Lineage-tracking of stem cell differentiation: a neutral model of hematopoiesis in rhesus macaque

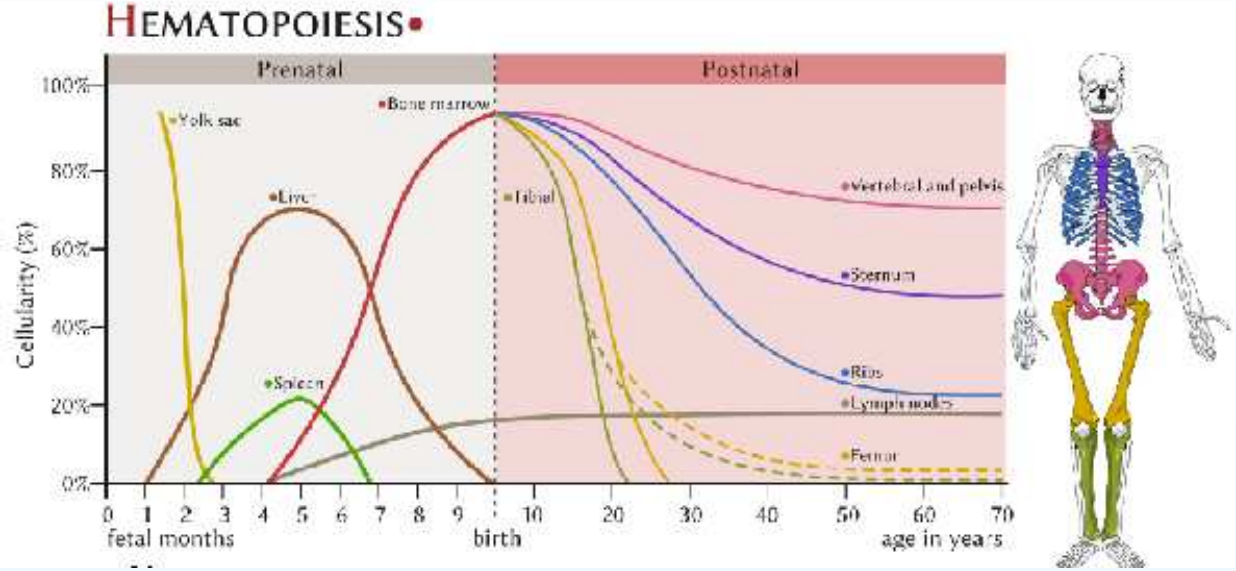
Tom Chou

UCLA

Depts. of Biomathematics & Mathematics

S. Goyal (UToronto), S. Kim (UCLA), I. S. Y. Chen (UCLA)

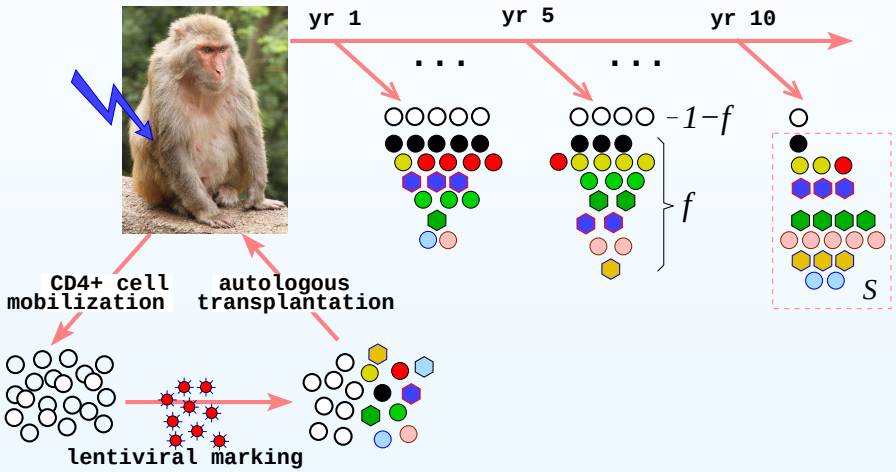
Blood generation & maintenance: some numbers

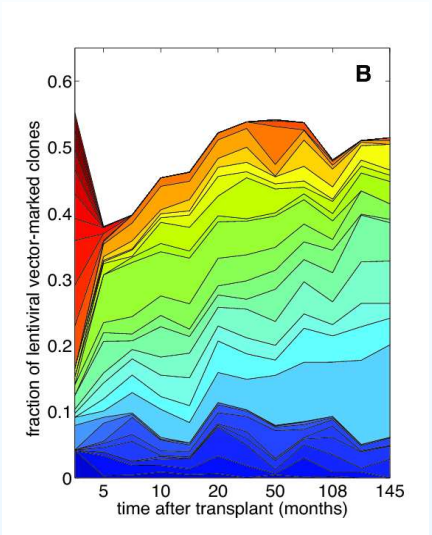


- $\sim 10^{11}$ new blood cells are generated daily
- HSCs are multipotent, producing first progenitors, then differentiated RBCs (10^{13}), platelets (10^{12}), and WBCs ($10^{10} - 10^{11}$)
- Number of HSCs unknown but $\sim 10^2 - 10^5$?
- When/are HSCs committed to lymphoid/myeloid?

Clonal repopulation studies

marking/barcoding:

