

# AltAnalyze and DomainGraph: analyzing and visualizing exon expression data

Dorothea Emig<sup>1,2,\*</sup>, Nathan Salomonis<sup>3,\*</sup>, Jan Baumbach<sup>2</sup>, Thomas Lengauer<sup>1</sup>, Bruce R. Conklin<sup>3,4</sup> and Mario Albrecht<sup>1</sup>

<sup>1</sup>Max Planck Institute for Informatics, 66123 Saarbrücken, Germany, <sup>2</sup>International Computer Science Institute, University of California, Berkeley, CA 94704, <sup>3</sup>Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158 and <sup>4</sup>Department of Medicine and Department of Molecular and Cellular Pharmacology, University of California, San Francisco, CA 94143, USA

Received January 31, 2010; Revised April 12, 2010; Accepted May 2, 2010

## ABSTRACT

Alternative splicing is an important mechanism for increasing protein diversity. However, its functional effects are largely unknown. Here, we present our new software workflow composed of the open-source application AltAnalyze and the Cytoscape plugin DomainGraph. Both programs provide an intuitive and comprehensive end-to-end solution for the analysis and visualization of alternative splicing data from Affymetrix Exon and Gene Arrays at the level of proteins, domains, microRNA binding sites, molecular interactions and pathways. Our software tools include easy-to-use graphical user interfaces, rigorous statistical methods (FIRMA, MiDAS and DABG filtering) and do not require prior knowledge of exon array analysis or programming. They provide new methods for automatic interpretation and visualization of the effects of alternative exon inclusion on protein domain composition and microRNA binding sites. These data can be visualized together with affected pathways and gene or protein interaction networks, allowing a straightforward identification of potential biological effects due to alternative splicing at different levels of granularity. Our programs are available at <http://www.altanalyze.org> and <http://www.domaingraph.de>. These websites also include extensive documentation, tutorials and sample data.

## INTRODUCTION

Alternative splicing is an important biological mechanism for producing a great variety of eukaryotic protein isoforms from a comparatively small number of genes. Recent studies indicate that 92–94% of all human multi-exon genes undergo alternative splicing (1). A large number of alternatively spliced genes and their protein products are identified by exon tiling microarrays, such as the Affymetrix Exon Array (2), as well as by deep sequencing of transcriptomes (3,4). Important functional implications of alternative splicing have been demonstrated for selected genes (5,6), but not yet for the large majority of splicing events discovered for many mammalian genes. Splice variants from a single gene might alter the composition of functional regions, such as protein domains and other sequence motifs (6) or prevent protein translation by introducing a premature stop codon (7). The functional impact of alternative splicing can be profound, ranging from the gain or loss of specific molecular interactions to changes of pathway dynamics (8). Recently, it was found that alternative splicing can also regulate the inclusion of microRNA (miRNA) binding sites as an important means of controlling protein expression (9). In addition to alternative splicing, transcription through alternative promoter selection and alternative 3'-end processing by selection of alternative polyadenylation sites are other critical modes of transcript regulation that effect protein composition and expression (10,11).

Several stand-alone programs, web services and Bioconductor packages have been developed to aid in the analysis of Affymetrix Exon Array data and to increase the accuracy and reliability of alternative exon

\*To whom correspondence should be addressed. Tel: +49 681 9325 329; Fax: +49 681 9325 399; Email: demig@mpi-inf.mpg.de

\*Correspondence may also be addressed to Nathan Salomonis. Tel: +1 415 734 2772; Fax: +1 415 355 0960; Email: nsalomonis@gladstone.ucsf.edu

Present address:  
Jan Baumbach, Saarland University, 66123 Saarbrücken, Germany.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

detection. Whereas the majority of currently available tools are principally focused on statistical methods for alternative exon detection, few report the absolute positions of regulated probesets within transcripts and exons or relative to other regulated probesets. Furthermore, none indicate whether these exons have prior evidence of alternative splicing or alternative promoter activity and how such events might alter the protein composition in terms of putative protein domains, motifs or other important sequence elements. The programs easyExon (12) and ExpressionConsole (EC) ([http://www.affymetrix.com/products\\_services/software/specific/expression\\_console\\_software.affx](http://www.affymetrix.com/products_services/software/specific/expression_console_software.affx)) add a few biological annotations to their expression statistics results such as gene and GO annotations for probesets. However, users have to perform advanced analyses manually. Other programs, such as APT ([http://www.affymetrix.com/partners\\_programs/programs/developer/tools/powertools.affx](http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx)), MADS (13), Exonmap (14) and FIRMA (15), concentrate on statistical computations only. They do not provide an easy-to-use graphical interface that guides the user through the analysis, and they require prior knowledge of statistical programming languages like R (16). Web services, such as ExonMiner (17), do not depend on additional tools or prior programming knowledge for the statistical analysis, but require users to upload their potentially confidential microarray data. To sum up, few of these tools provide methods for downstream interpretation of the experimental data, and none of them evaluates the effects of alternative splicing on biological functions that result from the protein domain composition, miRNA binding site inclusion and modified pathways.

