Novel multidimensional biomarkers for non-invasive medical diagnostics

M S Vonsky$^{1,2,3}$, A L Runov$^{1,2,3}$ and O M Moiseeva$^1$

$^1$ Almazov Medical Research Centre, St.-Petersburg, 197431, Russia
$^2$ D.I. Mendeleev Institute for Metrology, St.-Petersburg, 190005, Russia
$^3$ Institute of Cytology RAS, St.-Petersburg, 194064, Russia

E-mail: m.vonsky@gmail.com

Abstract. Value of clinical laboratory diagnostics results are often underestimated in medical practice due to the fact that result of laboratory analysis most cases represents a single biomarker of the organism homeostasis complex pathological alteration. New systemic approaches emerging on the basis of Omic’s knowledge and techniques allow development of multidimensional biomarkers, reflecting multiple pathobiological axes which characterize defined pathological condition. Combination of liquid biopsy technology with multidimensional biomarkers allows to obtain clinically valuable personalized information on the patient by non-invasive approach even for hard-to-reach organs and tissues.

1. Introduction

Biological markers, or biomarkers, are defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or therapeutic intervention” [1]. Even biomarkers cannot substitute clinical endpoints completely, their importance increases in medical laboratory diagnostics and medical research.

Laboratory diagnostics historically plays a secondary role in clinical practice. In most cases results of laboratory studies are taken into consideration, but they are not decisive in the diagnosis demanding supplementary instrumental examinations. This position of the clinical laboratory originated from the fact that each individual qualitative or quantitative result of laboratory analysis represents a single biomarker of the complex homeostasis of the organism alteration. A result, provided by laboratory, as a rule, does not carry, or carries indirect diagnostically significant information about the pathology, the causes of its development or localization. For example, elevated activity level of alanine transaminase (ALT) for a long time was concerned as markers of hepatitis or muscular disease. Nowadays, ALT is considered only as a biomarker of a violation of cells integrity - elevated ALT activity may be caused by different pathophysiological events. Therefore, a laboratory report on elevated ALT activity provides evidently insufficient information to explain the clinical implication underlying the pathology. Another cardiac biochemical marker – troponin I (cTn I) – previously was considered as the preferred biomarker for the diagnosis of myocardial infarction. By fact, cTn I is a contractile protein, a component of cardiac myocytes filaments. Release of cTn I into circulation may occur due to ischemia, necrosis or damage of cardiomyocyte and may be related to underlying cardiac or non-coronary and extracardiac disease, such as severe renal dysfunction. But in order to verify the acute
coronary syndrome diagnosis, combination of increased cTn I level, echocardiography and electrocardiography data are necessary.

The exceptions are biomarkers that carry the necessary and sufficient information for medical decision making. These include, for example, biomarkers of infectious diseases - nucleic acids and antigens of pathogenic microorganisms and viruses, identification of which prove the presence of the pathogen and allows to evaluate the microbial or viral load, to monitor the results of therapy in dynamics. Improving the information load of biomarkers could simplify the diagnosis algorithms and can be implemented through a systematic approach that uses a combination of biomarkers to target the noncommunicable disease.

The implementation of the Human Genome project has led to the explosive development of Omic’s, systemic approaches that consider biological processes in the complex, taking into account their direct and indirect influence on each other. Thus, genomics considers the genome not as a sum of genes, but as their organized system, proteomics — a complex of cell proteins in their interaction, and transcriptomics reflects the current pattern of gene expression due to the processes occurring in the cell. Further on, metabolomics deals with a comprehensive evaluation of metabolites, substrates and by-products of enzymatic reactions which reflect all functional activities ongoing in biological systems. Taken together, all this approaches allow to discuss patholomics, as the systemic study covering intertwining of genes, proteins and metabolites underlying, promoting and reflecting the pathology development Omic’s approach together with bioinformatics allows development of multiplex, or multidimensional biomarkers, reflecting different pathobiological axes which characterize defined pathological condition.