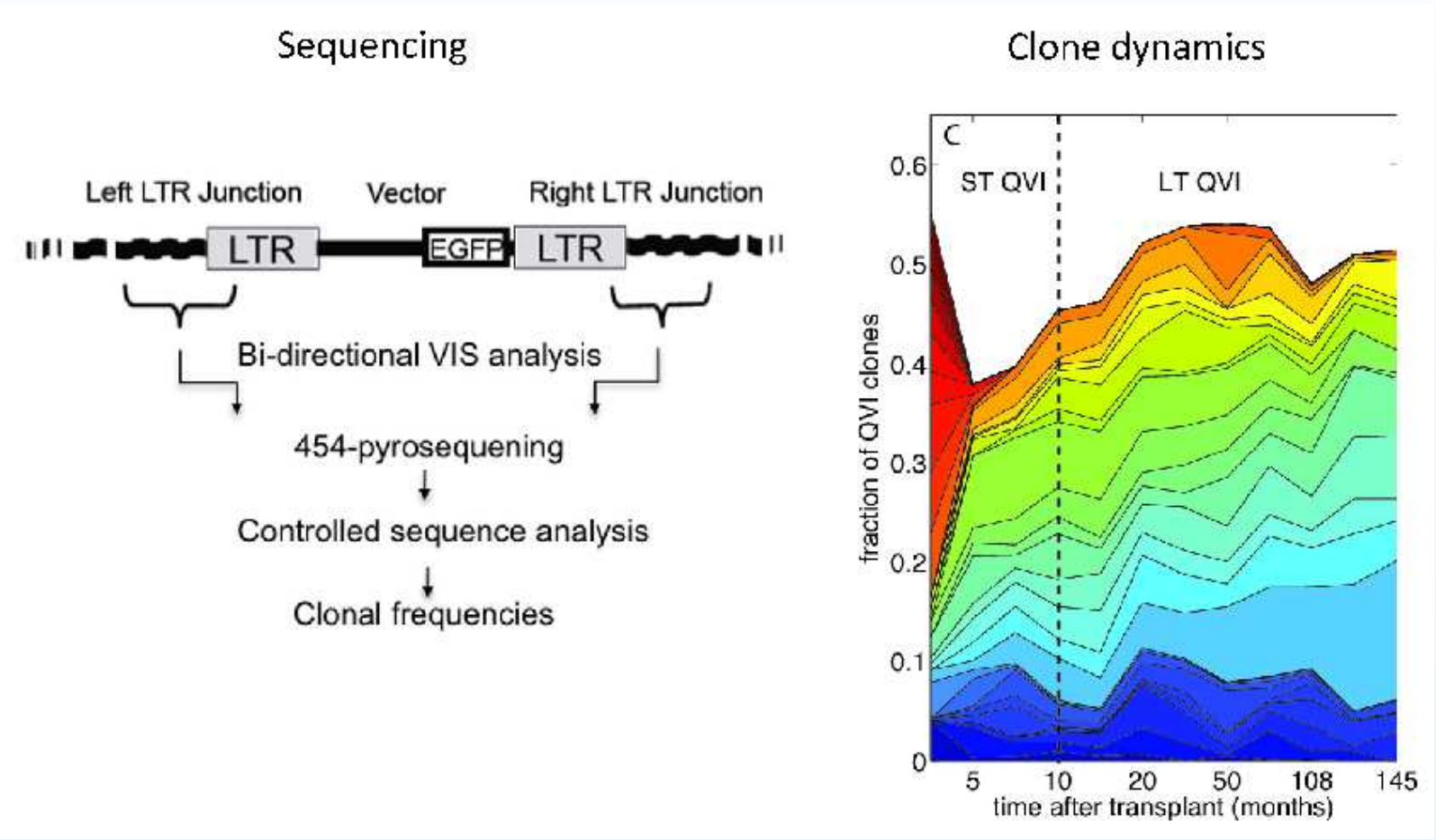




- isolate CD34+ cells from rhesus macaques
- irradiate animal
- mark (with lentivirus) extracted CD34+ cells
- transplant labeled cells into animal
- collect blood samples and count VIS

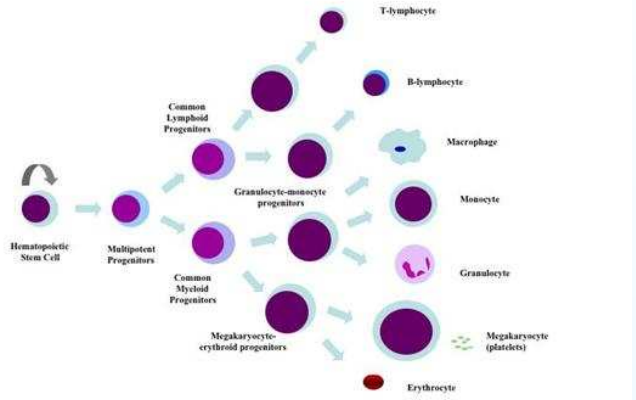
S. Kim *et al.* Cell Stem Cell, **14**, 473, (2014)

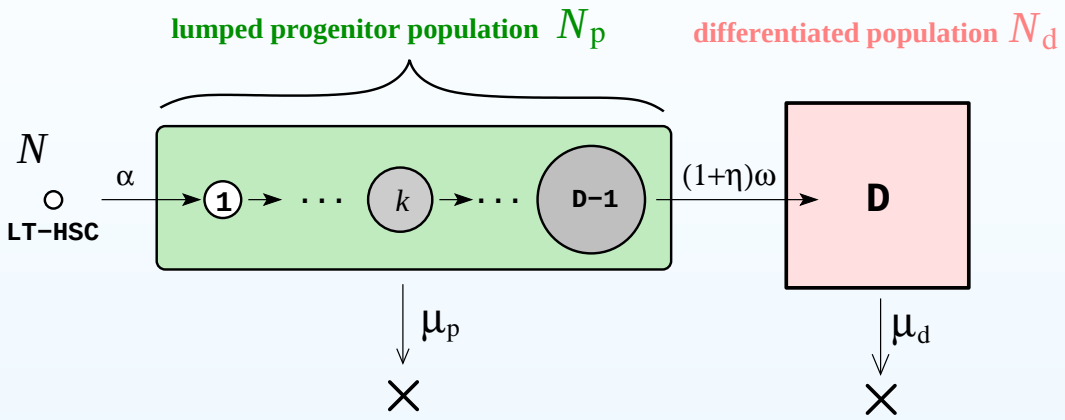
Tracking clone dynamics



smaller clones not shown

Simple lumped models





$\alpha \equiv$ asymmetric differentiation rate, $\omega \equiv$ terminal diff. rate

Progenitor and differentiated populations:

$$\frac{\mathrm{d}N_\mathrm{p}(t)}{\mathrm{d}t} = \alpha N_\mathrm{HSC} + \underbrace{r(N_\mathrm{p})}_{\frac{pK}{N_\mathrm{p}+K}} N_\mathrm{p} - \underbrace{(\mu_\mathrm{p} + \eta\omega)}_{\mu} N_\mathrm{p},$$

$$\frac{\mathrm{d}N_\mathrm{d}(t)}{\mathrm{d}t} = (1+\eta)\omega N_\mathrm{p} - \mu_\mathrm{d} N_\mathrm{d}$$

η: prob. of symmetric terminal diff., $1-\eta$: prob. of asymmetric diff.

Simple model for clones

Keep track of numbers of cells of each lineage i (clones)

$$\frac{\mathrm{d}n_i(t)}{\mathrm{d}t} = \alpha n_i^{(0)} + \underbrace{r(N_\mathrm{p})}_{\frac{pK}{N_\mathrm{p}+K}} n_i - \underbrace{(\mu_\mathrm{p} + \eta\omega)}_{\mu} n_i,$$

$$\frac{\mathrm{d}n_i^{(d)}(t)}{\mathrm{d}t} = (1+\eta)\omega n_i - \mu_\mathrm{d} n_i^{(d)}$$

