The EXPANDER Integrated Platform for Transcriptome Analysis

Tom Aharon Hait[1,2], Adi Maron-Katz[3], Dorit Sagir[1], David Amar[4], Igor Ulitsky[5], Chaim Linhart[1], Amos Tanay[5,6], Roded Sharan[1], Yosef Shiloh[2], Ran Elkon[2,7†] and Ron Shamir[1,†]

*1 - **The Blavatnik School of Computer Science,** Tel Aviv University, Tel Aviv, Israel*
*2 - **Department of Human Molecular Genetics and Biochemistry,** Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel*
*3 - **Department of Psychiatry and Behavioral Sciences,** Stanford University, Stanford, CA 94305, USA*
*4 - **Stanford Center for Inherited Cardiovascular Disease,** Stanford University, Stanford, CA 94305, USA*
*5 - **Department of Biological Regulation,** Weizmann Institute of Science, Rehovot 76100, Israel*
*6 - **Department of Computer Science and Applied Mathematics,** Weizmann Institute of Science, Rehovot 76100, Israel*
*7 - **Sagol School of Neuroscience,** Tel Aviv University, Tel Aviv, Israel*

***Correspondence to Ran Elkon and Ron Shamir:**Correspondence to: R. Elkon, Department of Human Molecular Genetics & Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel.Correspondence to: R. Shamir, Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, 69978, Israel. ranel@tauex.tau.ac.il, rshamir@tau.ac.il.*
https://doi.org/10.1016/j.jmb.2019.05.013

Abstract

Genome-wide analysis of cellular transcriptomes using RNA-seq or expression arrays is a major mainstay of current biological and biomedical research. *EXPANDER* (EXPression ANalyzer and DisplayER) is a comprehensive software package for analysis of expression data, with built-in support for 18 different organisms. It is designed as a "one-stop shop" platform for transcriptomic analysis, allowing for execution of all analysis steps starting with gene expression data matrix. Analyses offered include low-level preprocessing and normalization, differential expression analysis, clustering, bi-clustering, supervised grouping, high-level functional and pathway enrichment tests, and networks and motif analyses. A variety of options is offered for each step, using established algorithms, including many developed and published by our laboratory.

EXPANDER has been continuously developed since 2003, having to date over 18,000 downloads and 540 citations. One of the innovations in the recent version is support for combined analysis of gene expression and ChIP-seq data to enhance the inference of transcriptional networks and their functional interpretation. EXPANDER implements cutting-edge algorithms and makes them accessible to users through user-friendly interface and intuitive visualizations. It is freely available to users at http://acgt.cs.tau.ac.il/expander/.

Introduction

The ability to profile entire cellular transcriptomes, first by expression microarrays and subsequently by RNA-sequencing (RNA-seq) has transformed biological research over the last two decades, by turning the paradigm of systems-level analysis from a formidable task to one that is readily accessible to most experimental laboratories [1]. Consequently, transcriptome profiling is one of the most vastly used techniques in biomedical research, being routinely applied for multiple goals, including the elucidation of transcriptional networks that drive different biological processes and cellular responses to challenges and the identification of expression signatures for multiple pathological conditions [2–4]. The utility of transcriptomic experiments greatly depends on the availability of advanced, yet easy to use, bioinformatics tools for mining meaningful biological knowledge out of the raw data.

