

遺伝子発現の確率性：ばらつきの特徴・背景・機能との関係

栗津 暁紀¹, 永野 惇²

¹ 広島大学大学院理学研究科

〒739-8526 広島県東広島市鏡山 1-3-1

² 龍谷大学農学部

〒520-2194 滋賀県大津市瀬田大江町横谷 1-5

Akinori Awazu¹, Atsushi J. Nagano²

Stochastic behaviors of gene expressions:

Features, backgrounds, and relationships to functions of fluctuations

Keywords: Feedback, function, regulation, stochasticity, transcriptome

¹Dept. of Math. and Life Sciences, Hiroshima Univ., Higashi-hiroshima, 739-8526 Japan

²Faculty of Agriculture, Ryukoku University, Otsu, Shiga 520-2194, Japan

DOI: 10.24480/bsj-review.10a2.00149

1. 目的

本小論では、遺伝子発現に現れる確率性（個体間でのばらつき、または個体間での揺らぎ）が示す特徴、その特徴が現れる背景、及びその生命機能との関連について、少し前に出版された論文（Awazu et al. 2018。*以降「論文 X」と呼びます。）の内容に触れつつ紹介する。この論文 X では、20 個体以上の同一生育時期にあるシロイヌナズナの RNA シークエンスを行うという、従来ほぼ行われてこなかったサンプリングに基づいて考察が進められており、そのデータと得られた知見を是非多くの方に有効利用して頂きたいと思っている。しかしこの論文は数理的な色合いがやや強く、実験を主たる研究手段としている方々にはやや読み難いものになっている感がある。そこで詳細は論文 X をご参照頂くとして、本小論ではその内容のエッセンス、このような研究を行った背景、及び得られた結果の幅広い有用性を、少しでも伝えられたらと思っている。

2. 遺伝子発現量の「ばらつき」に注目する意義

同一の遺伝型を持つ個体間で遺伝子発現量（表現型）が大きくばらつく事は、よく見られる事である。そしてこれまでに、その「ばらつき方」の統計性やその背景について、様々な細胞・生物を用いた考察が進められている（Chang et al. 2008, Elowitz et al. 2002, Furusawa et al. 2005, Newman et al. 2006）。特に大腸菌では、1000 種類以上の遺伝子に対する網羅的な解析から、各遺伝子の個体（細胞）毎の発現量がガンマ分布に従ってばらつく事が明らかにされ、その確率性が、遺伝子の転写・翻訳に必然的に紛れ込む生化学反応過程の揺らぎに起因していると示唆されている（Taniguchi et al. 2010）。但し、この知見がそのままあらゆる生物、例えば多様な代謝産物を産出するために大腸菌よりも遥かに複雑な遺伝子制御系を内包する植

物，に対しても適用出来るのかには疑問の余地がある。また近年盛んに行われているトランスクリプトーム解析では，多くの場合，生化学的過程というより経験に基づいて，mRNA 量が負の二項分布等に従ってばらつくとして考察がなされている (Gierliński et al. 2015, Karthik et al. 2016, Marioni et al. 2008, Robinson et al. 2008, Schurch et al. 2016, Smith et al. 2016)。

一方，大腸菌，酵母，植物（イネ・シロイヌナズナ）を用いた研究より，発現量のばらつき方がその遺伝子のプロモーター配列，機能，環境変化（ストレス）に対する応答性（発現量変動），及び変異による機能（活性）の向上性，等と相関する事が報告されている (Hirao et al. 2015, Lehner 2010, Lehner et al. 2011, Nagano et al. 2012, Sato et al. 2003, Singh 2013, Tirosh et al. 2008)。この事は，発現制御から機能や進化可能性まで，各遺伝子の持つ様々な側面がその発現量のばらつきと密接に関連している事を示唆している。しかしこのような事実を包括的に議論する試みは，殆どなされていない（*現在，以下に述べる我々の考察（論文 X）や，その後に発表された研究（Cortijo et al. 2019）等が出てきている。）。

論文 X ではこのような背景を踏まえ，シロイヌナズナの遺伝子発現のばらつきに着目した考察を進めている。以下にその内容を端的に述べておく。この研究は，同一ゲノムを持つ同一条件で生育した 20 個体以上のシロイヌナズナに対し，RNA シークエンシングを行い，データを解析したものである。そして 8 種の生育時期でのデータを解析し，