Therefore, we developed an integrated software workflow for the statistical and visual analysis of exon expression data in the context of interaction networks, pathways, protein domains and miRNA binding sites. This workflow consists of our programs AltAnalyze and DomainGraph for exploring the functional impact of alternative splicing and other modes of transcript regulation in human, mouse and rat. It incorporates the large majority of analysis options available in other tools (e.g. DABG filtering, splicing-index, FIRMA and MiDAS) and provides one of the only end-to-end solutions, from the statistical analysis of both Exon and Gene Array files to probeset-level visualization. Our workflow has several novel features, including the assignment of alternative splicing annotations and potential miRNA binding sites, the analysis of protein domains and their interactions, the visualization of potentially affected biological pathways and corresponding overrepresentation analysis, and batch HTML export (see Supplementary Table S1 for a list of specific functionalities of AltAnalyze and DomainGraph and a comprehensive comparison to other programs).

## ANALYSIS WORKFLOW

Our software workflow consists of two main components: AltAnalyze, which performs alternative exon and

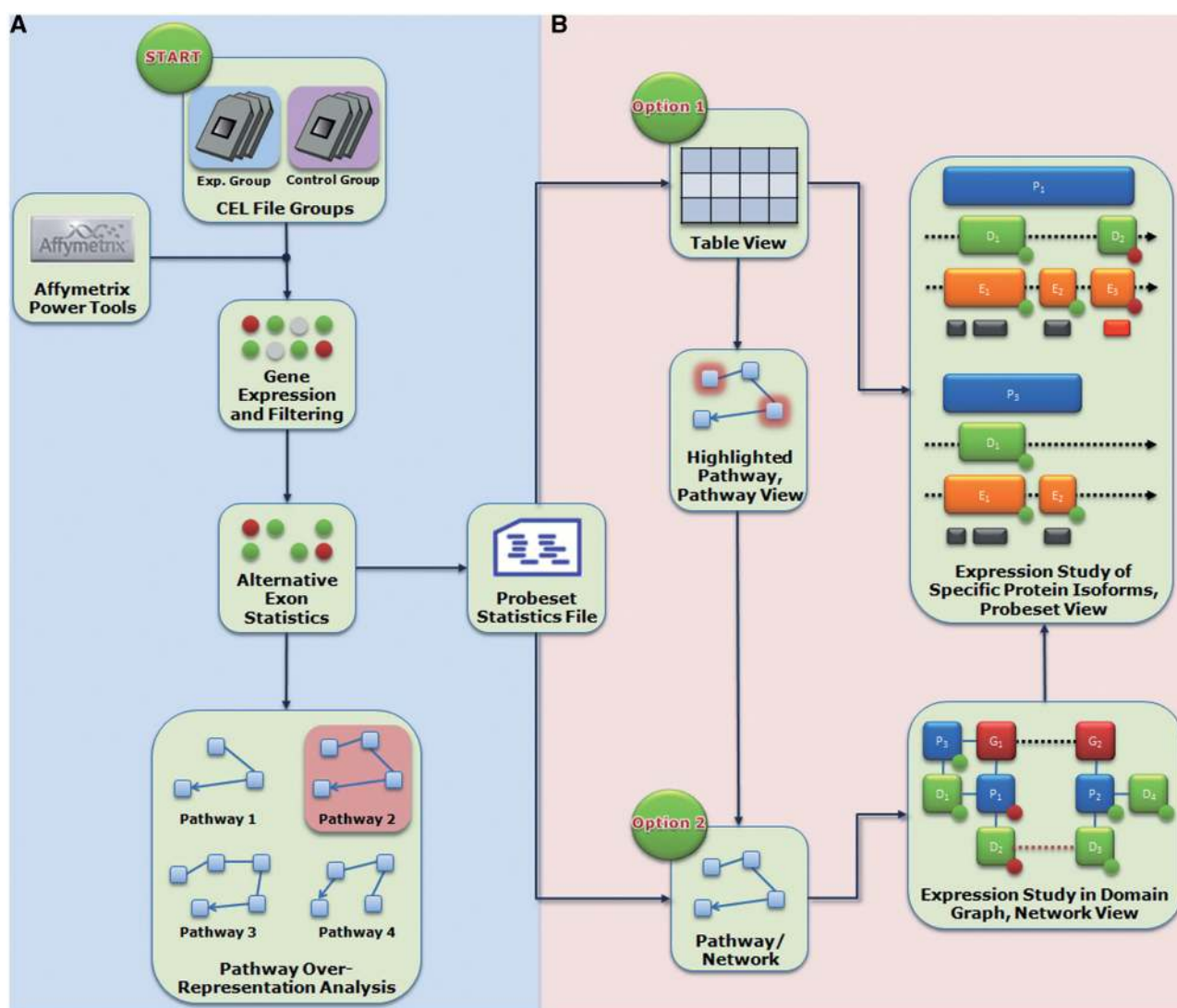
functional prediction analyses on Affymetrix Exon Array files, and DomainGraph, which is used for the visual investigation of potential biological effects of alternative splicing. The results file produced by AltAnalyze serves as interface between the two programs (Figure 1). This file contains alternative exon scores and *P*-values for all examined probesets from the microarray. These probeset statistics are used as input for DomainGraph, which annotates these statistically enriched probesets with gene and pathway information and allows for investigating the potential functional implications of regulated probesets.

