



The OncoFinder algorithm for minimizing the errors introduced by the high-throughput methods of transcriptome analysis

Anton A. Buzdin^{1,2,3*}, Alex A. Zhavoronkov^{2,3}, Mikhail B. Korzinkin^{3,4}, Sergey A. Roumiantsev², Alexander M. Aliper^{2,3}, Larisa S. Venkova^{3,4}, Philip Y. Smirnov^{3,4} and Nikolay M. Borisov^{3,4}

¹ Group for Genomic Regulation of Cell Signaling Systems, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

² Laboratory of Bioinformatics, D. Rogachyov Federal Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia

³ Pathway Pharmaceuticals, Wan Chai, Hong Kong

⁴ Laboratory of Systems Biology, A.I. Burnasyan Federal Medical Biophysical Center, Moscow, Russia

Edited by:

John Gregory Marshall, Ryerson University, Canada

Reviewed by:

Nikhil Tyagi, Mitchell Cancer Institute University of South Alabama, USA
Ana Cláudia Coelho, Universidade de Trás-os-Montes e Alto Douro, Portugal

*Correspondence:

Anton A. Buzdin,
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry,
Miklukho-Maklaya 16/10, Moscow 117997, Russia
e-mail: bu3din@mail.ru

The diversity of the installed sequencing and microarray equipment make it increasingly difficult to compare and analyze the gene expression datasets obtained using the different methods. Many applications requiring high-quality and low error rates cannot make use of available data using traditional analytical approaches. Recently, we proposed a new concept of signalome-wide analysis of functional changes in the intracellular pathways termed OncoFinder, a bioinformatic tool for quantitative estimation of the signaling pathway activation (SPA). We also developed methods to compare the gene expression data obtained using multiple platforms and minimizing the error rates by mapping the gene expression data onto the known and custom signaling pathways. This technique for the first time makes it possible to analyze the functional features of intracellular regulation on a mathematical basis. In this study we show that the OncoFinder method significantly reduces the errors introduced by transcriptome-wide experimental techniques. We compared the gene expression data for the same biological samples obtained by both the next generation sequencing (NGS) and microarray methods. For these different techniques we demonstrate that there is virtually no correlation between the gene expression values for all datasets analyzed ($R^2 < 0.1$). In contrast, when the OncoFinder algorithm is applied to the data we observed clear-cut correlations between the NGS and microarray gene expression datasets. The SPA profiles obtained using NGS and microarray techniques were almost identical for the same biological samples allowing for the platform-agnostic analytical applications. We conclude that this feature of the OncoFinder enables to characterize the functional states of the transcriptomes and interactomes more accurately as before, which makes OncoFinder a method of choice for many applications including genetics, physiology, biomedicine, and molecular diagnostics.

Keywords: signalome, RNA-Seq, intracellular signaling pathway activation, next generation sequencing, microarray hybridization, gene expression, transcriptome profiling, correction of errors

INTRODUCTION

The complex machinery of intracellular signaling determines the cell fate by governing all the most important processes including growth, differentiation, proliferation, migration, survival, and death. The molecular modeling of intracellular signaling pathways has been underway for more than two decades (Kholodenko et al., 1999; Hanahan and Weinberg, 2000). During that time a plethora of regulatory cascades have been discovered and cataloged (Nikitin et al., 2003; Mathivanan et al., 2006; Elkon et al., 2008; Bauer-Mehren et al., 2009; Haw and Stein, 2012; Nakaya et al., 2013). Each of these cascades contains dozens to hundreds of different types of molecules, mainly genomic DNA-encoded gene products. The information on the architecture of the signaling pathway can be used for the mathematical analysis of signal

transduction process (Nikitin et al., 2003; Mathivanan et al., 2006; Elkon et al., 2008; Bauer-Mehren et al., 2009; Haw and Stein, 2012; Nakaya et al., 2013; Yizhak et al., 2013).