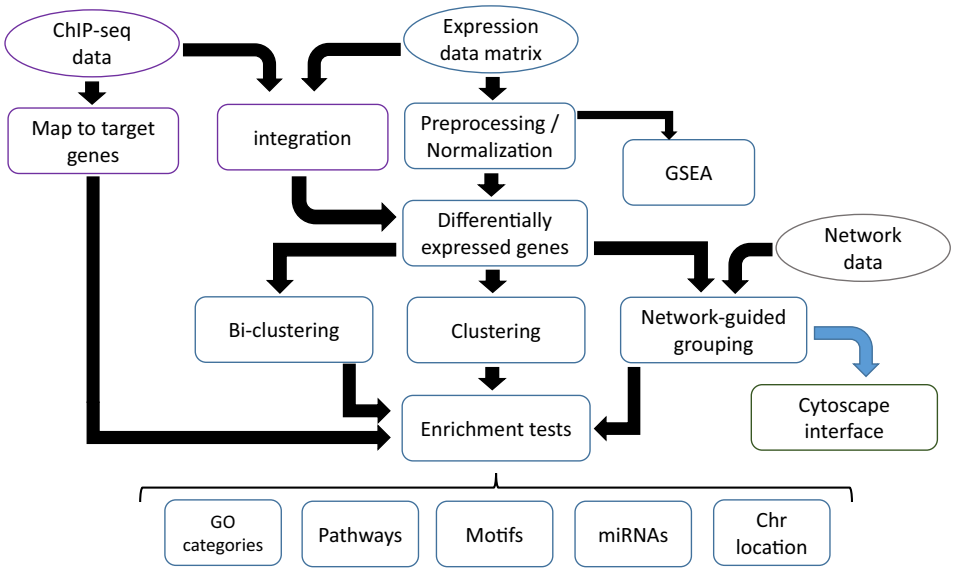

Fig. 1. A flowchart summarizing the main analysis modules that are implemented in *EXPANDER 8.0*. (analysis modules involving ChIP-seq data are detailed in Fig. 3).

Gene expression (GE) data analysis is a multi-step process, ranging from low-level analyses that generate a normalized data matrix to high-level network-guided analyses that dissect the transcriptional response into functional modules that carry out specific biological endpoints. While there are many tools and software packages that perform dedicated tasks, we reasoned that an integrative platform that covers the entire analysis pipeline would greatly boost the ability of nonspecialist experimental laboratories to effectively mine their transcriptomic data. With this guiding principle, we are continuously developing *EXPANDER* (EXPression ANalyzer and DisplayER) since 2003 [5,6] as a "one-stop shop" software tool for GE data analysis. Here we describe the current version of our tool—*EXPANDER 8.0*. Analysis modules provided by *EXPANDER 8.0* include low-level preprocessing and normalization, differential expression analysis, clustering, bi-clustering, supervised grouping, high-level functional and pathway enrichment tests, and networks and cis-regulatory motif analyses. A variety of established algorithms, including many developed and published by our lab, are offered for many steps. These algorithms are made accessible to users through user-friendly interface, and results are displayed in intuitive and interactive visualizations. As an interactive tool, by design, EXPANDER avoids computationally heavy processes. Hence, the majority of the implemented analysis modules typically take less than a minute or just a few minutes to complete on a standard PC.

Recent years also witnessed great advance in the application of epigenomics techniques, including genome-wide profiling of physical interactions between transcription factors (TFs) and the chromatin using ChIP-seq [7]. The true power of this method for delineating transcriptional networks lies in its integrated analysis with corresponding transcriptomic data sets. To meet this challenge, one of the innovations in *EXPANDER 8.0* is the support for combined analysis of GE and ChIP-seq data. Another aspect that significantly enhances inference of biological working models from transcriptomic data is its analysis in the context of gene and protein networks [8]. Therefore, in *EXPANDER 8.0*, we substantially augmented this analysis feature. The current version allows streamlined export of network-based results into the highly popular Cytoscape tool [9]. Once imported into Cytoscape, users can benefit from the multitude of analysis plugins and network visualization utilities provided by this platform.

EXPANDER is freely available to users at http://acgt.cs.tau.ac.il/expander/.

Results and Discussion

EXPANDER covers, under one roof, all the steps of GE data analysis, starting with GE data matrix recorded using either GE arrays or RNA-seq. The analysis modules provided by *EXPANDER* are summarized in Fig. 1, and the tools/algorithms implemented in each module are summarized in Supplementary Table S1. Some of these modules require species-specific annotation files [e.g., Gene Ontology (GO) annotations, promoter sequences], and *EXPANDER 8.0* has a built-in support for 18 different organisms (Table S2). Here, we demonstrate EXPANDER's analysis pipeline by its application to a data set that profiled GE, using RNA-seq,

in cells treated by ionizing irradiation (IR) at three time-points (0, 4, and 8 h) using two biological replicates (of CAL51 breast cancer cell line) [10] (Table S3). We focus here on the typical analysis pipeline. Description of alternative ones is given in EXPANDER's user manual.

