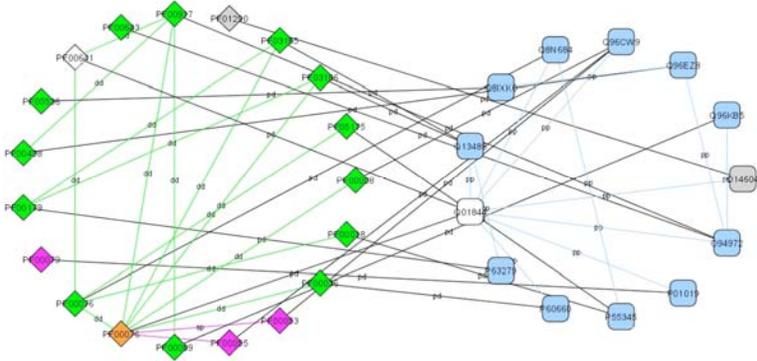


DomainGraph version 3.01 - Documentation

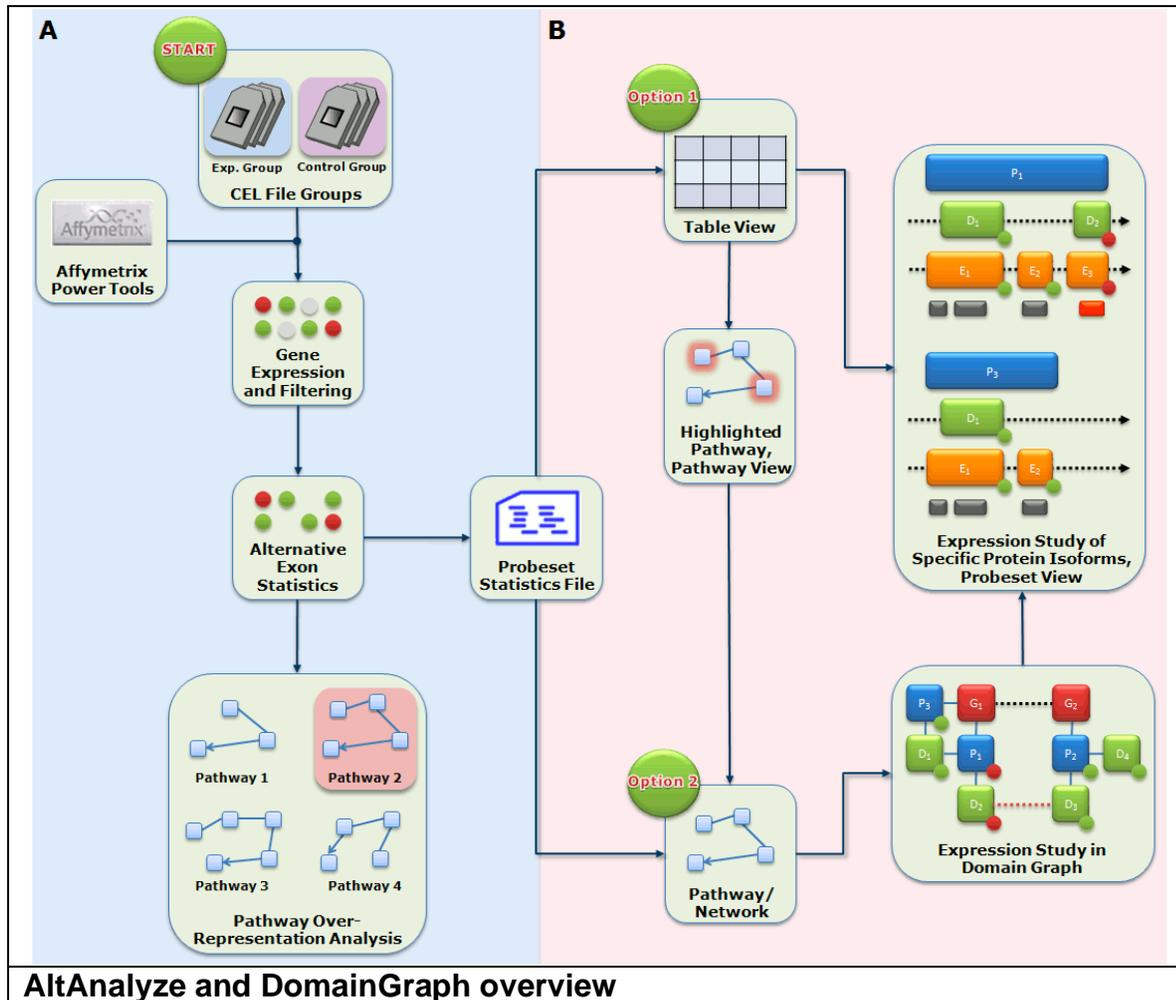


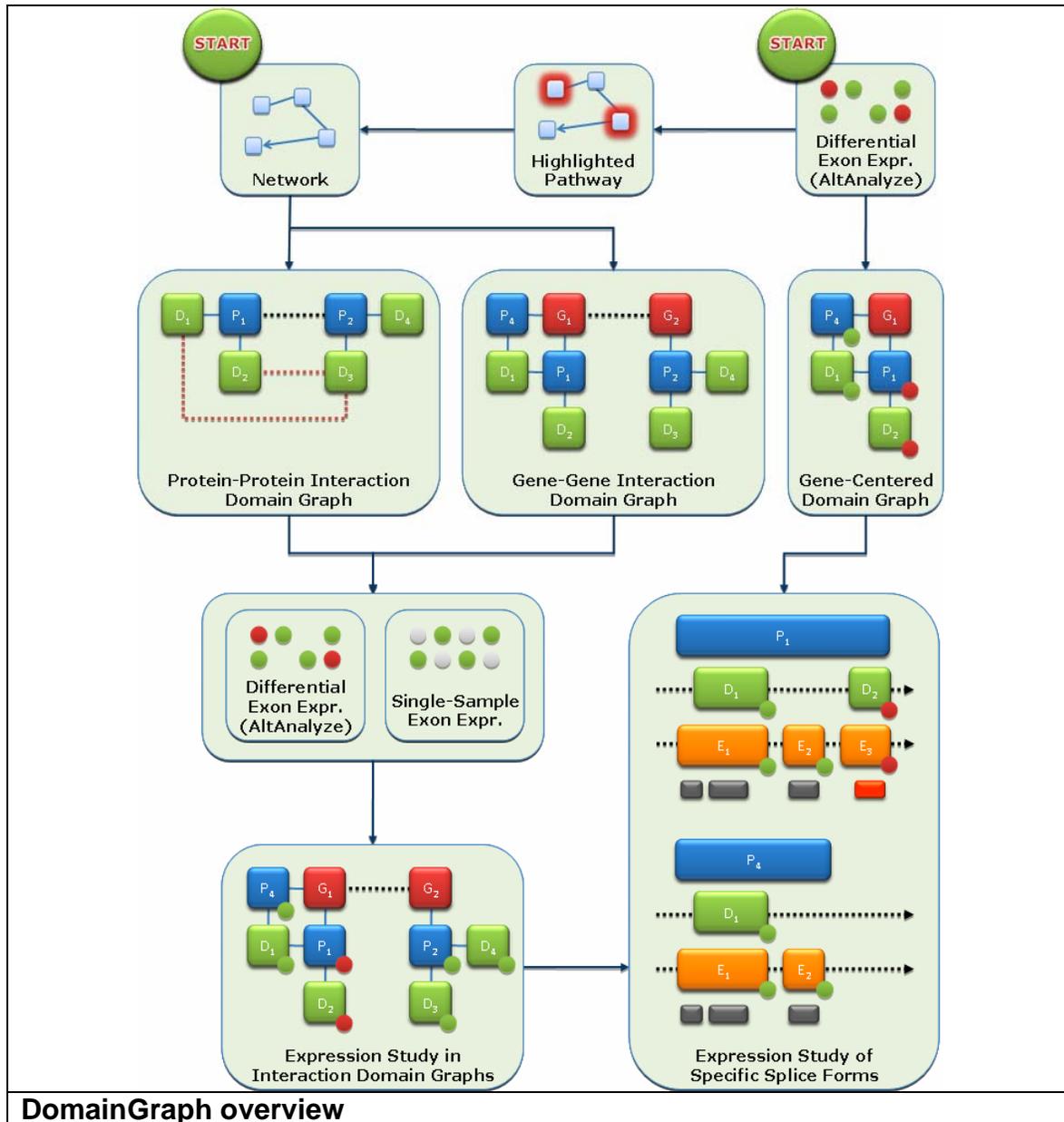
mp max planck institut
informatik

Table of Contents

DomainGraph Overview.....	3
Analyzing AltAnalyze Statistics.....	5
Analysis of Differential Exon Expression via AltAnalyze	5
Analyzing a particular network of interest	8
Interaction networks	8
Domain graph views	10
Integration of Affymetrix Exon Array Data	10
Visual Exploration of Expression Data.....	12
Preparing Affymetrix Exon Array data.....	15
How to obtain the appropriate file for Differential Expression Analysis.....	15
How to obtain the appropriate files for Single Experiments... ..	15
DomainGraph features	16
Visualization	16
Additional Information.....	19
Options	19
Analysis	19
Database Information.....	20
DomainGraph database versions	20
Embedded Database.....	21

DomainGraph Overview





The main features of DomainGraph include:

- the general analysis of Affymetrix Exon Array data to get an overview of genes, gene products, and pathways affected by alternative splicing.
