


High-Throughput Mutation Data Now Complement Transcriptomic Profiling: Advances in Molecular Pathway Activation Analysis Approach in Cancer Biology

Anton Buzdin^{1,2,3}, Maxim Sorokin^{1,2,3}, Elena Poddubskaya^{1,4} and Nicolas Borisov^{1,2} 

¹Institute for Personalized Medicine, I.M. Sechenov First Moscow State Medical University, Moscow, Russia. ²Omicsway Corp., Walnut, CA, USA. ³Group for Genome Analysis of Cell Signaling Systems, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia. ⁴Vitamed Oncological Clinics, Moscow, Russia.

Cancer Informatics
Volume 18: 1–4
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176935119838844



ABSTRACT: We recently reviewed the current progress in the use of high-throughput molecular “omics” data for the quantitative analysis of molecular pathway activation. These quantitative metrics may be used in many ways, and we focused on their application as tumor biomarkers. Here, we provide an update of the most recent conceptual findings related to pathway analysis in tumor biology, which were not included in the previous review. The major novelties include a method enabling calculation of pathway-scale tumor mutation burden termed “Pathway Instability” and its application for scoring of anticancer target drugs. A new technique termed Shambhala emerged that enables accurate common harmonization of any number of gene expression profiles obtained using any number of experimental platforms. This may be helpful for merging various gene expression data sets and for comparing their pathway activation characteristics. Another recent bioinformatics method, termed Floating-Window Projective Separator (FloWPS), has the potential to significantly enhance the value of pathway activation profiles as biomarkers of cancer response to treatments. It reduces the minimum required number of training samples needed to construct a machine-learning-based classifier. Finally, several documented clinical cases have been recently published, in which gene-expression-based pathway analysis was successfully used for personalized off-label prescription of target drugs to metastatic cancer patients.

KEYWORDS: bioinformatics, cancer, signaling pathways, mutation profiling, machine learning

RECEIVED: February 14, 2019. **ACCEPTED:** February 22, 2019.

TYPE: Commentary

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Russian Science Foundation under grant no. 18-15-00061.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Nicolas Borisov, I.M. Sechenov First Moscow State Medical University, Moscow, Russia. Email: nicolasborissoff@gmail.com

COMMENT ON: Buzdin A, Sorokin M, Garazha A, et al. Molecular pathway activation—new type of biomarkers for tumor morphology and personalized selection of target drugs. *Semin Cancer Biol.* 2018;53:110-124. doi:10.1016/j.semcancer.2018.06.003. PubMed PMID:29935311. <https://www.ncbi.nlm.nih.gov/pubmed/29935311>.

In the recent paper by Buzdin et al,¹ we reviewed methods of molecular pathway analysis and their applications in basic research and biomedicine, primarily as cancer biomarkers. These methods deal with high-throughput gene expression and even epigenetic data. Namely, both mRNA and proteomic profiles are considered for gene expression analysis; microRNA profiles and densities of transcription factor binding sites (TFBS) are the epigenetic landmarks reported as useful sources of primary data for pathway analyses. This may seem surprising, but no genetic data, such as cancer mutation profiles, were ever communicated as the material for high-throughput modeling of molecular pathway activation. Accordingly, no such approaches were mentioned in the review.¹

This has recently changed when Zolotovskaia et al^{2,3} published 2 research papers reporting cancer exome sequencing profiles being used to aggregate DNA mutation data in a value termed “pathway instability” (PI). PI reflects overall mutation burden of a pathway. It can be calculated for either total mutations or any specific group of them, such as truncating mutations that abrogate protein functions.³ On both levels, PI serves as a significantly better type of biomarker compared with mutations in individual genes.³ The study of 5956 individual

mutation profiles from 15 cancer types encompassing 2316670 mutations in 19872 genes and 1748 molecular pathways demonstrated unparalleled advantage of pathway-based mutation biomarkers (PI) over mutation profiles for the individual gene. This was reflected by more than 2 orders of magnitude greater numbers by high-quality (area under the receiver operating characteristic curve [AUC ROC] > 0.75) biomarkers. For example, the number of such biomarkers distinguishing between different cancer types in “one versus all” comparisons was only 6 (1 for truncating mutations only) for the gene mutations, but already 660 (21 for truncating mutations) for the PI values. Similarly, when doing pairwise comparisons for the above 15 cancer types, a total of 32594 (1024 for truncating mutations) good-quality PI biomarkers were identified versus only 226 (24 for truncating mutations) gene mutation biomarkers.³ Considering that the potential number of PI biomarkers (1748) was already 1 order of magnitude lower than the potential number of gene biomarkers (19872 for all mutations and 16760 for truncating mutations), this provides even stronger evidence for the superiority of the PI-based approach over mutations in individual genes.³