AltAnalyze and DomainGraph are both designed to run on Windows, Unix and Mac OS. AltAnalyze is a stand-alone, open-source application, while DomainGraph is designed as a plugin for the free, open-source network visualization software Cytoscape (18). Both AltAnalyze and DomainGraph rely on locally installed databases based on annotation files provided by Affymetrix and the corresponding builds of the Ensembl database (19). New database releases can be downloaded from within both applications whenever new annotation files are made available by Affymetrix. The databases contain all necessary gene and protein data for the analysis and visualization of Affymetrix Exon and Gene Array data. AltAnalyze database files are stored as tab-delimited text files in species- and release-specific local database directories. In case of DomainGraph, an embedded Apache Derby database (<http://db.apache.org/derby>) is employed, which maintains all required data as well as the user's exon expression data (see Supplementary Materials and Methods for details on database contents).

AltAnalyze and DomainGraph can be installed and run either consecutively or separately on the user's computer. The downloadable AltAnalyze package includes both Cytoscape and DomainGraph, allowing users to run the complete software workflow without a separate installation of the programs. Users can thus immediately continue analyzing potential functional implications after the statistical analysis has finished. For users who prefer to run AltAnalyze and DomainGraph separately, DomainGraph is included in the Cytoscape Plugin Manager and can be downloaded and installed directly from within Cytoscape.

## ALTANALYZE

Users begin the workflow by selecting their array type and species in AltAnalyze along with the option to process raw Affymetrix CEL files, already normalized expression files in order to identify alternative exon expression between pairs of biological groups (experimental group versus control group) or between multiple groups. Analyses can be performed in either the graphical user interface or by command-line instructions. Any number of CEL files for any number of biological groups and comparisons can be loaded. The biological groups and comparisons are established in the graphical user interface together with expression and 'detection above background' (DABG) *P*-value



**Figure 1.** Software overview. AltAnalyze is shown on the left (A, blue background) and DomainGraph on the right (B, red background). The alternative exon statistics computed by AltAnalyze form the interface between the two programs and are the starting point for DomainGraph. The analysis begins with AltAnalyze (START symbol): grouping of raw expression data, normalization, summarization, filtering and calculation of differential probeset expression. This results in a series of probeset and gene expression summary files along with overrepresentation statistics. The probeset statistics file can then be loaded into DomainGraph for further investigation. DomainGraph allows for the analysis of a given experiment (Option 1 symbol) or a given protein or gene interaction network (Option 2 symbol). Option 1 starts with a tabular view of the data with biological annotations. In Option 2, a given network is augmented with all known protein isoforms, protein domains and domain interactions.

thresholds and other alternative exon analysis parameters. Once the parameters are specified, AltAnalyze first performs a 'robust multi-chip analysis' (RMA) summarization on the selected CEL files, retaining only those probesets that align to a single Ensembl gene. The output of this step is the gene expression statistics and annotations for each Ensembl gene and each biological comparison. Next, the probeset expression values are filtered based on user-defined absent–present parameters according to the DABG *P*-values and absolute expression levels for each biological comparison, a procedure that is recommended for alternative exon analysis. These filtered expression values are used to calculate standard

alternative exon statistics [splicing-index (20), FIRMA (15), 'microarray detection of alternative splicing' (MiDAS) *P*-value, normalized *P*-value] for each analyzed probeset relative to the determined gene expression levels (see Supplementary Materials and Methods for further details).

Although AltAnalyze was designed for computing alternative exon statistics based on the widely used splicing-index, FIRMA and MiDAS methods, it is also able to handle Exon and Gene Array data that was pre-processed using other statistical algorithms. This option provides additional flexibility, allowing users to analyze and interpret potential functional implications of



their Exon Array data based on the statistical method of their choice.