- the analysis of a particular gene or protein interaction network of interest with regard to encoded protein isoforms, protein domain compositions, and domain-domain interactions.
- overlaying a particular network of interest with Affymetrix Exon Array data for the analysis of the effects of alternative splicing.

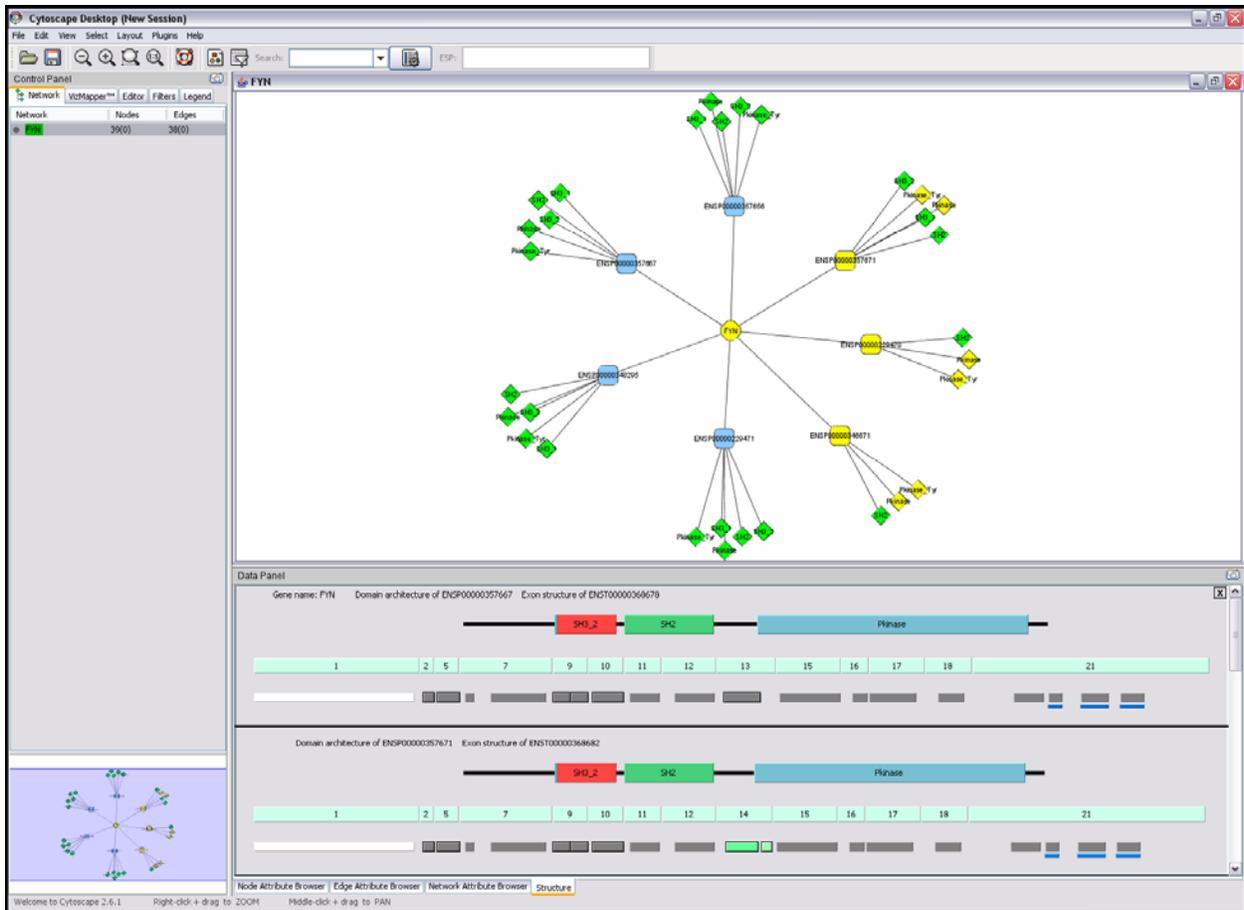
Analyzing AltAnalyze Statistics

Analysis of Differential Exon Expression via AltAnalyze

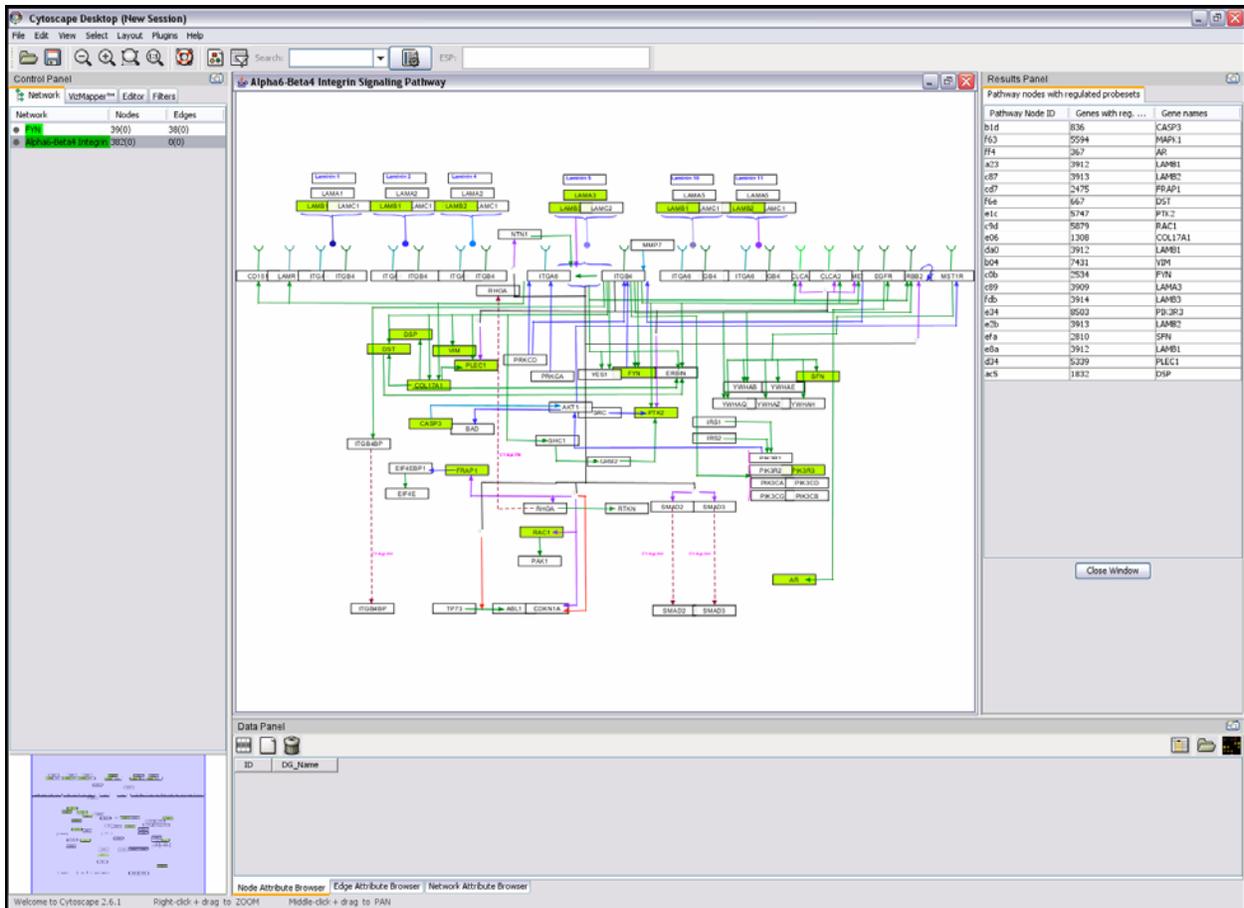
This option is intended for users who are not focused on a particular network but who would like to get a general overview of what is going on in the data. Therefore, the statistical results (containing the differentially expressed microarray probesets) computed via AltAnalyze can be loaded into DomainGraph. AltAnalyze statistically compares two groups of samples via the splicing index method to identify probesets that are differentially expressed in the two samples. These probeset data are automatically enriched with gene and pathway information as well as microRNA binding sites and exon information on known alternative splicing events. The results are provided in a table view.

