How multiple NFAT5 sites can impact to the NFAT5-inducible gene expression: an approximate solution of an ill-posed inverse problem

Petr M. Ponomarenko[1,*], Mikhail P. Ponomarenko[1], Nikolay A. Kolchanov[1,2,3],

[1]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia; [2]Novosibirsk State University, Novosibirsk, Russia; [3]National Research Centre "Kurchatov Institute", Moscow, Russia; [*]pon@bionet.nsc.ru

Birgit Eisenhaber[4] and Frank Eisenhaber[4,5,6]

*[4]Bioinformatics Institute; A*STA R; Singapore; [5]Department of Biological Sciences; National University of Singapore; [6]School of Computer Engineering; Nanyang Technological University; Singapore; franke@bii.a-star.edu.sg*

The transcription factor NFAT5 (Nuclear Factor of Activated T-cells 5; synonyms: TonEBP and OREBP as Tonicity Element-Binding Protein and Osmotic Response Element-Binding Protein, respectively) is a master-gene for the osmotic stress response. It is, therefore, critical for many developmental and regeneration processes such as wound healing and it is known that certain drugs interfere with this pathway [1]. Multiple response elements (RE) within the regulatory gene regions are characteristic of the glucocorticoid, steroid, thyroid, auxin (AuxRE), and many other hormones as well as of the heat shock, hypertonic, osmotic, and many other stresses. There is few data on the molecular mechanisms underlying how multiple REs contribute to gene expression whereas a bulk of the data on the only single RE is available.

In our previous work [2], we have introduced an approximate solution of an ill-posed inverse problem, of dissecting the expression, φ, of a given gene originating from different variants of multiple REs and to decode the sequence-activity relationships of each of them:

$$\varphi = \varphi_0 \left(1 + \sum\nolimits_{1 \le k \le N} P_k \sum\nolimits_{\substack{1 \le n_1 \le N-k \\ n_1 < n_2 \le N-1 \\ \cdots \\ n_{k-1} < n_k \le N}} \varphi_{n_1 n_2 \ldots n_{k-1} n_k} \right); \quad (1)$$

here: φ_0, a basal expression level; P_k, a probability of k from amount N REs, $\Sigma_{1 \le k \le N} P_k \equiv 1$; $N>1$; $\varphi_{\zeta\xi\ldots\varsigma}$ is an impact of a given set of ζ-th, ξ-th, … and ς-th REs into the induced expression level.

As for N REs Eq. (1) contains 2^N variables, the optimal size of independent experimental data for its estimation by the multiple linear regression is 2^{2N}, i.e. 16, 64, 256, … at $N \in \{2, 3, 4, \ldots\}$ that is unavailable. But two extreme cases, additive $\{P_1=1; P_{k>1}=0\}$ and multiplicative

{$P_{k<N}$=0; P_N=1}, are solvable by the standard tools STATISTICA, the last case by Sandwich theorem. By this way we found the insignificant additive and significant multiplicative impacts of the multiple AuxREs into the expression level magnitudes of three auxin-responsive genes in plants, and a repression of these genes without auxin [2] that is impossible to do otherwise.

In this work we analyzed six independent datasets on the multiple NFAT5 binding sites both wild-typed and subjected to genetic manipulations within promoters of six genes, namely: *HSP70-2* (heat shock protein 70-2) in human and in mouse, *AR* (aldose reductase) in human and in rat, *SMIT* (sodium/myo-inositol cotransporter) in human, *AQP2* (aquaporin) in mouse.

As an example, upon 19 probes of four TonREs - A, B, C, D, - of murine *HSP70-2* promoter, x, upstream *Luc* (luciferase) gene of pGL2 plasmid by which the murine IMCD cells were transfected [3], we found that both additive (a) and multiplicative (m) impacts of these TonREs into luciferase activity, φ, are significant: {$\varphi_0^{(a)}$=1.0, $\varphi_A^{(a)}$=5.2, $\varphi_B^{(a)}$=3.6, $\varphi_C^{(a)}$=−6.3, $\varphi_D^{(a)}$=2.7} and $\varphi_0^{(m)}$=0.5, $\varphi_A^{(m)}$=7.3, $\varphi_B^{(m)}$=3.8, $\varphi_C^{(m)}$=0.2, $\varphi_D^{(m)}$=2.3}. There are the significant correlations between $\varphi^{(a)}$ and $\varphi^{(m)}$ values (r=0.98, α<0.025), between reconstructions $\varphi(\varphi^{(a)}(x))$ and $\varphi(\varphi^{(m)}(x))$ based on each of them (r=0.96, α<0.05), and, also (r=0.98, α<0.025), between the magnitudes φ and their joint reconstruction φ(x) using Eq.(2), δ(true)=1 and δ(false)=0, as:

$$\varphi(x) = \frac{2}{3}\left(1 + \frac{1}{2}\sum\nolimits_{\xi \in \{A,B,C,D\}} \varphi_\xi^{(a)}\,\delta(\xi \in x) + \frac{1}{2}\prod\nolimits_{\xi \in \{A,B,C,D\}} \left(\varphi_\xi^{(m)}\right)^{\delta(\xi \in x)}\right). \qquad (2)$$

Both negative $\varphi_C^{(a)}$=−6.3 additive and fractional $\varphi_C^{(m)}$=0.2 multiplicative impacts mean the repressive affect of TonRE-C to murine *HSP70-2* that was consciously lost earlier [3] because it is immeasurable experimentally against three rest TonREs activating this gene. Finally, we confirmed the repression of *HSP70-2* using an independently derived human dataset [4].

We thank grants RFBR 11-04-01888, 13-04-00359, SciSh-5278.2012.4, Russian Ministry of Education & Science Project 8740, and RAS Programs “Molecular and cellular biology”.

1. L.K.Schenk et al. (2010) *Cell Physiol Biochem,* **126:**887-900.
2. V.V.Mironova et al. (2013) *J Bioinform Comput Biol.,* **11:**1340011.
3. S.K.Woo et al. (2002) *Mol Cell Biol.,* **22:** 5753-5760.
4. J.I.Heo et al. (2006) *Exp Mol Med.,* **38:**295-301.