Mathematical Biology and Bioinformatics
2016. V. 11. No. 2. P. 278–287. doi: 10.17537/2016.11.278

==========BIOINFORMATICS==========

UDC:573.2:519.2

Fractional-Stable Statistics of the Genes Expression in the Next Generation Sequence Results

©2016 Saenko V.V.[a1]

[a]*Technological Research Institute S.P. Kapitsa, Ulyanovsk State University, Ulyanovsk, Russia*

Abstract. As has been shown in the previous article [1] an application of class of the fractional-stable laws to the genes expression results obtained by DNA-microarrays leads to poor agreement between experimental and theoretical distributions. This difference can be explained by the imperfection of the technology of the gene expression determination. In this article the distributions of the gene expression obtained by Next Generation Sequence technology are investigated. In this technology the determination technique of the gene expression differs from the DNA-microarrays technology. This results to more qualitative results of an approximation. In particular, it is established that the probability density function of the gene expression has a form of shift-scale mixture of probability laws, where one of the components of the mixture is the fractional-stable distribution.

Key words: *gene expression, DNA-microarrays, Next Generation Sequence Technology, fractional-stable law, Levy's-stable law, shift-scale mixture.*

INTRODUCTION

There exist two main technology of genes expression determination today: DNA-microarrays [2, 3] and Next Generation Sequence (NGS) technology [4, 5, 6, 7, 8]. Historically the first technology was the technology of the DNA-microarray. In the basis of this technology the hybridization process lies. Oligonucleotides (probes) are attached to the solid substarte, each of which has predefined nucleotide sequence. The distance between the probes is 10^{-6} meters therefore on each square centimeter it is possible to place up to 10^8 probes. The samples under investigation are marked by fluorescent label and they are added on the microarray where the hybridization of the probe and target occurs. After the unhybridized nucleotides are washed away the fluorescent labels are excited by laser radiation. The relative nucleic acid content with defined sequence is determined by fluorescence radiation intensity. It should be noted the intensity of the fluorescence is defined by digital image of the microarray which is obtained by confocal microscope. Therefore the amount of the hybridized targets is proportional to the point brightness on the digital image.

The main principles of NGS technology the same as at DNA microarrays technology: the cyclic ferment reactions are used with following receiving of information about DNA structure in the form of image of fluorescent labels. However, the technology of the preparation of libraries targets and the method of differentiation of the different nucleotide sequences differ. In common case the sequencing process can be divided by four stages (see for example [5]). On the first stage the library of random sequences of the DNA is created. After this these sequences are ligated by adapter sequences containing universal primers templates. On the second stage the amplification of these random sequences DNA are fulfilled by polymerase chain reaction. On

[1]saenkovv@gmail.com

the third stage the primary structure of all samples is determined. The fourth stage consists in alignment of the obtained reads to the reference gene. The main advantage of NGS technology consists in possibility of determination individual nucleotides in a sequence. This allows to detect new nucleotide sequences.

A second advantage of NGS technology is possibility to determine a quantitative content of defined genes in the sample. This allows reduce an error at calculation of the gene expression levels. Really, as the result of the reads alignment is amount of some gene in the sample. Of course, there exists probability of wrong determine the reference gene. Nevertheless, the results of the gene expression calculation obtained by NGS technology will more qualitative than results obtained with the DNA-microarray technology. This due to the fact that in the DNA-microarray technology practically it is not possible to determine quantitative content of the genes since the value of the expression is determined with the intensity of the fluorescence radiation. At the same time it is unknown the amount fluorescent labels having incorporated to the sample and the radiation intensity of the single label. The background radiation is also problem. This radiation introduces the distortion into overall picture and it is not possible exclude one. All these facts lead to less qualitative results obtained with DNA microarrays in comparison with the results of NGS technology. The results of the article [1] confirm these facts. In this article the probability density functions of the gene expression levels obtained with DNA-microarrays of various manufactures are investigated. It was shown although in majority cases the distribution can be described by the fractional-stable law however empirical distributions deviate from the fractional-stable distributions. These deviation can be explained with distortions noted above.

In this article we consider the question about probability distribution of the gene expression obtained with NGS technology. It is well established that the empirical distributions are one-sided distributions, they manifest the power-law asymptotics and character of these distributions is the same for any tissues and organisms from bacteria to mammal [9]. Similar results have been reported by other authors [10, 11, 12, 13, 14].