Probeset	GeneID	SI	SI p-value	MIDAS p-value	Reactome Pathway(s)	Wikipathway(s)	mRNA Annotation	AS Events
2419236	FLBPI	-1.81724637681	0.0383884996656	0.039736	---	---	---	---
3130801	FUT10	-1.03066666667	0.0193513395033	0.019309	---	---	---	alt-5' bleedingExon
2969916	FYN	1.84888888889	0.00229723523524	0.010301	HIV	Alpha6-Beta4 Intein	---	cassette-exon cassetteExon
2969915	FYN	1.30888888889	0.00120981181658	0.014605	HIV	Alpha6-Beta4 Intein	---	cassette-exon cassetteExon
3012228	FZD1	1.24833333333	0.000558688137971	0.004022	---	Adipogenesis;Wnt Signaling	---	---
3012229	FZD1	1.03166666667	0.00234878213442	0.020344	---	Adipogenesis;Wnt Signaling	---	---
3012234	FZD1	-2.03166666667	0.000442437387692	0.008305	---	Adipogenesis;Wnt Signaling	---	---
3012237	FZD1	-2.91833333333	0.00808861571806	0.007533	---	Adipogenesis;Wnt Signaling	hsa-miR-611:mirbase,hsa...	---
3012236	FZD1	-2.95166666667	0.0101880844843	0.007642	---	Adipogenesis;Wnt Signaling	---	---
2378076	GOS2	-2.43666666667	0.00378508414526	0.026836	---	Diurnally regulated genes	hsa-miR-1:mirbase,hsa-m...	---
2745551	GAB1	1.34333333333	0.0405057662669	0.040326	Signaling by EGFR;	B Cell Receptor Signaling	---	---
2947946	GABBR1	1.35527777778	0.00851018660143	0.040624	---	GPCRs, Class C	---	---
2947934	GABBR1	1.08861111111	0.00965796856056	0.027475	---	GPCRs, Class C	---	bleedingExon
3585238	GABRAS	-1.31952380952	0.00572286302346	0.032988	---	---	---	---
3239730	GAD2	-1.005	0.010231093897	0.042696	---	Alanine and aspartate	---	intron-retention
3816523	GADD45B	1.56666666667	0.00570308168911	0.019086	---	Adipogenesis;	hsa-miR-886-5p:mirbase,...	exon-region-exclusion
3816513	GADD45B	1.21	0.0303127069603	0.03071	---	Adipogenesis;	---	alt-5' altPromoter bleeding...
3816522	GADD45B	1.06	0.00428825331703	0.02657	---	Adipogenesis;	hsa-miR-324-3p:mirbase,...	---
3957435	GAL3ST1	1.02555555556	0.0398794444337	0.037928	---	---	---	---
3957441	GAL3ST1	-1.39777777778	0.000133515798796	0.004752	---	---	---	---
3575170	GALC	1.02136363636	0.0014101382897	0.014744	---	---	---	---
2836520	GALNT10	1.38466666667	0.00228945846215	0.01068	---	---	---	---
2836581	GALNT10	1.328	0.0131121099257	0.03283	---	---	---	cassette-exon
2836521	GALNT10	1.12466666667	0.00102083233739	0.006426	---	---	---	---
2511112	GALNT13	-1.31181818182	0.00434685718828	0.027689	---	---	---	---
2751981	GALNT7	-1.55833333333	0.000847075419824	0.01182	---	---	---	alt-3'
2751979	GALNT7	-2.065	0.000569608585501	0.008443	---	---	---	alt-3' cassette-exon
3670728	GAN	-1.12044444444	0.0113638662261	0.031829	---	---	---	---
3561209	GARNL1	1.01833333333	0.00687232490897	0.025612	---	---	---	---
3561180	GARNL1	-2.375	0.00265181066632	0.01221	---	---	---	alt-C-term cassette-exon c...
3189747	GARNL3	1.38897435897	0.01415655482	0.013293	---	---	hsa-miR-16:mirbase,hsa-...	bleedingExon
3189780	GARNL3	-1.46102564103	0.0213340069698	0.022729	---	---	---	---
3189774	GARNL3	-1.73769230769	0.0145354287193	0.034039	---	---	---	---
3929763	GART	-1.24	0.020227942611	0.019095	Metablism of nucleotides;	One Carbon Metabolism;	---	alt-5' alt-N-term

The table provides an overview of all probesets that are differentially expressed in the two groups of samples. The gene, Reactome, and Wikipathway columns are clickable. Clicking on a gene, a network view is loaded containing the gene, all encoded protein isoforms and their domain compositions. The nodes are highlighted according to differentially expressed probesets. Additionally, the graphical representation of the proteins, domains, exons, probesets, and microRNA binding sites is displayed.



Clicking on the Reactome or Wikipathway column, the pathway, in which the gene is involved, is loaded. Furthermore, all genes/proteins are highlighted that contain differentially expressed probesets.