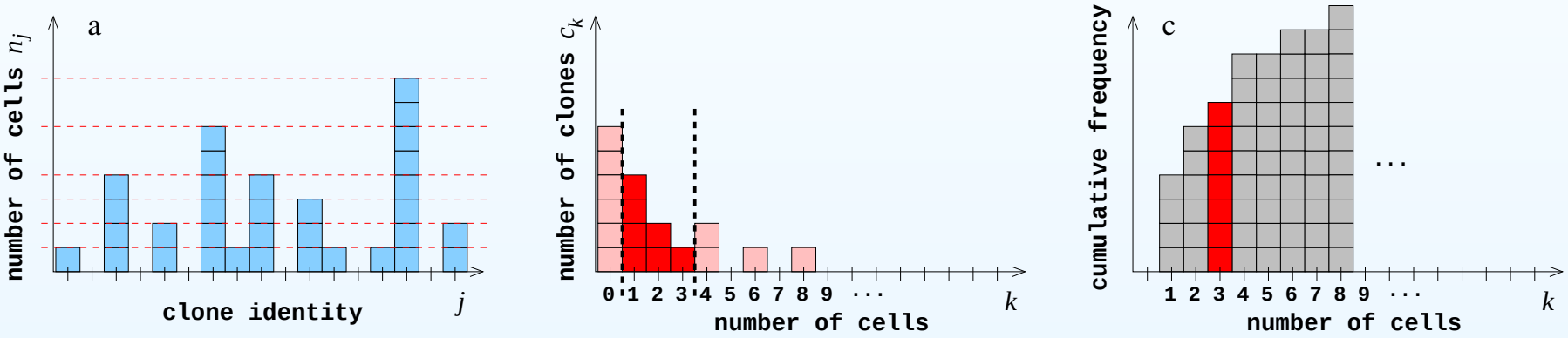
$n_i^{(0)}$, n_i, and $n_i^{(d)}$: HSCs, progenitors, and differentiated cells

These equations describe mean numbers and are identical for all i.

All populations equivalent: $n_i(t) = n_j(t)$

Clonal kinetics

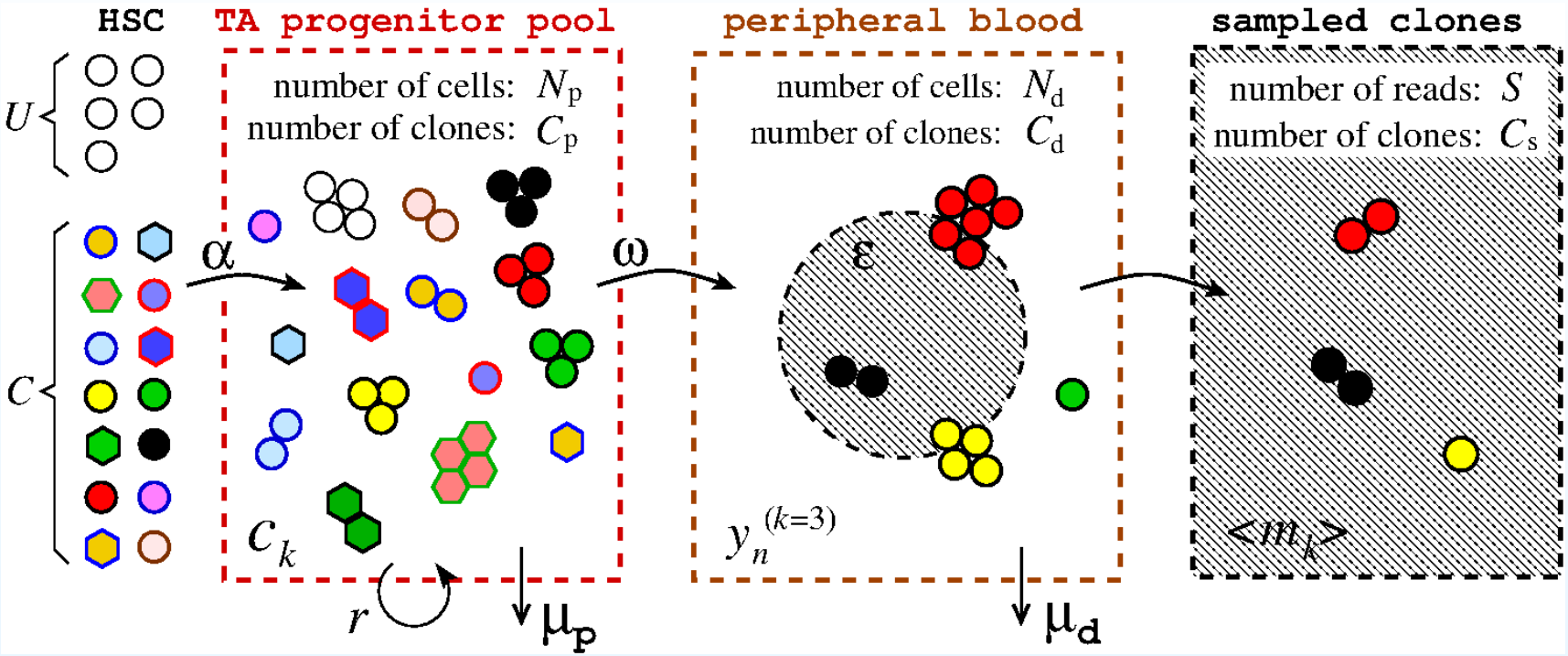
To model variations in clone sizes, look directly at “density of states:”



Instead of tracking numbers of cells $n_i(t)$, consider $c_k(t)$, the number of distinct *clones* that are represented by exactly k cells at time t.

$$c_k(t) = \sum_{i=1}^{C} \delta(n_i(t) - k)$$

Clone size distributions: three-pool model



$C \equiv$ number of labeled HSCs; $U \equiv$ number of unlabeled HSCs

S. Goyal, S. Kim, I. Chen, T. Chou, BMC Biology, **13**, (2015)

Clones in progenitor or “Transit amplifying” pool

$u(t)$: number of unlabeled progenitor cells
$c_k(t)$: number of distinct clones represented by k cells

$$\frac{\mathrm{d}u}{\mathrm{d}t} = ru - \mu u + \alpha U, \quad \mu \equiv \mu_{\mathrm{p}} + \eta\omega$$

$$\frac{\mathrm{d}c_1}{\mathrm{d}t} = -\alpha c_1 + \alpha c_0 - (r+\mu)c_1 + 2\mu c_2$$

$$\frac{\mathrm{d}c_k}{\mathrm{d}t} = \underbrace{\alpha(c_{k-1} - c_k)}_{\text{HSC asym. differentiation}} + \underbrace{r((k-1)c_{k-1} - kc_k)}_{\text{progenitor birth}} + \underbrace{\mu((k+1)c_{k+1} - kc_k)}_{\text{progenitor death}}$$

lost (unrepresented) clones $c_0 = C - \sum_{\ell=1}^{N_{\mathrm{p}}} c_\ell$

Clones in differentiated blood pool

Experiments actually sampled from *differentiated* blood.