This information has resulted in the accumulation of the large collections of public gene expression datasets including the Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>), Stanford Microarray Database (<http://smd.stanford.edu/>), Cancer Genome Atlas (<https://tcga-data.nci.nih.gov/>), EBI ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>), mAdb (<https://madb.nci.nih.gov/>), and the many commercial repositories providing analytical services. However, until very recently, these data repositories did not include pathway-based quantitative evaluation of the functional changes between different biological samples. Additionally, there were no technical

possibilities to analyze large-scale signalome signatures in high-throughput gene expression data.

Recently, we developed a set of bioinformatics algorithms and tools collectively called the OncoFinder (Buzdin et al., 2014), which encompasses the molecular signaling databases and enables performing the quantitative analysis of the signalome changes. This method digests the high throughput transcriptomic data and provides molecular signaling pathway activation (SPA) fingerprints for the individual samples.

The major distinction and novelty of the OncoFinder method is its unique algorithm that quantifies perturbations for each signaling pathway. This makes it possible to quantitatively estimate the extent of each SPA in a given sample relative to the control sample or a set of control samples (Buzdin et al., 2014).

In this study, we compared different gene expression datasets generated for the same biological samples using two different experimental techniques, *next generation sequencing* (NGS) and *microarray hybridization*. We showed that there is generally a very weak correlation between the initial NGS and microarray gene expression data. However, applying the OncoFinder method results in identifying pathway activation profiles that are highly correlated between the NGS and microarray datasets. We conclude that the OncoFinder algorithm efficiently removes errors introduced by the different experimental methodologies and yields accurate results that can be effectively used for various applications in biology, medicine and molecular diagnostics.

According to the International Aging Research Portfolio (<http://www.agingportfolio.org/>), over six billion dollars in government funding were spent on research projects involving microarray systems for gene expression analysis since 1993 (Zhavoronkov and Cantor, 2011) and resulted in tens of thousands of publications. And while the emerging NGS systems are providing for more-data rich and error-prone RNA sequencing methods, there is a significant installed base of the microarray systems generating new gene expression data sets. The high correlation of the pathway activation profiles of gene expression data from the same biological samples between the microarray platforms as well as NGS equipment presents the opportunity to compare the many microarray datasets generated over the past decades with the data sets obtained using the new NGS platforms.

RESULTS AND DISCUSSION

Recently, we proposed a new concept of signalome-wide analysis of functional changes in the intracellular pathways termed OncoFinder (Buzdin et al., 2014) and developed a bioinformatic instrument for quantitative estimation of SPA. The underlying algorithm of OncoFinder converts the results of transcriptome profiling into a quantitative and qualitative signalome profile, which characterizes the states called *pathway activation strength* (PAS), for each signaling pathway under investigation. The information is processed based on the “low-level” transcriptomic data represented by the aggregate of the so-called *case-to-normal ratio* values, CNR_n , i.e., the ratio of the expression levels of a gene n in the investigated sample (e.g., of an individual pathological tissue sample) to a control (e.g., average value for group of the healthy

tissue samples). The major distinction of our algorithm compared to other related approaches (Yizhak et al., 2013), is that it deals with the functional annotation of the gene product and its role in the individual pathway (e.g., activator or repressor of the signal transduction through the pathway). The absolute value of PAS is characterized by the degree of functional changes in the regulation of a signaling pathway, and the positive or negative sign of PAS indicates whether it is up- or down-regulated, respectively (Buzdin et al., 2014).

This algorithm was applied to the analysis of various human tissues and cell types (Buzdin et al., 2014; Vishnyakova et al., in press) including hematologic cancers (Spirin et al., 2014) and validated (Buzdin et al., 2014) on the previously established “low-level” kinetic protein interaction model of the EGFR pathway activation (Kuzmina and Borisov, 2011). While genetics of aging and longevity are complex, the knowledge base is rapidly increasing (Moskalev et al., 2014), cancer is related to aging (Blagosklonny and Campisi, 2008) and anti-cancer agents may act as geroprotectors (Blagosklonny, 2012) and geroprotectors may provide be used in cancer prevention (Blagosklonny, 2013). The ability of the OncoFinder algorithm to perform the cross-platform comparison of the gene expression data resulted in the first proposal to use PAS in aging research for screening and ranking of the geroprotective drugs (Zhavoronkov et al., 2014).