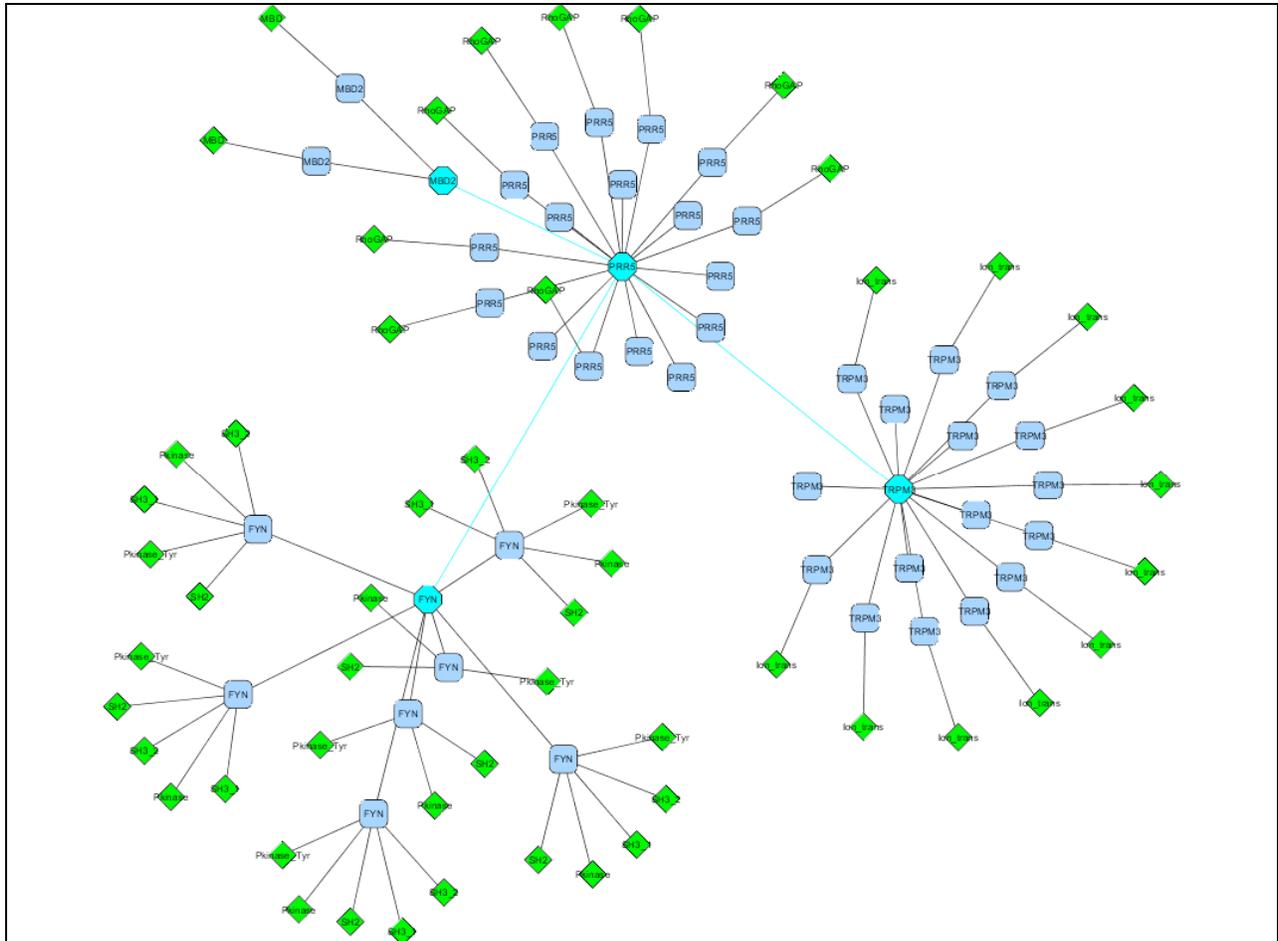


The pathway view is intended to provide an overview of the participating genes and proteins that might be affected by alternative splicing. For a more detailed analysis, the pathway can be loaded into DomainGraph (via the 'Analyzing a particular network of interest' option) and e.g. graphics/probeset annotations/microRNA annotations for all genes/proteins can be seen.

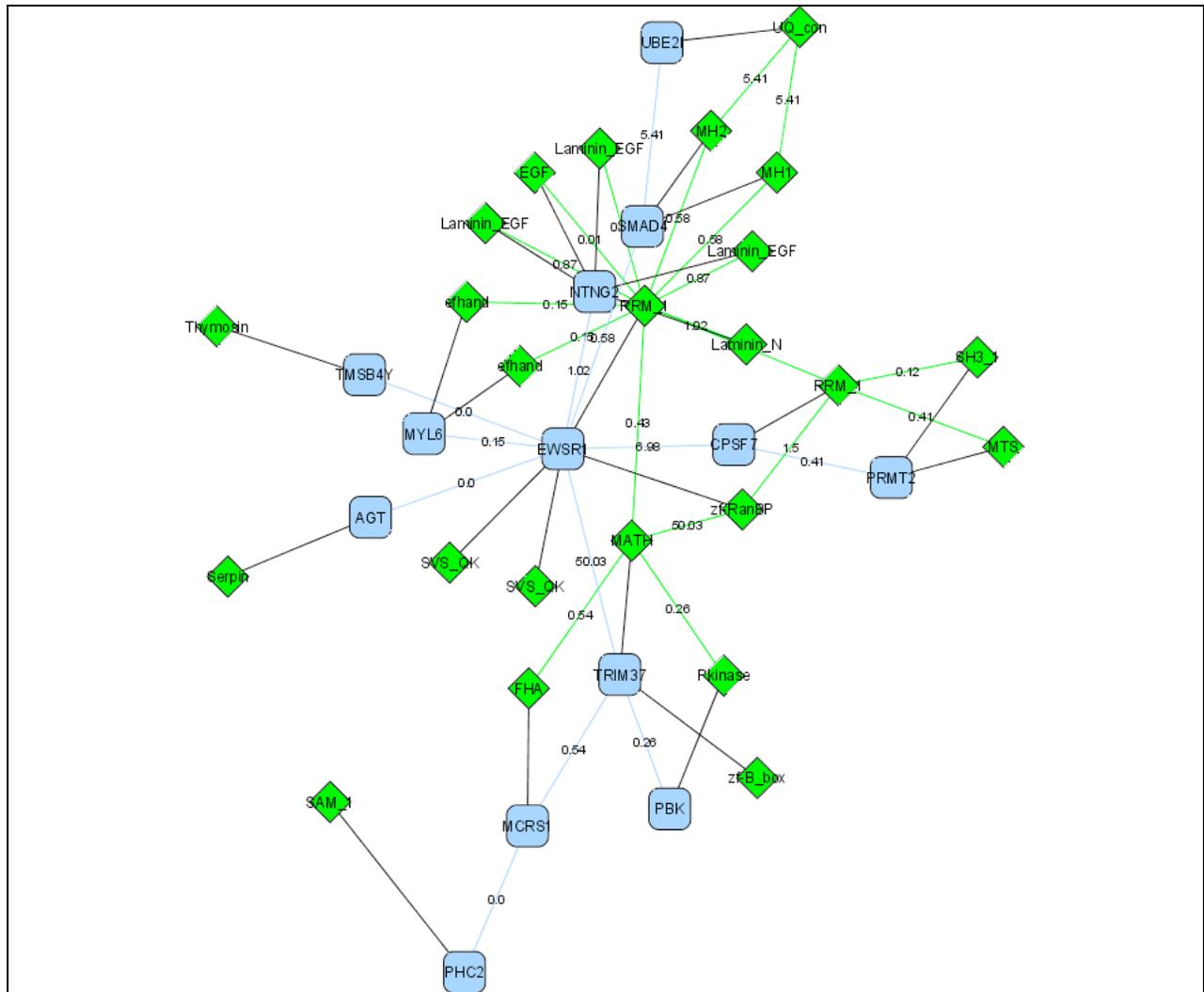
Analyzing a particular network of interest

Interaction networks

The user has to import a gene or protein interaction network of interest. DomainGraph supports the same file formats as Cytoscape for import, but users may also select a network imported via Cytoscape directly. For a gene interaction networks, the main focus is on the encoded protein isoforms and their domain compositions (see below).



Thus, the genes are visualized together with all encoded isoforms and their domains. For protein interaction networks, the focus is on the given protein isoforms and their underlying domain-domain interactions. The resulting network thus consists of the specific proteins, their domain compositions and the potential domain-domain interactions (see below):



Supported gene identifiers:

- Ensembl Gene ID (e.g. ENSG00000188681)
- Entrez Gene ID (e.g. 6310)

Supported protein identifiers:

- UniProt Accession Number (e.g. P29466)
- Ensembl Transcript ID (e.g. ENST00000260309)
- Ensembl Peptide ID (e.g. ENSP00000260309)

Supported species:

- Homo sapiens
- Saccharomyces cerevisiae
- Drosophila melanogaster
- Caenorhabditis elegans
- Mus musculus
- Rattus norvegicus
- Arabidopsis thaliana

- Escherichia coli
- Schizosaccharomyces pombe
- Plasmodium falciparum
- Danio rerio
- Gallus gallus

Domain-domain interaction datasets:

DomainGraph includes several structure-based and predicted datasets for inferring domain interactions. Please refer to [Database Information](#) for further information.