2. Transcriptomic biomarkers
Development of pathological processes is necessarily associated with changes in the transcriptome of the cells of the affected tissue, manifested through a decrease or increase of the expression level of a defined set of genes. The expression profiles of a diagnostic panel of genes determined in this way represent perspective biomarkers reflecting defined pathophysiologial processes. Microarray analysis (high density nucleic acids hybridization on microchip) is often performed to study transcription profile, but from analytical point of view this data can be considered as qualitative and preliminary, allowing to identify but not validate candidate biomarkers. Quantitative RNA measurements nowadays can be assessed after the reverse transcription by means of established DNA analysis methods – real-time and digital quantitative polymerase chain reaction (RT-qPCR and RT-dPCR) as well as RNA sequencing or next-generation sequencing (NGS).

2.1. RNA quantitative measurements
Comparability of RNA copy number ratio measurements for single and multiple gene targets were studied previously in the Consultative Committee for Amount of Substance (CCQM) international comparison studies CCQM P103 “Quantification of RNA transcript” and P103.1 “Measurement of multiple RNA transcripts” performed under the auspices of the CCQM Nucleic Acids Working Group [2]. Measurements of three human endogenous gene targets were performed in the background of human cell line RNA. Good agreement was observed between reported results obtained by means of RT-qPCR, RT-dPCR and NGS in this model study.

Capabilities to perform RNA copy number ratio measurements itself are not sufficient for transcriptomic biomarker development, validation and their application in medical purposes. Identification of internal control - reference genes, characterized by stable level of expression, unaffected by pathology, is a cornerstone of RNA quantitative measurements in biological material. These genes (usually in a combination of 2, up to 5) are used for normalization of RT-qPCR results, allowing to perform copy number ratio, or relative measurements of target genes expression with required accuracy. Usually a small subset of so called “housekeeping” candidate reference genes is tested in medical studies – glyceraldehydes 3-phosphate dehydrogenase (GAPDH), hypoxanthine phosphorybosyl transferase (HPRT), β-actin, tubulin, and ribosomal RNA genes are typical examples
of frequently used reference genes [3]. Stability of candidate gene expression usually evaluate applying software applications GeNorm [4], BestKeeper [5] or NormFinder [6] to process RT-qPCR experimental data.

Typical problems linked with the reference genes selection, include a limited number of candidate reference genes tested and requirements for comparable expression level of target and reference genes. To ensure comparability of RT-qPCR results, obtained in different laboratories, it is accepted to follow the guideline for minimum information for publication of quantitative real-time PCR experiments (MIQE) [7].

2.2. RNA biomarkers of cardiovascular disease
Cardiovascular disease is a major public health problem globally and causes in many countries twice more death than cancer, promoting development of new diagnostic and therapeutic approaches. A number of protein-based circulating biomarkers are well characterized and used in routine clinical practice for diagnosis and monitoring of cardiovascular disease. In case of known target it is possible to substitute analyte (protein or peptide) by corresponding gene transcript. This approach allow to measure target gene expression in tissues of interest directly, verifying localization of pathological process. For example, soluble ST2 protein – receptor of interleukin 1 – can be measured in peripheral blood and used in risk stratification of both acute and chronic heart failure. In the early study of Weinberg et al. it was shown, that expression of IL1RL1 (ST2 receptor coding gene) is increased in cardiac myocytes in case of stretch [8]. Even this study was performed by semiquantitative end-point PCR analysis, it result in a novel RNA biomarker of heart failure development and made one of the first steps on the way from single, “stand alone” to multidimensional biomarkers.