I) 遺伝子の発現量のばらつき方には大まかに，発現量の分布（ヒストグラム）がガウス分布（正規分布）に（近似的に）従うもの，ベキ分布に（近似的に）従うもの，その中間の分布型に従うものの 3 種類が存在する。ここで「近似的に」と付けたのは，実際にはガウス・ベキ混合分布と名付けられた分布に従うと考えられるからである。またこのばらつき方は，遺伝子毎，生育時期毎で変化する。

II) 発現量の分布がベキ分布に従っている遺伝子の発現制御は，確率的な（強いノイズにまみれた）フィードバックの影響を強く受けており，逆に発現量の分布がガウス分布に従う遺伝子では，そのようなフィードバックの影響を殆ど受けていないと考えられる。

III) 遺伝子の機能とその発現量の分布に相関がある（表 1）。例えば *Essential genes* と呼ばれる生存に必須な遺伝子群や代謝・エネルギー変換に関わる遺伝子群の多くで，その発現量がガウス分布に従う。一方，転写因子等をコードする遺伝子群の多くで，その発現量がベキ分布に従っている。また機能未知遺伝子の多くも，発現量がベキ分布に従っている。

といった事が見出されている。

これらの知見は多くの示唆に富んでいる。例えば II) と III) から，転写因子等は互いに活性・抑制をフィードバックし合う相互制御的なネットワークを形成しているのに対し，生存に必須な遺伝子群は上流遺伝子から *Passive* に制御されているといった，遺伝子ネットワークの全体像が容易に想像できる。また I) と II) から，各遺伝子が時期特異的にどのような制御（上流からか？フィードバックか？）をより強く受けているのかが推測出来る。更に I) と III) は，発現量の分布から機能未知の遺伝子の機能推定の手がかりが得られる事を示唆する。つまり，これが重要な事なのであるが，各遺伝子のその時々様々な側面，情報が，遺伝子発現量のばらつきとして滲み出ているのである。

このように，系の様々な性質がデータのばらつきとして滲み出ているケースは，遺伝子発

現量だけに留まらないと思われる。よって実験で複数の Replicate を取った時に、その値のばらつき方にも注目する事は、一つの実験からより多くの情報を得る上で有用と思われる。しかし次節で述べるように、この「データばらつき」の取り扱いには少々厄介な問題がついてまわるので、注意が必要である。

Ratio of	ベキ分布	中間の分布	ガウス分布
All genes	0.435357671	0.260434582	0.304207747
Function known genes (GO annotation)	0.322610702	0.276288934	0.401100364
Essential genes (GO slim)	0.208732694	0.248136315	0.54313099
mitochondria	0.464503043	0.249492901	0.286004057
extracellular	0.45	0.246206897	0.303793103
plastid	0.149706795	0.245946878	0.604346326
cytosol	0.205555556	0.283597884	0.510846561
ribosome	0.216624685	0.258186398	0.525188917
transcription factor activity	0.484303056	0.239849309	0.275847635
nucleic acid binding	0.458526875	0.262773723	0.278699403
other molecular functions	0.448415374	0.224544842	0.327039784
structural molecule activity	0.204889406	0.263096624	0.53201397
electron transport or energy pathways	0.154351396	0.238095238	0.607553366

表1. 日齢7日のシロイヌナズナの遺伝子の機能と発現量が従う分布型との関係。各機能グループにアノテーションされた遺伝子群に対し、各分布の型を示す遺伝子の割合を示している。遺伝子のアノテーションは Gene Ontology Slim (<http://www.arabidopsis.org>), Essential genes は SeedGenes Project (<http://www.seedgenes.org/GeneList>) (Meinke et al. 2008) に基づく。

3. 「外れ値」か? 「ばらつき」か?

例えば同一なゲノムを持つ生物個体を同一条件下で飼育し、各個体から2つの遺伝子（仮にAとBと呼ぶ）の発現量を計測し、個体順にプロットする（図1）。多くの人が経験する通り、計測値には「ばらつき」が現れる。

しかしここで注目する遺伝子AとBの発現量のばらつき方の間には、何か違いがあるように思える。Aの発現量はばらついているものの、全体として平均値の周りに対称に分布しており、この平均値はこの発現量をおよそ代表出来る値であるように思える。その一方で、Bの発現量は多くの個体で低い値を取るものの、目立って高い値も2つ（人によっては1つと思うかもしれない）現れている。実験の個体数の違いはあれ、このようなばらつき方をする結果に出会った経験を、多くの人が持っているのではと思われる。そしてこのような量をどのように特徴付けすべきか、悩んだのではないと思われる。