The output of AltAnalyze is a series of tab-delimited text files for each user-defined biological comparison. These can be opened in spreadsheet programs like Microsoft Excel or imported directly into DomainGraph for further analysis and visualization. Besides the alternative splicing statistics, these tabular data include domain-level predictions that are complementary to those found in DomainGraph as well as extended protein domain and motif predictions using an exhaustive alternative isoform analysis (21). This latter analysis also identifies the two most likely competitive isoforms for each alternative exon (one containing the exon and one that lacks it). Thus, it can be useful for identifying overall changes in domain/motif composition and predicted sequence lengths between these proteins. The predicted change in protein length and alignment to mRNAs that do not produce known proteins can be used to assess the likelihood of nonsense-mediated decay (21). The AltAnalyze result files additionally include overrepresented protein domains/motifs and miRNA binding sites affected by alternative exon inclusion. Furthermore, overrepresentation analyses can be performed for pathways annotated by WikiPathways (22) and Gene Ontology (23), using the integrated program GO-Elite (24).

## DOMAINGRAPH

The latest release 3.0 of the Cytoscape plugin DomainGraph (25,26) has been designed to directly load and analyze AltAnalyze alternative exon statistics, without requiring prior knowledge of genes or pathways potentially affected by alternative splicing. In addition, analyses of particular protein and domain interaction networks and their integration with AltAnalyze statistics are supported. For this purpose, DomainGraph includes a mapping between Exon Array probesets and Ensembl genes, transcripts, exons, proteins and Pfam domains (27). If a probeset shows significant up- or down-regulation according to the AltAnalyze results, biological annotations such as gene symbols, WikiPathway and Reactome (28) pathways, alternative splicing annotations and miRNA binding sites are automatically displayed. These biological data can then be easily visualized along with the effects of alternative splicing on pathways, genes, transcripts, exons, protein isoforms, protein domains and miRNA binding sites. Furthermore, users can start the analysis of AltAnalyze results with any interaction network or pathway that has been loaded into DomainGraph. The effects of alternative splicing can be comprehensively visualized and evaluated at different levels of granularity ranging from network- to exon-level perspectives.

### General analysis of AltAnalyze results

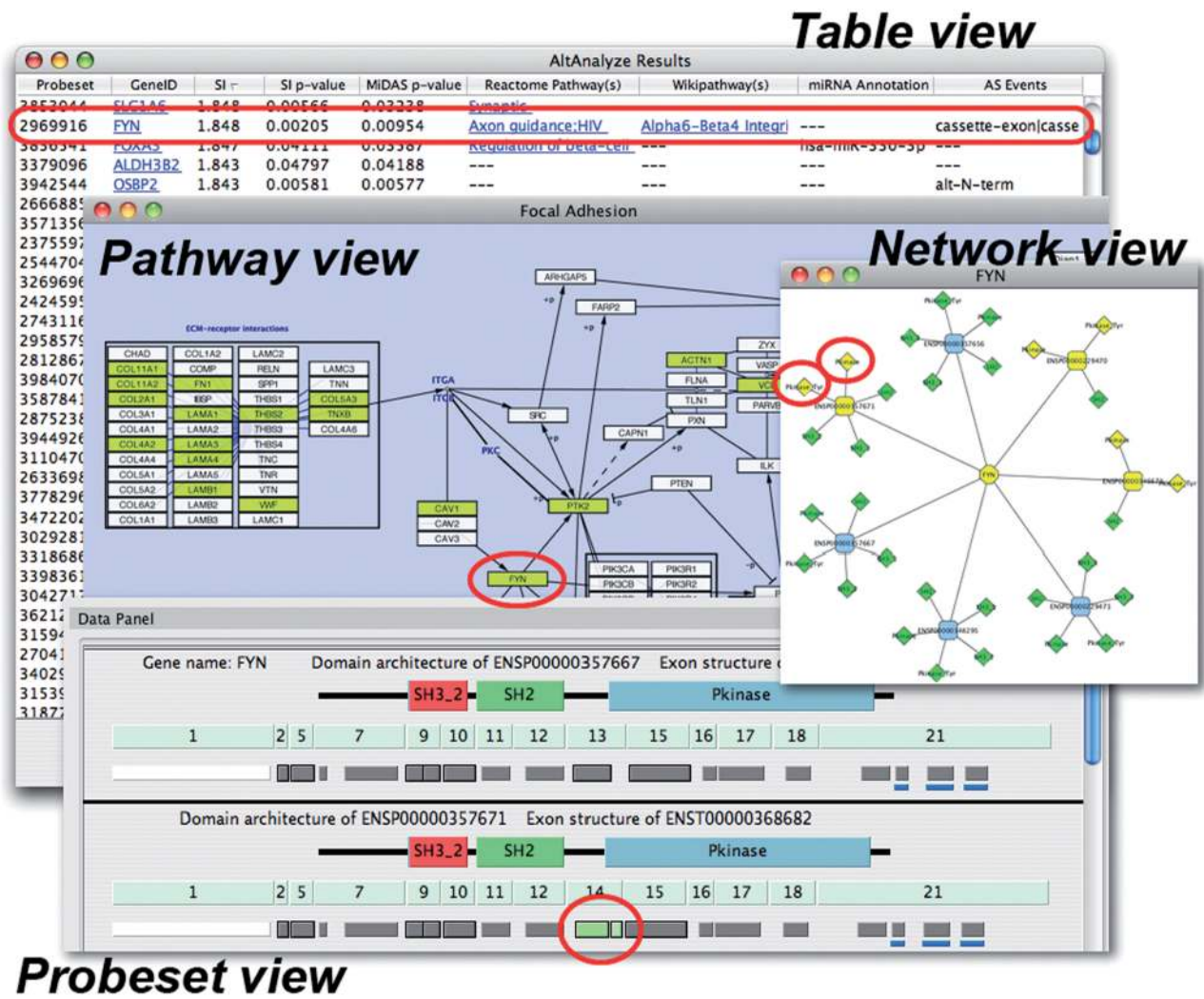
The most direct way to evaluate alternative exon statistics from AltAnalyze is to view significantly up- and down-regulated probesets in DomainGraph. After