Data preprocessing and low-level analysis

An EXPANDER session typically starts by uploading an expression data matrix (gene or probes in rows and samples in columns), which contains estimates for either absolute or relative expression levels. A matrix of read counts is supported for analysis of RNA-seq data. The user also indicates at this stage the organism of the analyzed data set. As *EXPANDER*'s species annotation files use specific gene accession IDs (e.g., Entrez Gene IDs for *Homo sapiens*), the user should provide a conversion file that maps the IDs used in the data file to IDs expected by *EXPANDER* for that organism. (Conversion files from probe ids to gene ids are provided by *EXPANDER* for selected widely used expression arrays). The example data file we analyzed here contains 14,992 genes measured over six biological samples. As initial preprocessing, *EXPANDER* offers multiple normalization schemes. For RNA-seq data, trimmed mean of M values [11], relative log expression [12], and upper quartile normalization [13] are offered. For microarray expression data or processed RNA-seq data (e.g., RPKM data) that requires further normalization, *EXPANDER* offers two normalization schemes: quantile [14] and lowess [15]. The overall structure of the data can then be inspected using hierarchical clustering and principal component analysis. Subsequent analyses are usually performed on a subset of genes that showed significant expression variation across the biological conditions [termed differentially expressed (DE) genes]. *EXPANDER* offers both generic tests for DE genes (e.g., moderated T test implemented by limma [16]) and methods specifically developed for data obtained by expression arrays (SAM [17]) or RNA-seq read counts data (edgeR [18] and DESeq2 [12,19]). In the analyzed data set, 637 DE genes were identified based on fold-change criteria.

Gene grouping based on expression patterns

DE genes in a data set exhibit various expression patterns (induction or repression, different response kinetics, etc.), where genes sharing an expression pattern (co-expressed genes) are expected to function together towards achieving certain biological endpoints [20]. Therefore, typically, the next step in the analysis pipeline is the partition of DE genes into groups according to their expression patterns. *EXPANDER* provides three approaches for this task: (a) clustering, (b) network-based grouping, and (c) biclustering.

Clustering divides the DE genes into distinct clusters such that genes assigned to the same cluster show highly similar patterns, while those assigned to different clusters show dissimilar patterns. *EXPANDER* implements three popular clustering algorithms: K-means [21], self-organizing maps [22], and CLICK, which was developed in our laboratory [23]. In the example data set, CLICK divided the set of DE genes into four clusters (Figs. S1 and 2A).

Network-based grouping seeks groups of genes that show highly correlated expression pattern and are located in close proximity to each other in some global gene network (e.g., protein–protein interaction network). For this task, *EXPANDER* provides protein–protein interaction network files for selected organisms (Table S2) and contains an implementation of MATISSE, our graph-theoretic algorithm that detects significant co-expressed connected subnetworks (*network modules*) [24]. Such modules, whose definition is based on the combined analysis of network and expression data, are often more biologically meaningful than co-expression clusters, as they account also for the network relationships between the genes. Furthermore, the genes in such a module are more likely to be functionally related and act together in the same pathway. Figure 2B shows a network module of 30 genes detected by MATISSE in our data set. Each module is composed of "forward" and "back" nodes. The forward nodes

Fig. 2. Expression-based gene grouping and enrichment tests. (A) Average expression patterns of the cluster of 367 genes, detected by CLICK in our example data set, showing a sustained induction of expression upon irradiation. In *EXPANDER*'s display, each cluster is represented by the average expression pattern calculated over all the genes in the cluster. Error bars indicate ± SD. Conditions are labeled over the X axis. Y axis shows standardized expression levels. (B) A network module detected by MATISSE in our example data set. This module contains 30 genes, consisting of 17 genes that showed a transient decrease in expression upon irradiation ("forward" nodes colored in blue) and 13 genes ("back" nodes colored in pink) added by the algorithm to induce connectivity of the forward nodes. (C) Results of GO enrichment analysis (using TANGO) for the cluster shown in panel A. This cluster is enriched for genes that function in the regulation of cell proliferation and cell death. (D) Results of motif enrichment analysis (using PRIMA) for the cluster shown in panel A. The promoters of the genes assigned to this cluster were found to be enriched for motifs of several TFs, including p53, which is a well-known key regulator of the transcriptional responses to irradiation. Each bar corresponds to an enriched TF. Clicking on a bar opens a window that lists the genes in the cluster whose promoters contain a hit for the TF motif.