これはあくまで筆者の見聞き・経験したことからの想像であるのだから、このBの発現量のような結果に対し多くの場合行われている処方方は、「目立って高い（低い）値を『外れ値』として無視した上で、各種統計量を評価する。」であろう。各種統計量の評価は、様々なモ

デル分布関数へのフィッティングを介して行われる事も多いが、その際にもこのような処理がまずなされているようである。(例えば Gierliński et al. 2015 等)。

上記のような対処は多くの場合、結果的に正しいものになっていると考えられる。実験・計測のプロセスで完全にエラーを排除するのは困難であり、他から大きく逸脱して見えるデータはエラーの結果と考えられる事が良くあるからである。しかし残念ながら、大抵の場合、その対処が正しいとする根拠を与える事は出来ない。更に少なくない場合で、この対処が正しくない可能性も疑われる。一般に計測された値が「外れ値」なのかどうかの判定は、あまり極端ではない常識的な値をパラメータとして与えた、経験的に適当と思われるモデル分布関数に基づいて行われる。しかしこの「常識的」「経験的に適当」という言葉は決して根拠を与えるものではない。つまり「外れ値」は我々の先入観の産物である可能性が否定出来ないのである。

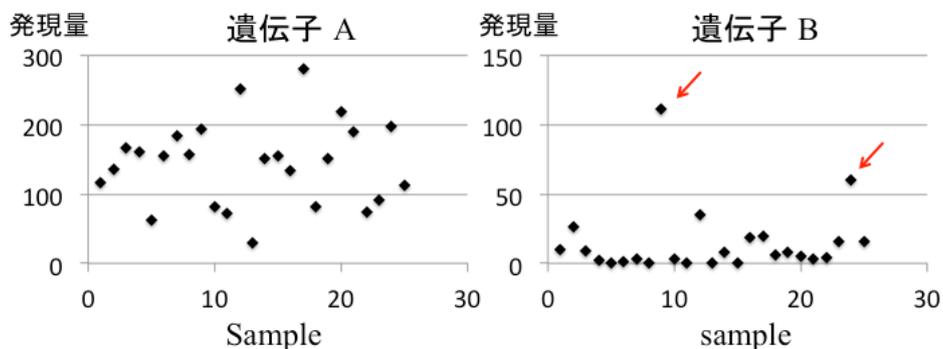


図 1. 同一条件で生育した同一のゲノムをもつシロイヌナズナ 25 サンプルから得られた、ある 2 つの遺伝子 (A, B) の発現量。赤矢印の示す点が、従来「外れ値」とされてしまう可能性が高いと考えられるサンプルの発現量。

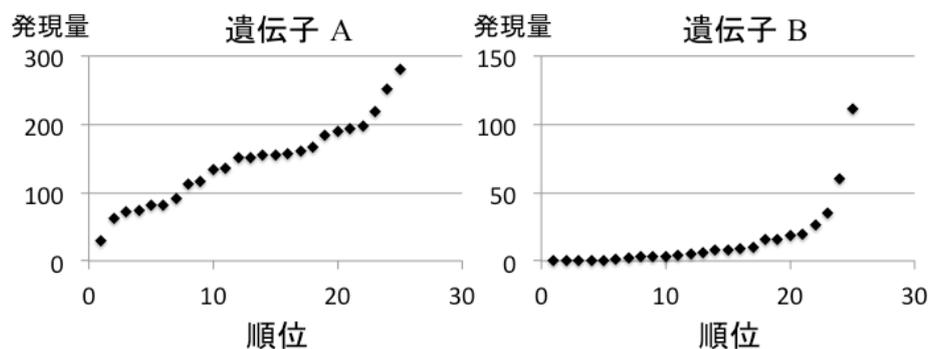


図 2. 2 つの遺伝子 (A, B) の発現量を小さい順に並べ直した「順位-発現量関係」。

ここで例えば、先ほどと少し視点を変えて、上記の各遺伝子発現量のデータに小さい順に順位をつけて、その順に並べてプロットした「順位-発現量関係」に着目する(図 2)。この場合、遺伝子 A では全体的に、順位が増加に対し発現量は最初の頃はやや急に増加しつつ途中緩やかになり、終盤やや急に増加するようなグラフが現れる。一方遺伝子 B では、全体と

して順位の増加に伴い発現量の増加の仕方が急になるグラフが現れる。更に重要な事は、後者では、順位の前半から最後の2つの大きな値までが、滑らかな曲線で繋がっているように見える事である。つまり殆ど発現していない状態から桁外れに発現量が上昇している状態のどれかが、「ある一つの規則に従って起きている」のではないかと疑いたくなるようなグラフの形状をしているのである。

