

# RAUSCHEN UND ROBUSTHEIT IN DER GENEXPRESSION

ROBERT LEHMANN

ABSTRACT. Literatur-Auswertung zum Thema Rauschen und Robustheit bei der Expression von Genen. Neben dem Theoretischen Hintergrund wird ein konkretes Experimenten-Design vorgestellt.

Verwendete Paper:

[1] [2] [3]

## CONTENTS

List of Figures	2
1. Einleitung	3
2. theoretischer Hintergrund	5
2.1. Modellieren des intrinsischen Rauschen	6
2.2. stochastische Simulation des intrinsischen Rauschen	6
2.3. analytische Lösung des Modells für intrinsisches Rauschen	7
3. Anwendung der Theorie	8
4. Zusammenfassung	11
References	12

## LIST OF FIGURES

1 Reporterger-Assay zur Rauschmessung	3
2 verschiedene Rauschquellen	4
3 Definition Rauschen in Expression	5
4 Modell des intrinsischen Rauschen	6
5 vereinfachtes Modell des intrinsischen Rauschen	7
6 Simulation des intrinsischen Rauschens	7
7 intrinsisches Rauschen im Zellzyklus	8
8 Reporterger-Assay zur Rauschmessung	9
9 Diagramm CFP/YFP Konzentration	10
10 Rauschanteile bei variabler Expression	11
11 Zellstamm mit erhöhtem Rauschen	12

## 1. EINLEITUNG

Das zentrale Thema dieser Ausarbeitung ist die Frage, ob bei einem Pool von genetisch identischen Zellen unter gleichen Bedingungen auch identische Konzentrationen der verschiedenen Proteine zu finden sind.

Das kann mit folgendem recht einfachen Experiment untersucht werden. Hierzu misst man bei einer genetisch identischen Zellkultur mit einem konstitutiv exprimierten Reportergen, das für ein fluoreszierendes Protein wie z.B. das grünfluoreszierende Protein (GFP) kodiert, die Helligkeit der einzelnen Zellen. Wie in Abb.1 zu erkennen, weisen die Zellen äußerst verschiedene Farben (hier wurden gleich 2 Reporter mit versch. Farben verwandt) und Helligkeiten auf, sodass die Expression der gleichen Gene also von Zelle zu Zelle differieren muss. Dieses differieren der Genexpression wird im Folgenden Rauschen genannt.

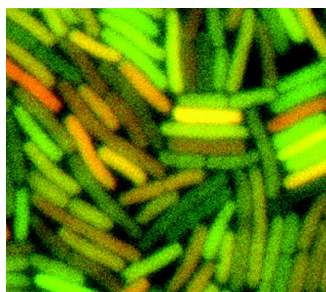


FIGURE 1. Reportergen-Assay zur Untersuchung der Fluktuationen in der Genexpressionsstärke bei genetisch identischen Zellen. Unterschiede in der Helligkeit einer Farbe (hier Cyan und Gelb) zwischen 2 Zellen zeigen Expressionsunterschiede zwischen diesen Zellen. Unterschiede in der Mischung der Farben zeigen Expressionsunterschiede innerhalb einer Zelle zwischen den beiden Genen. aus [2]

Abbildung 2 skizziert die beiden möglichen Quellen des Rauschen in der Genexpression. Als extrinsisch (zur Genexpression) werden in diesem Fall folgende Einflüsse benannt, bzw. alle stochastischen Einflüsse, die die Konzentration des Transkriptionsfaktors in der Nähe des Promotors beeinflussen.

- Zellen in versch. Zellstadien
- Zellen haben versch. Volumen
- unterschiedliche Anzahlen der Moleküle für Transkription/Translation
- Aktivität dieser Moleküle unterschiedlich groß

Dagegen gelten die stochastischen Ereignisse während Transkription und Translation wie z.B. die Bindung des Transkriptionsfaktors an den Promotor oder die Bindung der mRNA an ein Ribosom als Faktoren des intrinsischen Rauschens. Besonders wichtig hierbei ist jedoch, dass die Einteilung in extrinsisch oder intrinsisch lediglich davon abhängt, welches Gen oder welcher Mechanismus innerhalb der

Genexpression genauer Betrachtet wird, sodass bei jedem neuen Experimentdesign eine veränderte Einteilung verwendet werden kann oder muss.