In this study, we aimed to investigate whether the OncoFinder algorithm may be applied to minimize the error rates in gene expression data obtained using different experimental methods. We extracted the matching large-scale transcriptomic data obtained using the different experimental platforms for the same biological samples. Using the GEO repository of gene expression data (<http://www.ncbi.nlm.nih.gov/geo/>), we were able to extract three datasets corresponding to the simultaneous NGS and microarray profiling of the same human tissue samples. One of these datasets, GSE36244 represented human hepatocarcinoma HepG2 cells treated with benzopyrene, and the data were

Table 1 | Transcriptomic data deposited in the GEO database that were used for the current study.

Dataset ID	Origin	Case samples vs. control samples	Experimental platforms
GSE36244	HepG2 cells	Treated vs. untreated with benzopyrene	Affymetrix human genome U133 Plus 2.0 arrays and illumina genome analyzer sequencer
GSE41588	HT-29 cells	Treated vs. untreated with 5-aza-deoxy-cytidine	Affymetrix human genome U133 Plus 2.0 arrays and illumina genome analyzer sequencer
GSE37765	Lung adenocarcinoma	Tumor samples vs. matched samples of normal tissue	Agilent 1M CNV arrays and illumina genome analyzer sequencer

obtained by hybridization on the Affymetrix Human Genome U133 Plus 2.0 arrays and using the Illumina Genome Analyzer sequencing engine (van Delft et al., 2012). In the next dataset, GSE41588, published by another group of the authors (Xu et al., 2013) the same two experimental platforms were utilized for probing gene expression of the human colon cancer cell line HT-29 treated with 5-aza-deoxy-cytidine. Finally, the third dataset GSE37765 (Kim et al., 2013) included the information obtained using the Agilent 1M CNV arrays and the Illumina Genome Analyzer platform for the human female lung adenocarcinoma and for the normal lung samples. The overview of input data is presented in Table 1.

To apply the OncoFinder technique for signalome analysis of these datasets, we interpreted the untreated cell culture samples for the datasets GSE36244 and GSE4158, and healthy lung samples for the dataset GSE37765 to be the “normal” or control states.

Next, for each dataset we compared the transcriptomic signatures obtained by the NGS and microarray hybridization platforms. All the microarray data were quantile normalized. For further normalization of the transcription data to the control samples, we calculated the case-to-normal ratio (CNR). When comparing the normalized expression logarithms between the NGS and microarray expression data, we detected virtually no correlation for all the datasets under investigation, as reflected by the low correlation coefficients generally less than 0.1 for most of the samples (Figure 1, Table 2). These results suggest that there is a huge gap between the microarray and NGS expression data for all the investigated experimental platforms.

In contrast, for the OncoFinder-processed data and PAS we detected clear-cut correlations between the NGS and microarray gene expression datasets (Figure 1, Table 2). The correlation coefficients for PAS were significantly greater than for the CNR and varied in the interval 0.49–0.89 with a single outlier of 0.21.