Furthermore, a molecular pathway-based algorithm for the scoring of anticancer target drugs (ATDs) was developed



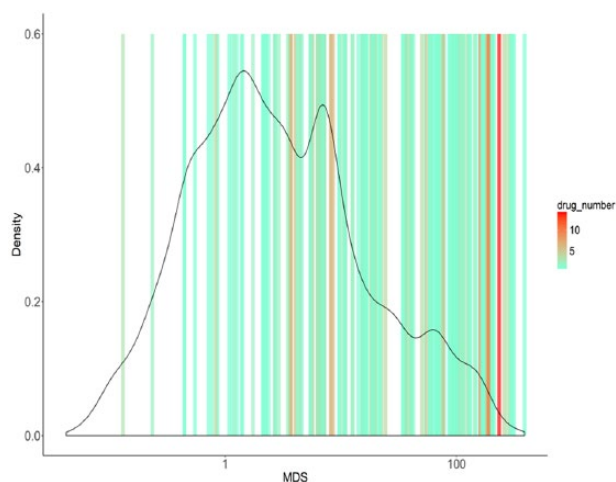


Figure 1. Dependence of MDS and occurrence of molecular targets in approved cancer drugs. Distribution of MDS values among the potential molecular drug targets. The color scale on the graph indicates densities of clinically approved cancer drugs exploiting the respective molecular targets. MDS indicates Mutation Drug Scoring.

using PI values to predict clinical efficacies of drugs. It deals with quantization of mutation enrichment for the molecular pathways having molecular targets of a drug under consideration. Here, the rationale used was that the greater the mutation level of the respective pathways, the higher the expected drug efficiency. The output value termed Mutation Drug Scoring (MDS) positively correlates with the expected efficacy of drugs for a specific tumor. Among 10 versions tested, the best version of this algorithm was selected and validated using 3800 exome mutation profiles for 128 ATDs by finding correlations of MDS values with the known drug efficiencies tested in clinical trials.² The same approach was next applied to simulate all known protein-coding genes as putative drug targets using 18 273 mutation profiles for 8 cancer types. Indeed, the MDS-predicted hits very frequently coincided with those already used as targets of the existing ATDs² (Figure 1). However, several novel candidate genes that can be considered promising targets for new drug development were identified. The MDS approach, therefore, is applicable to both the ranking of drugs and the identification of molecular targets for biopharmaceuticals. These applications of mutation data analysis are presented in detail in a recent methods paper.

A pioneering application of molecular pathway approach for TFBS data was reviewed previously^{1,4}; here, we just want to mention that it was used in a recent high-throughput study of human genes with rapidly or slowly evolving regulatory modules. Using a complete set of the ENCODE project TFBS data for 563 transcription factors in 13 human cell lines,⁵ the authors found that the major quickly reshaping processes in human evolution deal with gene regulation by microRNAs, olfaction, color vision, fertilization, cellular

immune response, detoxication, amino acids, and fatty acids metabolism. Genes from the “slow” group were involved in protein translation, RNA transcription and processing, chromatin organization, and in various aspects of intracellular molecular signaling.⁵

Another recent methods paper summarizes details of the study of molecular pathway activation using RNA sequencing data. At the first stage, this analysis compares gene expression in a sample with a pool of reference (“normal”) expression profiles. Several large collections of normal human tissue expression data sets that can be used for such analyses were published in the last year. Moreover, a novel version of the Oncobox pathway analysis algorithm⁷ adds an additional, previously unknown, step of transcriptomic data, termed “Shambhala harmonization,” to make gene expression profiles compatible for both normal and case bio samples.⁶ As previously shown for the cell culture gene expression data, molecular pathways activation data may be used to train the machine learning algorithms to predict efficiencies of ATDs.^{7–9} However, these algorithms have very limited applicability to clinical molecular data because, to avoid overtraining, they require large cohorts of the ATD responder/non-responder molecular profiles, which are not available for the overwhelming majority of cases. Although this major limitation still persists, a novel approach, termed FLOating-Window Projective Separator (FloWPS), that may significantly reduce the number of samples to be included in a statistically significant analysis was recently published.¹⁰ Briefly, this is a data trimming procedure tailored for personalized predictions based on molecular data. In its initial application, it uses support vector machines (SVMs), a machine learning method that is highly popular for the analysis of biomedical data.¹⁰ Its application prevents SVM from extrapolation by excluding non-informative features. FloWPS requires training on the data for the individuals with known clinical outcomes to create a clinically relevant classifier. Its unique property is that irrelevant features in validation data set that do not have significant number of neighboring hits in the training data set are removed from further analyses. Next, for each point of a validation data set, FloWPS takes into account only the proximal points of the training data set. Thus, for every point of a validation data set, the training data set is adjusted to form a floating window.¹⁰ The methods validation on 992 cancer patient expression data sets confirmed that FloWPS enables to significantly increase quality of a classifier.¹⁰ So far, FloWPS has been tested for single gene expression data only. We anticipate that using this method for molecular pathway activation levels will be fruitful for developing novel robust tools for the classification of responses on drugs, treatment methods, and their combinations.