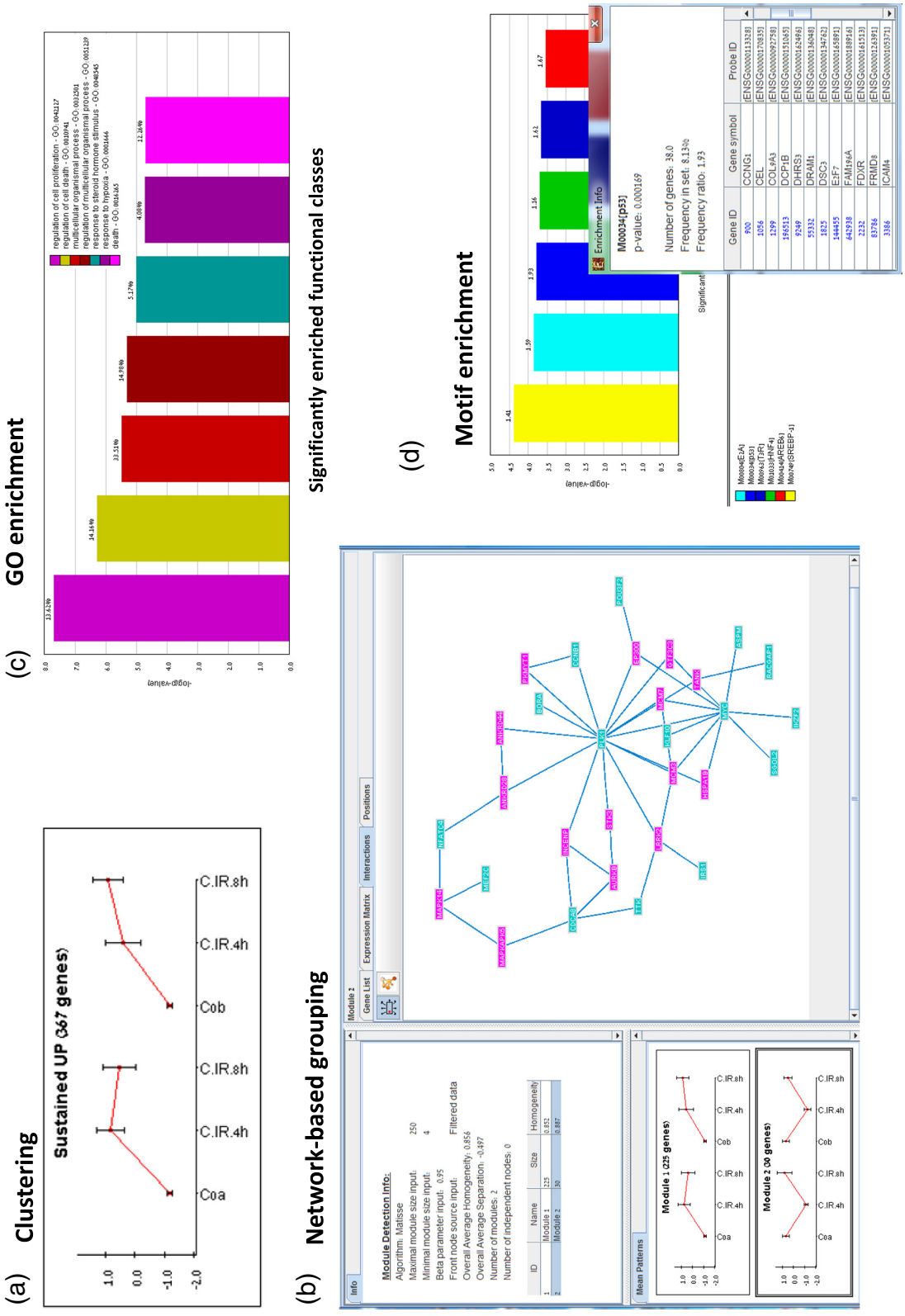

Fig. 2. (*legend on previous page*)

show highly correlated expression pattern in the data set, while the back nodes are added by the algorithm to keep high connectivity among the module's genes. Note that the robustness of network-based grouping building on correlated expression patterns tends to increase with the number of samples.

To further augment network-based analysis, *EXPANDER 8.0* implements a streamlined interface with the highly popular Cytoscape platform [9], which provides numerous visualization and analytical utilities for the combined analysis of expression and network data. Figure S2 shows the MATISSE module imported to Cytoscape. Once loaded into Cytoscape, users can benefit from the plethora of plugins offered by this rich resource.

Clustering assumes that co-expressed genes show a global similarity over all the examined conditions. This assumption is less likely to hold when a data set contains a large number of different conditions (e.g., above 20). In such cases, *biclustering* is a preferred alternative for grouping genes according to expression pattern. A bicluster is a set of genes that show high similarity over a *subset* of the conditions. A *biclustering algorithm* can detect a collection of biclusters, where genes or conditions can take part in more than one bicluster. *EXPANDER* implements the iterative signature algorithm [25] and our SAMBA biclustering algorithm [26], which can handle data sets with hundreds to thousands of conditions.

To increase flexibility, users can upload their own gene sets and use this partition for subsequent analyses.