$y_n^{(k)}(t) \equiv$ number of clones in peripheral blood represented by n copies *and* that originated from progenitor clones of size k:

$$\frac{\mathrm{d}y_n^{(k)}}{\mathrm{d}t} = (1+\eta)(\omega k)(y_{n-1}^{(k)} - y_n^{(k)}) + (n+1)\mu_{\mathrm{d}} y_{n+1}^{(k)} - n\mu_{\mathrm{d}} y_n^{(k)},$$

$y_0^{(k)} = c_k - \sum_{\ell=1}^{\infty} y_\ell^{(k)}$: no. of clones of size k that are not contributing

$y_n = \sum_{k=1}^{\infty} y_n^{(k)}$: no. of clones in peripheral blood of size n

Sampled blood pool

- $S \ll N_{\mathrm{d}}$: cells drawn from the animal *and* successfully sequenced.
- $s_{j\ell}$: number of cells sampled from the j^{th} clone among only those that have population ℓ.
- Mean number of clones of size k: $m_k = \sum_{\ell=1}^{N_{\mathrm{d}}} \sum_{j=1}^{y_\ell} \delta(s_{j\ell} - k)$,
- At any time, probability $s_{j\ell}$ is observed:

$$P(\{s\}; S) = \prod_{\ell=1}^{N} \prod_{j=1}^{y_\ell} \binom{\ell}{s_{j\ell}} \frac{S!}{N^{s_{j\ell}}} \delta \left(\sum_{\ell=1}^{N} \sum_{j=1}^{y_\ell} s_{j\ell} - S \right),$$

$$\langle m_k(t) \rangle = \sum_{\{s\}} P(\{s\}, t; S) \sum_{n=1}^{N} \sum_{m=1}^{y_n(t)} \delta(s_{mn}(t) - k),$$

Sampling - generating function calculation

$$G(z;S) = \sum_{k=0}^{\infty} \langle m_k \rangle z^k = \frac{\partial}{\partial \beta} \sum_{\{s\}} P(\{s\}) \exp\left[\beta \sum_{\ell=0}^{N} \sum_{j=1}^{y_\ell} z^{s_{j\ell}}\right]$$

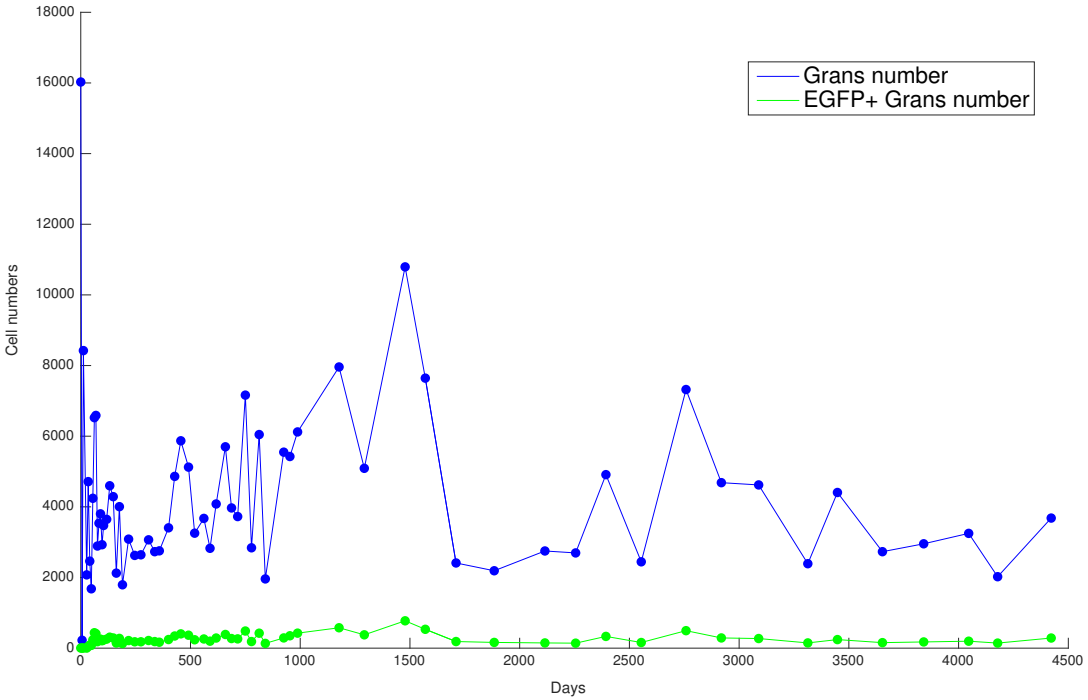
$$= \frac{\partial}{\partial \beta} \sum_{\{s\}} \prod_{\ell=0}^{N} \prod_{j=1}^{y_\ell} \binom{\ell}{s_{j\ell}} \frac{S!}{N^{s_{j\ell}}} \delta\left(\sum_{n=1}^{N} \sum_{m=1}^{y_n} s_{mn} - S\right) \exp\left[\beta \sum_{\ell=0}^{N} \sum_{j=1}^{y_\ell} z^{s_{j\ell}}\right]\Bigg|_{\beta=0}$$

$$\vdots$$

$$= \frac{\partial}{\partial \beta} \int_0^{2\pi} \frac{\mathrm{d}q}{2\pi} S! e^{-iqS} \prod_{\ell=0}^{N} \prod_{j=1}^{y_\ell} \left[\sum_{s=0}^{\ell} \binom{\ell}{s} \frac{e^{iqs}}{N^s} e^{\beta z^s}\right]\Bigg|_{\beta=0} = \sum_{\ell=0} y_\ell e^{\ell\varepsilon(z-1)}$$

$$\langle m_k(t) \rangle = \sum_{\ell=0}^{\infty} \left(\frac{\ell\varepsilon}{k}\right)^k e^{-\ell\varepsilon} y_\ell(t), \quad \varepsilon \equiv \frac{S}{N_\mathrm{d}} \ll 1 \quad \text{and} \quad Q(q,t) \equiv \sum_{k=0}^{q} \langle m_k(t) \rangle$$

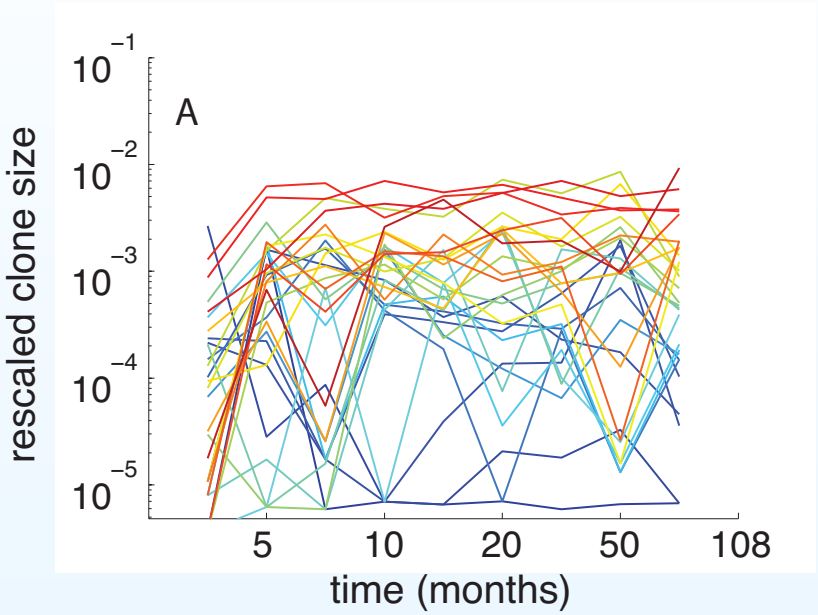
Noisy data – normalization?