2.3. Multidimensional RNA biomarkers
Development of real multidimensional transcriptomic biomarkers begins on the basis of microarray technology since 1999 - investigators at MIT identified a signature of 50 genes out of 6817 surveyed that could distinguish between acute myeloid leukemia and acute lymphoblastic leukemia [9]. Since then, in the area of oncology transcriptomic biomarkers – unique gene expression profiles – were applied for classification of hematologic malignancies, solid tumors, prediction of survival, clinical outcome, recurrence risk, and response to therapy. In the area of cardiology – another important group of noncommunicable disease, comparison of global gene expression profiles in endomyocardial biopsy (EMB) specimens from clinically and histopathologically characterized patients demonstrated ability of microarrays to facilitate the differential diagnosis of cardiomyopathies [10, 11]. It was shown, that about 45% of genes are differentially expressed between patients groups. In some cases transcriptomic biomarkers allow to detect pathology better than histological approaches regarded as “golden standards” in laboratory diagnostic [11].

3. Multidimensional biomarkers for myocarditis diagnosis
Myocarditis is commonly defined as inflammation of the myocardium followed by cardiomyocyte necrosis. Distribution of the myocarditis in human population reaches 4%, however, real prevalence of the disease suggested to be underestimated as it is often asimptomatic or has unclear clinical manifestations. Clinical diagnostic of myocarditis based on histolopathological analysis of EMB, so called Dallas criteria, proposed in 1986 by the European Society of Cardiology. It was suggested in 2006, that due to EMB sampling errors, variation in expert interpretation, variance with other markers of viral infection and immune activation in the heart, and variance with treatment outcomes the Dallas criteria are no longer adequate [12].

Microarray analysis on EMB samples from patients with histologically proven myocarditis (n=16) and dilated cardiomyopathy (n=32) reveal 9878 differentially expressed genes [13]. A multidimensional biomarker consists of 62 genes, which distinguished myocarditis with 100% sensitivity and 100% specificity, was identified. Results of microarray analysis were validated by of RT-qPCR; some candidate genes were not confirmed and the final panel included 13 genes involved
in inflammation and cardiac remodeling pathways. No data was presented in the paper on reference genes used or RT-qPCR normalization.

In our study a set of 6 candidate reference genes was assessed for normalization of expression in cardiac tissues from patients with myocarditis, a combination of GAPDH and HPRT1 was selected [14]. EMB samples (n=23, M:15, F:8) from myocardial patients were included into the study together with control samples – tissue fragments obtained during orthotopic heart transplantation (n=6, M:4, F:2). Target panel included 61 candidate genes. It was shown, that myocarditis induced statistically significant (p<0.035) modulation of gene expression in cardiac tissue of male and female patients (10 and 12 genes, correspondingly). Unexpectedly, we found significant gender-dependent difference in gene expression profiles: only 6 genes were co-modulated for both genders (FCER1G, IL2, NOTCH2, SEC24A, SIGLEC1, TMOD3). Even in this co-modulated subset expression of FCER1G was up-regulated 7-fold higher in males, expression TMOD3 was down-regulated 3-fold lower in females. Expression of ADCY7 was modulated for both genders, up-regulated for males and down-regulated for females. Three genes were differentially expressed only in male patients (GLIPR, ITGB2, NFKB), five – only in female (ACOX1, CD1D, CD14, CD8B1, HRH3). A correlation was found between expression modulation and disease progress for 4 genes, so this subset can be used for therapeutic monitoring purpose.

Unfortunately, this perspective approach includes invasive endomyocardial biopsy step, but it can be used as a supplementary study to improve histopathological examinations of EMB samples.

4. Non-invasive multidimensional biomarkers
Combination of liquid biopsy technology with transcriptional multidimensional biomarkers allows to realize non-invasive differential diagnostic of hard-to-reach organs and tissues pathologies, including pathologies of cardiovascular system and oncopathologies. Potential correlation of simultaneous gene expression in the cardiac tissue and peripheral blood mononuclear cells (PBMC) is widely discussed: blood considered as a surrogate tissue, a liquid organ, being in an intimate proximity to the heart. Up to 80% of known genes are expressed in PBMCs; blood cells, being in constant contact with numerous exo- endogenous factors, have to respond to them via differential gene expression.