The power-law asymptotics of the experimental distribution means that the theoretical distribution must have the asymptotics of the following form $p(x) \propto x^{-\alpha-1}, x \to \infty$. In the above work of [9], the same distribution was applied for approximation of the profiles of the gene expression in various organisms. The parameter α was shown to vary within the limits from 0.69 to 1.09. In the work [12] the authors have investigated more than 40 tissues for six organisms and the power-law distribution was obtained for all samples with value of the parameter $\alpha \in (0.86, 1)$. In the work [14] the discrete Pareto distribution was used. There were processed more than 50 human tissues, mouse tissues and yeast tissues including 30 samples of the human cancer tissues and 30 normal human tissues. In this article reported that the best approximation among the Poisson distribution, exponential distribution, logarithmic distribution, power-law distribution, paretolike distributions and mixtures of the logarithmic and the exponential distributions is the discrete Pareto distribution $p(m) = (m + b)^{-\alpha-1}/z$, where α is varying within the limits from 0.974 to 1.88. In the article [11], the authors use the double Pareto-log-normal distribution. Along with the Pareto-log-normal distribution the authors tested the Zipf-Pareto distribution, log-normal distribution, log-gamma distribution, log logistic distribution, right-side Pareto-distribution. Finally, the authors conclude that the double Pareto-lognormal distribution provides the best results.

However the authors of these articles had not considered one more sufficiently significant class of distributions the Fractional-Stable Distributions (FSD). The fractional-stable laws are limit distributions of sums of independent identical distributed random variables (see [15]). At working with these distributions the main difficulty consists in absence of explicit expressions for densities. Probably this fact is main reason excluding this distribution class from the list

of the potential candidates of theoretical densities for approximation of the genes expression. However there exist some methods of calculation of the densities which allow circumvent this difficulty. Moreover this class of the distributions possesses the characteristic property which inherent to the experimental distributions. These distributions possess by power-law asymptotics. All above noted facts make the fractional-stable distribution good candidate for approximation of the experimental data. In this article we consider the question about possibility of the approximation of the genes expression obtained with NGS technology by the fractional-stable laws.

2 FRACTIONAL-STABLE LAWS AND ESTIMATORS OF THEIR PARAMETERS

Fractional-stable laws are limit distributions of sums the independent identically distributed random variables. For the first time, FSD was introduced by Kotulski [16]. The name FSD was introduced in 2001 [17]. FSD are expressed through Mellins transformation of two stable distributions

$$q(x;\alpha,\beta,\theta,\lambda)=\int_0^\infty g(xy^{\beta/\alpha};\alpha,\theta,\lambda)g(y;\beta,1,1)y^{\beta/\alpha}dy. \tag{1}$$

Here $g(x;\alpha,\theta,\lambda)$ is the density function of the strictly stable law and $g(y;\beta,1,1)$ is the density of the one-sided strictly stable law with the characteristic function (see [18])

$$\hat{g}(k;\alpha,\theta,\lambda)=\exp\{-\lambda|k|^\alpha\exp\{-i\alpha\theta(\pi/2)\mathrm{sign}(k)\}\}. \tag{2}$$

As we can see from (1), FSD is defined by four parameters. α and β are two characteristic parameters. They vary in the limits $\alpha\in(0,2]$ and $\beta\in(0,1]$. The variation domain of the parameters θ and λ coincides with the domain of the variation respective parameters of the stable distribution and they have the same meaning, θ is the asymmetry parameter ($|\theta|\leqslant\min(1,2/\alpha-1)$) and $\lambda>0$ is the scale parameter. FSDs have the power-law asymptotics $q(x;\alpha,\beta,\theta,\lambda)\propto x^{-\alpha-1}$, $x\to\infty$. When $\beta=1$, the class of FSD passes into the class of strictly stable distributions. Indeed, when $\beta=1$ and $\theta=1$ the strictly-stable law $g(y;1,1,1)$ is the singular distribution at the point $y=1$. Hence, from (1) we obtain $\int_0^\infty g(xy^{\beta/\alpha};\alpha,\theta,\lambda)\delta(y-1)y^{\beta/\alpha}dy=g(x;\alpha,\theta,\lambda)$, where $\delta(y-1)$ is Dirac's function. In the case when $\alpha=2$, $\beta=1$ and $\theta=0$ from (1) and (2) we obtain that FSD passes into the normal distribution. Hence, the class of fractional-stable laws involves the class of stable distributions. Since the FSD are fully described by their four parameters then knowledge of these parameters allows to calculate the distribution. One of the main tasks in the work is parameter estimation upon experimental data. Shortly one can formulate this task as follows. Let $Z_1,Z_2,\ldots,Z_n$ is the sample of independent identically distributed random variables. Each random variable Z_i has the FSD. It is necessary by this sample to calculate the estimators $\hat{\alpha},\hat{\beta},\hat{\theta},\hat{\lambda}$ of the parameters $\alpha,\beta,\theta,\lambda$ of the FSD.

