Identifying networks of estrogen-responsive genes in breast cancer cells

ID - 116

Altobelli Gioia[1,2], Caselle Michele[3,2], De Bortoli Michele[1,2]
[1]Department of Oncological Sciences, University of Turin
[2]Complex Systems in Molecular Biology and Medicine, University of Turin
[3]Department of Theoretical Physics, University of Turin

Motivation

Estrogens regulate a variety of physiological processes in human body. In the genomic pathway, they act through transcription factors, ERs, which recognize either specific nucleotide sequences, EREs, or other proteins already bound to DNA. The transcriptional result (activation or repression) is determined by type of co-factor complexes recruited locally, which in turn is gene- and cell type-specific. This raises a key question: which is the syntax of these complexes recruitment? Or, with the traditional words of biology, how are these genes regulated? An answer would be of great interest in medicine, as endocrine therapies are widely employed in breast cancer treatment. Indeed, anti-estrogenic therapies fail in 40% cases, likely due to the cellular context which may convert the action of estrogenic antagonists into an agonistic one. Expression data from microarray experiments show patterns which suggest a selective co-regulation of the estrogen-responsive genes, e.g. repressed vs. stimulated genes, and early vs. late responders. We collected and integrated available genome-wide data in order to classify the myriad of estrogen-responsive genes functionally, a valuable first step towards an understanding of the molecular syntax and hence a better molecular therapy.

Methods

We built a database of genes which are unambiguously up/down-regulate at different kinetics (early, intermediate, late, or unknown) after estrogen stimulation, in immortalized breast cancer cells and mainly considering micro-arrays expression data.
Each gene has been given a score which reflects number of independent experiments and experimental assessment of ERE presence. The database, EREGLON, also contains additional information such as 5-flanking regions and number of ERE elements found in their (-2000,+500) DNA regions with a computational tool (Lazzarato et al.
Bioinformatics, 2003). Afterwards, we used a combination of DNA sequence analyses in order to assess features of down-regulated, as compared to up-regulated, genes 5-flanking regions of the different classes of genes stored in EREGLON. In addition to the traditional approaches based both on alignment matrices and enumerative algorithms, we employed a method that combines phylogenetic conservation with motif over-representation (Corï¿½et al. BCM Bioinformatics, 2004).

Results

We obtained two distinctive collections of overrepresented motifs, out of highly conserved and aligned 15kb-upstream regions of those genes possessing a mouse orthologue. The upstream regions of early up-regulated genes remarkably differ from the ones of early down-regulated genes. We are extending our comparison to all gene lists, including late vs. early for both regulation classes. A screening of the (-2000,+500) DNA regions for TATA box presence with a hidden Markov model (Frith et al., 2001), revealed that the percentage of TATA+ genes is noticeably higher within the up-regulated set as compared to both down-regulated and random control ones.
Besides, we found 30% of the down regulated genes and 43% of up-regulated ones can be classified as TATA+ promoters according to a published study, where, over 9,010 core promoter sequences in human genome, the ones with TATA box are 20.5% as compared to 13% in randomized sequences (Jin et al. BCM Bioinformatics, 2006). We are validating our results with the RIKEN database. Our final goal is to infer regulatory combinations of factors to be tested in-vivo by biomolecular tools. Responsiveness to estrogens and anti-estrogens in real tumors shall be included in EREGLON, for the comparison between clinical and model data must provide a valuable insight into breast cancer pathogenesis. The fully integrated database may become a publicly accessible resource.

Email: gioia.altobelli@unito.it