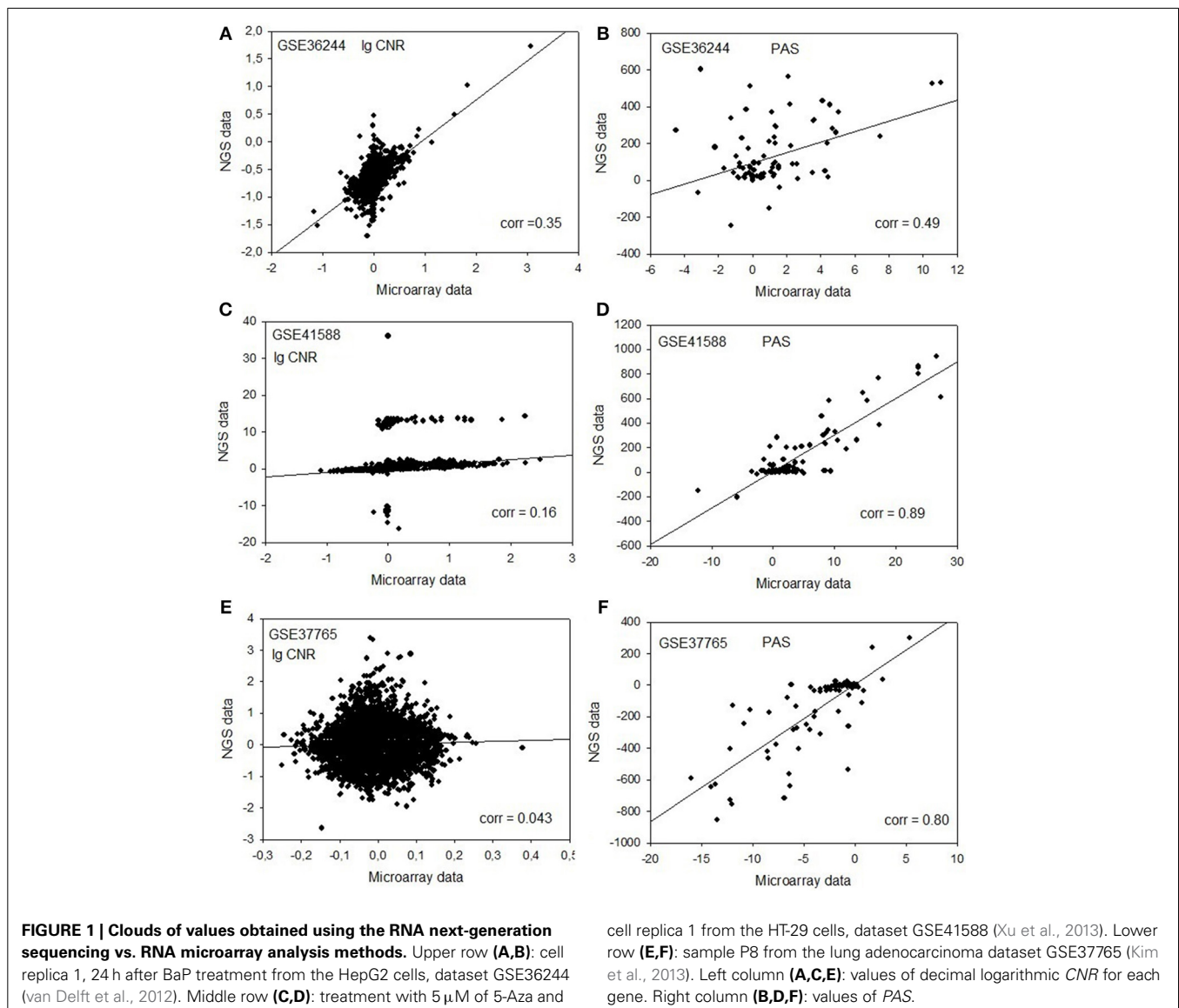


Table 2 | Correlation coefficients between values obtained using the RNA microarray analysis and RNA sequencing methods for the HepG2 cells dataset GSE36244 (van Delft et al., 2012), HT-29 cells dataset GSE41588 (Xu et al., 2013), and lung adenocarcinoma dataset GSE37765 (Kim et al., 2013).

Sample		Transcriptome level (logarithmic <i>CNR</i> for different genes)	Signalome level (<i>PAS</i> value for different pathways)
GSE36244, 24 h after BaP treatment	Replica 1	0.35	0.49
	Replica 2	0.10	0.47
	Averaged over 2 samples above	0.22	0.49
GSE41588, 5 μ M of 5-Aza	Replica 1	0.16	0.89
	Replica 2	0.049	0.88
	Replica 3	0.047	0.80
	Averaged over 3 samples above	0.082	0.87
GSE37765	P1	0.18	0.79
	P3	0.098	0.75
	P4	0.12	0.80
	P5	-0.029	0.21
	P8	0.043	0.80
	Averaged over 5 samples above	0.068	0.77

This finding evidences that the *PAS* calculation algorithm produces significantly more congruent results compared to the initial gene expression signatures between the microarray and NGS datasets.

