# Relationships between total, relative and absolute molecular abundance in aging genomics studies: Simple theoretical consideration.

Daigo Okada[1, *]

**1 Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Nanbusogo-Kenkyu-To-1, 5F, 53 Syogoin-Kawaramachi, Sakyo-ku, Kyoto 606-8507, Japan * dokada@genome.med.kyoto-u.ac.jp**

## Abstract

In most of the studies using Next Generation Sequencing, the normalization step is performed to adjust the total number of short reads to quantitate the expression of biomolecules such as mRNA. This preprocessing is based on the assumption that the total number of the sequenced reads is constant across different biological conditions and conducted to remove technical variability. Normalized expression values such as count per million are relative expression values, losing information on absolute amounts. However, in aging and cancer research, it has been reported the assumption that the total number of reads is constant across conditions does not hold. In this study, based on a simple theoretical analysis, we present that log fold-change in absolute abundance can be expressed by the observed log fold-change of relative abundance plus log fold-change of total abundance. The logic presented in this study focuses on the relationship between global and local change in omics research. It can be used as a quick check on how a differentially expressed gene in an omics study based on relative abundance should be interpreted in terms of absolute abundance.

**Keywords:** Aging, Bioinformatics.

## Introduction

Next Generation Sequencing (NGS) is widely used in the life sciences to obtain comprehensive biomolecular information. By comprehensively sequencing RNA and DNA, this technology can quantify various biological information such as transcriptome and epigenome as short read counts. In particular, the approach of comparing gene expression levels between groups to identify Differentially Expressed Genes (DEG) is a standard RNA-seq experimental design [5]. In most cases, Next Generation Sequencing (NGS) measurements are not designed to quantify absolute abundances, so a normalization step is performed to adjust the total number of reads [11]. This preprocessing is based on the assumption that the total number of the sequenced reads is constant across different biological conditions and conducted to remove technical variability. Normalized expression values such as count per million (CPM) are relative expression values, losing information on absolute amounts. Under this assumption, the changes in relative molecule count can be directly interpreted as changes in absolute molecule counts.

However, in aging and cancer research, cases have been reported in which the assumption that a total number of reads is constant across conditions does not hold.

With aging, total RNA has been reported to decrease in several species [7–9]. This can affect the interpretation of RNA-seq interpretation in aging study [3, 6]. It has been suggested that total RNA may be altered in cancer cells compared to normal cells [1]. Although there are some experimental and computational methods designed to capture the absolute abundance of mRNAs in RNA-seq [1, 4], most published studies identify the age-related gene based on relative abundance using NGS technology. The difference between aging-related genes based on relative abundance and those based on absolute abundance has not received much attention.

In this study, based on a simple theoretical analysis, we present the simple relationship established between total, relative, and absolute molecule abundance. The logic presented in this study can be used as a quick check on how a differentially expressed gene in an omics study based on relative abundance should be interpreted in terms of absolute abundance.

## The relationship between total, relative and absolute abundance

Using standard RNA-seq as an example, we derived an equation relating relative abundance to absolute abundance. The graphical examples of absolute, total, and relative abundance are shown in Figure 1. Let $X_i$ and $Y_i$ be the absolute mRNA abundance of the molecules of $i$-th gene in two groups, respectively. The total RNA abundances $T_x$ and $T_y$ are calculated as follow:

$$T_X = \Sigma_{i=1}^{N} X_i$$
$$T_Y = \Sigma_{i=1}^{N} Y_i$$

where $N$ is the number of all genes. Then, the observed relative abundance values $x_i$ and $y_i$ in the genomics study are expressed by the following equation.

$$x_i = \frac{X_i}{T_x}$$
$$y_i = \frac{Y_i}{T_y}$$

The following relationship equation holds between the log fold-change of the genes reported in the paper and the log fold-change in absolute abundance, which cannot be directly observed.

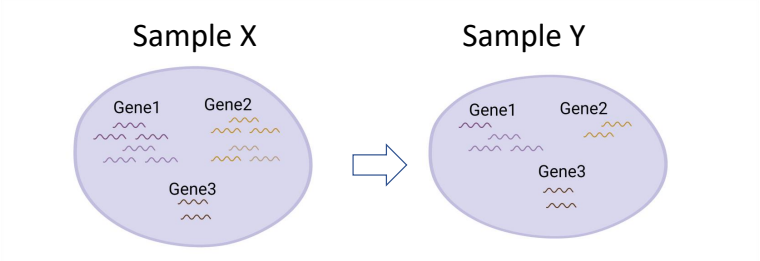
$$\log(Y_i/X_i) = \log(y_i/x_i) + \log\frac{T_y}{T_x}$$

Log fold-change in absolute abundance ($\log(y_i/x_i)$) can be expressed by the observed log fold-change of relative abundance ($\log(y_i/x_i)$ plus log fold-change of total abundance ($\log\frac{T_y}{T_x}$). The log fold-change of total abundance represents the global change in that molecule. The logic presented in this study focuses on the relationship between global and local change in omics research.