もし上記の疑いが本当であれば、高い発現量値を「外れ値」として除外するのは言語同断である。ここで注意すべき事は、論文 X で扱われているシロイヌナズナ遺伝子発現量データには、遺伝子 A に似た様相のものも遺伝子 B に似た様相のものもそれぞれ数千種類ずつ存在しており、これらは決して例外的なものではないという事である。では実際どのような対処をしたら良いのだろうか？

ご存知の通りデータが理想的に十分多く取られていれば、これまで提案されてきているモデル分布関数に頼らずに、原理的にデータだけから発現量の分布関数型を推定できる。しかし多くの人が経験している通り、多くの場合（特に生物を対象とした実験の場合）そのような理想的な状況は期待出来ない。それに対し論文 X では、I) 各生育時期の遺伝子発現データに 20 以上の Replicate が存在する事、II) 順位-発現量関係の形状の類似性から遺伝子が（数十から数千の遺伝子からなる）グループに分けられる事、の2点を利用して、データの非理想性を部分的にもクリアできる試みを行っている。そして次節で述べるような遺伝子発現量の分布関数（ヒストグラム）型を、定性的にはあるが推定している（推定手法や導出の詳細は論文 X を参照のこと）。

4. 「ばらつき」の分布関数（ヒストグラム）の型の推定とその特徴

論文 X でなされた発現量データのみからの分布の推定の結果、シロイヌナズナの各遺伝子の発現量の分布が、順位-発現量関係の形状の類似性で分けられた時に属するグループに依存して、i) ガウス分布（正規分布）、ii) ベキ分布、iii) 中間の分布、の何れかに近い分布になるが示唆されている（図 3）。例えば上記の遺伝子 A と同じグループに属する遺伝子の発現量はガウス分布、遺伝子 B と同じグループに属する遺伝子の発現量はベキ分布に従うと推定される。

ところでガウス分布は多くの人に馴染みがあるかと思われるが、ベキ分布に関してはそうでないかも知れない。そこでガウス分布とベキ分布の違いをここで大まかに述べておく。発現量を x とすると、ガウス分布は $G(x) \propto \exp(-a(x - x_0)^2)$ という関数型で表され、平均発現量 x_0 を中心に対称な釣り鐘型の形をしている（図 4）。それに対しベキ分布は $P(x) \propto x^{-b}$ ($b > 0$) という関数型で表され、 $x = 0$ から単調に（急激に）減少する関数である（図 4）。どちらの分布も十分大きな x （ガウス分布の場合は x_0 より十分大きな x ）が起こる頻度は非常に低い事が分かる。しかし実は、この部分での両者の振る舞いは全く異なり、またこの事が発現量のばらつきの特徴、もしくは「ばらつき」そのものに対する見方、を考える上で重要になる。

ガウス分布とベキ分布の違いは、発現量 (x) とその頻度 ($G(x)$, $P(x)$) の対数 ($\log x$, $\log G(x)$, $\log P(x)$) を軸とした両対数プロットを見ると、顕著になる（図 4）。ガウス分布を

両対数プロットで見ると、 x_0 よりある程度 ($\sim\sqrt{1/a}$ 程度) 大きい値まで x が増大すると、頻度 $G(x)$ が急激に ($-\infty$ に向けて加速度的に「ストーン」という感じで) 減少しているのが分かる。つまり x の分布がガウス分布に従うという事は、 x が x_0 と比べ $\sqrt{1/a}$ 程度より大きな桁の値を取る頻度がほぼ無視できる事を意味している。このような場合、発現量 x は、例えばらつきがあったとしても、大体 x_0 程度の値を取ると考えて差し支えない。