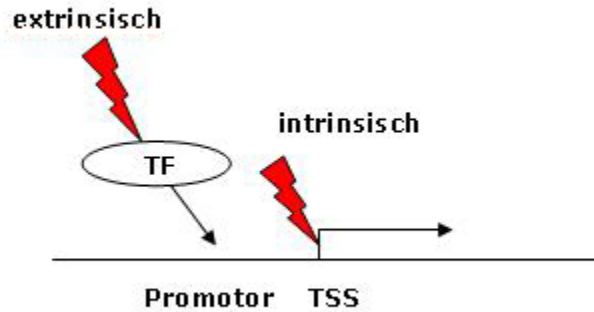


FIGURE 2. Verschiedene stochastische Einflüsse (sog. Rauschen) beeinflussen die Expressionsstärke. Einflüsse auf die Signalweiterleitung zwischen Signalentstehung bis hin zum Transkriptionsfaktor am Promotor verursachen das extrinsische Rauschen. Intrinsisches Rauschen tritt dagegen während der Bindung des Transkriptionsfaktors an den Promotor bis hin zur Bildung des fertigen Proteins auf.  
aus [3]

Im oben beschriebenen Experiment sind die Rausch-Arten folgendermaßen zu erkennen. Unterscheiden sich die Expressionsstärken der beiden Reportergene, dann weist die Zelle also eine andere Farbmischung auf als 50% Cyan und 50% Gelb und erscheint in einer anderen Farbe. Da aber das extrinsische Rauschen sich auf die gesamte Zelle und deren besondere Expressions-Eigenschaften bezieht und somit für beide Gene gleich groß sein muss, so kann dieser Unterschied einzig auf das intrazelluläre Rauschen zurückzuführen sein.

Das gesamte Rauschen ist dagegen als Helligkeitsunterschied in einer Farbe zwischen zwei Zellen zu erkennen, da einerseits wieder das intrinsische Rauschen auf beide Gene wirkt. Des weiteren fällt nun aber der extrinsische Rauschanteil mit ins Gewicht. Wie im Folgenden geschildert, kann nun aus diesen experimentellen Beobachtungen das extrinsische Rauschen berechnet werden.

## 2. THEORETISCHER HINTERGRUND

Das gesamte Rauschen in der Genexpression ist vor dem Hintergrund des o.g. Experimentes mit fluoreszierenden Reporter-gen-Produkten folgendermaßen definiert:

$$\eta^2(t) = \frac{\langle P(t)^2 \rangle - \langle P(t) \rangle^2}{\langle P(t) \rangle^2},$$

FIGURE 3. Definition des Rauschens in der Genexpression als normalisierte Varianz.  $P(t)$  - Proteinkonzentration zum Zeitpunkt  $t$ ,  $\langle \rangle$  - Mittelwert über die Verteilung, aus [3]

Hierbei ist das Rauschen des Proteins als normalisierte Varianz, auch coefficient of variance genannt, definiert.  $\langle P(t) \rangle$  ist der Mittelwert über die Verteilung der Proteinexpression  $P$ , also dessen Konzentration.

Der Mittelwert der Reporterprotein-Konzentration der Zellpopulation wird folgendermaßen aus den Helligkeitswerten der einzelnen Zellen berechnet.

$$\frac{1}{N} \sum_{k=1}^N P_k^m \approx \int d\mathbf{E} d\mathbf{I} P^m(\mathbf{E}, \mathbf{I}) p(\mathbf{E}, \mathbf{I}).$$

