Transcriptome-wide noise controls lineage choice in mammalian progenitor cells

Hannah H. Chang,
Martin Hemberg,
Mauricio Barahona,
Donald E. Ingber
Sui Huang

Background on cellular noise

- in Eukariotic systems: dominated by bursty transcription
- phenotypic cell to cell variability
- beautifully measured, quantified, modelled in E. coli and Yeast

- **stable** phenotypic variants with the same genome (memory)
- in human cells: some proteins fluctuate between the low and high range on the timescale of 2 cell cycles!
- proteins in the same pathway **fluctuate in correlated fashion**
- **adaptive advantage** upon environmental challenge

---

Problems in differentiation

- External signals impose a genetic program
- Master regulators of lineage
- Tissue specific promoters

- Complicated interplay between many regulators
- Directing differentiation is very difficult... (<50% efficiency)

- Mouse embryonic stem cells: remove stem cell medium => spontaneous differentiation to many lineages
- Single cell plating of hematopoietic stem cells => macrophage, erythrocyte, platelet

- Spontaneously differentiated subpopulations “selected” by external factors
Does cellular noise play a role in differentiation?

★ Lets find this noise...
★ immunofluorescence flow cytometry of cell surface Sca-1 protein
★ clonal population
★ 1000-fold range!
★ stable over time
★ much larger than measurement noise
★ NOT just cell size or cycle

Figure 1 | Robust clonal heterogeneity. a, b, Heterogeneity among clonal cells in Sca-1 protein expression, detected by immunofluorescence flow cytometry (a), was significantly larger than the resolution limit of flow cytometry approximated by measurement of reference fluorescent MESF beads (b). The dashed lines show the difference in spread of the distributions as explained in the text. c, Stability of clonal heterogeneity in Sca-1 over three weeks.
How does the heterogeneity arise?

Let's FACS sort the subpopulations!
- Low, Med, High Sca-1
- culture the fractions
- slow restoration of full heterogeneity (> 12 cell doublings)!
- also works with single-cell clones (very slowly...):

Figure 2 | Restoration of heterogeneity from sorted cell fractions. **a**, Clonal cells with the highest (Sca-1<sup>high</sup>), middle (Sca-1<sup>mid</sup>) and lowest (Sca-1<sup>low</sup>) 15% Sca-1 expression independently re-established the parental extent of clonal heterogeneity after 216 h in separate culture. As an example, each cell in the Sca-1<sup>high</sup> experiment was theoretically partitioned into one of two GMM subpopulations (blue and red, right).
What drives the heterogeneity?

A few things to exclude:

- NO differential growth of Sca-1 subpopulations
- mutations are too slow (9 days, 12 cell divisions)
- widening of distribution too fast for uneven partitioning of Sca-1 protein during cell division
- Not much difference in transcription

What governs Sca-1 expression?

- circuitry not known
- explicit modeling unfeasible

Phenomenological approach

- find class of stochastic processes that can explain the data

1. Mean-reverting process (Ornstein–Uhlenbeck process)

- noisy relaxation process towards an equilibrium

Gaussian distribution
What drives the heterogeneity?

2. Gaussian mixing model

- rugged landscape
- multiple meta-stable states
- relaxation within basins (sub-populations!)
- stochastic transition between states

Overlap of 2 Gaussians!

Figure 2 | Restoration of heterogeneity from sorted cell fractions. a, Clonal cells with the highest (Sca-1<sup>high</sup>), middle (Sca-1<sup>mid</sup>) and lowest (Sca-1<sup>low</sup>) 15% Sca-1 expression independently re-established the parental extent of clonal heterogeneity after 216 h in separate culture. As an example, each cell in the Sca-1<sup>high</sup> experiment was theoretically partitioned into one of two GMM subpopulations (blue and red, right). b, c, The temporal evolution of the means $\mu_{1,2}$ (b) and weights $w_{1,2}$ (c) for the Sca-1<sup>high</sup> GMM subpopulations 1 and 2. The evolution of the weights was fitted to a sigmoidal function (c, dotted curves). Black dashed lines, equilibrium values for $\mu_i$ and $w_i$. 