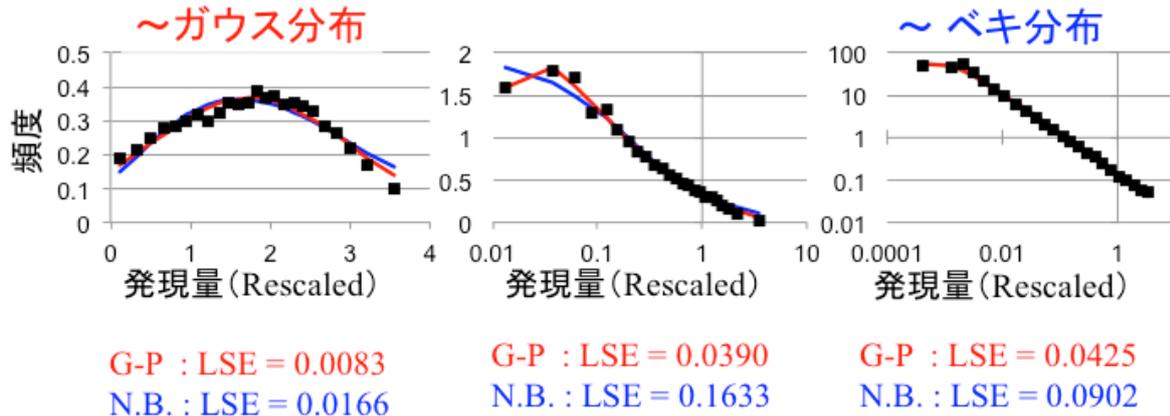


図 3. 遺伝子発現量の典型的な分布関数型。左からガウス分布，中間の分布，ベキ分布。中間の分布は横軸を，ベキ分布では横軸縦軸双方を対数軸としてプロットしている。グラフ中の赤と青の曲線はそれぞれ G-P 分布と（一般化した）負の二項分布（N.B.）でのフィッティング曲線。各グラフ下の数値は，G-P 分布と負の二項分布でのフィッティングした際の最小自乗誤差を表す。

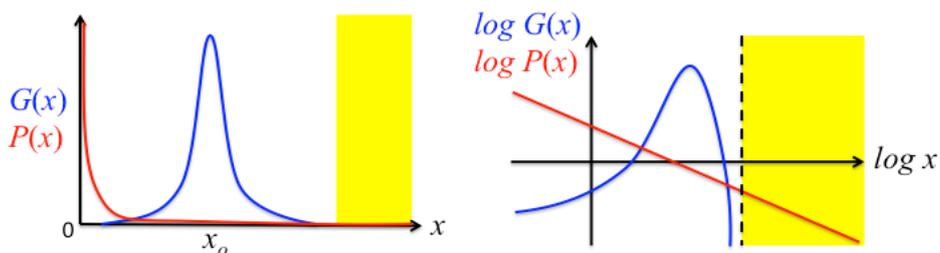


図 4. ガウス分布 ($G(x) \propto \exp(-a(x - x_0)^2)$) : 青) とベキ分布は ($P(x) \propto x^{-b}$ ($b > 0$)) : 赤) の概形。左図が通常スケールのプロット，右図が両対数プロットでの各分布の概形。黄色の領域は $x = x_0$ よりも十分大きな x の領域を表す。

それに対しベキ分布を両対数プロットで見ると、 x の増大に対し頻度 $P(x)$ は減少するが、その割合は一定で（「だらだら」と線形で減少）、ガウス分布のような急激な減少をすることがない。これはガウス分布での経験を踏まえると、「この程度以上は無視してよい」という値が存在しない事を意味しており、発現量 x はどんなに大きな桁の値でも（稀ではあるが）有限の頻度でとりうる事を意味している。これはまた、例えばある時点で得られている発現量 x の値からその平均値や分散値等の特徴量を求めたとしても、それよりも何桁も大

きな x の値が後から有限の頻度で出てきて、これらの特徴量を大幅に変更してしまう可能性がある事を意味している。つまり平均値等が安易に発現量 x の特徴量となり得ない事を意味しているのである。（実は地震のエネルギーとそのような地震が起こる頻度の関係もベキ分布になっている（グーテンベルグ・リヒター則（Gutenberg et al. 1941）。この事は近年我々も体験している通り、地震は小さなものから桁違いに大きなエネルギーのものまで、有限の頻度で起こりうる事を示している。）

勿論、遺伝子発現量は無限大になる事はなく何らかの上限があるため、発現量の分布が理想的なベキ分布になる事は無く、近似的にある発現量の大きさの範囲でだけベキ分布になっている、というのが現実である。しかし発現量の分布の中にそのようなベキ分布になっている部分がある程度含まれていると、上記で述べたように、平均値等を発現量の特徴量としてみる事が的外れになる。そして論文 X で発現量の分布がベキ分布と推定されたものは正に、発現量が「何桁にもわたってばらつく」「平均値等が特徴量にならない」ものとなっている。

5. 「ばらつき」の背景と機能との関係

前節までの話で、シロイヌナズナの RNA シークエンスデータから、遺伝子毎に発現量の「ばらつき方」が異なる事、少なくない遺伝子で発現量の分布が近似的に「ベキ分布」に従うという、「たちの悪いばらつき方」をする事などが推定されている事を紹介した。すると、実際にはどのような分布の型に従っていると考えられるのであろうか？またどうしてこのような「ばらつき方」の多様性が現れるのであろうか？等という疑問が自然に出てくるのではと思われる。