Dabei ist  $\mathbf{E}$  der Vektor aller auf diese Zelle wirkender extrinsischer Einflüsse,  $\mathbf{I}$  der entspr. intrinsische Vektor,  $P^m(E, I)$  die Konzentration bei fest gewählten  $\mathbf{E}, \mathbf{I}$  und  $p(\mathbf{E}, \mathbf{I})$  die Wahrscheinlichkeit der Einflüsse aus  $\mathbf{E}$  und  $\mathbf{I}$ . Das kann durch festhalten eines  $\mathbf{E}$  und anschliessendes integrieren über alle  $\mathbf{I}$  umgeformt werden zu:

$$\frac{1}{N} \sum_{k=1}^N P_k^m \approx \int d\mathbf{E} p(\mathbf{E}) \int d\mathbf{I} P^m(\mathbf{E}, \mathbf{I}) p(\mathbf{I} | \mathbf{E}).$$

Das Wählen eines festen  $\mathbf{E}$  und integrieren über alle möglichen  $\mathbf{I}$  wird dann neu benannt als  $\langle \rangle$ :

$$\langle P^m(\mathbf{E}) \rangle \equiv \int d\mathbf{I} P^m(\mathbf{E}, \mathbf{I}) p(\mathbf{I} | \mathbf{E}).$$

Benennt man nun das integrieren über alle möglichen  $\mathbf{E}$  als Balken, dann lässt sich Formel 3 schreiben als:

$$\frac{1}{N} \sum_{k=1}^N P_k^m = \overline{\langle P^m \rangle},$$

Man kann nun zeigen, das sich diese Formel in die Summe der beiden Rauscharten umformen lässt:

$$\eta_{\text{tot}}^2 = \frac{\overline{\langle P^2 \rangle} - \langle P \rangle^2}{\langle P \rangle^2} + \frac{\overline{\langle P \rangle^2} - \langle \overline{\langle P \rangle} \rangle^2}{\langle \overline{\langle P \rangle} \rangle^2} \equiv \eta_{\text{int}}^2 + \eta_{\text{ext}}^2 \cdot 1$$

Bleibt lediglich die Frage, welches Experiment-Design diese Werte liefert. Misst man die Helligkeiten zweier Kopien ( $P_k^{(1)}$   $P_k^{(2)}$ ) des selben Gens in einer Zelle (intrinsische Variablen versch./ extrinsische gleich), so entspricht deren Mittelwert:

$$\begin{aligned} \frac{1}{N} \sum_{k=1}^N P_k^{(1)} P_k^{(2)} &\approx \int d\mathbf{E} d\mathbf{I}_1 d\mathbf{I}_2 P(\mathbf{E}, \mathbf{I}_1) P(\mathbf{E}, \mathbf{I}_2) p(\mathbf{E}, \mathbf{I}_1, \mathbf{I}_2) \\ &= \int d\mathbf{E} p(\mathbf{E}) \left[ \int d\mathbf{I} P(\mathbf{E}, \mathbf{I}) p(\mathbf{I} | \mathbf{E}) \right]^2 \\ &= \overline{\langle P \rangle^2}, \end{aligned}$$



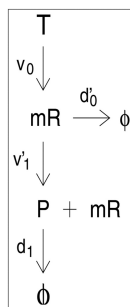


FIGURE 5. Vereinfachtes und daher analytisch lösbares Modell der konstitutiven Genexpression. T - Transkriptionsstart, mR - mRNA, P - fertiges Protein,  $\phi$  - Abbau des Proteins. aus [3]

des Wachstums auch die Verdopplung der DNA zum Zeitpunkt  $t_d = 0.4T$  ( $T$  = Dauer eines Zellzyklus) und die Zellteilung zum Zeitpunkt  $T$  berücksichtigt. Dies zeigt sich z.B. im Verlauf der Proteinanzahl, die vom Zeitpunkt  $t_d$  an etwas stärker steigt und nach der Zellteilung wieder auf das alte Maß zurückgeht. Da mRNA-Bildung und -Abbau im Vergleich zu Zellzyklus, Protein-Bildung und Abbau als viel schneller modelliert wurden, zeigen sich Veränderungen in der Gen-Anzahl (z.B. während Replikation oder Teilung) als schnelle, steile Zacken im Verlauf der mRNA-Anzahl.