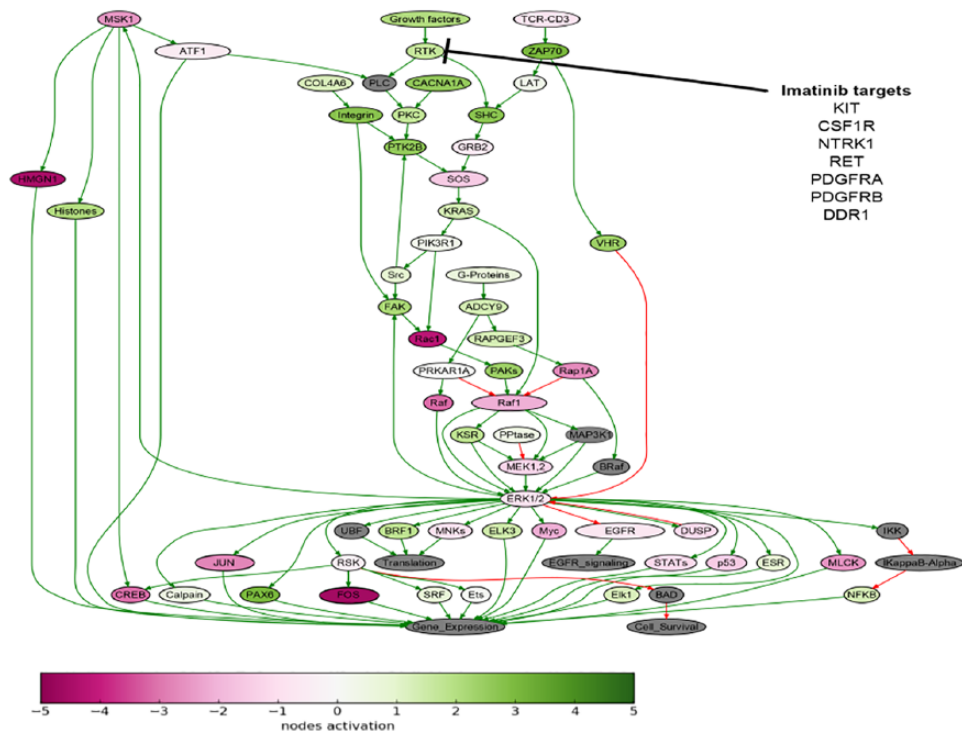


Figure 2. ERK signaling pathway was hyperactivated in the patient's tumor tissue. Visualization was provided by Oncobox software. The pathway is shown as an interacting network, where green arrows indicate activation and red arrows indicate inhibition. Color depth of each node of the network corresponds to the logarithms of the case-to-normal (CNR) expression rate for each node, where "normal" is a geometric average between normal tissue samples, and the scale represents extent of up-/down-regulation. The molecular targets of Imatinib are shown by black arrows. ERK indicates extracellular signal-regulated kinase.

Finally, a second-opinion platform termed Oncobox has been generated for clinical oncologists, which uses the original pathway activation scoring algorithm working with RNA-seq data. It personalizes selection of ATDs for individual cancer patients. Pathway activation levels are calculated and, along with the concentrations of molecular target gene products, are used as predictors of tumor response on ATDs, reflected by the Balanced Drug Efficiency Score (BES) index. Previously, a pilot prospective clinical study was performed for a cohort of 23 recurrent/metastatic solid tumor patients using microarray gene expression data.¹¹ The objective response rate for the Oncobox-guided ATD prescriptions was ~61% (complete + partial response, RECIST). Since April 2018, a new trial started using RNA-seq data for recurrent/metastatic solid tumors that already includes 239 patients (trial ID NCT03724097). RNAs were extracted from the formalin-fixed, paraffin-embedded (FFPE) tissue blocks. Following the test, 130 ATDs were rated according to their predicted effectiveness. After appointment of therapy, patients are divided into 3 observation groups. The first group consists of patients receiving target drugs in agreement with the Oncobox drug efficiency prediction as monotherapy or in combination(s); the second group of patients only receives drugs not recommended according to the Oncobox tests; and the third group of patients receives palliative care. The trial is ongoing and, so far, there are no preliminary results available, but several recently published