If the global log fold-change decreases in Sample Y, genes with decreased relative abundance in Y also should have decreased absolute abundance. Given that total RNA expression decreases with age, genes identified as decreased in the old sample should also be decreased in absolute abundance. In addition, the recent single-cell RNA-seq analysis study reported that the number of genes that decrease with aging is greater than the number of genes that increase with aging [10], and this proposition is also true for

absolute abundance. On the other hand, genes identified as elevated in the old sample may not necessarily be elevated in absolute abundance if the fold-change is small.

This argument can also be applied to the interpretation of the results of epigenomic data analysis using NGS. For example, a recent study examined the global aging changes in various histone marks associated with aging using mass cytometry [2]. These results about the total molecular abundance of histone marks will be useful in interpreting the other results of ChiP-seq studies to detect specific age-related genomic regions.




| | Gene1 | Gene2 | Gene3 | Total abundance |
|---|---|---|---|---|
| Absolute abundance | 6 | 6 | 2 | 14 |
| Relative abundance | 6/14 | 6/14 | 2/14 | 1 |

| | Gene1 | Gene2 | Gene3 | Total abundance |
|---|---|---|---|---|
| Absolute abundance | 4 | 2 | 2 | 8 |
| Relative abundance | 6/8 | 6/8 | 2/8 | 1 |

**Figure 1.** Graphical example of the absolute, total and relative abundance. The absolute abundance is the number of molecules in the mRNA sample. The total abundance is the sum of the absolute abundance for all genes and is examined in molecular biology experiments that differ from omics experiments. The relative abundance is the value observed in omics experiments using NGS. In this study, we have shown the log fold-change in the absolute abundance of each molecule can be expressed as the sum of the log fold-change in the relative abundance and the log fold-change in the total abundance. The figure was created with Biorender (https://www.biorender.com/).

# Conclusion

Log fold-change in absolute abundance can be expressed by the observed log fold-change of relative abundance plus the log fold-change of total abundance. The logic presented in this study can be used as a quick check on how a differentially expressed gene in an omics study based on relative abundance should be interpreted in terms of absolute abundance.

# Declarations

## Funding

Not applicable.

## Conflict of interest

The author has no conflict of interest to be declared.

### Ethics approval

Not applicable.

## References


1. H. Cai, X. Li, J. He, W. Zhou, K. Song, Y. Guo, H. Liu, Q. Guan, H. Yan, X. Wang, et al. Identification and characterization of genes with absolute mrna abundances changes in tumor cells with varied transcriptome sizes. *BMC genomics*, 20(1):1–12, 2019.

2. P. Cheung, F. Vallania, H. C. Warsinske, M. Donato, S. Schaffert, S. E. Chang, M. Dvorak, C. L. Dekker, M. M. Davis, P. J. Utz, et al. Single-cell chromatin modification profiling reveals increased epigenetic variations with aging. *Cell*, 173(6):1385–1397, 2018.

3. K. Davie, J. Janssens, D. Koldere, M. De Waegeneer, U. Pech, Ł. Kreft, S. Aibar, S. Makhzami, V. Christiaens, C. B. González-Blas, et al. A single-cell transcriptome atlas of the aging drosophila brain. *Cell*, 174(4):982–998, 2018.

4. J. Lovén, D. A. Orlando, A. A. Sigova, C. Y. Lin, P. B. Rahl, C. B. Burge, D. L. Levens, T. I. Lee, and R. A. Young. Revisiting global gene expression analysis. *Cell*, 151(3):476–482, 2012.

5. R. Rodriguez-Esteban and X. Jiang. Differential gene expression in disease: a comparison between high-throughput studies and the literature. *BMC medical genomics*, 10:1–10, 2017.

6. T. Stoeger, R. A. Grant, A. C. McQuattie-Pimentel, K. R. Anekalla, S. S. Liu, H. Tejedor-Navarro, B. D. Singer, H. Abdala-Valencia, M. Schwake, M.-P. Tetreault, et al. Aging is associated with a systemic length-associated transcriptome imbalance. *Nature Aging*, 2(12):1191–1206, 2022.

7. N. M. Tahoe, A. Mokhtarzadeh, and J. W. Curtsinger. Age-related rna decline in adult drosophila melanogaster. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59(9):B896–B901, 2004.

8. E. Uemura. Age-related changes in neuronal rna content in rhesus monkeys (macaca mulatta). *Brain research bulletin*, 5(2):117–119, 1980.

9. V. Wulff and M. Freshman. Age-related reduction of the rna content of rat cardiac muscle and cerebellum. *Archives of biochemistry and biophysics*, 95:181–182, 1961.

10. M. J. Zhang, A. O. Pisco, S. Darmanis, and J. Zou. Mouse aging cell atlas analysis reveals global and cell type-specific aging signatures. *Elife*, 10:e62293, 2021.

11. Y. Zhao, M.-C. Li, M. M. Konaté, L. Chen, B. Das, C. Karlovich, P. M. Williams, Y. A. Evrard, J. H. Doroshow, and L. M. McShane. Tpm, fpkm, or normalized counts? a comparative study of quantification measures for the analysis of rna-seq data from the nci patient-derived models repository. *Journal of translational medicine*, 19(1):1–15, 2021.