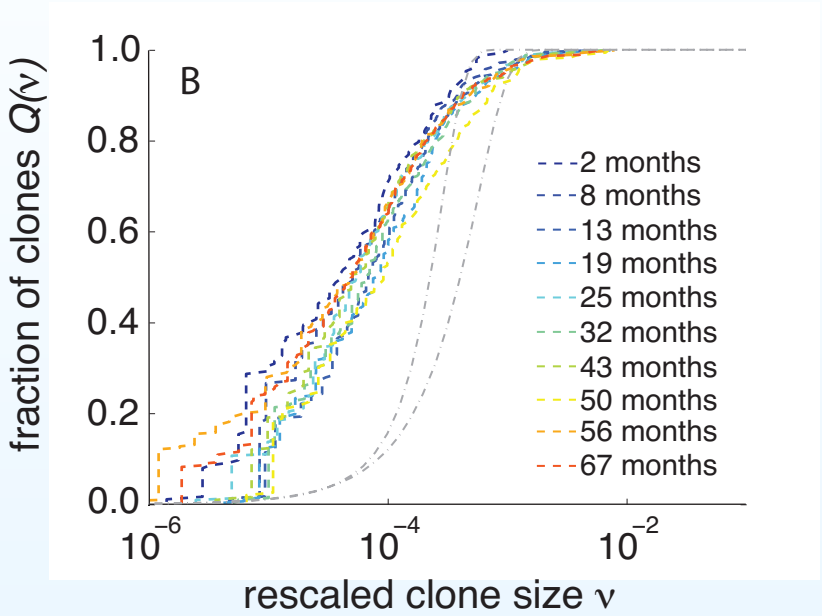


consider relative clone frequencies

$$\bar{Q}(\bar{q}) = \frac{Q(\bar{q}S) - Q(0)}{Q(S) - Q(0)}, \quad 0 < \bar{q} \leq 1.$$

Replotting Clone size distributions $\longrightarrow$ steady-state





Each clone varies in size, but normalized size-distributions fairly static

Assume steady-state: $r(N_{\mathrm{p}}^{*}) = \dfrac{2p}{\frac{\alpha}{\mu}\frac{U+C}{K}+\frac{p}{\mu}+1+\sqrt{\left(\frac{\alpha}{\mu}\frac{U+C}{K}+\frac{p}{\mu}-1\right)^{2}+\frac{4\alpha}{\mu}\frac{U+C}{K}}} \lesssim \mu$

Analytic solutions at steady-state

Define $a \equiv \alpha/r$ and $\bar{r} \equiv r/\mu = r/(\mu_{\rm p} + \eta\omega)$:

- Steady-state progenitor cell population

$$c_{k\geq 1} = \frac{c_0}{k!\mu^k} \prod_{\ell=1}^{k} (\alpha + (\ell - 1)r)$$

$$c_0 = C(1-\bar{r})^a, \text{ number of lost clones.}$$

- Steady-state differentiated cell population:

$$y_n^{(k)} = \frac{(wk)^n}{n!} e^{-wk} c_k, \quad w \equiv \frac{(1+\eta)\omega}{\mu_{\rm d}}$$

Number of observed clones in steady-state

Total expected number of clones in each compartment:

$$C_{\mathrm{p}} = \sum_{k=1}^{\infty} c_k = C\left[1-(1-\bar{r})^a\right],$$

$$C_{\mathrm{d}} = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} y_n^{(k)} = \sum_{k=1}^{\infty}(1-e^{-wk})c_k = C\left[1-\left(\frac{1-\bar{r}}{1-\bar{r}e^{-w}}\right)^a\right],$$

$$C_{\mathrm{s}} = \sum_{j=1}^{\infty}\langle m_j\rangle = \sum_{j=1}^{\infty}\sum_{\ell=1}^{\infty} e^{-\ell\varepsilon}\frac{(\ell\varepsilon)^j}{j!}y_\ell = \sum_{k=1}^{\infty}(1-e^{-wke^{-\varepsilon}})c_k$$

$$= C\left[1-\left(\frac{1-\bar{r}}{1-\bar{r}e^{-w(1-e^{-\varepsilon})}}\right)^a\right],$$

$$C \gtrsim C_{\mathrm{p}} \gtrsim C_{\mathrm{d}} > C_{\mathrm{s}} \text{ (clone losses)},$$

Controlling parameters

Exact results depend on many parameters: $\alpha, K, U, C, p, \omega, \mu_{\mathrm{p}}, \mu_{\mathrm{d}}, S$.

Approximations:

$$Q(q) \equiv \sum_{k=0}^{q} \langle m_k \rangle \approx \sum_{k=0}^{\infty} \frac{\Gamma(q+1, \varepsilon w k)}{\Gamma(q+1)} c_k \approx c_0 + \int_1^{(q+1)/(\varepsilon w)} c_k \mathrm{d}k$$

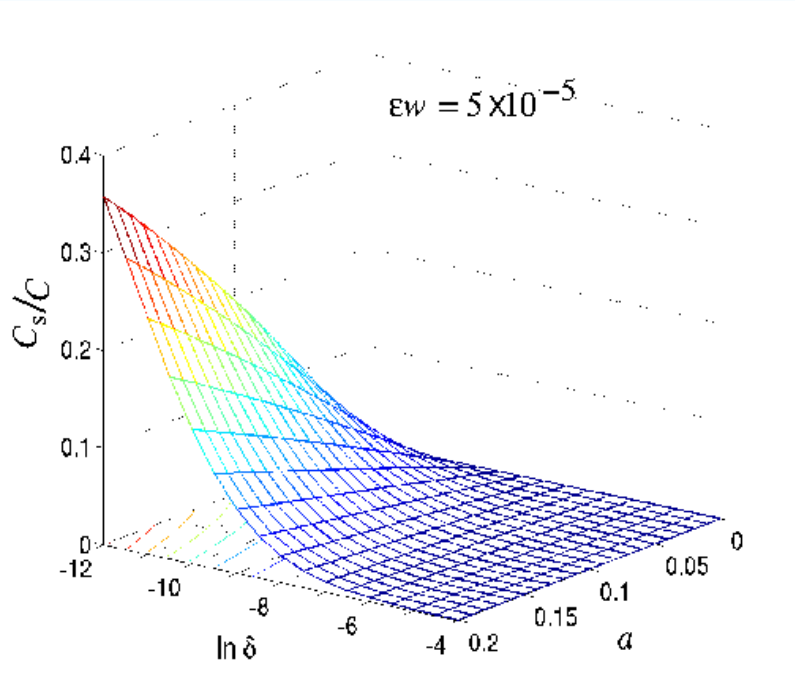
and

$$\langle m_k \rangle \approx aC \left[\Gamma\left(a, \frac{k}{R}\right) - \Gamma\left(a, \frac{k+1}{R}\right) \right].$$