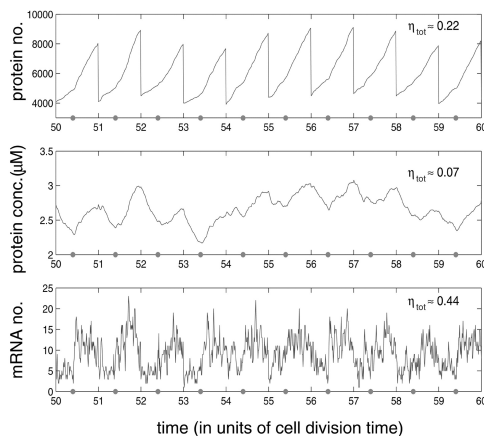


FIGURE 6. Simulation des intrinsischen Rauschens über mehrere Zellzyklen hinweg. Oben - Anzahl des modellierten Proteins, Mitte - Konzentration des modellierten Proteins, Unten - Anzahl der mRNA des modellierten Proteins. Der Punkt auf der X-Achse markiert die Zellteilung. aus [3]

**2.3. analytische Lösung des Modells für intrinsisches Rauschen.** Das in Abbildung 5 gezeigte vereinfachte Modell für das intrinsische Rauschen während der Genexpression kann auch analytisch mittels Differentialgleichungen berechnet werden, was den Vorteil bietet, keine Wahrscheinlichkeiten benutzen zu müssen. Da das System aus Differentialgleichungen mit wachsender Komplexität des Modells jedoch sehr schnell äußerst rechenaufwendig wird, wurde lediglich das vereinfachte Modell aus Abbildung 5 analytisch berechnet. Die Abbildung 7 zeigt den Verlauf des simulierten bzw. berechneten intrinsischen Rauschens während eines Zellzykluses, wobei die Simulations-Ergebnisse über 5000 Zyklen gemittelt wurden. Hierbei ist eine gute Übereinstimmung der 3 Berechnungsarten des intrinsischen Rauschens erkennbar.

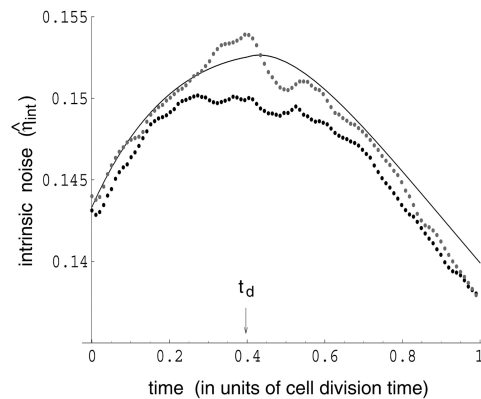


FIGURE 7. Vergleich des analytisch aus dem vereinfachten Modell (Abb.5) berechneten intrinsischen Rauschens (glatte Kurve) während des Zellzyklus mit den Simulationen des vereinfachten (obere gepunktete Kurve) und vollen (Abb.4) Modells (untere gepunktete Kurve).

aus [3]

### 3. ANWENDUNG DER THEORIE

Wie zu Beginn erwähnt, kann die Theorie auch im Experiment angewandt werden. Vergleicht man die Konzentration des Reportergens (d.h. Helligkeit) zweier unterscheidbarer Reporter in jeweils einer Zelle, so sind die Unterschiede aufgrund des identischen extrinsischen Rauschanteils bei beiden Genen pro Zelle auf das intrinsische Rauschen zurückzuführen. Dagegen wird der Unterschied des gleichen Reportergens zwischen zwei beliebigen Zellen durch die Summation intrinsischen wie extrinsischen Rauschens verursacht und ist somit Ausdruck des gesamten Rauschens. Bei einer Zellkultur, wie sie in Abbildung 8 dargestellt ist, besitzt jede Zelle eine bestimmte Zahl Kopien des Cyan-fluoreszierenden Proteins (cfp) und die gleiche Zahl Kopien des Gelb-fluoreszierenden Proteins (yfp), was die gleichzeitige Untersuchung des intrinsischen wie auch des gesamten Rauschens ermöglicht. Dabei ist zu beachten, dass beide Gene jeweils den gleichen Promotortyp besitzen müssen, um exakte Rauschmessungen zu bekommen. Die Helligkeit der Reporter



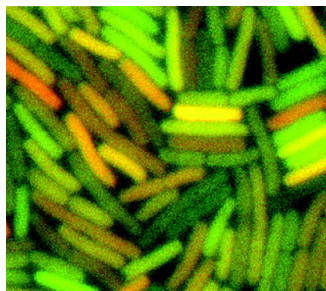


FIGURE 8. siehe 1  
aus [2]

pro Zelle kann dann per Computer mittels einem Bildanalysesystem recht exakt festgestellt werden.

