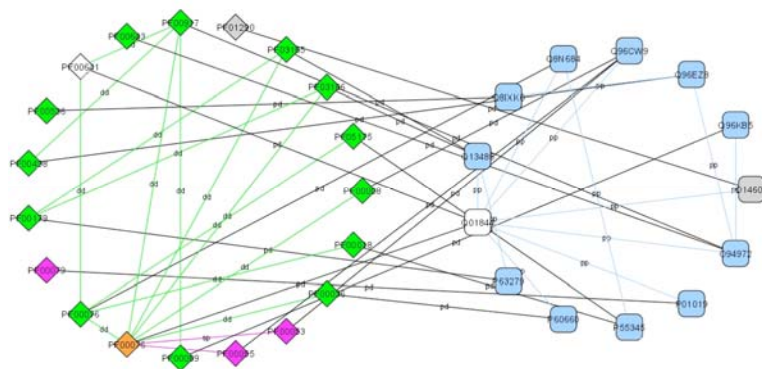


DomainGraph version 3.01 - Tutorials



max planck institut
informatik

Table of Contents

Workflow 1: Analyzing the statistical results of AltAnalyze.....	3
Installing the DomainGraph	3
Installing the DomainGraph database (WITH internet connection)	3
Installing the DomainGraph database (on a machine WITHOUT internet connection)	3
General analysis of differential expression computed by AltAnalyze.....	3
Analysis of a single gene	4
Analysis of a WikiPathway pathway.....	5
Transforming a WikiPathway pathway into a gene interaction Domain Graph	6
Analysis of a Reactome pathway.....	7
Transforming a Reactome pathway into a protein interaction Domain Graph	8
Workflow 2: Analyzing a protein interaction network.....	10
Installing DomainGraph	10
Installing the DomainGraph database (WITH internet connection)	10
Installing the DomainGraph database (on a machine WITHOUT internet connection)	10
Importing a protein interaction network into DomainGraph	10
Integrating Affymetrix Exon Array data for Single Experiments.....	11
Viewing proteins, domains, exons, probesets, and miRNA binding sites as graphics	12
Workflow 3: Analyzing a gene interaction network.....	14
Installing DomainGraph	14
Installing the DomainGraph database (WITH internet connection)	14
Installing the DomainGraph database (on a machine WITHOUT internet connection)	14
Importing a gene interaction network into DomainGraph	14
Integrating Affymetrix Exon Array data produced by AltAnalyze for analyzing differential exon expression	15
Viewing proteins, domains, exons, probesets, and miRNA binding sites as graphics	16

Workflow 1: Analyzing the statistical results of AltAnalyze

Installing the DomainGraph

- Download and install Cytoscape 2.6
 - Go to the plugins manager in the Cytoscape menu, 'Network inference' and select DomainGraph for install.
 - Start DomainGraph via the plugins menu.
 - Upon first usage you need to register and accept the DomainGraph license.
-

Installing the DomainGraph database (WITH internet connection)

- Go to the plugins menu, select DomainGraph -> Manage DomainGraph database -> Import data for selected species into database.
 - Choose a local folder to install the database. Make sure to have write permission for this folder.
 - Select 'Homo sapiens'.
 - Select database version ENS_52.
 - Click on the import button.
 - Wait until a window pops up telling you the database is complete.
-

Installing the DomainGraph database (on a machine WITHOUT internet connection)

- You first have to install the database on a computer that is connected to the internet, as described above.
 - Once you have installed the database, you can copy it to any other machine as follows:
 - Copy the database directory (the folder that you chose for the database) to the non-connected machine.
 - Example: You installed the database to "C:\Cytoscape\Database\" on the internet computer.
 - You copy the "Database" folder anywhere to the non-internet computer.
 - Go to the plugins menu, select DomainGraph -> Manage DomainGraph database -> Change path to database.
 - Choose the "Database" directory as new path.
 - The database can now be used on your non-internet computer.
-

General analysis of differential expression computed by AltAnalyze

This step requires a probeset statistics file produced by AltAnalyze (the appropriate file can be found in the results folder of AltAnalyze, called "YOUR_DATASET_NAME-DomainGraph.txt"). You can download a sample results file via right-click this link: http://domaingraph.bioinf.mpg.de/sampleddata/AltAnalyze-results.Hs_Exon_CP_vs_hESC-splicing-index-DomainGraph.txt. This sample file compares human embryonic stem cells (control group) to cardiac progenitor cells (experimental group). Save the file to your disk.