Domain graph views

For a domain graph, you can choose one of two different views for gene networks and three different views for protein networks (and you can switch between them at any time):

Extended view (Gene network):

Every gene is shown together with all its proteins and their constituent Pfam domains.

Gene network view:

Start with the gene network only, *no proteins or domains shown at the beginning*.

Extended view (Protein network):

Every protein is shown together with all its constituent Pfam domains.

Compact view (Protein network):

Each Pfam domain is shown only once, and all proteins containing the domain are linked to this domain node.

Protein network view:

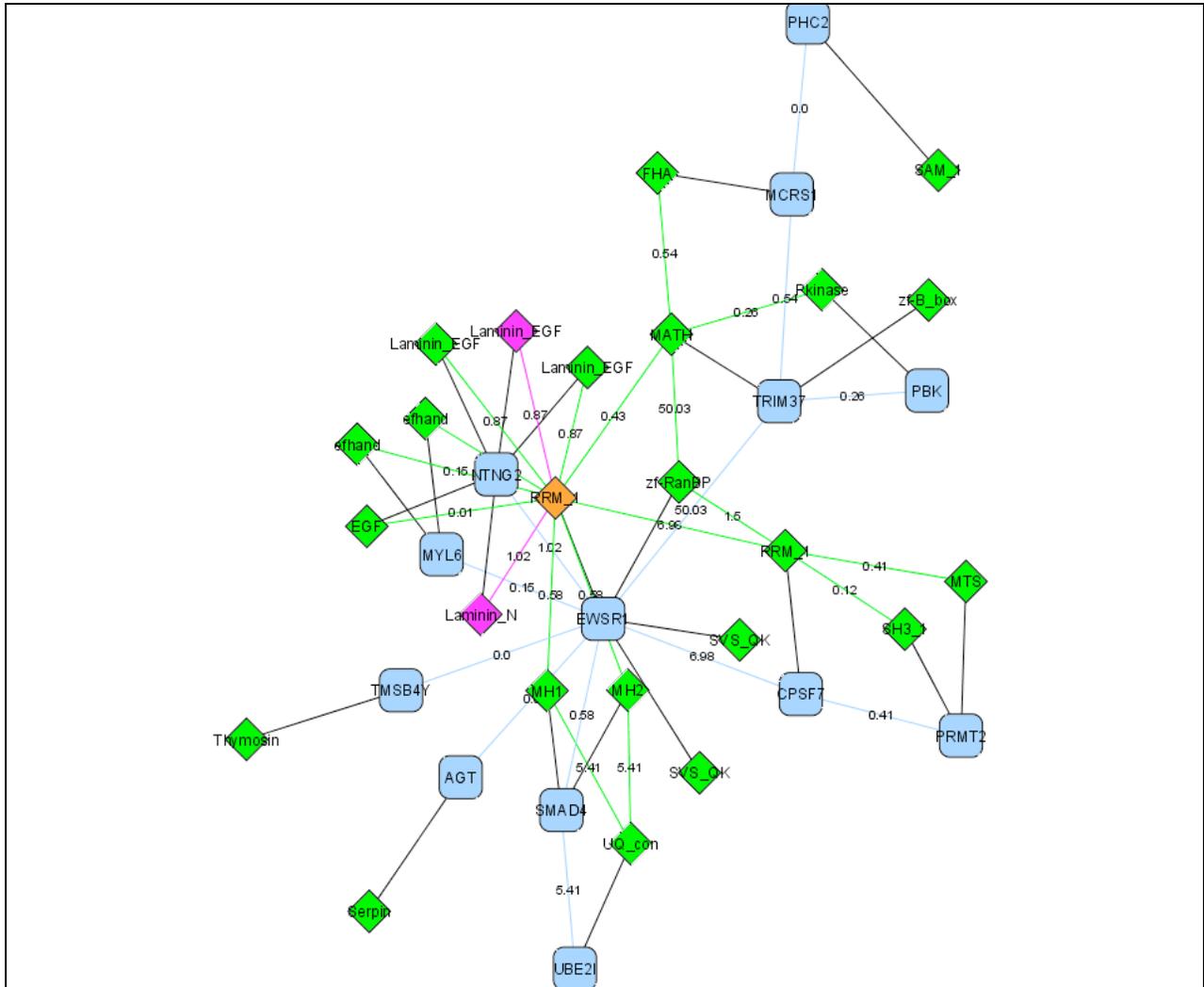
Start with the protein network *without showing any domains*. It is possible to individually select proteins, for which you want to see the domain composition and the domain interactions.

Integration of Affymetrix Exon Array Data

DomainGraph supports the analysis of 'Single Experiment' data as well as 'Differential Expression' among groups of samples. The Affymetrix Exon Expression data can be imported into the local DomainGraph database and the data is automatically mapped

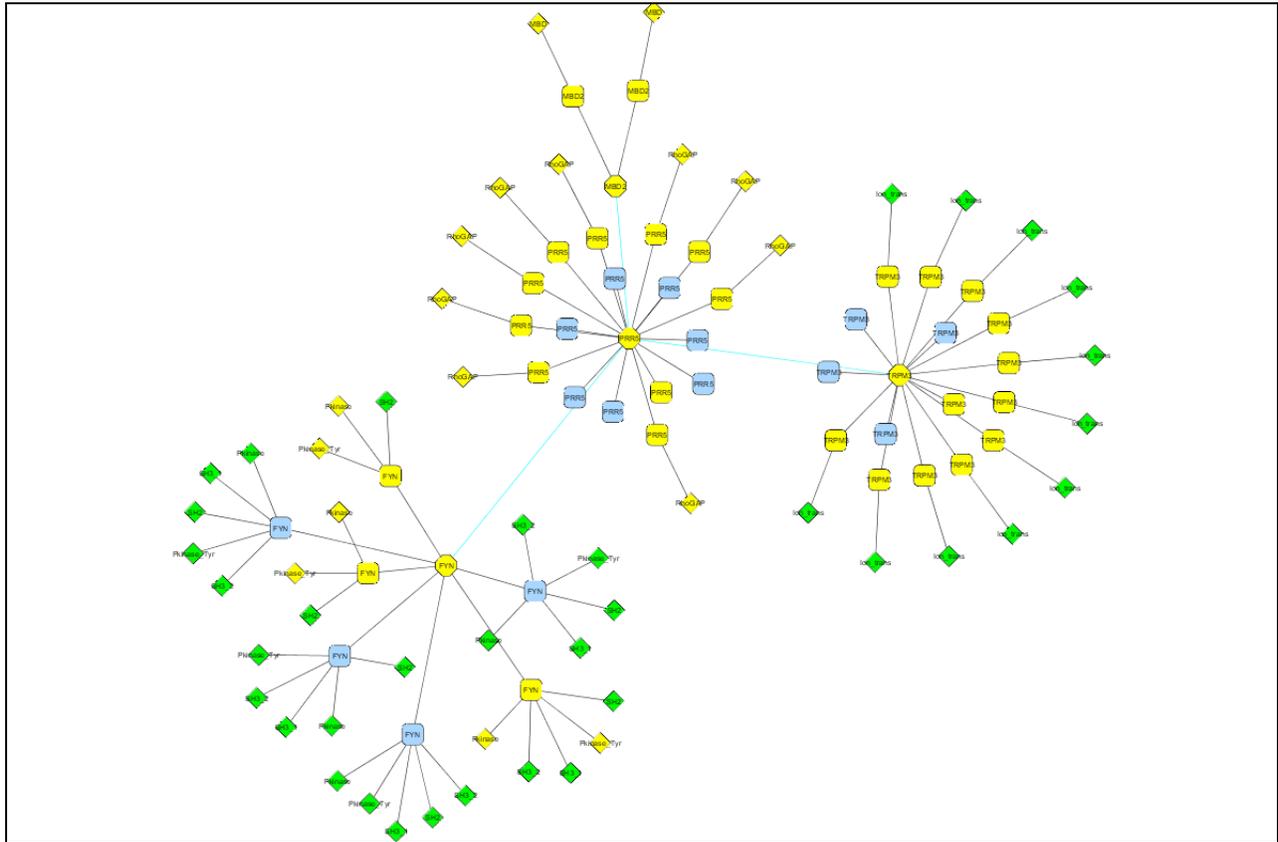
onto the network. Potential alternative splicing events and differential exon expression is highlighted in the network.