Enrichment tests

Once sets of co-expressed genes are identified, the next challenge is to ascribe each set with some biological function and to discover key regulators that control the observed co-expression. To this end, *EXPANDER* implements five enrichment tests (Fig. 1): (a) GO tests using our TANGO algorithm [6], which takes into account the dependencies between GO terms that stem from the hierarchical structure of this ontology; (b) *Pathway DB* tests (including KEGG [27] and WikiPathways [28]), which examine, respectively, whether each gene set is enriched for particular GO terms or canonical pathways (Fig. 2C shows the results of GO enrichment analysis for the cluster of genes that were induced in response to IR in our data set, and Fig. S3A shows the results for all the clusters; Fig. S3B shows enriched GO terms found on MATISSE's module; Fig. S4 shows an enriched KEGG pathway detected in our data set); (c) *Chromosome location* tests, evaluating if the genes in any gene set are enriched for any specific chromosomal region (this test could, for example, detect mega-base scale deletions and amplifications in cancer cells); (d) *Motif analysis* aimed at detecting cis-regulatory DNA motifs that are over-represented in the promoter sequences of the co-expressed genes, under the assumption that co-expression implies transcriptional co-regulation. *EXPANDER* implements both our PRIMA method [29], which uses prior information on TF PWMs, and our AMADEUS algorithm for *de novo* motif analysis [30]. Over the last decade, we demonstrated the power of this reverse engineering approach to reveal key regulators of numerous biological processes in human and mouse, including cell-cycle progression [29], DNA damage response [31], immune responses [32], and cellular differentiation [33] (Fig. 2D shows the over-represented motifs detected by PRIMA in the cluster of IR-induced genes); and (e) *miRNA analysis* that searches for enriched miRNA target sites in 3′ UTRs of co-expressed genes (reasoning that co-expression could stem not only from transcriptional co-regulation but also from co-regulation of mRNA stability controlled by miRNAs). *EXPANDER* implements our FAME algorithm [6], which uses TargetScan predictions [34], utilizing target site context scores and accounting for possible biases owing to 3′-UTR sequence length.

The above enrichment tests are applied to clusters of genes that passed some test for differential expression. However, transcriptomic experiments typically include only a small number of replicate samples (mostly 1–3 replicates per condition), which limits the statistical power of tests for DE genes. An alternative statistical approach provided by

Fig. 3. EXPANDER's modules for ChIP-seq analysis. (A) A flowchart summarizing the modules implemented in *EXPANDER* for analysis of ChIP-seq data. (B) Results of de novo motif analysis (using AMADEUS) applied to the 1830 peaks detected in the p53 ChIP-seq data set. Reassuringly, the top enriched motif corresponds to the known binding motif of p53 (accession M00272 in Transfac DB). (C) Mapping of the p53 peaks to their nearest genes, within a distance limit, defined a set of 376 putative p53 target genes. Gene sets derived from ChIP-seq peaks can be tested for differential expression between two selected conditions. Shown here is the fold change in expression upon irradiation (IR 8 h) compared to the control untreated cells (C0). The putative p53 target genes (n = 367) show significant induction compared to a background (Bg) gene set in the data set (n = 11,697; *p*-value calculated using one-sided Wilcoxon's test). (D) Gene sets derived from ChIP-seq peaks can also be utilized for enrichment tests applied to gene groups defined by *EXPANDER* based on their expression patterns. In the example shown here, the cluster of 367 genes that showed a sustained induction of expression upon irradiation (shown in Fig. 2A) was significantly enriched for the 376 putative p53 target genes (93 overlapping genes; enrichment factor = 8.35; $p = 2.75 \times 10^{-61}$ (hypergeometric test)).