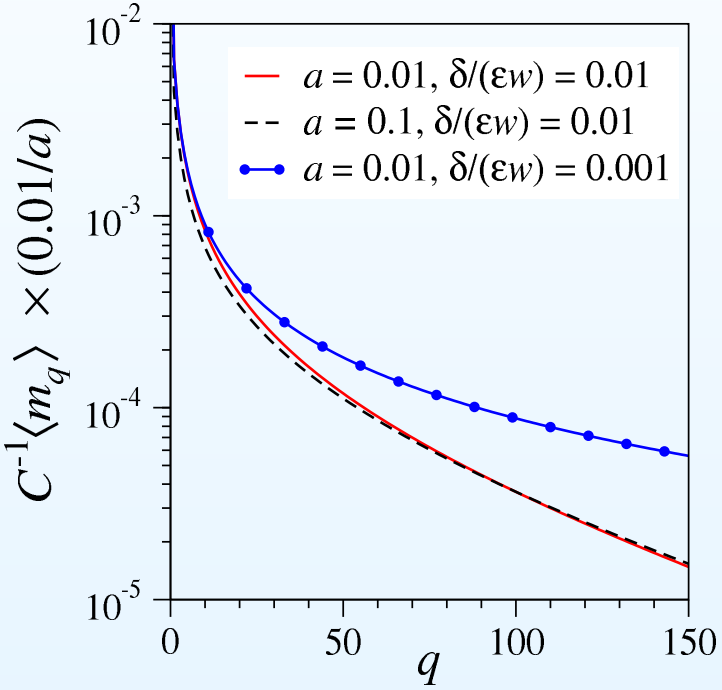
Controlling parameters: $a \equiv \frac{\alpha}{r}$ and $R \equiv \frac{\varepsilon w}{\ln(\mu/r)}$

Analytic results

Main dependence: $a \equiv \frac{\alpha}{r}$ and $R \equiv \frac{\varepsilon w}{\ln(\mu/r)} \approx \frac{\varepsilon w}{\delta} = \frac{(1+\eta)\omega S}{N_{\rm d}\mu_{\rm d}\delta}$ $(\delta = 1 - \frac{r}{\mu})$:



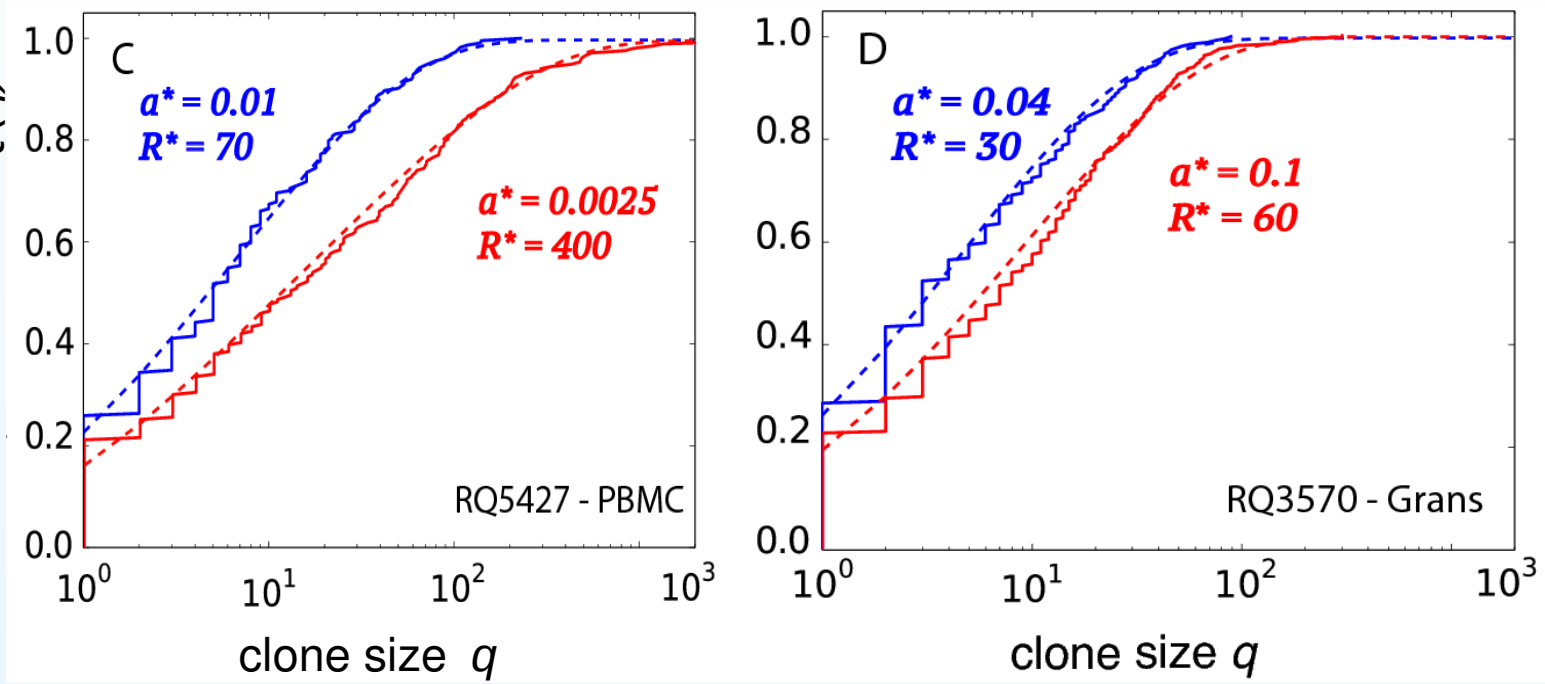
$C_{\rm s}/C$ *vs* a and $\ln\delta$



$C^{-1}\langle m_q\rangle$ *vs* a and $\delta/(\varepsilon w)$

Data fitting: Maximum Likelihood Estimation

Maximum likelihood fitting of $\langle m_q \rangle$ (plotting raw $Q(q)$):



to find MLE values $a^* \sim 0.01 - 0.1$ and $R^* \sim 50 - 500$.