importing the AltAnalyze statistics file, the user is automatically provided with a 'Table view' containing the AltAnalyze results with information on gene symbols, Reactome and WikiPathway pathway occurrences, miRNA binding site disruption and alternative splicing annotations for each probeset identified as differentially expressed by AltAnalyze (Figure 2, 'Table view'). Gene and pathway annotations immediately provide an overview of the biological context in which the regulation event occurs. Furthermore, the user can directly get an idea about the up- and down-regulated probesets mapping to putative miRNA binding sites and the genes they belong to. Additionally, several types of alternative exons are annotated in the table, e.g. exon skipping and alternative splice sites (see Supplementary Materials and Methods for all types of alternative exons that can be detected and the connection between probeset and exon annotations).

The selection of a gene in the table will display a 'Network view' with the gene and all known Ensembl protein isoforms and their domain compositions (Figure 2, 'Network view'). Additionally, a 'Probeset view' shows all these protein isoforms together with constituent Pfam domains, corresponding mRNA transcripts and exon structures, Affymetrix Exon Array probesets and miRNA binding sites (Figure 2, 'Probeset view'). In this view, probesets are colored according to their differential expression, pointing users to probesets with a significant up- or down-regulation in one of the biological groups.

The 'Probeset view' allows users to directly compare and analyze alternative exon expression between different protein isoforms produced by one gene. DomainGraph does not predict new protein isoforms or transcripts, but uses all information given by Ensembl. This allows users to view all gene products at once and to determine, which probesets show up- or down-regulation events and to which exons (and thus to which transcripts and protein isoforms) they map. Of course, a single isoform can be affected by multiple regulation events, which should be reflected by multiple up- or down-regulated probesets annotated to the respective isoform. As shown in Figure 3A, observing which alternative exons are present in which isoforms can lead to the identification of the affected isoforms rather than just a single alternative exon. By visualizing probesets overlapping with protein domains or miRNA binding sites, users can visually assess how an alternative exon may translate into altered protein function, expression or truncation. Tooltips provide additional information for the user, including the splicing-index fold change of probesets, *P*-values, alternative splicing annotations and cross-hybridization types. The latter indicates if a probeset matches to one or several genomic locations (see Supplementary Materials and Methods for details on annotations).

Furthermore, users can select Reactome or WikiPathways annotations from the table to load and visualize pathways of interest (Pathway view). These pathways are automatically overlaid with the AltAnalyze probeset statistics, and all network nodes associated with



**Figure 2.** DomainGraph visualization options for AltAnalyze alternative exon statistics. The ‘Table view’ contains all differentially expressed probesets with biological annotations. Gene and pathway annotations are clickable and lead to the ‘Pathway’, ‘Network’ and ‘Probeset views’. The ‘Pathway view’ shows the ‘Focal Adhesion’ WikiPathway; yellow colored gene boxes indicate the presence of alternative exons (see red frame in ‘Table view’). The ‘Probeset view’ displays two protein isoforms (black lines) with constituent domains (red, green and blue rectangles) and mRNA transcripts (light green, subdivided into exons; identical numbers correspond to the same exons, including 3’ and 5’ UTRs) produced by *FYN*. Probesets are shown below the mRNA transcripts (white boxes are probesets that did not meet the minimum expression thresholds in AltAnalyze; gray boxes for no alternative expression among the experimental and control groups; green boxes for decreased exon inclusion in the experimental group; red boxes for increased exon inclusion in experimental group; black frames around gray boxes for alternative splicing annotation). Potential miRNA binding sites are drawn as blue lines below probesets if present. The ‘Network view’ depicts all known protein isoforms (rectangle nodes) encoded by *FYN* together with their constituent domains (diamond nodes). Yellow nodes indicate the node overlaps with regulated probesets. The protein kinase domain of *FYN* is highlighted here because their interactions might be altered due to alternative splicing (red circle).