There exist several methods of estimation of the FSD parameters. The first method has been described in the work [19] and the estimators were obtained on basis of the method of moments. However, it has been established that usage of these estimators lead to wrong parameters values. The second method [20] is based on the maximum likelihood method. However this method is also unhandy for approximation of the gene expression. Really, in this method it is necessary to calculate the probability density function at the points defined by the value of the random variables $Z_1,Z_2,\ldots,Z_n$. The main difficulty here consists exactly this since there does not exist explicit expression for the FSD. In the article by [20] the author avoided this difficulty by usage of the local estimator of the Monte Carlo method for calculation of the symmetric (at $\theta=0$) FSD. However, it is not possible to apply this method for describing the expression data since the $\theta\neq0$. Even in the case $\theta=0$ to use this method would be a very tedious task since in

this case it would be necessary to estimate the density at each point $Z_1, Z_2, \ldots, Z_n$. Taking into account that expression data contain tens thousand values then estimation will takes big time span.

The third method [21] is based on minimization of the distance between theoretical and empirical distributions. As the distance between two distributions the χ^2 distance was chosen

$$d(\mathbf{P}_{\mathbf{\Theta}}, \mathbf{Q}) = \sum_{k=1}^{n} \frac{(N\mathbf{P}_{\mathbf{\Theta}}(\Delta_k) - \nu_k)^2}{N\mathbf{P}_{\mathbf{\Theta}}(\Delta_k)}. \tag{3}$$

Here $\mathbf{Q}$ is experimental distribution, $\mathbf{P}_{\mathbf{\Theta}}$ is theoretical distribution, $\mathbf{P}_{\mathbf{\Theta}}(\Delta_k)$ is the theoretical probability corresponding to the interval Δ_k, $\mathbf{\Theta}$ is a vector of parameters, Δ_k is the partition of the domain under consideration $R \equiv \{x : a \leqslant x \leqslant b\}$ on n disjoint intervals $\bigcup_{k=1}^{n} \Delta_k = R$, ν_k is the number of observations fallen into interval Δ_k. As a result, the estimator $\hat{\mathbf{\Theta}}$ of the parameters $\mathbf{\Theta}$ obtains the values of the $\hat{\mathbf{\Theta}}$ at which the distance $d(\mathbf{P}_{\hat{\mathbf{\Theta}}}, \mathbf{Q})$ possesses the most minimal value. We will use this estimator for estimation of the FSD parameters.

Applying this algorithm to the considered task the distribution $\mathbf{Q}$ in (3) corresponds to the experimental distribution of the gene expression and distribution $\mathbf{P}_{\mathbf{\Theta}}$ corresponds to the FSD with parameters $\mathbf{\Theta} = (\alpha, \beta, \theta, \lambda)$. The FSD is estimated by Monte Carlo method. Namely, the histogram of the density is constructed. The fractional-stable random variables are simulated with algorithm

$$Z(\alpha, \beta, \theta, \lambda) = \lambda Y(\alpha, \theta)/[S(\beta, 1)]^{\beta/\alpha}, \tag{4}$$

where $Y(\alpha, \theta)$ $S(\beta, 1)$ and $S(\beta, 1)$ are stable and one-sided stable random variables with characteristic function (2). For simulation of the random variable $Y(\alpha, \theta)$ is used the algorithm described in [22] and for simulation the random variable $S(\beta, 1)$ was used the algorithm [23].

3 EXPERIMENTAL DATA AND RESULTS

As a source of the experimental data the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) database was chosen. Several experimental series were chosen from the database. They were obtained with the NGS technology (the results are expressed in FPKM units). In particular, the experimental data for the human (series GSE44875 and GSE50760), macaque (series GSE53690), mouse (series GSE53690 and GSE53110) and for drosophila (series GSE54600 and GSE41487) were processed. The results of processing are shown below. It is established that the experimental distributions have fractional-stable character within the range of $1 \div 1000$ FPKM units only. Beyond this interval deviations of the experimental distribution from the FSD are observed.

Methods of the analysis of experimental data are as follows. The expression data are divided into three disjoint domains $\mathcal{R} = \{x : E_{\min} \leqslant x \leqslant E_{\max}\}$, $\mathcal{R}_1 = \{x : x \leqslant E_{\min}\}$ $\mathcal{R}_2 = \{x : x \geqslant E_{\max}\}$, and behaviour of the probability density function within each domain is considered. It is visually established that in the domain $\mathcal{R}$ the behaviour of the empirical distribution is characteristic for FSD. Therefore the expression values satisfying to the condition $X \in \mathcal{R}$ are chosen for the approximation of the experimental data within this domain. The obtained sample is considered as a sample $X_i, \ldots, X_n$ of independent identically distributed random variables. Further parameters of the FSD are estimated with obtained sample $X_i, i = 1, \ldots, n$.