Importantly, both NGS and microarray hybridization strategies may produce a large number of errors through the stages of RNA purification, library preparation and amplification, hybridization and sequencing, and finally mapping and annotation of the reads and reading the array (Chalaya et al., 2004; Buzdin and Lukyanov, 2007; Shugay et al., 2014). It is hard to identify the errors and to find out what type of experimental assay provides more accurate data for each individual gene. It is important to minimize the errors in the transcriptomic data and, theoretically, quantitative real-time PCR might provide a solution as a reference gene expression measure. However, existing PCR platforms do not allow for making high throughput, transcriptome-scale experiments. Our approach makes it possible to surmount this obstacle as, unlike the original data, the outgoing *PAS* values are highly congruent among the NGS and microarray data. This effect of the OncoFinder algorithm is most likely due to its cumulative nature. The *PAS* value is formed by the addition of multiple individual members each representing a gene product involved in the pathway. The concentration of each individual gene product can be measured with an error, which is seen when untreated NGS vs. array data are compared, but a combination of a large number of these concentration members into a signalome-oriented network apparently diminishes an overall error, as reflected by the good correlation records.

We conclude that this feature of *PAS* makes it possible to more accurately measure the changes in the functional states of the cellular/tissue transcriptome and interactome across the many microarray and NGS platforms, which makes OncoFinder a method of choice for many applications including genetics, physiology, biomedicine, and molecular diagnostics.

MATERIALS AND METHODS

SOURCE DATASETS

Gene expression data used in this study were downloaded from the GEO repository of transcriptomic information (<http://www.ncbi.nlm.nih.gov/geo/>). The following datasets were used: GSE36244 (van Delft et al., 2012) where transcriptomes were from HepG2 cells treated with benzopyrene (four samples for the treated and four for the untreated cells); GSE37765 (Kim et al., 2013) where female normal lung and lung adenocarcinoma samples were tested (six samples for the normal and 6 for the cancer samples), and GSE41588 (Xu et al., 2013) where HT-29 cells were treated with 5-aza-deoxy-cytidine (six samples for the treated and three for the untreated cells).

For the dataset GSE36244 the two transcriptome datasets were generated from the Illumina Genome Analyzer Iix sequencer and also from the Affymetrix Human Genome U133 Plus 2.0 GeneChip arrays. We took untreated HepG2 cells as the controls for further calculations. The GSE41588 data set was generated from both the Illumina sequencing platform and the Affymetrix Human Genome U133 Plus 2.0 arrays. For this dataset the untreated HT-29 cells were used as the controls. For the third dataset, GSE37765 (Kim et al., 2013), the data were obtained from the Illumina Genome Analyzer Iix sequencer and the Agilent 1M CNV microarray hybridization device. The normal lung samples were used as the controls.

The signalome knowledge base developed by SABiosciences (<http://www.sabiosciences.com/pathwaycentral.php>) was used to determine structures of intracellular pathways, which was used for the computational algorithm OncoFinder exactly as described previously (Buzdin et al., 2014; Spirin et al., 2014; Zhavoronkov et al., 2014).

Functional annotation of gene expression data

We applied the OncoFinder algorithm (Buzdin et al., 2014) for the functional annotation of the primary microarray and NGS genome-wide expression data and for the calculation of

the regulatory SPA scores. The extracted raw microarray expression data were quantile normalized (Bolstad et al., 2003). Our approach to the transcriptome-wide gene expression analysis applies processing of these results with the signalome knowledge base developed by SABiosciences (<http://www.sabiosciences.com/pathwaycentral.php>). Our algorithm utilizes a scheme that takes into account the overall impact of each gene product in the signaling pathway but ignores its position in the pathway graph. The formula used to calculate the PAS for a given sample and a given pathway p is as follows:

$$PS_p = \sum_n ARR_{np} \cdot BTIF_n \cdot \lg(CNR_n)$$

Here the *case-to-normal ratio*, CNR_n , is the ratio of expression levels for a gene n in the sample under investigation to the same average value for the control group of samples. The Boolean flag of $BTIF$ (*beyond tolerance interval flag*) equals zero when the CNR value has passed simultaneously the two criteria that demark the significantly perturbed expression level from essentially normal. The first criterion is the expression level for the sample lies within the tolerance interval, where $p > 0.05$. The second criterion is the discrete value of ARR (*activator/repressor role*) equals to the following fixed values: -1 , when the gene/protein n is a repressor of pathway excitation; 1 , if the gene/protein n is an activator of pathway excitation; 0 , when the gene/protein n can be both an activator and a repressor of the pathway; 0.5 and -0.5 , respectively, if the gene/protein n is rather an activator or repressor of the signaling pathway p , respectively. The results for the 90 pathways were obtained for each sample (listed in the Supplementary file 1). Statistical tests were done using the R software package.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research grant 12-04-33094 and the Program of the Presidium of the Russian Academy of Sciences “Dynamics and Conservation of Genomes.” The authors thank The First Oncology Research and Advisory Center (Moscow, Russia) and “UMA Foundation” (Moscow) for their support in preparation of the manuscript and Dr. Kristen Swithers from Yale University for her assistance with editing the manuscript. We would like to thank Alex Kim and ASUSTek for equipment and support of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/journal/10.3389/fmolb.2014.00008/abstract>