In Abbildung 9 sind die Messwerte des oben beschriebenen Experiments in einem Diagramm aufgetragen. Jeder Punkt steht für eine Zelle, der X-Wert für die Konzentration an CFP und der Y-Wert für die YFP-Konzentration in der Zelle. Die im theoretischen Teil angesprochenen intrinsischen und extrinsischen Rauschanteile lassen sich nun hierin wiederfinden. Eine Zelle ohne jegliches Rauschen würde exakt auf der Winkelhalbierenden zwischen den beiden Achsen liegen, da beide Reporter-Gene gleich häufig abgelesen würden, sodass es nur von Effektivität der Proteinsynthese abhängt, wo auf dieser Gerade der Punkt wäre.

Eine zweite Zelle mit ausschließlich extrinsischem Rauschen, das ja auf alle Gene einer Zelle gleichermaßen einwirkt, hätte ebenfalls stets gleich viel CFP wie YFP und würde daher auch auf dieser Winkelhalbierenden liegen. Jedoch wäre die Position dieser Zelle zur rauschfreien Zelle entlang der Geraden verschoben, da nun zufällig gleichzeitig mehr oder weniger CFP und YFP gebildet wird. Dies deutet der mit extrinsic beschriftete Pfeil in Abb. 9 an.

Dagegen wäre eine Zelle mit ausschließlich intrinsischem Rauschen, das auf jedes einzelne Gen einer Zelle einwirkt, genau orthogonal zur Winkelhalbierenden beider Achsen von der rauschfreien Zelle verschoben, da nun CFP und YFP-Produktion unabhängig voneinander durch Rauschen beeinflusst werden können und so das Verhältnis beider gestört wird (symbolisiert durch intrinsic - Pfeil).

Des weiteren ist in Abbildung 9 zu erkennen, dass der E.choli-Stamm D22 eine stärkere Streuung und also ein stärkeres Rauschen der Zellen aufweist als der Stamm M22. Diese Beobachtung wurde in [2] näher untersucht, wobei zunächst beide Reportergene in jeweils einem Versuch mit den starken, konstitutiv exprimierten Promotoren  $\lambda P_R$  und dem lac-sensitiven Promotor (lac-Inhibitor deletiert) versehen wurden. Bei beiden Versuchen wurde etwa die gleiche Streuung und also das gleiche Maß Rauschen festgestellt wie in Abb. 9. Hieraus folgern die Autoren, das starke konstitutive Expression erstaunlich uniform sein kann und das das Rauschniveau nicht vom verwendeten Promotor abhängt.

Daraufhin wurde die Frage untersucht, ob die reduzierte Expression mittels einem teilweise gehemmten Promotor ein verändertes Rauschniveau aufweist. Dazu wurden die Reportergene wiederum mit dem lac-sensitiven Promotor versehen und in Wildtyp - e.choli transportiert, in der ständig eine geringe Menge lac-Inhibitor gebildet wird. Dies führte zu einer Reduktion der Expression der Reporter auf ca.