For checking the assumption about fractional-stale distribution of the random variables X_i the hypothesis H_0 is suggested which can be formulated as follows: *the random variables* $X_i, i = 1, \ldots, n$ *have the fractional-stable distribution with density* (1). Since the parameters of this distribution are unknown then the hypothesis H_0 is a complex one. To obtain the estimators $(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\lambda})$ of the parameters $(\alpha, \beta, \theta, \lambda)$ the statistical algorithm of the estimation of the fractional stable parameters based on the minimum distance method [21] is used. The

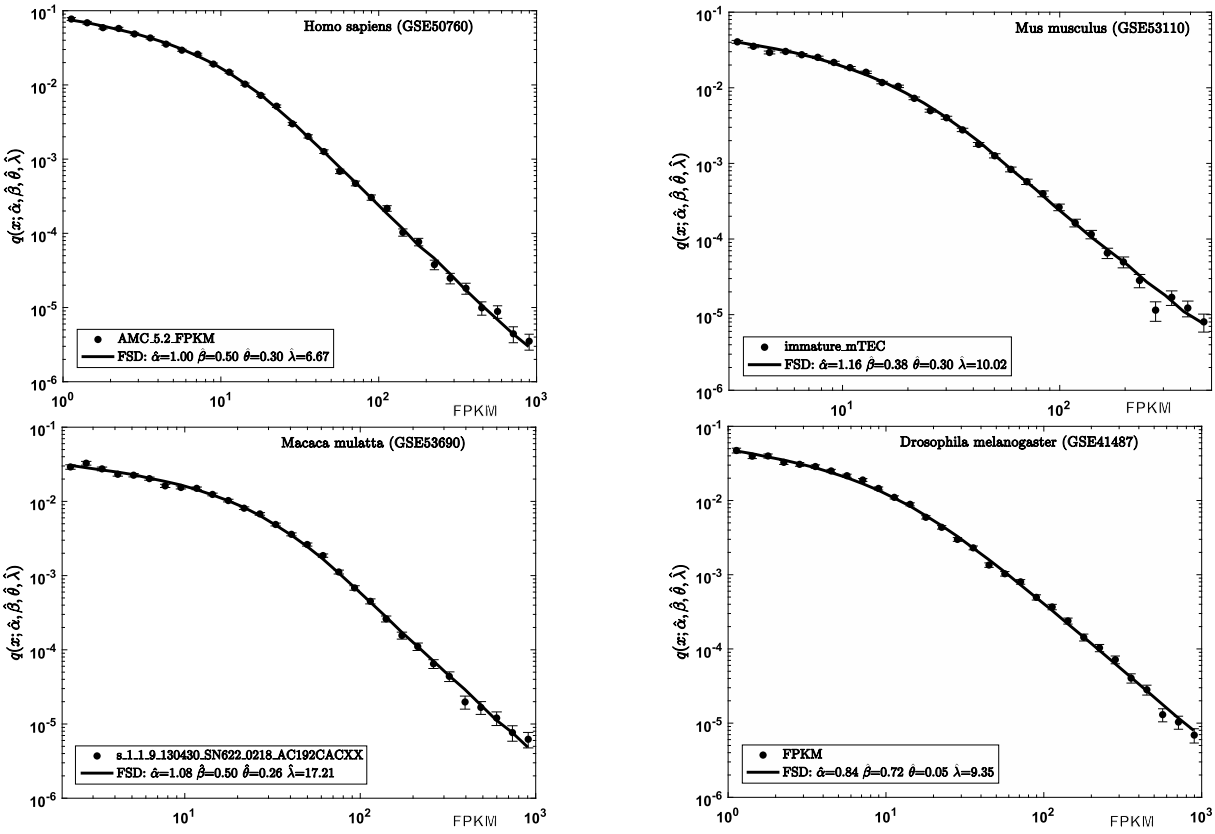

Figure 1. Probability density distribution of gene expression for human (*Homo sapiens*), mouse (*Mus musculus*), macaque (*Macaca mulata*) and drosophila (*Drosophila melanogaster*) tissues. Solid circles are experimental data, solid lines are FSDs for the estimated values of the parameters $\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\lambda}$. The values of the parameters are shown on the figure.

hypothesis H_0 was chi-square Fisher tested. As it was noted above, the main difficulty of using the FSD consists in the absence of explicit expression for densities. Therefore a numerical method is used for calculating the density. In this work the histogram estimation of the distribution is used for calculation of the assumed distribution $q(x; \hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\lambda})$. The sample of random variables $Z(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\lambda})$ was simulated and the histogram of the density was computed. The volume of the modeled sample is 10^6 random variables for reducing the statistical error.