differentially expressed probesets are highlighted to facilitate the identification of potentially modified pathways (Figure 2, ‘Pathway view’). The ‘Table’, ‘Pathway’ and ‘Probeset’ views can be exported as an HTML web archive, which can be used to publish the data for all affected genes on a web server.

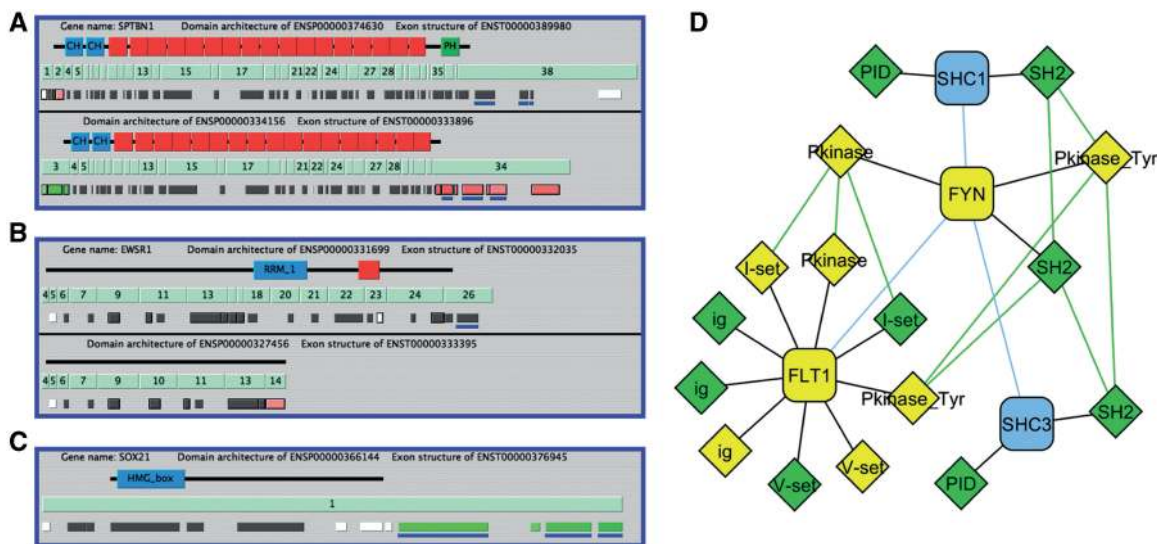
### Overlaying a particular network or pathway with AltAnalyze results

If a user is interested in a particular interaction network or pathway, AltAnalyze results can be integrated in order to evaluate protein isoforms or putative protein domain interactions and disruptions thereof. To this end, the user can import either gene or protein interactions into

Cytoscape from a flat file or by using another Cytoscape plugin. Interactions can also be obtained from external pathway resources, such as WikiPathways. If AltAnalyze data are integrated into DomainGraph, genes, proteins and domains associated with differentially expressed probesets are automatically highlighted (see Supplementary Materials and Methods).

When importing gene interactions, the focus lies on the encoded protein isoforms and their domain compositions to identify those isoforms potentially affected by alternative splicing and those remaining unchanged. The protein isoforms and their domains are extracted from the DomainGraph database (based on Ensembl annotations) and automatically added to the gene interaction network.





**Figure 3.** Using the DomainGraph 'Probeset view' to determine the mode and effect of alternative exon inclusion. (A) Alternative promoter and C-terminal exon expression for two protein isoforms of *STBN1*. Reciprocal expression of two alternative N-terminal exons is shown, along with the up-regulation of an alternative C-terminal exon, adding several predicted miRNA binding sites and eliminating the presence of a PH domain (green rectangle). (B) Truncation of *EWSR1* by up-regulation of an internal cassette exon. The up-regulation of this exon eliminates several critical protein domains for this transcription factor. (C) Alternative 3'-end processing of *Sox21* resulting in a shorter 3'UTR and loss of predicted miRNA binding sites. (D) Impact of alternative exon expression on protein and domain interactions. Five genes of the 'Focal Adhesion' WikiPathway are shown. *FYN* and *FLT1* contain at least one alternative exon overlapping with a protein domain, while *SHC1* and *SHC3* do not have any regulated probesets annotated (yellow nodes indicate regulated probesets). Domain interactions (obtained from iPfam) for one protein isoform per gene are shown. Protein isoforms are depicted as rectangle nodes and constituent domains as diamond nodes. Protein interactions are drawn as blue edges and domain interactions as green edges.