clinical case studies using the same system look promising and suggest that the Oncobox system may be a useful tool for predicting efficiencies of tyrosine kinase inhibitor ATDs in metastatic advanced tumors. For example, Oncobox-guided off-label prescription of imatinib (Figure 2) was efficient for a patient with recurrent granulosa cell ovarian tumor,¹² and (sequential) prescription of sorafenib and pazopanib (Figure 3) to a patient with unresectable metastatic cholangiocarcinoma also showed promising results.¹³

An alternative line of biomedical investigations may deal with pathway activation-guided personalized selection of combinations of ATDs. In theory, the same pathway approaches, which were shown effective with single ATDs, may also work for their combinations. However, these studies are today at an early stage with 1 recent report published predicting and experimentally validating various effective ATD combinations for cancer cell lines based on molecular pathway activation data.¹⁴ We expect intensive development of this field, assuming that ATDs may work synergistically in a patient-specific manner.

We conclude, therefore, that the field covered by the topic "Molecular pathway activation—New type of biomarkers for tumor morphology and personalized selection of target drugs" is evolving rapidly so that the landscape depicted in our previous review¹ already looks hardly recognizable in less than a year.

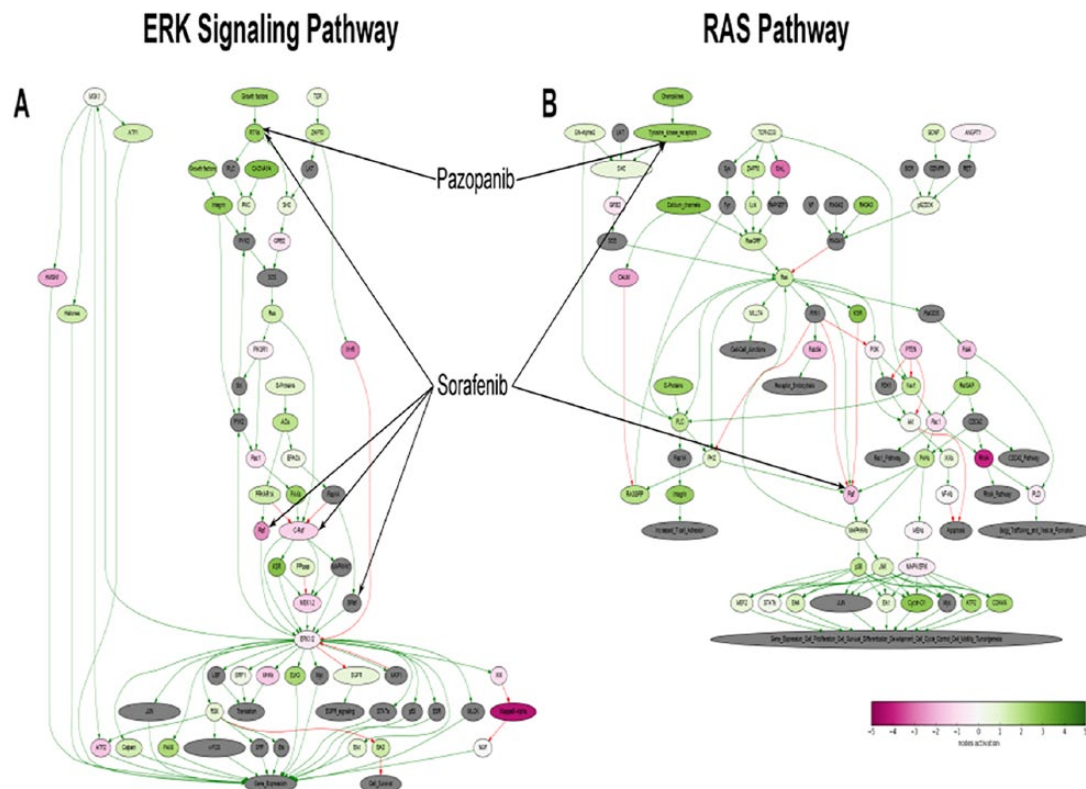


Figure 3. (A) ERK and (B) Ras signaling pathways were hyperactivated in the biopsy CCA tissue. Visualization was provided by Oncobox software. The pathways are shown as an interacting network, where green arrows indicate activation and red arrows indicate inhibition. Color depth of each node of the network corresponds to the logarithms of the case-to-normal (CNR) expression rate for each node, where “normal” is a geometric average between normal tissue samples, and the scale represents extent of up-/down-regulation. The molecular targets of sorafenib and pazopanib are shown by black arrows. ERK indicates extracellular signal-regulated kinase; CCA, cholangiocarcinoma.