The approximation results for experimental densities of the gene expression within domain $\mathcal{R}$ are presented in the Figure 1. It was established that the boundaries of the domain $\mathcal{R}$ are defined by the values $E_{\min} \gtrsim 1, E_{\max} \lesssim 1000$ for the processed data. As it is seen from the figures the FSD proves good to describe experimental data within this domain. The chi-square Fisher's test confirms this conclusion. For the experimental data shown on the figures this test does not reject the H_0 hypothesis at the significance level of 1%. More complete results of the parameter estimation and checking the H_0 hypothesis are presented in the Table 1 (See Appendix 1). The results of chi-square Fisher testing are presented in the column H_0 (where "0" means that the test rejects the hypothesis H_0 and "1" means the test does not reject the hypothesis). As we can see from the table the hypothesis about the fractional-stable nature of the experimental data is not rejected almost for all processed data. Since we know the boundaries of the domain $\mathcal{R}$ we can calculate the probability contained inside: $p = \mathsf{P}\{E_{\min} \leqslant X \leqslant E_{\max}\} = (1/N) \sum_{i=1}^{N} \mathbf{I}(E_{\min} \leqslant X_i \leqslant E_{\max})$ where $\mathbf{I}(A)$ is the indicator of the event A, N is the total sample volume. The values of the probability p are also presented in the Table 1. As we can see from 36% to 55% of the data in the initial sample have fractional-stable law of distribution.

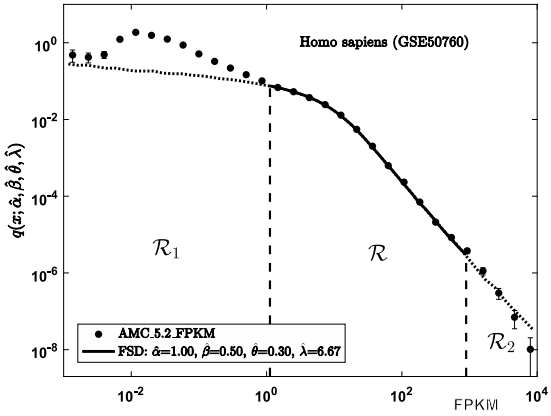

Figure 2. The probability density function of the gene expression obtained with the NGS technology within the full range of data variation. Solid circles are experimental data for human tissues (series GSE50760), solid curve is FSD within the domain $\mathcal{R}$ for shown values of the parameters, dashed curve is FSD for the same parameter values and $x \notin \mathcal{R}$.

Let us consider the behaviour of the experimental density within the domains $x \notin \mathcal{R}$. In the Figure 2 the probability density function is shown within the full range of experimental data variation. It is seen from the figure that the deviation of the empirical from theoretical distribution within the domain $x \in \mathcal{R}_2$ is weak. Indeed, within the domain $\mathcal{R}_2$ the experimental distribution has power-law dependence $x^{-\alpha-1}$ well described by FSD extended into this domain. However, attempts to check the hypothesis H_0 for the domains $\mathcal{R} + \mathcal{R}_2$ lead to the rejecting of this hypothesis. The reason underlies in fluctuations of the empirical and theoretical (since the theoretical distribution is estimated by Monte Carlo method) distributions. These fluctuations are caused by statistical error. Such increasing of statistical error $\mathcal{R}_2$ is caused by the small sample volume. Indeed, in Table 1 the probability $p_2 = \mathsf{P}\{X \in \mathcal{R}_2\}$ is presented. As it is seen from the Table the domain $\mathcal{R}_2$ contains no more than 2% of all data. Therefore, the increasing of the fluctuation leads to the increasing of the calculated value of χ^2 statistics. This can lead to false-negative result of checking the hypothesis H_0. Therefore, without the loss of generality, we can claim that within the domain $\mathcal{R}_2$ experimental data also follow the fractional-stable distribution, but only in case of power-law dependence.

Within the domain $\mathcal{R}_1$ the empirical distribution has absolutely different behaviour from that of FSD (see Fig. 2). One of possible reasons of such behaviour can consist in appearing of an additional component in the distribution of the experimental data. This means that within this domain one part of the data has FSD while another part of the data is distributed according to some unknown law of distribution $f(x;\mathbf{\Phi})$, where $\mathbf{\Phi}$ is the vector of parameters of this distribution. Such models are well known and named shift-scale mixture of distributions. In common case such shift-scale mixtures have the following form:

$$F(x) = \sum_{j=1}^{M} w_j f_j((x-\mu_j)/\lambda_j;\Phi_j), \tag{5}$$

where w_j are weight coefficients ($w_1 + w_2 + \cdots + w_M = 1$), $f_j(x;\Phi_j)$ are components of this mixture, M is the total number of components, Φ_j are parameters of the j-th component of the mixture, μ_j and λ_j are shift and scale parameters of the j-th component of the mixture.