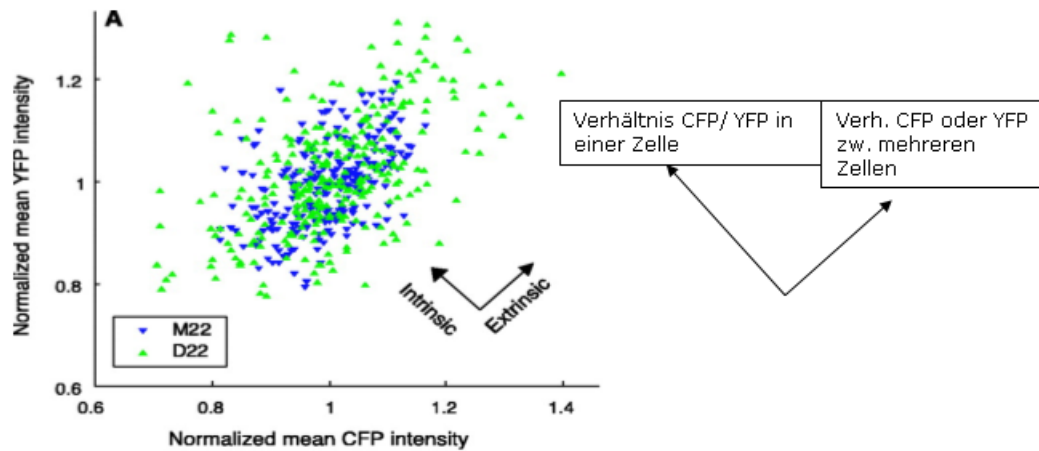


FIGURE 9. Diagramm der CFP-Konzentrationen (X-Achse) gegen YFP-Konzentrationen (Y-Achse) zweier *e.coli* Stämme (blau - M22, grün - D22). Die beiden Pfeile zeigen die Richtungen, in die extrinsisches bzw. intrinsisches Rauschen den Datenpunkt einer Zelle gegenüber einer hypothetischen rauschfreien Zelle verschiebt. Beide Reportergene haben den *lac*-sensitiven Promotor und *lac*-Inhibitor deletiert.

aus [2]

3-6% der ursprünglichen Expression aus dem vorigen Versuch und damit einhergehend zu einem Anstieg des Rauschniveaus um den Faktor 5.

Als Kontrolleexperiment wurde der im Wildtyp produzierte *lac*-Inhibitor anschließend mittels IPTG gehemmt, sodass der Promotor wieder ungehemmt arbeiten und die Expression wieder auf 100% ansteigen konnte. Das nun auch das Rauschniveau wieder auf das alte Maß zurückging, zeigt den kausalen Zusammenhang zwischen Hemmung und Rauschzunahme.

Da das Rauschen bei unterschiedlicher Promotor-Aktivität verschieden stark ist, wurden daraufhin die Stärke der Rauschanteile bei wechselnder Promotor-Aktivität untersucht. Abbildung 10 zeigt das Ergebnis der Messreihe, bei der beide Reportergene mit dem *lac*-sensitiven Promotor ausgestattet sind. Zusätzlich wurde ein Plasmid mit einem stark konstitutiv exprimierten *lac*-Inhibitor-Gen in die Zellen eingefügt, um die Reportergen-Expression im Normalzustand vollständig zu hemmen. Eine stufenweise zunehmende Expression wurde nun erzielt, indem steigende Konzentrationen an IPTG zugegeben wurde, das seinerseits den *lac*-Inhibitor hemmt und Reportergen-Expression ermöglicht. Bei jeder Konzentration an zugegebenem IPTG wurden dann die beiden Rauschanteile bestimmt und in der Abbildung auf der Y-Achse gegen die beobachtete Expressionsstärke auf der X-Achse aufgetragen. Besonders auffällig ist, dass das extrinsische Rauschen ein Maximum durchläuft und deshalb nicht eindeutig aus dem intrinsischen Rauschen berechnet werden kann. In [3] wurde als Ursache für das Maximum das Design des Experimentes genannt, da die Expression des *lac*-Inhibitors ebenfalls dem Rauschen unterliegt und bei starker Aktivierung (also geringerer Reporter-Expression!) kleiner wird. Dem ist jedoch entgegen zuhalten, das im Experiment in [2] ein gleichbleibend stark konstitutiv