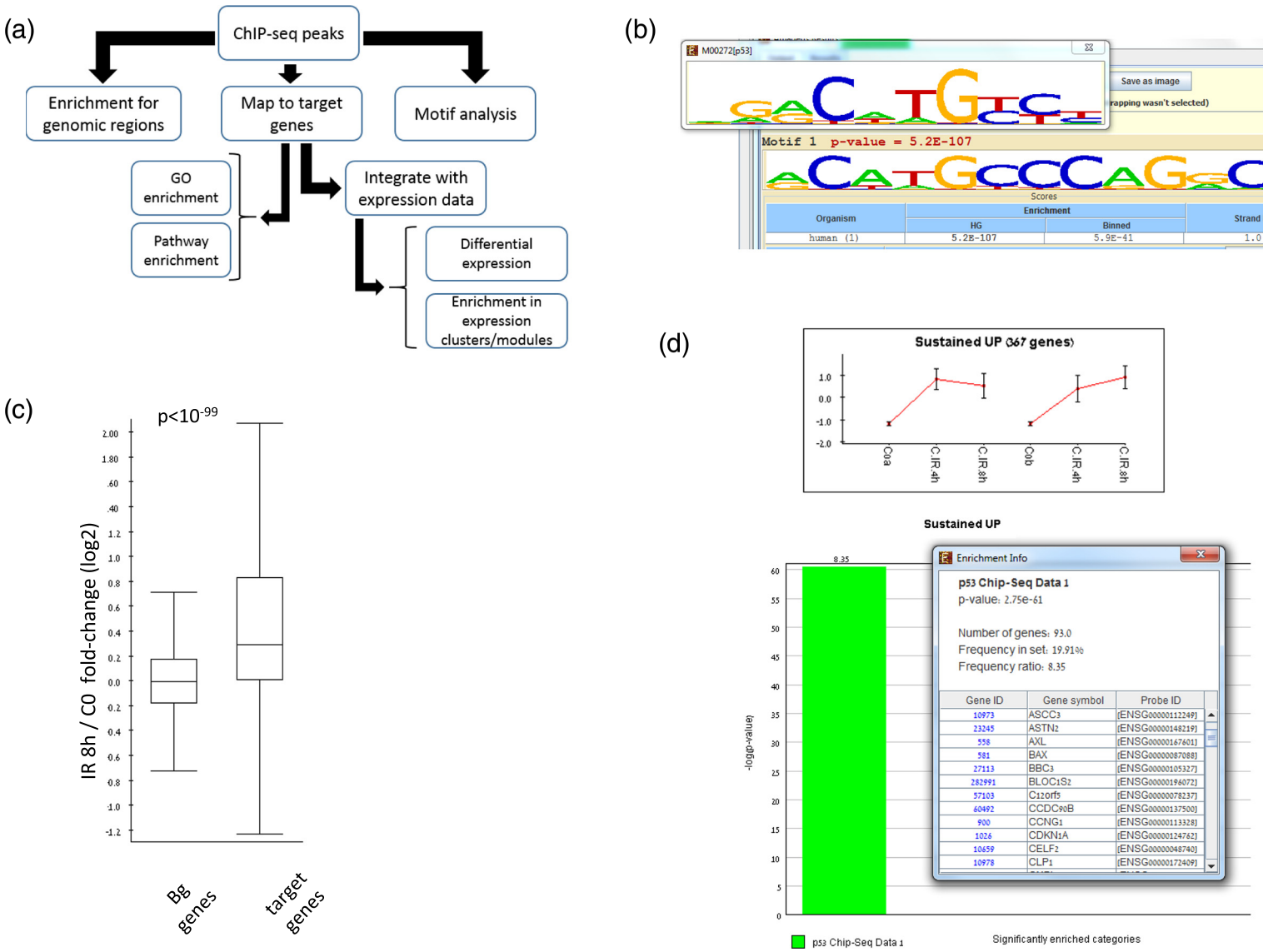

Fig. 3. (*legend on previous page*)

EXPANDER for functional interpretation of transcriptomic data is based on implementation of gene set enrichment analysis (GSEA) [35]. Instead of focusing on the set of DE genes, GSEA considers all the genes that are expressed in a data set and ranks them based on a score of differential expression between the compared samples (e.g., fold-change or T score calculated between treated and control samples). The ranked gene list is then tested against a large number of curated gene sets, seeking those whose genes are significantly concentrated at either end of the expression list (these ends represent, respectively, induced and repressed genes). This sensitive method builds on the amplification of weak signals, achieved by considering the coordinated response of many genes that function in the same process, where individually most of them show only mild change in expression, not sufficient to reach statistical significance in per-gene tests. Figure S5 shows a few results of GSEA on our data set.