In contrast, for protein interactions, the focus is on the underlying domain interactions of specific protein isoforms. Domain interactions are automatically derived from various data sources and interactions potentially disrupted by alternative splicing according to the AltAnalyze results can be readily identified (Figure 3D, domain interactions obtained from iPfam; 29). By double-clicking on any gene or protein in the network, DomainGraph automatically displays the 'Probeset view' for that gene.

### APPLICATION OF THE ALTANALYZE AND DOMAINGRAPH WORKFLOW

As an exemplary application of our workflow, we chose a previously described Exon Array dataset (GEO accession GSE13297) for human embryonic stem cells (hESCs) differentiated to cardiac precursors (21). In AltAnalyze, the Affymetrix CEL files were processed using default parameters, yielding 4660 alternative probesets for 2477 genes (Supplementary Data File 1). Over half of these probesets, 2353, aligned to a known alternative splicing event or alternative promoter, and the remainder occurred in either previously unannotated alternative exons or introns. The majority of alternatively expressed probesets, 3438 in total, were predicted to alter the inclusion of protein domains or protein sequence motifs, while 354 probesets overlapped with putative miRNA binding sites. A large subset of alternative probesets, 1647, was predicted to result in the absence of protein translation due to the introduction of a premature stop codon or missing

protein annotation. The WikiPathways 'Focal Adhesion' (WP306) and 'mRNA processing' (WP411) were the most statistically enriched pathways among all genes with alternative exons, based on the GO-Elite results in AltAnalyze (Supplementary Figure S1).

Importing these results (Supplementary Data File 2) into DomainGraph displays all alternatively expressed probesets from AltAnalyze in the 'Table view', along with alternative splicing annotations, affected miRNA binding sites, associated pathways and gene annotations (Figure 2). Selection of any gene symbol will build the 'Network view' for that gene together with the 'Probeset view', in which the alignment of Exon Array probesets to exons, transcripts, proteins, domains and miRNA binding sites is displayed.

For the exemplary dataset, the effect of alternative probeset expression is diverse in terms of the apparent mechanism of action and its functional impact. This includes alternative splicing, alternative promoter selection, alternative 3'-end processing, protein truncation, protein domain disruption and removal of predicted miRNA binding sites (Figure 3A–C). While AltAnalyze provides these predictions for each probeset, only DomainGraph visualizes the combination of alternative exon changes and the specificity of their effect at the transcript level. How these alternative exons affect larger biological processes can be assessed further by examining the interactions between these alternative genes in biological pathways or by examining domain interactions between these genes. For regulated genes associated with the 'Focal Adhesion' WikiPathway, the respective interactions

were imported into DomainGraph. DomainGraph automatically adds putative domain interactions and highlights potentially affected domain interactions (Figure 3D). This interaction network specifically demonstrates that alternative exon inclusion within the domain of both binding partners has the potential to significantly alter the corresponding pathway.

## CONCLUSIONS

Our new computational analysis workflow provides a convenient way of studying the biological impact of alternative splicing for mammalian genomes at large scale. It is more amenable even to biologists with limited knowledge and experience of bioinformatics and programming. The statistical data processing with AltAnalyze aims at extracting significant regulation events from large experimental datasets in order to support biologists in identifying potentially interesting alternative splicing events. The visual network analytics provided by DomainGraph can subsequently be used to investigate the potential biological implications of such splicing events and related mechanisms in the context of molecular interactions and binding sites. In the future, this analysis software may be extended to incorporate data by other high-throughput exon and splicing detection methods, for example, deep-sequencing techniques.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank Alexander R. Pico, Andreas Schlicker, Melissa S. Cline and Gary Howard for their valuable advice on the development of AltAnalyze and DomainGraph, and on the article.

## FUNDING

Boehringer Ingelheim Fonds (to D.E.); German National Genome Research Network (NGFN; 01GS0817 to D.E., M.A.); German Research Foundation (DFG; KFO 129/1-2 to T.L., M.A.); National Institutes of Health (NIH; GM080223, HG003053, HL66621 to N.S., B.R.C.). The work in Saarbrücken was conducted in the context of the DFG-funded Cluster of Excellence for Multimodal Computing and Interaction. Funding for open access charge: Max Planck Society.

*Conflict of interest statement.* None declared.