Implications on parameters

Express unknowns in terms of better-known parameters:

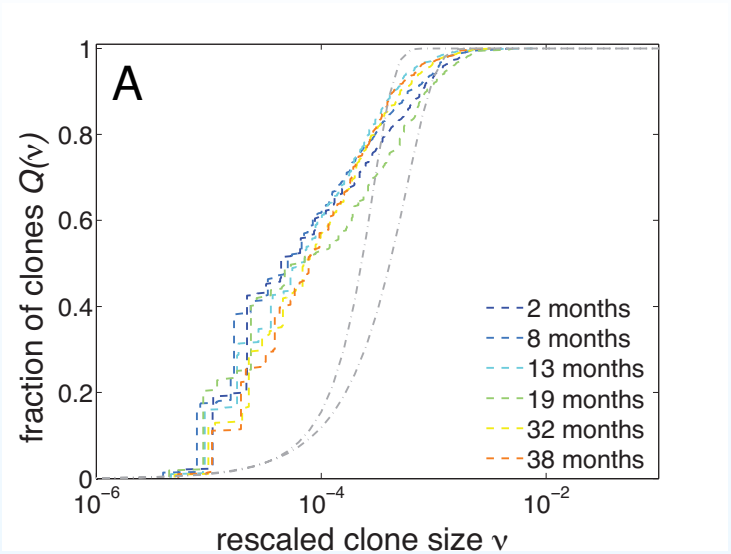
$$U + C \approx \frac{S}{a^* R^*} \quad \text{and} \quad \alpha \approx \mu a^*$$

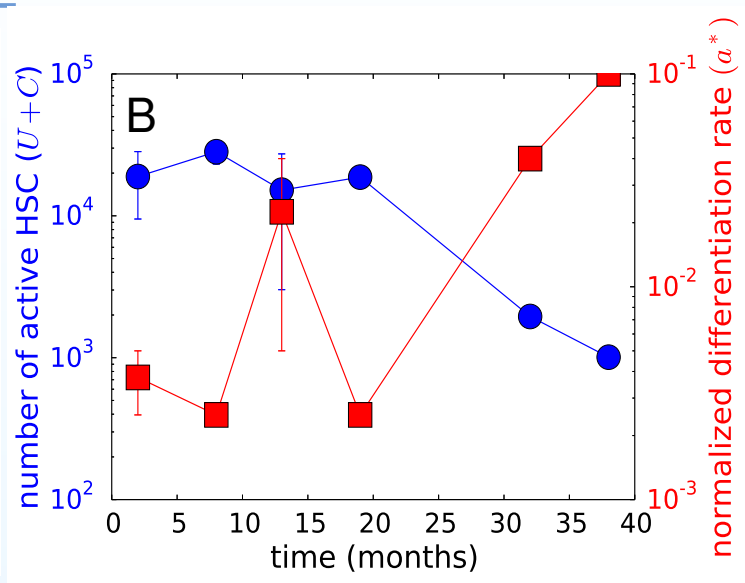
- We find $U + C \sim 10^3 - 10^4$.
- Fitting of a^* is ill-conditioned for $a^* \ll 1$.
- Determine p, ω, and/or μ_p from $N_\mathrm{d}(t)$ and

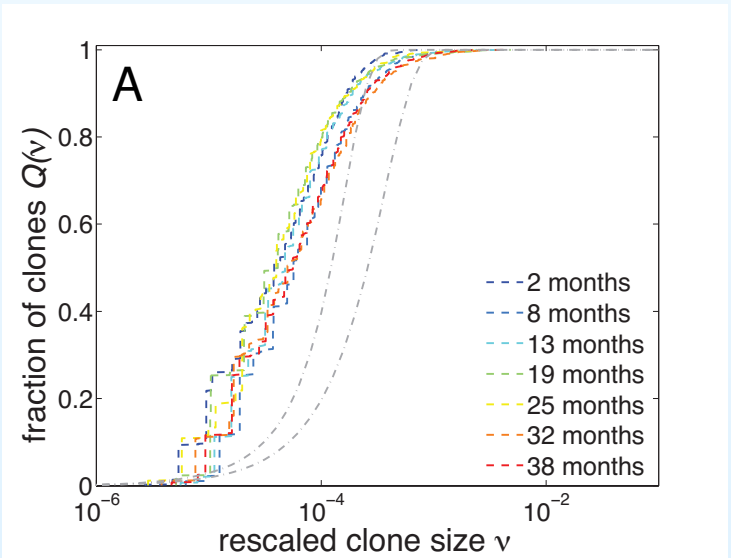
$$\frac{N_\mathrm{p}}{K} = \frac{1}{2}\left[\frac{\alpha}{\mu}\frac{U+C}{K} + \frac{p}{\mu} - 1 + \sqrt{\left(\frac{\alpha}{\mu}\frac{U+C}{K} + \frac{p}{\mu} - 1\right)^2 + \frac{4\alpha}{\mu}\frac{U+C}{K}}\right]$$

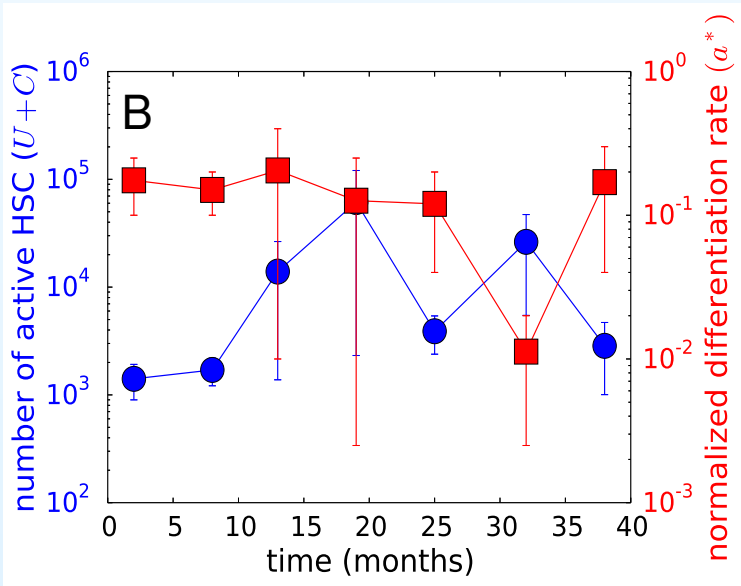
$$N_\mathrm{d}^* = \frac{(1+\eta)\omega N_\mathrm{p}}{\mu_\mathrm{d}}$$