Analysis of ChIP-seq data

One of the main novel features implemented in *EXPANDER 8.0* is the support for integrated analysis of GE and ChIP-seq data (Table S2). For this task, users should upload ChIP-seq peaks (detected by a peak calling tool, e.g., MACS2 [36]) in the standard .bed format, and indicate the version of the genome assembly in which peak coordinates are given (e.g., hg19 or hg38 for *H. sapiens*). As an example, we uploaded ChIP-seq results for p53 profiled 2 h after irradiation in the same cell line that was used in the expression data set described above (Table S4) [10]. A total of 1830 p53 binding sites were detected in this ChIP-seq experiments. Once the peaks .bed file is uploaded, *EXPANDER* provides three utilities (Figs. 3A; S6A): (a) *Enrichment for genomic regions*. Based on genome annotation files, *EXPANDER* provides statistical summary for the prevalence of peaks located in different genomic categories (e.g., intergenic, upstream TSS/promoter, introns) and tests for enrichment of the peaks for any category (Fig. S6B–C). (b) *Motif analysis*. Enriched DNA motifs are sought in the peaks' sequences (compared to flanking sequences) using AMADEUS. (The sequences of the peaks and their control flanking regions are extracted from the corresponding genome using our implementation of twoBitToFa utility from UCSC (https://github.com/ENCODE-DCC/kentUtils)). This forms an important quality control step for TF ChIP-seq experiments, as the called peaks are expected to be enriched for the motif of the examined TF (and potentially, also for motifs of its cofactors) (Fig. 3B). (c) *Map peaks to nearest genes*. To allow integrated analysis with GE data, the peaks are mapped to their putative target genes. Currently, *EXPANDER* implements a naïve approach in which each peak is mapped to its closest k genes located within a distance of L bp for it (k and L are set by the user; by default, $k = 1$ and $L = 1$ Mbp) (Fig. S6A).

The mapping of ChIP-seq peaks to genes produces a gene set of putative targets of the analyzed TF. This gene set can be combined into the analysis of GE data using two modes (Fig. 3A): (a) *differential expression analysis*, which tests if the genes in the set are significantly up- or down-regulated in the comparison between two conditions selected by the user (Fig. 3C), and (b) *enrichment tests*, which examine if the gene-set defined by the ChIP-seq peaks is enriched for any gene cluster detected based on expression data (as described in the gene grouping section above) (Fig. 3D). Collectively, these features enhance users' ability to gain insights into the function of TFs and delineate the pathways by which these biological endpoints are mediated and executed.

In conclusion, the widespread scope of the analysis modules provided by *EXPANDER*, the strength of the algorithms it implements, and the ease of their access to users make our package a unique software suit for analysis of transcriptome data.

Methods

EXPANDER is implemented in Java programming language and makes use of components of the R statistical language. EXPANDER works under Windows and Linux operating systems (OSs). EXPANDER requires at least 2 GB RAM in order to functionally work and offers two options to launch the application: EXPANDER_2GB.bat (2GB RAM) and EXPANDER_4GB.bat (4GB RAM). Users are expected to install the following tools according to their OS and 32/64-bit computer architecture: (1) Java run-time environment for running Java applications (https://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html), (2) R statistical language tool (https://cran.r-project.org/), and (3) Cytoscape tool (https://cytoscape.org/download-platforms.html). Additional information can be found in our online user manual: http://acgt.cs.tau.ac.il/expander/help/ver8.0Help/html/.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.05.013.

Acknowledgments

The study is supported in part by the DIP German–Israeli Project cooperation (to R.S. and R.E.), Israel Science Foundation (Grant No. 1339/18 to R.S.), and Len Blavatnik and the Blavatnik Family foundation (to R.S). T.A.H. is supported in part by a

fellowship from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University. R.E. is a Faculty Fellow of the Edmond J. Safra Center for Bioinformatics at Tel Aviv University.

Received 7 February 2019;
Received in revised form 7 May 2019;
Available online xxxx

Keywords Gene expression analysis;
Software package;
Transcriptomic analysis;
RNA-seq analysis

Equal contribution.

Abbreviations used:
RNA-seq, RNA-sequencing; GE, gene expression; *EXPANDER*, EXPression ANalyzer and DisplayER; TF, transcription factor; IR, ionizing irradiation; DE, differentially expressed; GO, Gene Ontology; GSEA, gene set enrichment analysis.