4.1. Cardiac pathology biomarkers
Doxorubicin (DOX) is a commonly used antibiotic for malignancies treatment. It may cause unpredictable doze-dependent cardiotoxicity begins with the first doze: patients, receiving chemotherapy, are at increased risk of developing cardiac dysfunction. Differential 235 genes expression was revealed in PBMC of breast cancer patients treated with single dose of DOX by RT-qPCR method with single 18S mRNA reference gene [15]. Out of these 235 genes, Todorova et al. identified 63 genes, differentially expressed in patients with abnormal heart left ventricle ejection fraction as a marker of cardiotoxicity. Candidate non-invasive multidimensional biomarker was developed as a result of the study, proving applicability of PBMC gene expression profiling for diagnosis of cardiac pathologies.

In the ongoing in our laboratory study we are developing candidate transcriptomic biomarkers of myocarditis examining gene expression in PBMC of patients and healthy donors. No correlation up to the moment was found between gene expression in EMB samples and PBMCs. Gender-dependent target gene expression difference is not as significant in PBMC, as in cardiac tissue. More then 20 differentially expressed genes (p<0.03) were identified at the moment by RT-qPCR, which allows us to hope for a successful completion of the project.

5. Conclusion
Multidimensional biomarkers are one of the key tools of personalized medicine. Considering liquid biopsy as non-invasive source of biomaterial, we can develop RNA biomarkers using coding and non-coding nucleic acids. Application of Omic’s approach together with bioinformatics allows to develop
not only diagnostic biomarkers, but to monitor disease and treatment progress taking into account the individual properties of pathophysiological processes.

**Ethics Statement**
This study was carried out in accordance with the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the ethics committee of the Almazov Medical Center. All subjects signed an IRB approved informed consent where they were informed for the use of their EMB, blood samples and medical records for research purposes.

**References**
[1] FDA-NIH Biomarker Working Group 2016 *BEST (Biomarkers, EndpointS, and other Tools) Resource* (Silver Spring (MD): Food and Drug Administration (US); Bethesda (MD): National Institutes of Health (US))
[2] Devonshire A et al. 2016 *Biomol. Detect. Quantif.* 8 15
[3] Suzuki T, Higgins P and Crawford D 2000 *BioTechniques* 29 332
[4] Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T, Igout A and Heinen E 1999 *J. Biotechnol.* 75 291
[5] Vandesompele J, De P, Pattyn F, Poppe B, Van R, De P and Speleman F 2002 *Genome Biol.* 3 RESEARCH0034
[6] Pfaffl M, Tichopad A, Prgoet C and Neuvians T 2004 *Biotechnol. Lett.* 26 509
[7] Andersen C, Jensen J and Orntoft T 2004 *Cancer Res.* 64 5245
[8] Bustin S et al 2009 *Clin. Chem.* 55 611
[9] Weinberg E, Shimpo M, De Keulenaer G, MacGillivray C, Tominaga S, Solomon S, Rouleau J and Lee R 2002 *Circulation* 106 2961
[10] Ginsburg G, Seo D and Frazier C 2006 *JACC* 48 1618
[11] Barth A et al. 2006 *J. Am. Coll. Cardiol.* 48 1610
[12] Ruppert V, Meyer T, Pankuweit S, Möller E, Funck C, Grimm W, Maisch B and German Heart Failure Network 2008 *J. Thorac. Cardiovasc. Surg.* 136(2) 360
[13] Baughman K 2006 *Circulation* 113 593
[14] Heidecker B, Kittleson M, Kasper E, Wittstein I, Champion H, Russell S, Hruban R, Rodriguez E, Baughman K and Hare J 2011 *Circulation* 123 1174
[15] Runov A, Kurchakova E, Khaschevskaya D, Moiseeva O and Vonsky M 2015 *Cell Tissue Biol.* 9 330.
[16] Todorova V, Makhoul I, Siegel E, Wei J, Stone A, Carter W, Beggs M, Owen A and Klimberg °V 2016 *PLoS One* 11 e0160224