On this assumption we can conclude that the experimental data of the gene expression within the domain $\mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}$ most probably are described by the shift-scale mixture with two components

$$F(x) = w_1 q(x;\alpha,\beta,\theta,\lambda) + w_2 f(x;\mathbf{\Phi}).$$

The first component $q(x;\alpha,\beta,\theta,\lambda)$ is FSD (1), the second component $f(x;\mathbf{\Phi})$ is the law of distribution which is unknown for the time being. The influence of the component $f(x;\mathbf{\Phi})$ dominates within the domain $\mathcal{R}_1$ and, at the same time, the influence of this component within the domain $\mathcal{R} + \mathcal{R}_2$ is negligibly small. This leads to the fact that the FSD describes the experimental data within the domain $\mathcal{R} + \mathcal{R}_2$ appropriately.

4 CONCLUSION

Possibly, one of the reasons of poor agreement of the experimental and model distributions in [1, 24] consists in distortions introduced during the stage of determining the amount of the hybridized material. Indeed, within the DNA microarray technology gene expression is determined with the intensity of the fluorescence radiation of the material joined to probes. On this stage the radiation of the substrate overlaps the required signal from the hybridized material, being parasitic. Moreover, the distortions are also introduced on the next stage which is the digitization of the image. Since the intensity of the fluorescence is determined not by direct measurements but by the brightness of dots in the digital image, this process also introduces the distortion to the expression data. All these facts lead to the problems of probability distribution law detection.

Another principle underlies the method of calculating gene expression in the NGS technology (see [5, 4]). In this technology the DNA-polymerase attaches only one fluorescently modified nucleotide being complementary to the base template. Since each type of the nucleotide is labeled with own color, this allows to identify it unambiguously. As a result only one nucleotide is recognized for one cycle. After multiple cycle repetition a file is created containing recognized DNA fragments and consisting of nucleotides codes. Then the alignment of these fragments to a reference genome is fulfilled and gene expression in FPKM or RPKM units is calculated. Thus, the NGS technology of gene expression is derived by direct calculation of the synthesised amount of the cell molecules. Hence, more qualitative results in approximation of the experimental data are expected.

This assumption is confirmed by the obtained results. It is established that the distribution of gene expression has the form of a shift-scale mixture of distributions defined by the formula (5) where one of the components is defined by FSD (1). The domain boundaries where each of the components dominates are established. In particular, within the domain $\mathcal{R} + \mathcal{R}_2$ the FSD dominated. Influence of other mixture components are negligibly small. It is possible to assert that within this range of values gene expression is described by the fractional-stable distributions. This assumption is confirmed by the chi-square Fisher test. Within the domain $\mathcal{R}_1$ the influence of other components of the mixture becomes significant. This fact is expressed by the appearance of the ”hump” in the experimental distribution (see Fig. 2). We failed to determine the type of this mixture component currently. However taking into account that influence of these components in the domain $\mathcal{R} + \mathcal{R}_2$ are negligibly small we can claim that these components rapidly decrease with the increasing of x.

This work was supported by the Ministry of Education and Science of the Russian Federation (No. 6.1617.2014/K) and Russian Foundation for Basic Research (No. 16-04-00504).

APPENDIX 1.

Table 1. The results of gene expression approximation obtained with NGS technology by the FSD and chi-square Fisher testing of the hypothesis H_0. The results are presented for human (series GSE50760), mouse (series GSE53110), macaque (series GSE53690), drosophila (series GSE54600)