Single Experiment Data: This option can be used to analyze e.g. a single tissue with regard to potentially occurring alternative splicing events. Users simply import their expression and p-value files (see [Preparing Affymetrix Exon Array data](#)). These data are used to compute expression or non-expression of exons and domains. Potential alternative splicing events are highlighted in the network.



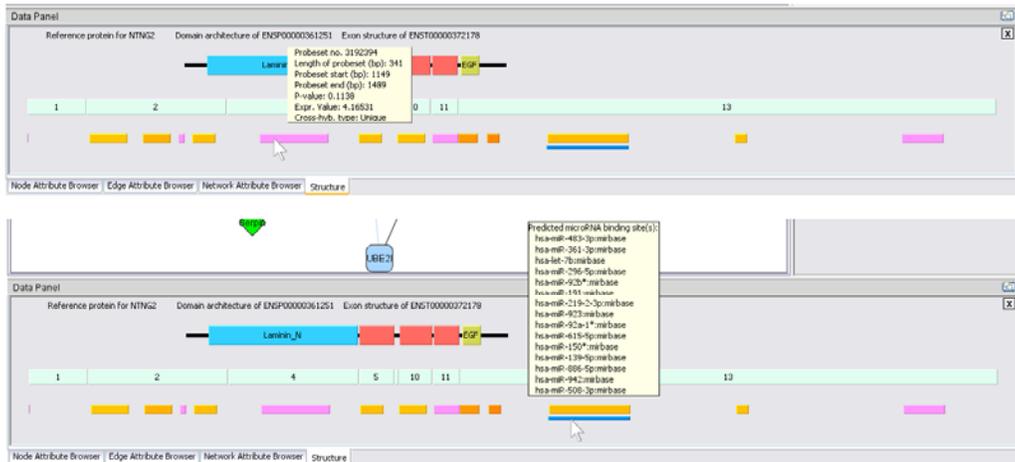
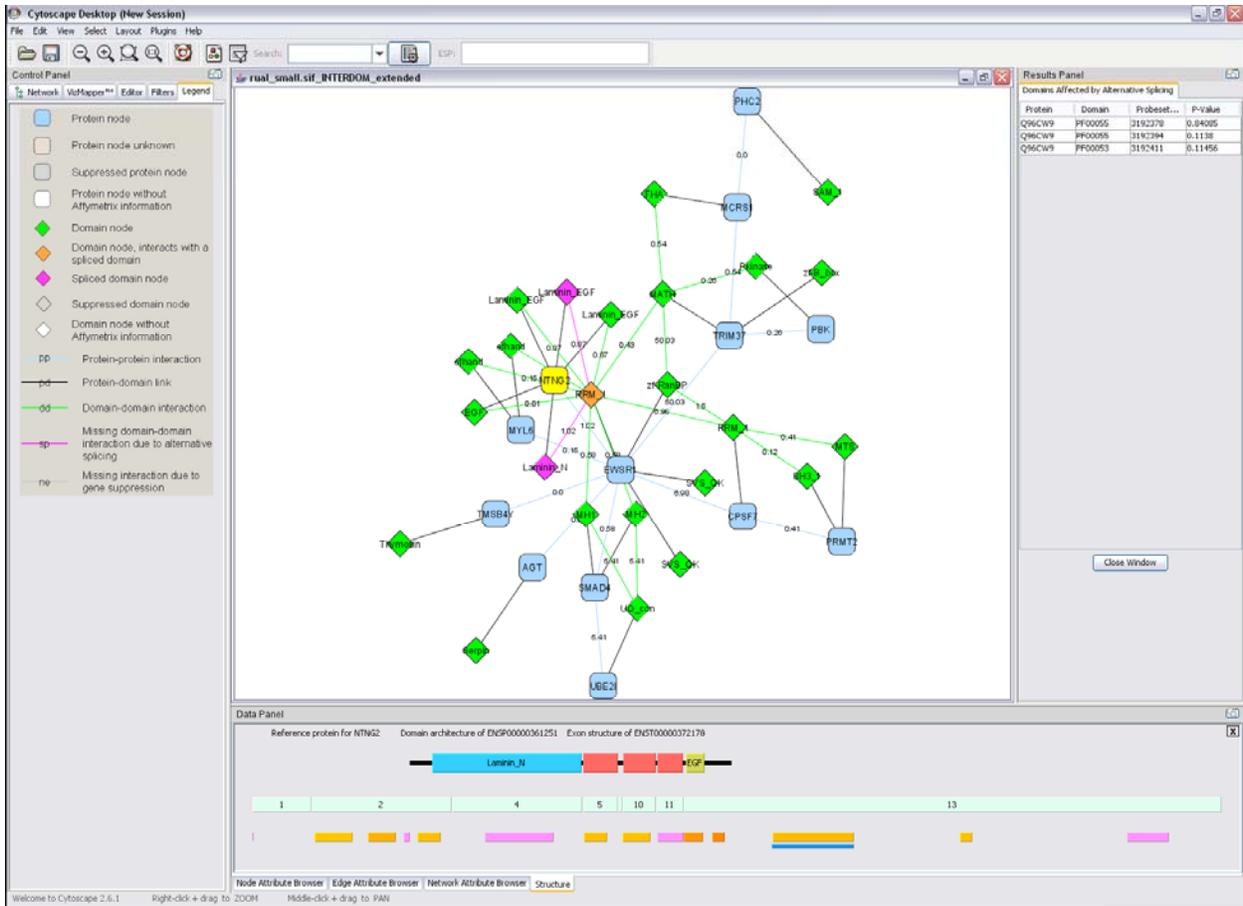
Differential Expression Data: This option is used to compare groups of samples to detect up- and downregulation events of certain exons. Protein isoforms favored in a certain sample can thus be detected easily. Users just need to import the statistics file produced by AltAnalyze, which is specifically designed for DomainGraph (see [Preparing Affymetrix Exon Array data](#)). The differential expression data are mapped onto the network. Proteins and domains potentially affected by alternative splicing are

highlighted.



Visual Exploration of Expression Data

DomainGraph provides a graphical representation of the proteins with domain architectures, the underlying exon structures, the Affymetrix Exon Array probesets, and microRNA binding sites. The graphics are available for proteins and for genes, allowing for the analysis of a single protein isoform as well as the comparison of all transcripts encoded by one gene. Tooltips provide additional information on the domains, exons, probesets, and microRNA binding sites. For 'Single Experiment' data, the probesets are colored according to their presence and absence, and expressed probesets are additionally colored according to the expression strength.



For 'Differential Expression' data, the probesets are colored according to whether they were analyzed or filtered out by AltAnalyze. Moreover, differentially expressed probesets are colored according to their up- or downregulation in the samples.

Preparing Affymetrix Exon Array data

How to obtain the appropriate file for Differential Expression Analysis...

... using AltAnalyze

AltAnalyze generates a file named *-DomainGraph.txt, which is specifically designed for further analysis using DomainGraph. This file can be used to analyze a particular network of interest. It can also be used for the general analysis of differential expression within the data.