詳細は論文 X に譲るとして、その問いに対する一つの考察を紹介する。まず結果を述べると、ある遺伝子の発現量 x の従う分布 $Q(x)$ は、次のような式に従うと考えられる。

$$Q(x) = A \frac{K+x}{fx} x^{\frac{2gK-K^2}{f^2}} e^{-\frac{1}{f^2} \left[\frac{gK^2}{x} + (2K-g)x + \frac{x^2}{2} \right]}$$

ここで f, g, K , は、後で説明するこの遺伝子の発現制御に関わるパラメータから決まる定数で、 A は規格化（ヒストグラムの度数の和を 1 にスケールする）定数である。この $Q(x)$ を眺めてみると x^c (c は定数) と表されるベキ分布部分と $\exp(-d(x-x_0)^2)$ と表されるガウス分布部分の積を含んでいる事が分かる。つまり $Q(x)$ はガウス分布とベキ分布を掛け合わせて混合させたような分布となっているのである。そこで論文 X ではこの分布を「ガウス-ベキ混合分布 (G-P 分布)」と名付けている。

ではこの G-P 分布は、シロイヌナズナの遺伝子発現量の分布をどれほど表現できるのか？そこで、RNA シークエンスのデータ解析でしばしば仮定される負の二項分布（例えば Gierliński et al. 2015, Marioni et al. 2008, Robinson et al. 2008, 等）と G-P 分布で、前節で得られたデータのみから推定された各遺伝子の発現量分布のフィッティングを行い、双方の分布の妥当性を比較する（図 3）。ここで負の二項分布については、パラメータを一つ加え、整数値に限らない拡張されたものを用いる。その結果、パラメータ数はどちらも 3 つと同じであるが、多くの場合で負の 2 項分布より G-P 分布 $Q(x)$ の方がより良いフィッティングが可能である事が分かる。

また従来よく使用されていた負の二項分布と異なり、G-P 分布 $Q(x)$ に含まれる各パラメータは、以下のような遺伝子の発現制御に関わるパラメータと関連している (図 5。 $Q(x)$ の導出は論文 X を参照の事)。まずある遺伝子に対するその上流遺伝子からの寄与 (転写活性化) の強さを G , その転写・翻訳の産物の分解速度を C , この遺伝子から下流の遺伝子 (群) への活性化を特徴付ける (平均的な, ミカエリス定数のような) 量を K , そしてその下流遺伝子 (群) からのフィードバックに伴うノイズの強さを F とする。このとき f, g はそれぞれ $f = F/C, g = G/C$ と与えられる。

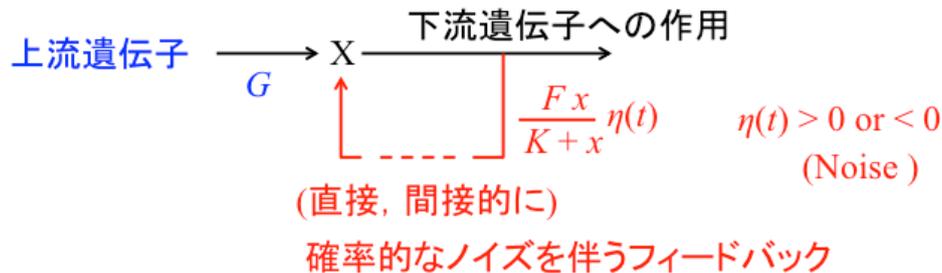


図 5. ある遺伝子 X の発現制御ネットワークのモデル図。一般に遺伝子発現量は、上流遺伝子からの作用による発現活性 (G) と下流遺伝子への作用からの確率性を伴うフィードバック ($Fx\eta(t)/(K+x)$ に比例) を受ける。ここで $\eta(t)$ は時刻 t に依存してランダムに変化する平均 0 のノイズを表す。

この事から次の事が考えられる。ある遺伝子が下流遺伝子を活性化するがそのフィードバックの影響が無視出来る場合、つまり f や K が小さいとき、この遺伝子の発現量分布 $Q(x)$ はガウス分布に漸近し、逆にこれらが g に比べ大きいとき、 $Q(x)$ のベキ分布部分が顕著になってくる。逆に言えば、発現量の分布がベキ分布に従っている遺伝子は確率的なフィードバック制御の影響を強く受けており、逆に発現量の分布がガウス分布に従う遺伝子では、フィードバック制御の影響を殆ど受けていない事が示唆される。つまり、発現量のばらつきの分布が分かれば、その遺伝子があるとき実際に受けている制御の様相を推測出来るのである。