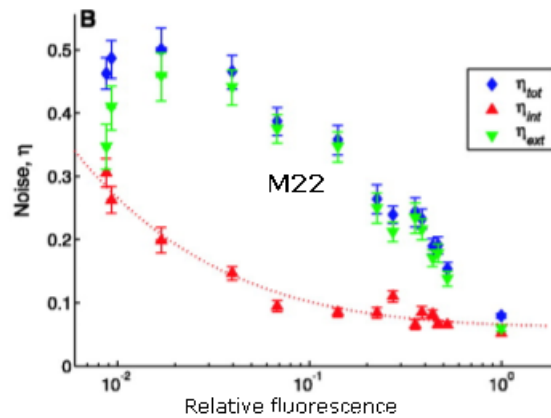


FIGURE 10. Das Verhältnis der Rauschanteile sowie die gesamte Rauschstärke hängen davon ab, wie stark der Promotor gehemmt oder aktiviert ist. Mit zunehmender Hemmung steigt das intrinsische wie auch das Gesamtrauschen. Das extrinsische Rauschen dagegen steigt mit zunehmender Hemmung, um dann nach einem Maximum wieder geringer zu werden.  
aus [2]

exprimierendes lac-I Gen auf einem Plasmid verwendet wurde und die Repression der Reporter durch eine variable Repression des Inhibitors mittels IPTG erreicht wurde.

Das selbe Experiment wurde auch beim E.coli Stamm D22 durchgeführt (Abb. 11), bei dem das Reparaturgen recA für abgerissene Replikationsgabeln nicht vorhanden ist. Das hier beobachtete erhöhte Rauschniveau ist, so die Autoren, möglicherweise auf eine erhöhte Variabilität bei der Zahl der Genkopien aufgrund des Fehlens von recA zurückzuführen.

Da bisher lediglich Rauschen in Abwesenheit regulativer Signale untersucht wurde, wurde in [2] ein Regulations-Netzwerk mit Namen Repressilator entwickelt, welches die Expression des lac-Inhibitors wiederholt an und aus schaltet. Wie erwartet wurde bei der Untersuchung ein deutlich erhöhtes Rauschniveau während des Erreichens des Regulationsziels festgestellt, sodass gefolgert werden kann, dass Dynamik in der Genexpression die Genauigkeit der Regulation stark beeinflusst.

#### 4. ZUSAMMENFASSUNG

- (1) Zufällige Schwankungen in der Genexpression können in extrinsisches und intrinsisches Rauschen unterteilt werden
- (2) Diese beiden Rauschanteile sind unterscheidbar und addieren sich zum gesamten Rauschen
- (3) Je stärker die Expression gehemmt wird, desto größer wird das gesamte Rauschen
- (4)

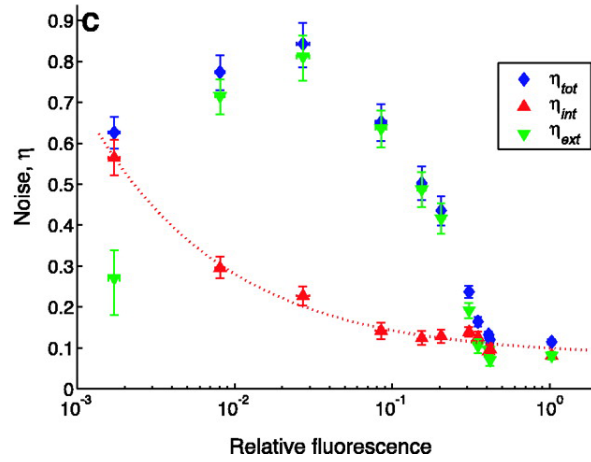


FIGURE 11. Bei Versuchen mit verschiedenen E.choli Stämmen fiel D22 mit erhöhtem Rauschniveau auf. Ursache hierfür könnte das fehlende Reparatur-Gen recA sein.  
aus [2]

#### REFERENCES

1. Alejandro Colman-Lerner, Andrew Gordon, Eduard Serra, Tina Chin, Orna Resnekov, Drew Endy, C Gustavo Pesce, and Roger Brent, *Regulated cell-to-cell variation in a cell-fate decision system.*, Nature **437** (2005), no. 7059, 699–706.
2. Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain, *Stochastic gene expression in a single cell.*, Science **297** (2002), no. 5584, 1183–6.
3. Peter S Swain, Michael B Elowitz, and Eric D Siggia, *Intrinsic and extrinsic contributions to stochasticity in gene expression.*, Proc Natl Acad Sci U S A **99** (2002), no. 20, 12795–800.

E-mail address: lehmannr@molgen.mpg.de