Experiment

<table>
<thead>
<tr>
<th>Time (h)</th>
<th>Sca-1&lt;sup&gt;low&lt;/sup&gt;</th>
<th>Sca-1&lt;sup&gt;mid&lt;/sup&gt;</th>
<th>Sca-1&lt;sup&gt;high&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 h</td>
<td>[histogram]</td>
<td>[histogram]</td>
<td>[histogram]</td>
</tr>
<tr>
<td>9 h</td>
<td>[histogram]</td>
<td>[histogram]</td>
<td>[histogram]</td>
</tr>
<tr>
<td>48 h</td>
<td>[histogram]</td>
<td>[histogram]</td>
<td>[histogram]</td>
</tr>
<tr>
<td>96 h</td>
<td>[histogram]</td>
<td>[histogram]</td>
<td>[histogram]</td>
</tr>
<tr>
<td>144 h</td>
<td>[histogram]</td>
<td>[histogram]</td>
<td>[histogram]</td>
</tr>
<tr>
<td>216 h</td>
<td>[histogram]</td>
<td>[histogram]</td>
<td>[histogram]</td>
</tr>
</tbody>
</table>
Hidden surprise in Supplementary Data!

does the fitted line mean anything?

Best fit: relative population size AFFECTS transition rates!
A cell population with two states!

Is it biologically relevant?

How about differentiation potential?

Similar in secondary clonal populations

低 Sca-1

高 Sca-1

Pre-stimulation

Sca-1\textsubscript{low}
Sca-1\textsubscript{mid}
Sca-1\textsubscript{high}

Myeloid differentiation rate

Erythroid differentiation rate

0 d
7 d
14 d
21 d

Time after sorting in culture before Epo stimulation

NOISE

pro-erythrocyte

myeloid cell

Is it biologically relevant?

how about differentiation potential?
Persistent but reversible lineage preference

The low Sca-1 subpopulation is not pre-committed!

preference lost in 3 weeks (distributions are no longer distinguishable at day 7!)

more than Sca-1 to the difference!

how about master regulators of lineage?

![Diagram showing lineage preference and regulation](image-url)
Lineage-specific markers drive a broad expression program!

These two master regulators do not DECIDE lineage!!!

Figure 4 | Clonal heterogeneity of Sca-1 expression reflects transcriptome-wide noise. Self-organizing maps of global gene expression for a subset of 2,997 genes from each of the Sca-1low (L), Sca-1mid (M), Sca-1high (H) fractions at 0 and 6 d after FACS isolation and for a control sample. Pixels in the same location within each GEDI map contain the same mRNA expression value of genes (Supplementary Table 1). Thus, the outlier populations reconstituted large, transcriptome-wide differences! >3,900 genes L ↔ H
Instructive AND Selective

- **Instructive**
  - terminal differentiation only happens upon stimulation
  - master regulator expression is part of the commitment process

- **Selective**
  - noise can switch cells between functional states with different differentiation potentials
  - master regulators do not always determine lineage choice

**Toggle switch**

**Noise drives preference but not pre-commitment of lineage**
Outlook

**Strengths:**
- I admit I am biased: it’s beautiful (inspiring is a more accepted way of putting it...)
- it really asks us to keep in mind the complexity and non-linear nature of the regulatory network
- supports the idea of cell states as stable attractors

**Weaknesses:**
- ??? (I have no expertise to judge the experimental techniques)

**More coming ...**
- Siu Huang’s lab: working on switching cancer cell lines back to normal
- Jim Collins’s lab (bioengineer, leader in cellular noise control) with first author Hanna Chang: noise-assisted embryonic stem cell differentiation
Thank you!