このように発現量の多様なばらつき方の背景には、その遺伝子がある時点で受けている実効的な制御の違いが存在している。更に各遺伝子の発現量のばらつき方の傾向は、その遺伝子が持つ機能とも関連している (表 1)。先の節でも述べたが、例えば生存に必須な遺伝子群や代謝・エネルギー変換に関わる遺伝子群の多くは、発現量がガウス分布に従う。この事から、これらの遺伝子の制御にはフィードバックが殆ど影響せず、その結果として平均値で特徴付けられる安定的な発現量を維持する傾向にあると考えられる。一方、転写因子等をコードする遺伝子群の多くは、発現量がベキ分布に従っている。この事から、これらの遺伝子は (相互) フィードバックの影響により、状況に応じて数桁に渡る大きな発現量変化を実現出来るようになっている、と推測される。更に別の研究から、同一条件下において発現量のばらつきが大きい遺伝子は、環境の変動に対し大きな発現量の変動を示す傾向にある事が知られている (Hirao et al. 2015)。この知見に加え、一般的に発現量がベキ分布を示す遺伝子の方

がガウス分布を示すものより発現量のばらつきが大きい事を踏まえると、ベキ分布を示す転写因子等をコードする遺伝子群の多くは、フィードバックの影響により環境変動に対し敏感な発現量変化を可能にしているとも推測される。つまり発現量のばらつき方は、各遺伝子の機能とその遺伝子が受けている制御の双方と密接に関係し、またそれによって、各遺伝子が受ける発現制御と機能を結びつけているのである。

6. まとめと願望：本考察の知見と得られたデータを有効活用して頂けたら

ここまで、遺伝子発現量に現れるばらつき（確率性）の特徴、その特徴が現れる背景、及びその生命機能との関連について、論文 X の内容を参照しつつ紹介した。ここまで詳しくは述べなかったが、上記の考察では、よく実験で用いられる生育条件下（明暗が変化は 12 時間毎）でシロイヌナズナを育て、7 日目及び 22 日目の、明るくなってから 1 時間後、7 時間後、13 時間後（暗くなってから 1 時間後）、19 時間後、の計 8 生育時期に、各時期につき 20 以上の Replicate から RNA が採取され、解析に利用されている。このようなサンプリングは従来ほぼ行われてこなかったであろう。同時にそれらのデータの解析も、従来踏襲されてきたものから大きく逸脱した方法で試みられている。そういった事を踏まえ、ここまで読んで下さった方々に向けた願望を、最後に幾つか述べさせて頂きたいと思う。

論文 X では、時期依存的に各遺伝子の揺らぎ方を分類し、表にまとめている。勿論全ての遺伝子について解析出来ている訳ではないが、各条件につき 1 万遺伝子種程度の情報は得られている。そこで、ある遺伝子に着目して実験をする状況になったとき、一度まずこの表をご参照頂ければと思う。そうする事でこの遺伝子の制御の背景やその変化を推測出来るだけでなく、実際に発現量の計測の際どの程度結果がばらつくものなのかも予測出来るので、「外れ値」なのか「考慮すべき結果」なのかの判定にも役立つのではと考えられる。

また、ストレス等の何らかの処理を加える場合に、注目する遺伝子の処理前（無し） - 処理後（有り）での発現量に対し、その平均だけでなく「ばらつき方の変化」にも注目して頂ければと思う。発現量のばらつき方の変化は、背景にある発現制御の変化（例えばフィードバックの ON-OFF）を意味すると予測されるため、未知の制御関係を探せる可能性があるからである。これら 2 つの事は、これまでの実験研究の道筋から少し寄り道をして頂くだけで、オプションに有益な情報を提供してくれるのではと思われる。

最後に、これは見方に依っては一部敗北宣言になるのかもしれないが、このようなデータの更なる有効な活用法をお考え頂けたらと考えている。発現量データは論文 X の Supplement のファイルとして誰でも自由に取得・利用出来るようになっているので、ここで紹介した知見を遥かに凌駕する考察が現れる事を期待したい。

謝辞

本小論の執筆に当たり、その基となる論文の共著者である神谷麻梨氏、手塚あゆみ氏、田邊章洋氏に感謝致します。また本小論の執筆の機会を頂きました川出健介氏、北沢美帆氏に感謝致します。