## REFERENCES

- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
- Clark, T.A., Schweitzer, A.C., Chen, T.X., Staples, M.K., Lu, G., Wang, H., Williams, A. and Blume, J.E. (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.*, **8**, R64.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D. *et al.* (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, **321**, 956–960.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A. *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods*, **6**, 377–382.
- Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A. and Soreq, H. (2005) Function of alternative splicing. *Gene*, **344**, 1–20.
- Resch, A., Xing, Y., Modrek, B., Gorlick, M., Riley, R. and Lee, C. (2004) Assessing the impact of alternative splicing on domain interactions in the human proteome. *J. Proteome Res.*, **3**, 76–83.
- McGlincy, N.J. and Smith, C.W. (2008) Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem. Sci.*, **33**, 385–393.
- Leeman, J.R. and Gilmore, T.D. (2008) Alternative splicing in the NF-kappaB signaling pathway. *Gene*, **423**, 97–107.
- Duursma, A.M., Kedde, M., Schrier, M., le Sage, C. and Agami, R. (2008) miR-148 targets human DNMT3b protein coding region. *RNA*, **14**, 872–877.
- Millevoi, S. and Vagner, S. (2010) Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res.*, **38**, 2757–2774.
- Mayr, C. and Bartel, D.P. (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, **138**, 673–684.
- Chang, T.Y., Li, Y.Y., Jen, C.H., Yang, T.P., Lin, C.H., Hsu, M.T. and Wang, H.W. (2008) easyExon—a Java-based GUI tool for processing and visualization of Affymetrix exon array data. *BMC Bioinformatics*, **9**, 432.
- Xing, Y., Stoilov, P., Kapur, K., Han, A., Jiang, H., Shen, S., Black, D.L. and Wong, W.H. (2008) MADS: a new and improved method for analysis of differential alternative splicing by exon-tiling microarrays. *RNA*, **14**, 1470–1479.
- Yates, T., Okoniewski, M.J. and Miller, C.J. (2008) X:Map: annotation and visualization of genome structure for Affymetrix exon array analysis. *Nucleic Acids Res.*, **36**, D780–D786.
- Purdom, E., Simpson, K.M., Robinson, M.D., Conboy, J.G., Lapuk, A.V. and Speed, T.P. (2008) FIRMA: a method for detection of alternative splicing from exon array data. *Bioinformatics*, **24**, 1707–1714.
- Okoniewski, M.J. and Miller, C.J. (2008) Comprehensive analysis of affymetrix exon arrays using BioConductor. *PLoS Comput. Biol.*, **4**, e6.
- Numata, K., Yoshida, R., Nagasaki, M., Saito, A., Imoto, S. and Miyano, S. (2008) ExonMiner: web service for analysis of GeneChip Exon array data. *BMC Bioinformatics*, **9**, 494.
- Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protocols*, **2**, 2366–2382.
- Hubbard, T.J., Aken, B.L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L. *et al.* (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**, D690–D697.
- Srinivasan, K., Shiue, L., Hayes, J.D., Centers, R., Fitzwater, S., Loewen, R., Edmondson, L.R., Bryant, J., Smith, M., Rommelfanger, C. *et al.* (2005) Detection and measurement of alternative splicing using splicing-sensitive microarrays. *Methods*, **37**, 345–359.
- Salomonis, N., Nelson, B., Vranizan, K., Pico, A.R., Hanspers, K., Kuchinsky, A., Ta, L., Mercola, M. and Conklin, B.R. (2009) Alternative splicing in the differentiation of human embryonic stem cells into cardiac precursors. *PLoS Comput. Biol.*, **5**, e1000553.
- Pico, A.R., Kelder, T., van Iersel, M.P., Hanspers, K., Conklin, B.R. and Evelo, C. (2008) WikiPathways: pathway editing for the people. *PLoS Biol.*, **6**, e184.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.

- et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
24. Krishnan,K., Salomonis,N. and Guo,S. (2008) Identification of Spt5 target genes in zebrafish development reveals its dual activity in vivo. *PLoS ONE*, **3**, e3621.
25. Emig,D., Cline,M.S., Klein,K., Kunert,A., Mutzel,P., Lengauer,T. and Albrecht,M. (2008) Integrative visual analysis of the effects of alternative splicing on protein domain interaction networks. *J. Int. Bioinformatics*, 101–115.
26. Emig,D., Cline,M.S., Lengauer,T. and Albrecht,M. (2008) Integrating expression data with domain interaction networks. *Bioinformatics*, **24**, 2546–2548.
27. Finn,R.D., Tate,J., Mistry,J., Coggill,P.C., Sammut,S.J., Hotz,H.R., Ceric,G., Forslund,K., Eddy,S.R., Sonnhammer,E.L. *et al.* (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**, D281–D288.
28. Matthews,L., Gopinath,G., Gillespie,M., Caudy,M., Croft,D., de Bono,B., Garapati,P., Hemish,J., Hermjakob,H., Jassal,B. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, **37**, D619–D622.
29. Finn,R.D., Marshall,M. and Bateman,A. (2005) iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, **21**, 410–412.