Organism	Channel name	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\lambda}$	H_0	p	p_2
Human	AMC_2.2_FPKM	1.01	0.51	0.30	6.72	1	0.543	0.002
	AMC_3.2_FPKM	0.98	0.46	0.30	6.91	1	0.536	0.003
	AMC_5.2_FPKM	1.00	0.50	0.30	6.67	1	0.540	0.002
	AMC_6.2_FPKM	0.99	0.40	0.35	6.52	0	0.541	0.002
	AMC_7.2_FPKM	0.98	0.50	0.30	6.82	1	0.545	0.002
	AMC_8.2_FPKM	0.97	0.40	0.31	7.33	0	0.540	0.003
	AMC_9.2_FPKM	0.98	0.50	0.30	6.87	1	0.545	0.002
	AMC_10.2_FPKM	1.00	0.45	0.31	6.98	1	0.542	0.003
	AMC_12.2_FPKM	0.99	0.44	0.30	7.25	1	0.546	0.003
	AMC_13.2_FPKM	0.95	0.39	0.30	7.42	1	0.537	0.003
	AMC_17.2_FPKM	1.01	0.36	0.34	7.89	1	0.567	0.003
	AMC_18.2_FPKM	0.96	0.39	0.30	7.43	0	0.539	0.003
	AMC_19.2_FPKM	0.99	0.54	0.31	7.21	0	0.555	0.002
	AMC_20.2_FPKM	0.99	0.49	0.30	7.25	1	0.546	0.003
	AMC_21.2_FPKM	0.99	0.38	0.30	7.35	1	0.544	0.003
Mouse	cTEC	1.20	0.54	0.15	12.00	1	0.440	0.001
	mTEC	0.91	0.23	0.65	7.70	0	0.515	0.001
	immature_mTEC	1.16	0.38	0.30	10.02	1	0.515	0.001
	mature_mTECE	1.06	0.17	0.45	9.53	1	0.470	0.001
	Aire_neg_mTEC	1.05	0.35	0.32	9.62	1	0.508	0.001
	Aire_pos_mTEC	0.94	0.21	0.60	8.61	0	0.500	0.001
	Aire_knockout_mTEC	1.05	0.37	0.32	9.79	1	0.462	0.001
Macaque	s_1_1_7_130430	1.10	0.38	0.28	17.85	1	0.554	0.005
	s_1_1_9_130430	1.08	0.50	0.26	17.21	1	0.548	0.005
	s_2_1_20_130607	0.97	0.57	0.40	16.32	1	0.599	0.005
	s_1_1_12_130430	1.07	0.66	0.27	16.58	0	0.556	0.006
	s_1_1_22_130607	1.10	0.61	0.30	17.27	0	0.557	0.005
	s_2_1_25_130607	1.11	0.49	0.28	17.31	1	0.558	0.006
	s_1_1_11_130430	1.13	0.52	0.29	17.76	1	0.559	0.005
	s_1_1_10_130430	1.13	0.59	0.25	17.70	1	0.555	0.006
	s_1_1_16_130430	1.11	0.63	0.25	18.19	1	0.554	0.005
	s_1_1_18_130430	1.19	0.51	0.23	19.10	1	0.550	0.005
	s_1_1_15_130430	1.16	0.63	0.19	19.31	1	0.535	0.004
	s_1_1_19_130430	1.12	0.36	0.30	18.84	1	0.552	0.006
	s_1_1_5_130430	1.14	0.46	0.21	20.34	1	0.541	0.005
	D_Yki_1	1.04	0.46	0.27	18.68	1	0.415	0.010
	D_Yki_2	1.00	0.37	0.39	18.38	1	0.369	0.011
	D_Yki_3	1.10	0.59	0.20	21.95	1	0.374	0.012

Continuation on the next page

Organism	Channel name	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\lambda}$	H_0	p	p_2
Drosophila	M_Yki_1	1.07	0.41	0.30	20.11	1	0.364	0.011
	M_Yki_2	0.93	0.47	0.30	16.27	0	0.361	0.011
	M_Yki_3	1.10	0.60	0.26	21.80	1	0.407	0.011
	M_Yki_T_1	1.15	0.48	0.20	22.60	1	0.398	0.008
	M_Yki_T_2	1.08	0.48	0.28	24.72	0	0.383	0.012
	M_Yki_T_3	1.13	0.56	0.20	22.21	1	0.510	0.008
	Tr_Yki_13	1.22	0.69	0.15	23.00	1	0.512	0.008
	Tr_Yki_23	1.14	0.74	0.15	22.61	1	0.509	0.008
	Tr_Yki_33	1.24	0.75	0.11	25.54	1	0.431	0.008