- Download AltAnalyze at <http://www.altanalyze.org>.
- Unpack and start the program.
- Follow the steps for Exon Expression Analysis.
- You will receive a *-DomainGraph.txt file, which is stored in the AltAnalyze output folder (under the AltAnalyze program folder).

How to obtain the appropriate files for Single Experiments...

... using the Affymetrix Power Tools (APT)

- Download APT and the library files from Affymetrix.
- Install APT and start the APT Command Prompt.
- Type the following command (using correct pathnames) in the apt console (this example uses RMA for normalization):
`apt-probeset-summarize -a rma-sketch -a dabg -p HuEx-1_0-st-v2.r2.pgf -c HuEx-1_0-st-v2.r2.clf -b HuEx-1_0-st-v2.r2.antigenomic.bgp --qc-probesets HuEx-1_0-st-v2.r2.qcc -o OutputDirectory --cel-files YourCelfiles.txt`
- You will receive a DABG file (containing the probeset ids and corresponding p-values) and an expression file (containing probeset ids and corresponding expression values), which you can import into DomainGraph.

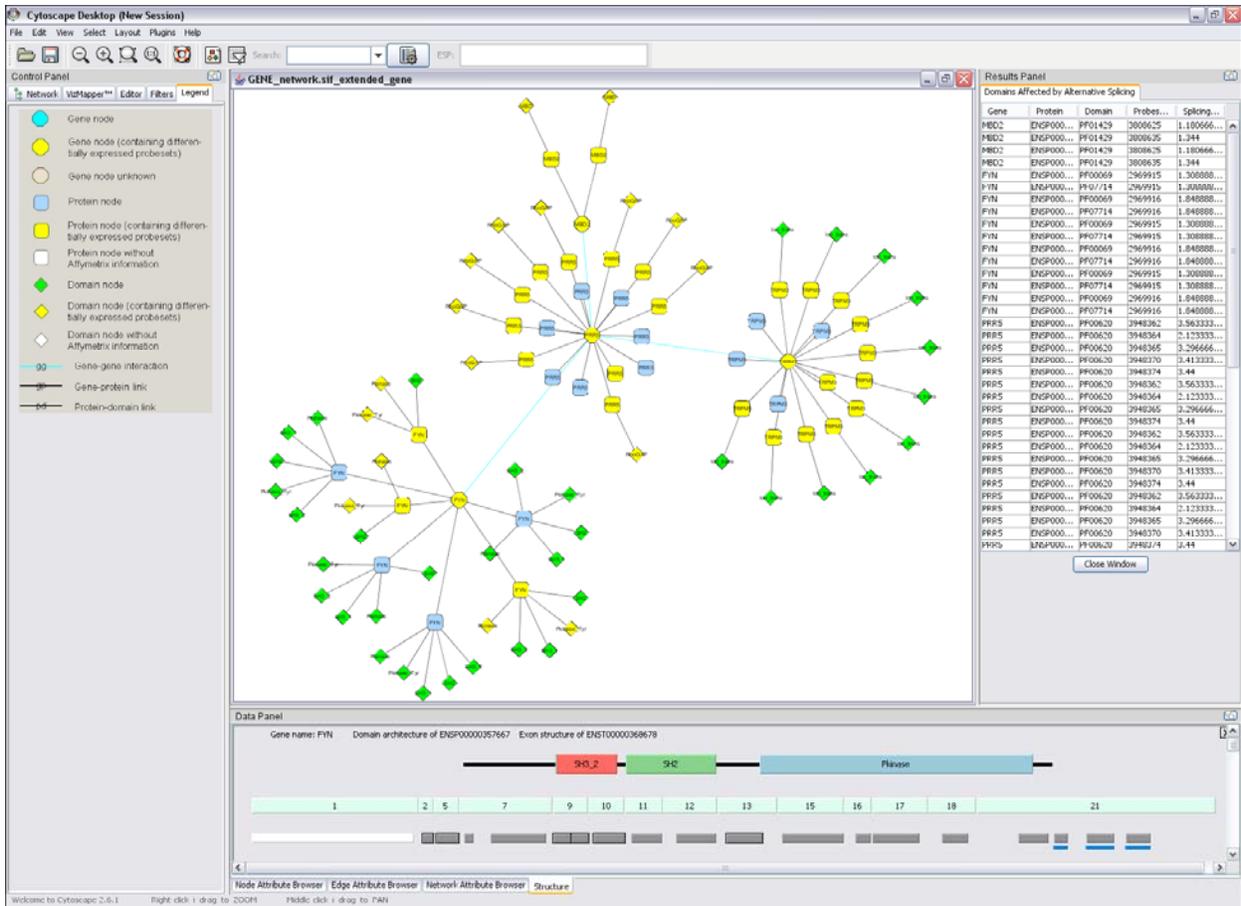
... using Affymetrix' Expression Console

- Download the Expression Console from Affymetrix.
- Install Expression Console and start the program.
- Set up a new study
- Press the 'Add Intensity Files' button to select the CEL files you want.
- Go to the Menu, select 'Analysis -> Exon Level -> Core RMA'.
- You will receive a DABG file (containing the probeset ids and corresponding p-values) and an expression file (containing probeset ids and corresponding expression values) in the folder where your CEL files are stored.