References

[1] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (2009) 57–63.
[2] M. Cieslik, A.M. Chinnaiyan, Cancer transcriptome profiling at the juncture of clinical translation, Nat. Rev. Genet. 19 (2018) 93–109.
[3] K.F. Au, V. Sebastiano, The transcriptome of human pluripotent stem cells, Curr. Opin. Genet. Dev. 28 (2014) 71–77.
[4] T.G. Rubin, J.D. Gray, B.S. McEwen, Experience and the ever-changing brain: what the transcriptome can reveal, Bioessays 36 (2014) 1072–1081.
[5] R. Shamir, A. Maron-Katz, A. Tanay, C. Linhart, I. Steinfeld, R. Sharan, Y. Shiloh, R. Elkon, EXPANDER—an integrative program suite for microarray data analysis, BMC Bioinformatics 6 (2005), 232.
[6] I. Ulitsky, A. Maron-Katz, S. Shavit, D. Sagir, C. Linhart, R. Elkon, A. Tanay, R. Sharan, Y. Shiloh, R. Shamir, Expander: from expression microarrays to networks and functions, Nat. Protoc. 5 (2010) 303–322.
[7] T.S. Furey, ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions, Nat. Rev. Genet. 13 (2012) 840–852.
[8] K. Mitra, A.R. Carvunis, S.K. Ramesh, T. Ideker, Integrative approaches for finding modular structure in biological networks, Nat. Rev. Genet. 14 (2013) 719–732.
[9] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504.
[10] S. Rashi-Elkeles, H.J. Warnatz, R. Elkon, A. Kupershtein, Y. Chobod, A. Paz, V. Amstislavskiy, M. Sultan, H. Safer, W. Nietfeld, et al., Parallel profiling of the transcriptome, cistrome, and epigenome in the cellular response to ionizing radiation, Sci. Signal. 7 (2014) rs3.
[11] M.D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data, Genome Biol. 11 (2010) R25.
[12] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (2010) R106.
[13] J.H. Bullard, E. Purdom, K.D. Hansen, S. Dudoit, Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments, BMC Bioinformatics 11 (2010) 94.
[14] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, Bioinformatics 19 (2003) 185–193.
[15] J.A. Berger, S. Hautaniemi, A.K. Jarvinen, H. Edgren, S.K. Mitra, J. Astola, Optimized LOWESS normalization parameter selection for DNA microarray data, BMC Bioinformatics 5 (2004) 194.
[16] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (2015) e47.
[17] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 5116–5121.
[18] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
[19] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.
[20] W. Tian, L.V. Zhang, M. Tasan, F.D. Gibbons, O.D. King, J. Park, Z. Wunderlich, J.M. Cherry, F.P. Roth, Combining guilt-by-association and guilt-by-profiling to predict *Saccharomyces cerevisiae* gene function, Genome Biol. 9 (Suppl. 1) (2008) S7.
[21] S. Tavazoie, J.D. Hughes, M.J. Campbell, R.J. Cho, G.M. Church, Systematic determination of genetic network architecture, Nat. Genet. 22 (1999) 281–285.
[22] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, T.R. Golub, Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 2907–2912.
[23] R. Sharan, A. Maron-Katz, R. Shamir, CLICK and EXPANDER: a system for clustering and visualizing gene expression data, Bioinformatics 19 (2003) 1787–1799.
[24] I. Ulitsky, R. Shamir, Identification of functional modules using network topology and high-throughput data, BMC Syst. Biol. 1 (2007) 8.
[25] S. Bergmann, J. Ihmels, N. Barkai, Iterative signature algorithm for the analysis of large-scale gene expression data, Phys. Rev. E Stat. Nonlinear Soft Matter Phys. 67 (2003), 031902.
[26] A. Tanay, R. Sharan, M. Kupiec, R. Shamir, Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 2981–2986.
[27] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, Nucleic Acids Res. 44 (2016) D457–D462.
[28] D.N. Slenter, M. Kutmon, K. Hanspers, A. Riutta, J. Windsor, N. Nunes, J. Melius, E. Cirillo, S.L. Coort, D. Digles, et al., WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research, Nucleic Acids Res. 46 (2018) D661–D667.

[29] R. Elkon, C. Linhart, R. Sharan, R. Shamir, Y. Shiloh, Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells, Genome Res. 13 (2003) 773–780.
[30] C. Linhart, Y. Halperin, R. Shamir, Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets, Genome Res. 18 (2008) 1180–1189.
[31] R. Elkon, S. Rashi-Elkeles, Y. Lerenthal, C. Linhart, T. Tenne, N. Amariglio, G. Rechavi, R. Shamir, Y. Shiloh, Dissection of a DNA-damage-induced transcriptional network using a combination of microarrays, RNA interference and computational promoter analysis, Genome Biol. 6 (2005) R43.
[32] R. Elkon, C. Linhart, Y. Halperin, Y. Shiloh, R. Shamir, Functional genomic delineation of TLR-induced transcriptional networks, BMC Genomics 8 (2007) 394.
[33] R. Elkon, B. Milon, L. Morrison, M. Shah, S. Vijayakumar, M. Racherla, C.C. Leitch, L. Silipino, S. Hadi, M. Weiss-Gayet, et al., RFX transcription factors are essential for hearing in mice, Nat. Commun. 6 (2015) 8549.
[34] V. Agarwal, G.W. Bell, J.W. Nam, D.P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs, Elife 4 (2015).
[35] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B. L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.
[36] T. Liu, Use model-based analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein–DNA interactions in embryonic stem cells, Methods Mol. Biol. 1150 (2014) 81–95.