Plotting rescaled $\bar{Q}(\bar{q})$ and fits









Dynamics near steady-state

Linearizing the eqns about steady-state: $c_k = c_k^* + \delta c_k$,

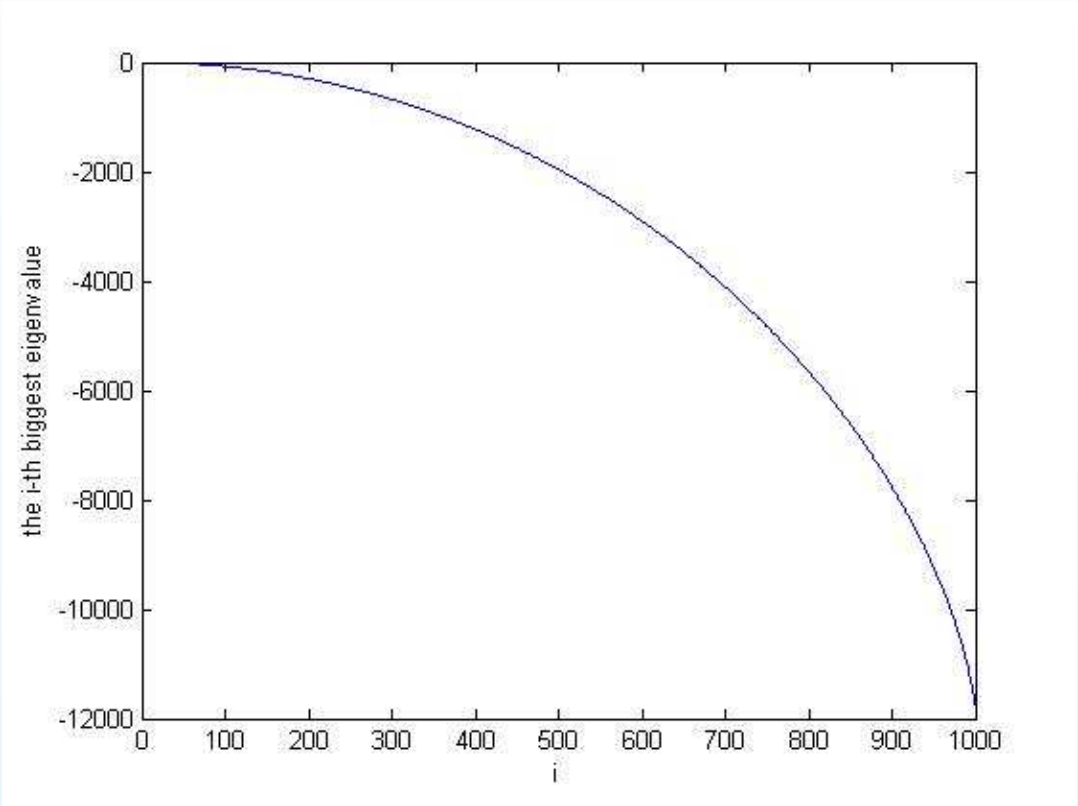
$$\delta \dot{c}_1 = -(\alpha + r + \mu_{\mathrm{p}})\delta c_1 + 2\mu_{\mathrm{p}}\delta c_2 + \frac{p}{K} c_1 \delta N_{\mathrm{p}}$$

$$\delta \dot{c}_k = (\alpha + (k-1)r)\delta c_{k-1} - (\alpha + rk + \mu_{\mathrm{p}}k)\delta c_k + \mu_{\mathrm{p}}(k+1)\delta c_{k+1} + \frac{p}{K}(kc_k - (k-1)c_{k-1})\delta N_{\mathrm{p}},$$

$$\delta \dot{N}_{\mathrm{p}} = (r - \mu)\delta N_{\mathrm{p}} - \frac{p}{K} N_{\mathrm{p}} \delta N_{\mathrm{p}}$$

Consider eigenvalue spectrum of this system:

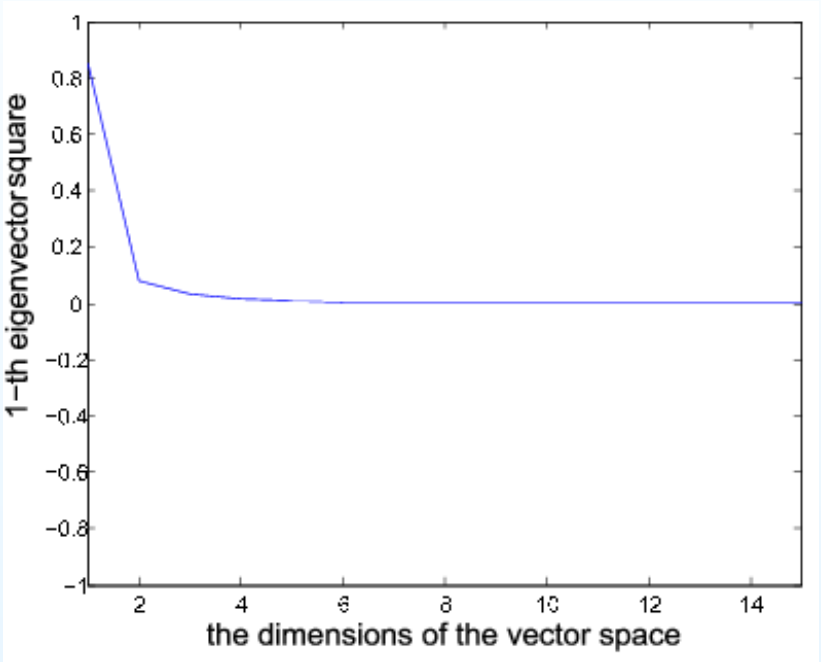
Time-dependence: eigenvalue spectrum

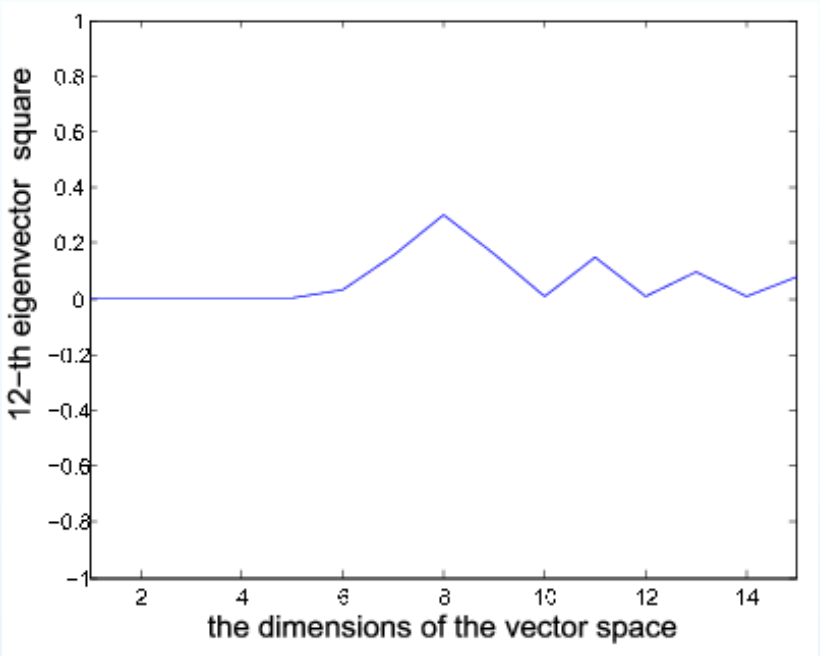


Eigenvalues span many values, including small negative values (slow time scales)

Time-dependence: eigenvectors

Eigenvectors for the 1^{st} and 12^{th} eigenvalues:





Slow modes are those involving clones represented by very few cells.
Rest of distribution reaches steady-state faster.

Individual lineages highly variable

Other possibilities:

- intrinsic stochasticity (numbers too large)
- extrinsic noise
- self-renewal of HSCs?
- heterogeneity
- HSC aging?
- Progenitor *lineage* aging...

Conclusions and Future Directions

- Homogeneous model can be consistent with observed distribution
- Model predicts reasonable physiological parameters upon fitting
- More complex models needed to explain clone extinctions (aging?)
- Time dependence of process: depletion experiments?