Author Contributions

AB designed the research, integrated and analyzed the results obtained by other workers, and wrote the manuscript. MS performed the research on mutation pathway instability and mutation-based scoring of anticancer target drugs. EP performed the clinical parts of the research and provided case histories. NB performed the research on harmonization of expression profiles and machine learning for personalized prediction of drug efficiency.

ORCID iD

Nicolas Borisov  <https://orcid.org/0000-0002-1671-5524>

REFERENCES

- Buzdin A, Sorokin M, Garazha A, et al. Molecular pathway activation—new type of biomarkers for tumor morphology and personalized selection of target drugs. *Semin Cancer Biol.* 2018;53:110–124. doi:10.1016/j.semcancer.2018.06.003.
- Zolotovskaia MA, Sorokin MI, Emelianova AA, et al. Pathway based analysis of mutation data is efficient for scoring target cancer drugs. *Front Pharmacol.* 2019;10:1. doi:10.3389/fphar.2019.00001.
- Zolotovskaia MA, Sorokin MI, Roumiantsev SA, Borisov NM, Buzdin AA. Pathway instability is an effective new mutation-based type of cancer biomarkers. *Front Oncol.* 2019;8:658. doi:10.3389/fonc.2018.00658.
- Nikitin D, Penzar D, Garazha A, et al. Profiling of human molecular pathways affected by retrotransposons at the level of regulation by transcription factor proteins. *Front Immunol.* 2018;9:30. doi:10.3389/fimmu.2018.00030.
- Nikitin D, Garazha A, Sorokin M, et al. Retroelement-linked transcription factor binding patterns point to quickly developing molecular pathways in human evolution. *Cells.* 2019;8:E130. doi:10.3390/cells8020130.
- Borisov N, Shabalina I, Tkachev V, et al. Shambhala: a platform-agnostic data harmonizer for gene expression data. *BMC Bioinformatics.* 2019;20:66. doi:10.1186/s12859-019-2641-8.
- Borisov N, Tkachev V, Muchnik I, Buzdin A. Individual drug treatment prediction in oncology based on machine learning using cell culture gene expression data. Paper presented at: International Conference on Computational Biology and Bioinformatics; October 18–20, 2017:1–6; Newark, NJ. New York, NY: ACM Press. doi:10.1145/3155077.3155078.
- Borisov N, Tkachev V, Buzdin A, Muchnik I. Prediction of drug efficiency by transferring gene expression data from cell lines to cancer patients. In: Rozonoer L, Mirkin B, Muchnik I, eds. *Braverman Readings in Machine Learning. Key Ideas from Inception to Current State*, Vol 11100. Cham, Switzerland: Springer; 2018:201–212. doi:10.1007/978-3-319-99492-5_9.
- Borisov N, Tkachev V, Suntsova M, et al. A method of gene expression data transfer from cell lines to cancer patients for machine-learning prediction of drug efficiency. *Cell Cycle.* 2018;17:486–491. doi:10.1080/15384101.2017.1417706.
- Tkachev V, Sorokin M, Mescheryakov A, et al. FLOating-Window Projective Separator (FloWPS): a data trimming tool for support vector machines (SVM) to improve robustness of the classifier. *Front Genet.* 2019;9:717. doi:10.3389/fgenet.2018.00717.
- Buzdin A, Sorokin M, Glusker A, et al. Activation of intracellular signaling pathways as a new type of biomarkers for selection of target anticancer drugs. *J Clin Oncol.* 2017;35:e23142. doi:10.1200/JCO.2017.35.15_suppl.e23142.
- Poddubskaya E, Baranova M, Allina DM, et al. Personalized prescription of imatinib in recurrent granulosa cell tumor of the ovary: case report [published online ahead of print January 17, 2019]. *Cold Spring Harb Mol Case Stud.* doi:10.1101/mcs.a003434.
- Poddubskaya EV, Baranova MP, Allina DO, et al. Personalized prescription of tyrosine kinase inhibitors in unresectable metastatic cholangiocarcinoma. *Exp Hematol Oncol.* 2018;7:21. doi:10.1186/s40164-018-0113-x.
- Sorokin M, Kholodenko R, Suntsova M, et al. Oncobox bioinformatical platform for selecting potentially effective combinations of target cancer drugs using high-throughput gene expression data. *Cancers (Basel).* 2018;10:E365. doi:10.3390/cancers10100365.