REFERENCES

- Bauer-Mehren, A., Furlong, L. I., and Sanz, F. (2009). Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol. Syst. Biol.* 5:290. doi: 10.1038/msb.2009.47
- Blagosklonny, M. V. (2012). Rapalogs in cancer prevention. *Cancer Biol. Ther.* 13, 1349–1354. doi: 10.4161/cbt.22859
- Blagosklonny, M. V. (2013). Selective anti-cancer agents as anti-aging drugs. *Cancer Biol. Ther.* 14:1092. doi: 10.4161/cbt.27350
- Blagosklonny, M. V., and Campisi, J. (2008). Cancer and aging. *Cell Cycle* 7, 2615–2618. doi: 10.4161/cc.7.17.6626
- Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. *Bioinformatics* 19, 185–193. doi: 10.1093/bioinformatics/19.2.185
- Buzdin, A. A., and Lukyanov, S. A. (eds.). (2007). *Nucleic acids Hybridization Modern Applications*. Dordrecht: Springer. ISBN: 978-1-4020-6039-7 (Print) 978-1-4020-6040-3
- Buzdin, A. A., Zhavoronkov, A. A., Korzinkin, M. B., Venkova, L. S., Zenin, A. A., Smirnov, P. H., et al. (2014). OncoFinder, a new method for the analysis of intracellular signaling pathway activation using transcriptomic data. *Front. Genet.* 5, 55. doi: 10.3389/fgene.2014.00055
- Chalaya, T., Gogvadze, E., Buzdin, A., Kovalskaya, E., and Sverdlov, E. D. (2004). Improving specificity of DNA hybridization-based methods. *Nucleic Acids Res.* 32:e130. doi: 10.1093/nar/gnh125
- Elkon, R., Vesterman, R., and Amit, N. (2008). SPIKE – a database, visualization and analysis tool of cellular signaling pathways. *BMC Bioinformatics* 9:110. doi: 10.1093/nar/gkq1167
- Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* 100, 57–70. doi: 10.1016/S0092-8674(00)81683-9
- Haw, R., and Stein, L. (2012). Using the reactome database. *Curr. Protoc. Bioinformatics*. chapter 8, Unit 8.7. doi: 10.1002/0471250953.bi0807s38
- Kholodenko, B. N., Demin, O. V., Moehren, G., and Hoek, J. B. (1999). Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.*, 274, 30169–30181. doi: 10.1074/jbc.274.42.30169
- Kim, S. C., Jung, Y., Park, J., Cho, S., Seo, C., Kim, J., et al. (2013). A high-dimensional, deep-sequencing study of lung adenocarcinoma in female never-smokers. *PLoS ONE* 8:e55596. doi: 10.1371/journal.pone.0055596
- Kuzmina, N. B., and Borisov, N. M. (2011). Handling complex rule-based models of mitogenic cell signaling (On the example of ERK activation upon EGF stimulation). *Intl. Proc. Chem. Biol. Environ. Eng.* 5, 76–82.
- Mathivanan, S., Periaswamy, B., Gandhi, T., Kandasamy, K., Suresh, S., Mohmood, R., et al. (2006). An evaluation of human protein-protein interaction data in the public domain. *BMC Bioinformatics* 7:S19. doi: 10.1186/1471-2105-7-S5-S19
- Moskalev, A. A., Aliper, A. M., Smit-McBride, Z., Buzdin, A., and Zhavoronkov, A. (2014). Genetics and epigenetics of aging and longevity. *Cell Cycle* 13, 1063–1077. doi: 10.4161/cc.28433
- Nakaya, A., Katayama, T., Itoh, M., Hiranuka, K., Kawashima, S., Moriya, Y., et al. (2013). KEGG OC: a large-scale automatic construction of taxonomy-based ortholog clusters. *Nucleic Acids Res.* 41, D353–D357. doi: 10.1093/nar/gks1239
- Nikitin, A., Egorov, S., Daraselia, N., and Mazo, I. (2003). Pathway studio – the analysis and navigation of molecular networks. *Bioinformatics* 19, 2155–2157. doi: 10.1093/bioinformatics/btg290
- Shugay, M., Britanova, O. V., Merzlyak, E. M., Turchaninova, M. A., Mamedov, I. Z., Tuganbaev, T. R., et al. (2014). Towards error-free profiling of immune repertoires. *Nat. Methods* 6, 653–655. doi: 10.1038/nmeth.2960
- Spirin, P. V., Ledbedev, T. D., Orlova, N. N., Gornostaeva, A. S., Prokofjeva, M. M., Nikitenko, N. A., et al. (2014). Silencing AML1-ETO gene expression leads to simultaneous activation of both pro-apoptotic and proliferation signaling. *Leukemia*. doi: 10.1038/leu.2014.130. [Epub ahead of print].
- van Delft, J., Gaj, S., Lienhard, J., Albrecht, M. W., Kirpiy, A., Brauers, M., et al. (2012). RNA-seq provides new insights in the transcriptome responses induced by the carcinogen benzo[a]pyrene. *Toxicol. Sci.* 130, 427–439. doi: 10.1093/toxsci/kfs250
- Vishnyakova, K. S., Babizhayev, M. A., Aliper, A. M., Busdin, A. A., Kudryavzeva, A. V., and Yegorov, Y. E. (in press). Stimulation of proliferation by carnosine: cellular and transcriptome approaches. *Mol. Biol. (Moscow)*.
- Xu, X., Zhang, Y., Williams, J., Antoniou, E., McCombie, W. R., and Wu, S. (2013). Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. *BMC Bioinformatics* 14(Suppl. 9):S1. doi: 10.1186/1471-2105-14-S9-S1
- Yizhak, K., Gabay, O., Cohen, H., and Rupin, E. (2013). Model-based identification of drug targets that, revert disrupted metabolism and its application to ageing. *Nat. Commun.* 4:2632. doi: 10.1038/ncomms3632
- Zhavoronkov, A., Buzdin, A. A., Garazha, A. V., Borisov, N. M., and Moskalev, A. A. (2014). Signaling pathway cloud regulation for in silico screening and ranking of the potential geroprotective drugs. *Front. Genet.* 5:49. doi: 10.3389/fgene.2014.00049

Zhavoronkov, A., and Cantor, C. R. (2011). Methods for structuring scientific knowledge from many areas related to aging research. *PLoS ONE* 6:e22597. doi: 10.1371/journal.pone.0022597

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 June 2014; paper pending published: 24 July 2014; accepted: 04 August 2014; published online: 26 August 2014.

Citation: Buzdin AA, Zhavoronkov AA, Korzinkin MB, Roumiantsev SA, Aliper AM, Venkova LS, Smirnov PY and Borisov NM (2014) The OncoFinder algorithm for

minimizing the errors introduced by the high-throughput methods of transcriptome analysis. Front. Mol. Biosci. 1:8. doi: 10.3389/fmolb.2014.00008

This article was submitted to *Molecular Diagnostics*, a section of the journal *Frontiers in Molecular Biosciences*.

Copyright © 2014 Buzdin, Zhavoronkov, Korzinkin, Roumiantsev, Aliper, Venkova, Smirnov and Borisov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.