REFERENCES

1. Saenko V., Saenko Yu. Application of the fractional stable distributions for approximation of gene expression profiles. *Statistical Applications in Genetics and Molecular Biology*. 2015. V. 14. No. 3. P. 295–306. doi: 10.1515/sagmb-2014-0094
2. Knudsen S. *Guide to Analysis of DNA Microarray Data*. Hoboken, NJ, USA : John Wiley & Sons, Inc., 2004. doi: 10.1002/0471670278
3. McLachlan G.J., Do K.-A., Ambroise C. *Analyzing Microarray Gene Expression Data*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA : John Wiley & Sons, Inc., 2004. doi: 10.1002/047172842X
4. Goodwin S., McPherson J.D., McCombie W.R. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*. 2016. V. 17. No. 6. P. 333–351. doi: 10.1038/nrg.2016.49
5. Metzker M.L. Sequencing technologies - the next generation. *Nature reviews. Genetics*. 2010. V. 11. No. 1. P. 31–46. doi: 10.1038/nrg2626
6. Ozsolak F., Milos P.M. RNA sequencing: advances, challenges and opportunities. *Nature reviews. Genetics*. 2011. V. 12. No. 2. P. 87–98. doi: 10.1038/nrg2934
7. Ansorge W.J. Next-generation DNA sequencing techniques.*New Biotechnology*. 2009. V. 25. No. 4. P. 195–203. doi: 10.1016/j.nbt.2008.12.009
8. Wang Z., Gerstein M., Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*. 2009. V. 10. No. 1. P. 57–63. doi: 10.1038/nrg2484
9. Ueda H.R., Hayashi S., Matsuyama S., Yomo T., Hashimoto S., Kay S.A., Hogenesch J.B., Iino M. Universality and flexibility in gene expression from bacteria to human. *Proceedings of the National Academy of Sciences*. 2004. V. 101. No. 11. P. 3765–3769. doi: 10.1073/pnas.0306244101
10. Liebovitch L.S., Jirsa V.K., Shehadeh L.A. Structure of genetic regulatory networks: evidence for scale free networks. In: *Complexus Mundi Emergent Patterns in Nature* / Ed. by UK Novak M.M. (Kingston University). Singapore : World Scientific Publishing Co. Pte. Ltd., 2006. P. 1–8. doi: 10.1142/9789812774217_0001
11. Lu Ch., King R.D. An investigation into the population abundance distribution of mRNAs, proteins, and metabolites in biological systems. *Bioinformatics (Oxford, England)*. 2009. V. 25. No. 16. P. 2020–2027. doi: 10.1093/bioinformatics/btp360
12. Furusawa Ch., Kaneko K. Zipf's Law in Gene Expression. *Physical Review Letters*. 2003. V. 90. No. 8. P. 8–11. doi: 10.1103/PhysRevLett.90.088102
13. Hoyle D.C.,Rattray M.,Jupp R.,Brass A. Making sense of microarray data distributions. *Bioinformatics*. 2002. V. 18. No. 4. P. 576–584. doi: 10.1093/bioinformatics/18.4.576
14. Kuznetsov V.A., Knott G.D., Bonner R.F. General statistics of stochastic process of gene expression in eukaryotic cells. *Genetics*. 2002. V. 161. No. 3. P. 1321–1332.

15. Bening V.E., Korolev V.Yu., Sukhorukova T.A., Gusarov G.G., Saenko V.E., Uchaikin V.V., Kolokoltsov V.N. Fractionally stable distributions. In: *Stochastic Models of Structural Plasma Turbulence.* Eds. by Korolev V.Yu., Skvortsova N.N. Utrecht : Brill Academic Publishers, 2006. P. 175–244. doi: 10.1515/9783110936032.175
16. Kotulski M. Asymptotic distributions of continuous-time random walks: A probabilistic approach. *Journal of Statistical Physics*. 1995. V. 81. No. 3–4. P. 777–792. doi: 10.1007/BF02179257
17. Kolokoltsov V.N., Korolev V.Yu., Uchaikin V.V. Fractional Stable Distributions.*Journal of Mathematical Sciences*. 2001. V. 105. No. 6. P. 2569–2576. doi: 10.1023/A:1011359219202
18. Zolotarev V.M. *One-dimensional stable Distributions*. Providence, RI : Amer. Mat. Soc., 1986.
19. Bening V.E., Korolev V.Yu., Kolokol'tsov V.N., Saenko V.V., Uchaikin V.V., Zolotarev V.M. Estimation of parameters of fractional stable distributions. *Journal of Mathematical Sciences*. 2004. V. 123. No. 1. P. 3722–3732. doi: 10.1023/B:JOTH.0000036313.13319.cf
20. Saenko V.V. Maximum likelihood algorithm for approximation of local fluctuational fluxes at the plasma periphery by fractional stable distributions. *arXiv:1209.2297 [physics.plasm-ph]*. 2012. URL: http://arxiv.org/abs/1209.2297.
21. Saenko V.V. Estimation of the Parameters of Fractional-Stable Laws by the Method of Minimum Distance. *Journal of Mathematical Sciences*. 2016. V. 214. No. 1. P. 101–114. doi: 10.1007/s10958-016-2760-y
22. Chambers J.M., Mallows C.L., Stuck B.W. A method for simulating stable random variables. *Journal of the American Statistical Association*. 1976. V. 71. No. 354. P. 340–344. doi: 10.2307/2285309
23. Kanter M. Stable Densities Under Change of Scale and Total Variation Inequalities. *The Annals of Probability*. 1975. V. 3. No. 4. P. 697–707. doi: 10.1214/aop/1176996309
24. Saenko V., Saenko Yu. Approximation of Microarray Gene Expression Profiles by the Stable Laws. *International Journal of Environmental Engineering*. 2015. V. 2. No. 1. P. 98–102. URL: http://seekdl.org/journal_page_papers.php?jourid=125&issueid=149

Received 20.06.2016. Published 25.11.2016.