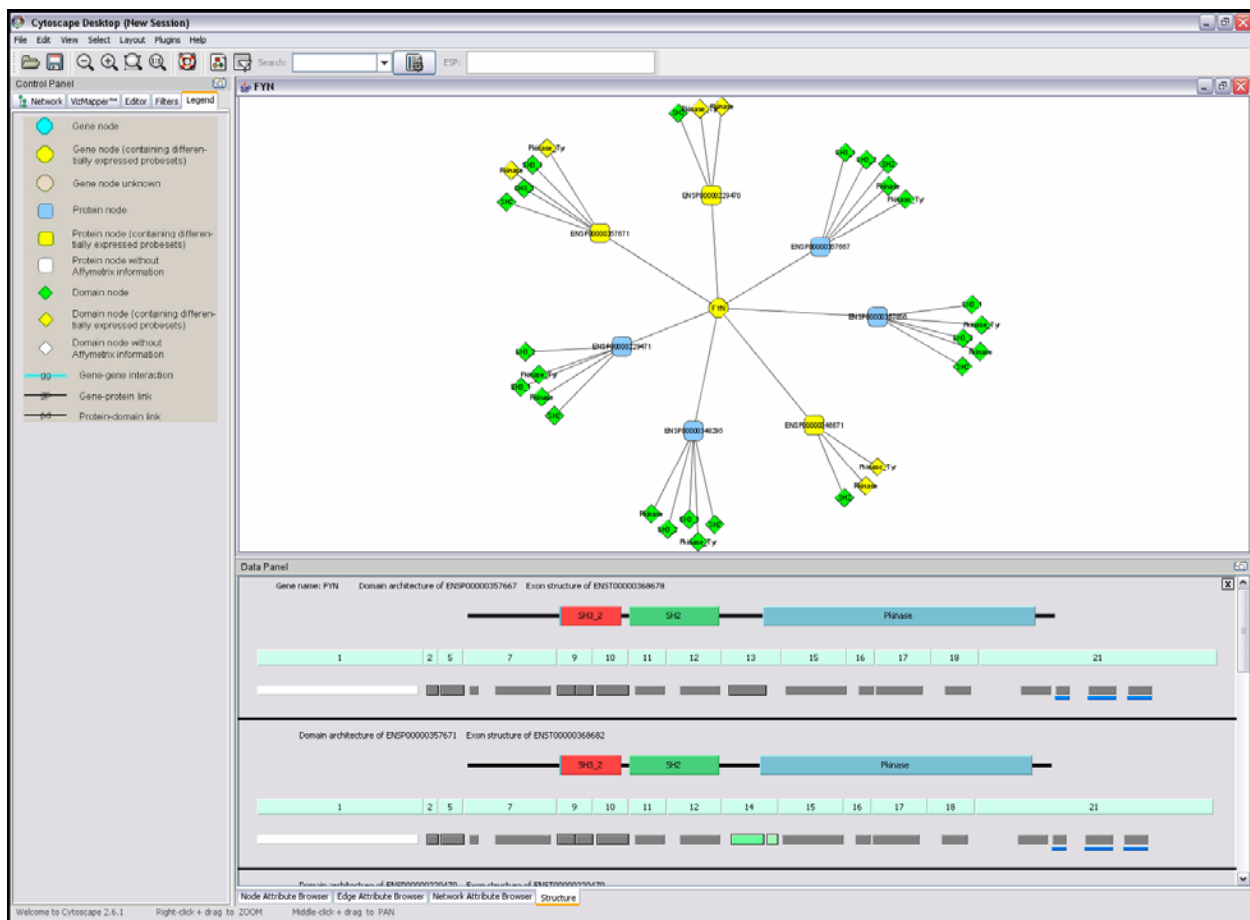
- Go to the plugins menu, select DomainGraph -> Start DomainGraph.
- Select 'Import AltAnalyze results for further analysis'.
- Click the Start button.
- Select the AltAnalyze results file and click the Open button.
- Wait for the file to be processed.
- The resulting table should look similar to this **screenshot**:

Probeset	GeneID	SI	SI p-value	MIDAS p-value	Reactome Path...	WikiPathway(s)	miRNA Annotation	AS Events
3808623	MBD2	4.67733333333	0.00586491796138	0.005585	---	---	---	alt-C-term
3209235	TRPM3	3.72555555556	0.000342560948685	0.008562	---	---	---	alt-N-term
3521084	SOX21	3.71555555556	0.00582713571632	0.005231	---	---	---	---
3948362	ARHGAP8;PRRS	3.56333333333	0.0048922774077	0.004428	Signaling by Rho	---	---	---
3030197	CNTNAP2	3.53314814815	0.0246858091982	0.023973	---	---	---	---
3030198	CNTNAP2	3.49981481481	0.000862191266274	0.005324	---	---	---	---
3948374	ARHGAP8;PRRS	3.44	0.00200184707636	0.009032	Signaling by Rho	---	---	---
3948370	ARHGAP8;PRRS	3.41333333333	0.00497540749503	0.01561	Signaling by Rho	---	---	cassette-exon
3808622	MBD2	3.32066666667	0.000403883771849	0.003247	---	---	---	alt-C-term
3948365	ARHGAP8;PRRS	3.29666666667	0.00014410216395	0.004866	Signaling by Rho	---	---	---
3253439	RPS24	3.25	7.5851716971e-005	0.00353	3'-UTR-mediated	Cytoplasmic	---	---
3948345	ARHGAP8;PRRS	3.14	0.00243874868736	0.002334	Signaling by Rho	---	---	---
3397633	ETS1	3.13166666667	0.000211570253159	0.005479	---	IL-2 Signaling	---	alt-N-term
3948338	ARHGAP8;PRRS	3.11	0.00040831736918	0.003164	Signaling by Rho	---	---	cassette-exon ca...
3030199	CNTNAP2	3.10314814815	0.0104111481058	0.010632	---	---	hsa-miR-512-5p:m...	---
3217131	TRIM14	3.08666666667	0.0163664063161	0.01479	---	---	---	bleedingExon
3030190	CNTNAP2	3.06981481481	0.000264888060628	0.006837	---	---	---	---
3016024	MUC3A;MUC3B	3.01666666667	0.00826488356767	0.007059	---	---	---	---
3267416	INPP5F	3.00133333333	0.000836020045189	0.011365	---	---	---	---
2743124	LARP2	2.9475	0.0270193122931	0.022077	---	---	---	cassette-exon
3521085	SOX21	2.88888888889	0.00429542867754	0.014182	---	---	hsa-miR-663:mirb...	---
3016026	MUC3A;MUC3B	2.85333333333	0.000592161416464	0.010226	---	---	hsa-miR-483-5p:m...	---
3030200	CNTNAP2	2.84314814815	0.000136834455788	0.004868	---	---	hsa-miR-204:Targ...	---
3948348	ARHGAP8;PRRS	2.81	0.00315395021369	0.002886	Signaling by Rho	---	---	---
3262461	SLK	2.75857142857	0.000103395842964	0.004067	---	---	---	cassette-exon ca...
3948322	ARHGAP8;PRRS	2.75333333333	3.8639064416e-005	0.000664	Signaling by Rho	---	---	---
3030202	CNTNAP2	2.75314814815	0.00459511805907	0.00467	---	---	hsa-miR-27b:picta...	---
4002108	MAP7D2	2.72	0.00384504220909	0.010793	---	---	---	cassette-exon
2547475	NLRC4	2.72	0.00201849874868	0.004446	---	---	---	---
4019500	SEPT6	2.68555555556	0.00250017857695	0.010389	---	---	hsa-miR-491-3p:m...	cassette-exon ca...
3233073	AIKRIC3	2.67333333333	0.0165673850661	0.018017	---	---	---	---
3881904	ASXL1	2.66492063492	0.000277846862745	0.002521	---	---	---	exon-region-exclu...
4000719	AP1S2	2.66	0.000174966260899	0.005621	HIV	---	---	alt-C-term
3922490	ABCG1	2.65606060606	0.00590030629184	0.006202	Metabolism of lipids	Nuclear receptors in	---	cassette-exon

- The table shows the differentially expressed probesets detected by AltAnalyze. The table is enriched with information on the genes containing these probesets, the Splicing Indices (SI), the SI p-values, the MiDAS p-values, the Reactome and WikiPathway pathway(s) the genes participate in, miRNA binding sites associated with the probesets and alternative splicing (AS) annotations..

