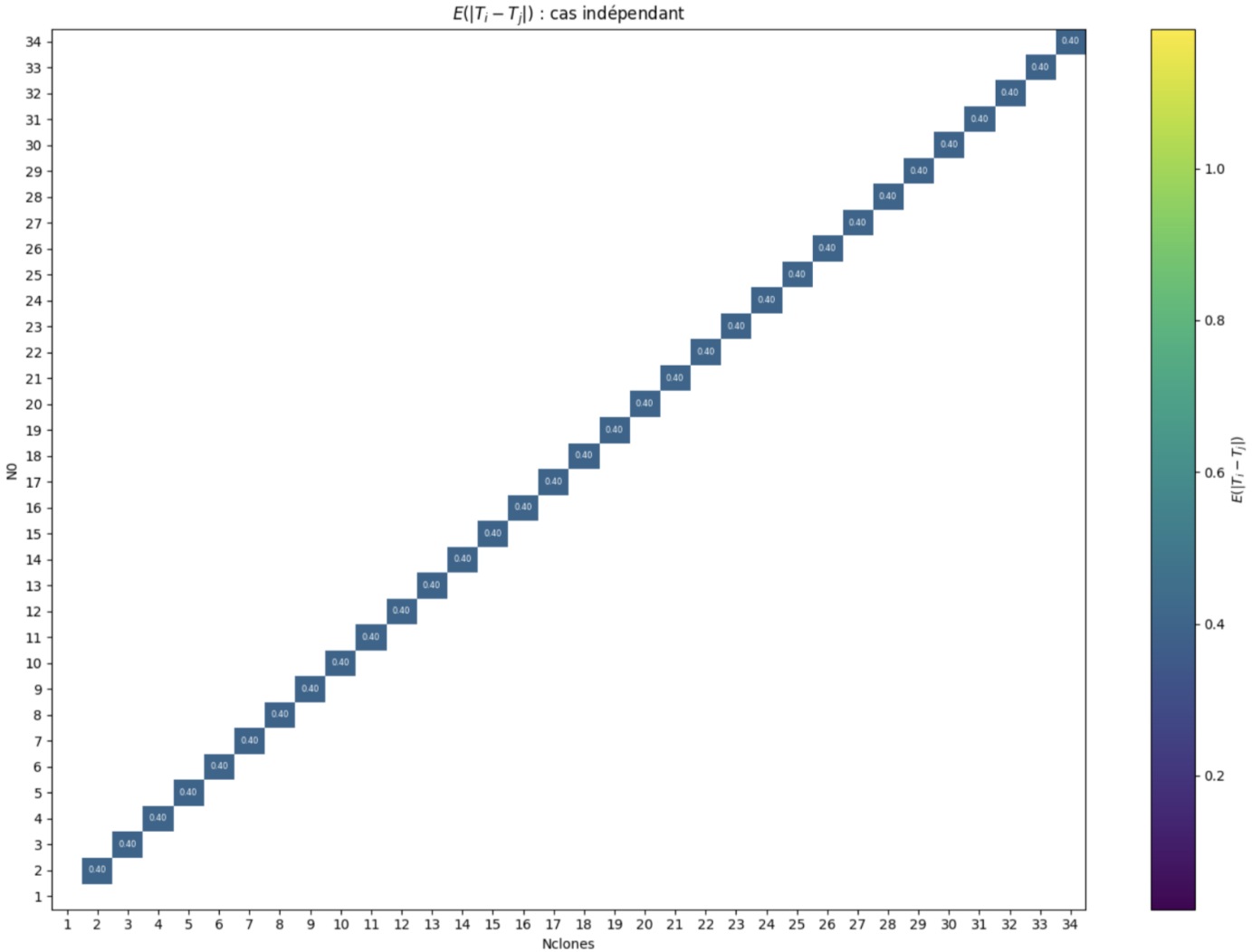
N_{clones}



Gamma distributed lifetimes and Galton-Watson processes



$E(|T_i - T_j|)$



$E(|T_i - T_j|)$