引用文献

- Awazu, A., Tanabe, T., Kamitani, M., Tezuka, A., Nagano, J. A., 2018. Broad distribution spectrum from Gaussian to power law appears in stochastic variations in RNA-seq data. *Sci. Rep.* **8**, 8339.
- Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E. & Huang, S. 2008. Transcriptome-Wide Noise Controls Lineage Choice in Mammalian Progenitor Cells. *Nature* **453**, 544–547.
- Cortijo, S., Aydin, Z., Ahnert, S., & Locke, J. C. (2019). Widespread inter-individual gene expression variability in *Arabidopsis thaliana*. *Molecular systems biology*, **15**, e8591.
- Elowitz, M.B., Levine, A.D., Siggia, E.D. & Swain, P.S. 2002. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183–1186.
- Furusawa, C., Suzuki, S., Kashiwagi, A., Yomo, T. & Kaneko, K. 2005. Ubiquity of Log-Normal Distributions in Intra-Cellular Reaction Dynamics. *BIOPHYSICS* **1**, 25–31.
- Gierliński, M., Cole, C., Schofield, P., Schurch, N.J., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G., Owen-Hughes, T., et al. 2015. Statistical Models for RNA-Seq Data Derived from a Two-Condition 48-Replicate Experiment. *Bioinformatics* **31**, 3625-3630.
- Gutenberg, B. & Richter, C.F. 1941. Seismicity of the earth. *Geol. Soc. Am. Sp. Pap.* **34**, 131
- Hirao, K., Nagano, A.J. & Awazu, A. 2015. Noise–Plasticity Correlations of Gene Expression in the Multicellular Organism *Arabidopsis Thaliana*. *J. Theo. Biol.* **387**, 13–22.
- Karthik, D., Stelzer, G., Gershanov, S., Baranes, D. & Salmon-Divon, M. 2016. Elucidating Tissue Specific Genes Using the Benford Distribution. *BMC Genomics* **17**, 595.
- Lehner, B. 2010. Conflict Between Noise and Plasticity in Yeast. *PLoS Genet.* **6**, e1001185.
- Lehner, B. & Kaneko, 2011. K. Fluctuation and Response in Biology. *Cell. Mol. Life Sci.* **68**, 1005–1010.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M. & Gilad, Y. 2008. RNA-Seq: an Assessment of Technical Reproducibility and Comparison with Gene Expression Arrays. *Genome Res.* **18**, 1509–1517.
- Meinke, D., Muralla, R., Sweeney, C. & Dickerman, A. 2008. Identifying Essential Genes in *Arabidopsis Thaliana*. *Trends Plant Sci.* **13**, 483–491.
- Nagano, A.J., Sato, Y., Mihara M., Antonio, B.A., Motoyama, R., Itoh, H., Nagamura, Y. & Izawa T. 2012. Deciphering and Prediction of Transcriptome Dynamics Under Fluctuating Field Conditions. *Cell* **151**, 1358–1369.
- Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L. & Weissman, J.S. 2006. Single-Cell Proteomic Analysis of *S. cerevisiae* Reveals the Architecture of Biological Noise. *Nature* **441**, 840–846.
- Robinson, M.D. & Smyth, G.K. 2008. Small-Sample Estimation of Negative Binomial Dispersion, with Applications to SAGE Data. *Biostatistics* **9**, 321–332.
- Sato, K., Ito, Y., Yomo, T. & Kaneko, K. 2003. On the Relation Between Fluctuation and Response in Biological Systems. *Proc. Natl. Acad. Sci. USA* **100**, 14086–14090.
- Singh, G.P. 2013. Coupling Between Noise and Plasticity in *E. Coli*. *G3 (Bethesda)* **3**, 2115.
- Schurch, N.J., Schofield, P., Gierliński, M, Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G., Owen-Hughes, T., et al. 2016. How Many Biological Replicates Are Needed in an RNA-Seq Experiment and Which Differential Expression Tool Should You Use? *RNA* **22**, 839-851.

- Smith, G.R. & Birtwistle, M.R. 2016. A Mechanistic Beta-Binomial Probability Model for mRNA Sequencing Data. *PLoS One* **11**. e0157828.
- Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A. & Xie, X.S. 2010. Quantifying E. Coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329**. 533–538.
- Tirosh, I. & Barkai, N. 2008. Two Strategies for Gene Regulation by Promoter Nucleosomes. *Genome Res.* **18**. 1084–109.