DomainGraph features

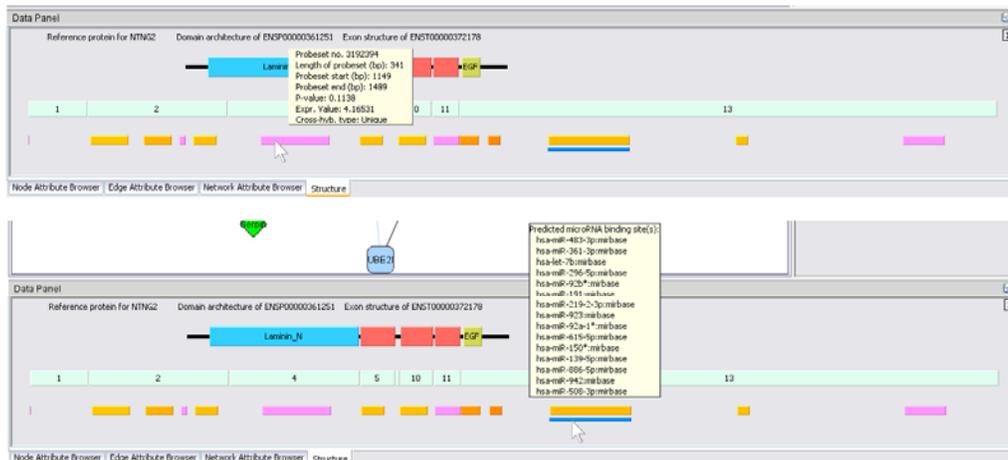
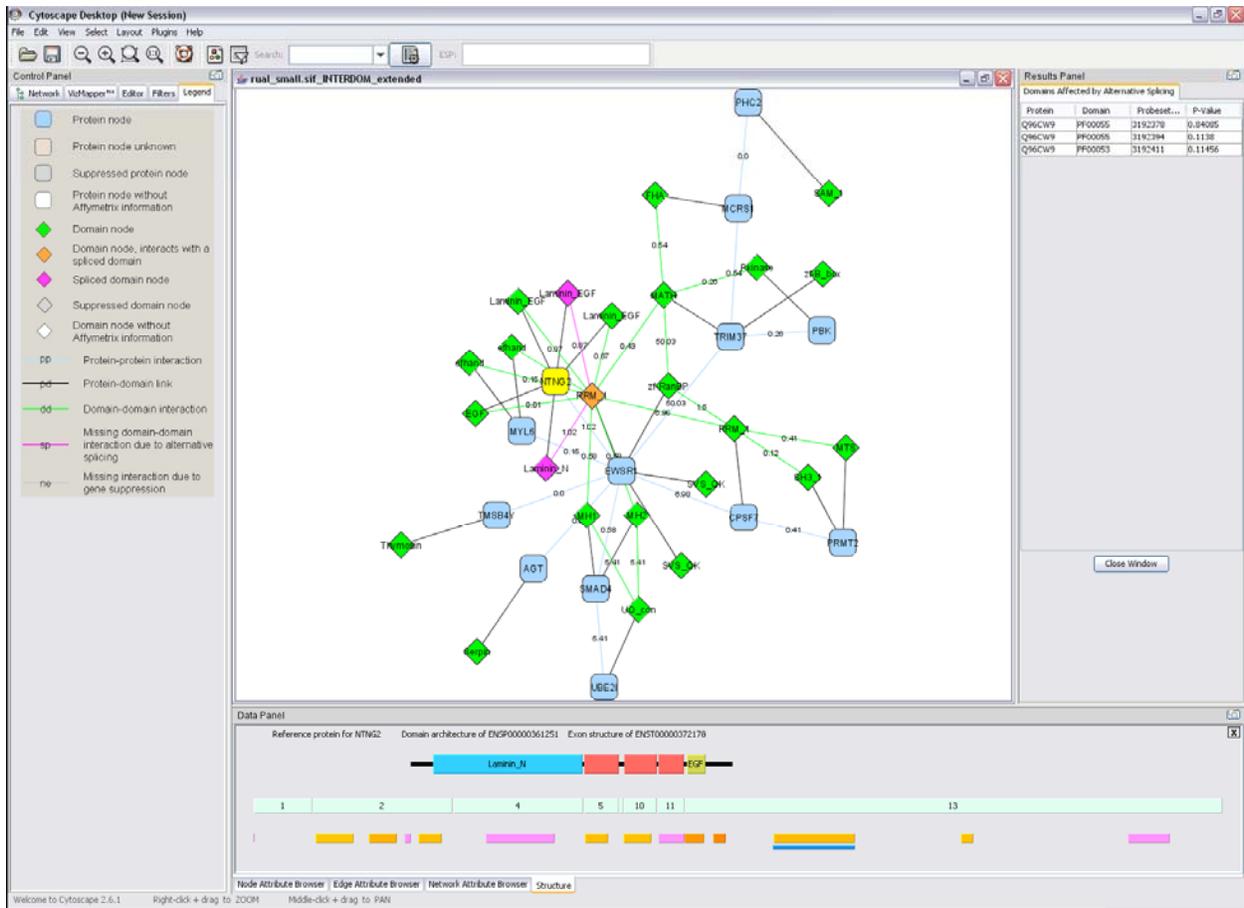
Visualization

A main feature of DomainGraph is the graphical representation, which can be obtained by double-clicking on a gene or protein node. This feature is especially useful for visualizing the Affymetrix probesets annotated to protein isoforms. Using the graphics, users are directly pointed to exons affected by alternative splicing, which protein isoforms are favored under a certain condition and if the alternative splicing event affects a domain-coding region or a microRNA binding site. **Tooltips** provide additional information, e.g. if the probeset is known to align to an alternative or constitutive exon or if the probeset is associated with an microRNA binding site. Furthermore, cross-hybridization information for the probesets is given, indicating whether the probeset is known to align to more than one genomic position (which would mean, the measured expression could be caused by a transcript to this other genomic location).



The first screenshot shows a toy network (gene interaction network) with integrated AltAnalyze data. The graphical representation of 'FYN' is shown in the Data Panel. The black line with the red, blue and green rectangles represents the protein and its contained domains. Below, the exon structure is shown including the 3' and 5' UTRs. Next, the Affymetrix probesets are displayed: white color indicates that the probeset was not statistically analyzed by AltAnalyze but filtered out before. Gray stands for

probesets that do **not** show differential expression. Probesets colored in red or green show differential expression in the two samples, with green indicating down-regulation in the experimental group and red indicating up-regulation in the experimental group.



The second screenshot shows a toy network (protein interaction network) with integrated Single Experiment data. As can be seen from the network, the NTNG2 protein contains 2 domains that are indicated to be spliced out (colored in pink). A closer look at the graphical representation makes this even clearer: expressed

probesets are colored in yellow to red (color gradient according to expression strength) while not expressed probesets are colored in pink. As can be seen from the graphics, most of the Laminin-N domain is not expressed as is the 3rd Laminin-EGF domain (the 3rd domain colored in red). The tooltips provide additional information, e.g. on the probeset p-values, on cross-hybridization events, and on the microRNA binding sites.

Additional Information

DomainGraph provides some basic information on the proteins, which consists of the common names, contained Pfam domains, GO annotations, and OMIM annotations (for human proteins). They can be used to obtain a quick overview of the function of a specific protein. The information is available via right-click on a protein node.

Options

There are several options available to customize the visual appearance of a domain graph, such as showing or hiding node and edge labels, or adjusting the edge width to the confidence scores. Furthermore, domain interactions can be filtered according to their confidence scores and low-confidence edges can be removed from the network. Additionally, the stepwise exploration of protein networks and their underlying domain interactions is coordinated in the options.

Analysis

The analysis methods are mainly intended for domain graphs with integrated 'Single Experiment' data. Users can choose to compute the intersection, union, and difference of two graphs. These graph methods do not only take identical and different nodes and edges into account but also the biological types (i.e. an expressed domain or a spliced domain, for example) and highlight nodes and edges that are contained in both networks but with different biological types.

Database Information

DomainGraph database versions

[ENS_52](#)

Based on Ensembl 52, UniProt version 15.0, and Pfam version 23.
Affymetrix probeset data were taken from the NetAffx release of 03/09.
Reactome data were downloaded 04/09/09.
WikiPathway data were downloaded 06/17/09.
Omim data were downloaded 04/03/09.
Mappings between Ensembl, Entrez, and UniProt were downloaded via BioMart (using Ensembl 52).
DDI dataset version 1.
Structural domain-domain interaction datasets: [iPfam](#) version 21; [3did](#) version of November 2007
Predicted domain-domain interaction datasets: [InterDom](#) version 2.0; [LLZ](#); [DPEA](#); [DIMA](#) version 2.0, consisting of 3 subsets (domain profiling, DPEA, and string); [LDSC](#) consisting of 2 sets: all interactions and a core subset; [RDFF](#); [RCDP50](#); [APMM](#) consisting of 2 sets: APMM1 and APMM2; [LP](#); [PINS](#); [IPPRI](#)

[ENS_54](#)

Based on Ensembl 54, UniProt version 15.10, and Pfam version 24.
Affymetrix probeset data were taken from the NetAffx release of 07/09.
Reactome data were downloaded 11/13/09.
WikiPathway data were downloaded 11/13/09.
Omim data were downloaded 11/12/09.
Mappings between Ensembl, Entrez, and UniProt were downloaded via BioMart (using Ensembl 54).
DDI dataset version 2 (same as 1, but includes UniDomInt).
Structural domain-domain interaction datasets: [iPfam](#) version 21; [3did](#) version of November 2007
Predicted domain-domain interaction datasets: [InterDom](#) version 2.0; [LLZ](#); [DPEA](#); [DIMA](#) version 2.0, consisting of 3 subsets (domain profiling, DPEA, and string); [LDSC](#) consisting of 2 sets: all interactions and a core subset; [RDFF](#); [RCDP50](#); [APMM](#) consisting of 2 sets: APMM1 and APMM2; [LP](#); [PINS](#); [IPPRI](#); [UniDomInt](#)

Please note: the Affymetrix probesets are mapped to Ensembl transcripts according to their genomic coordinates as given by the respective NetAffx release.

Embedded Database

The DomainGraph plugin uses an embedded database by [Apache Derby](#). The database will be installed locally in the user's Cytoscape directory upon first usage. Exon Array data will also be stored in this local database to remain confidential data locally.