Analysis of a single gene

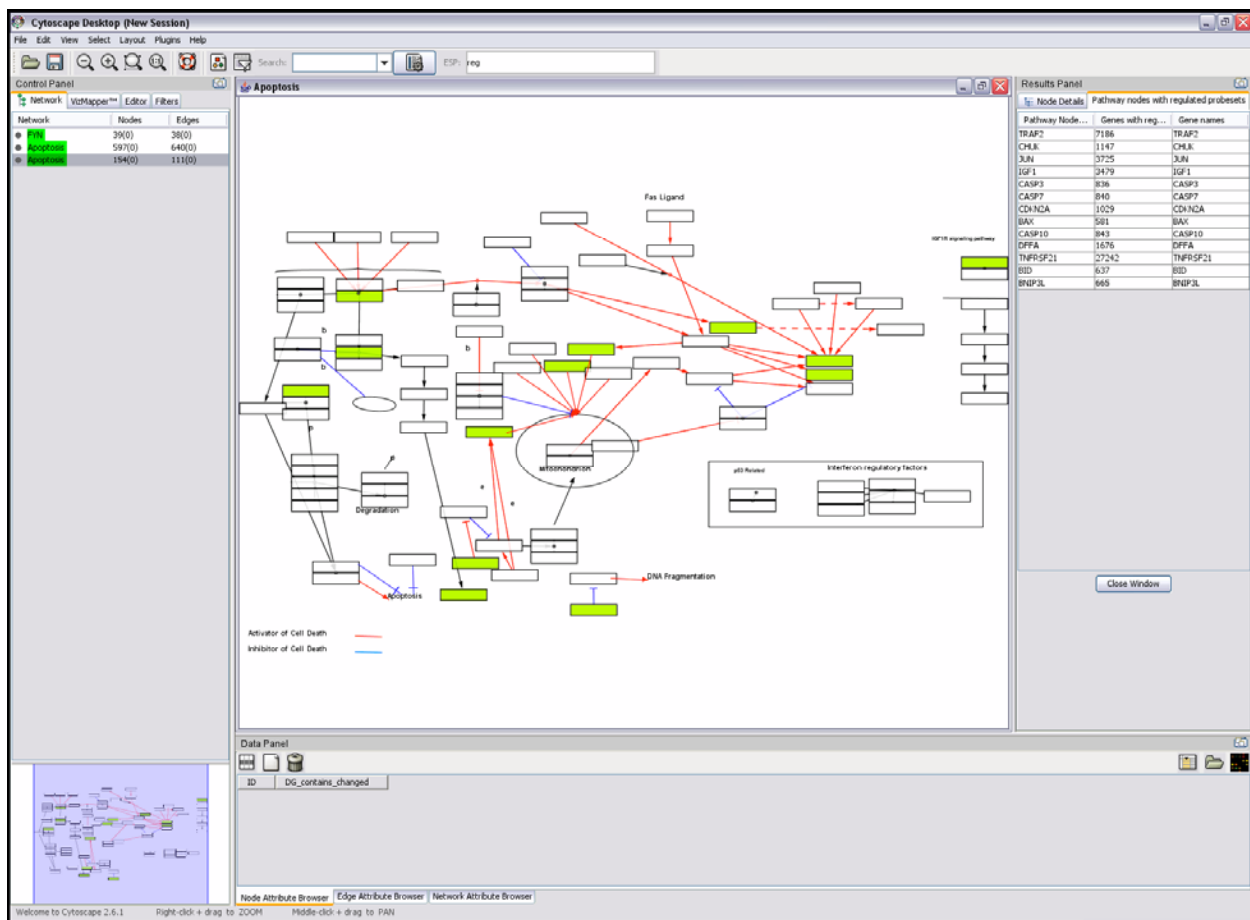
- Sort the table according to gene names by clicking on the header of the 'GeneID' column.
- Scroll to the entry 'FYN' in the gene column.
- Click on 'FYN'.
- Wait for the graphics to be created.
- You will receive a 'single gene'- domain graph. This network consists of the gene, all encoded proteins and their domain compositions. Nodes highlighted in yellow indicate the presence of differentially expressed probesets.
- The data panel shows the graphical representation of the protein domain architecture, the exon structure, the probeset and miRNA annotations.
- The results should look similar to this **screenshot**:



- Please note exon no 14 in the graphics: it shows significant down-regulation in the experimental group (cardiac progenitors) compared to the control group (human embryonic stem cells) as can be seen by the annotated probesets colored in green. This result indicates that transcripts containing exon no 14 are not favored in cardiac progenitor cells.
- Move the mouse over a domain, exon, probeset, miRNA block in the graphics. The tooltips show some additional information on the domain.
- Move the mouse over the probesets below exon no 14. The tooltip shows e.g. the Splicing Index (dl), alternative splicing annotations and cross-hybridization information.

Analysis of a WikiPathway pathway

- Sort the table according to wikiPathway pathways by clicking on the 'WikiPathway(s)' header.
- Scroll to an entry containing 'Apoptosis' in the Wikipathway column.
- Click on 'Apoptosis'.
- A new window opens displaying all pathways associated with the probeset's gene. Click on 'Apoptosis' in this window and then on the 'Submit' button.
- Wait for the WikiPathway pathway to be loaded. Please note that you need to have an internet connection to load pathways!
- The resulting pathway should look similar to this **screenshot**:

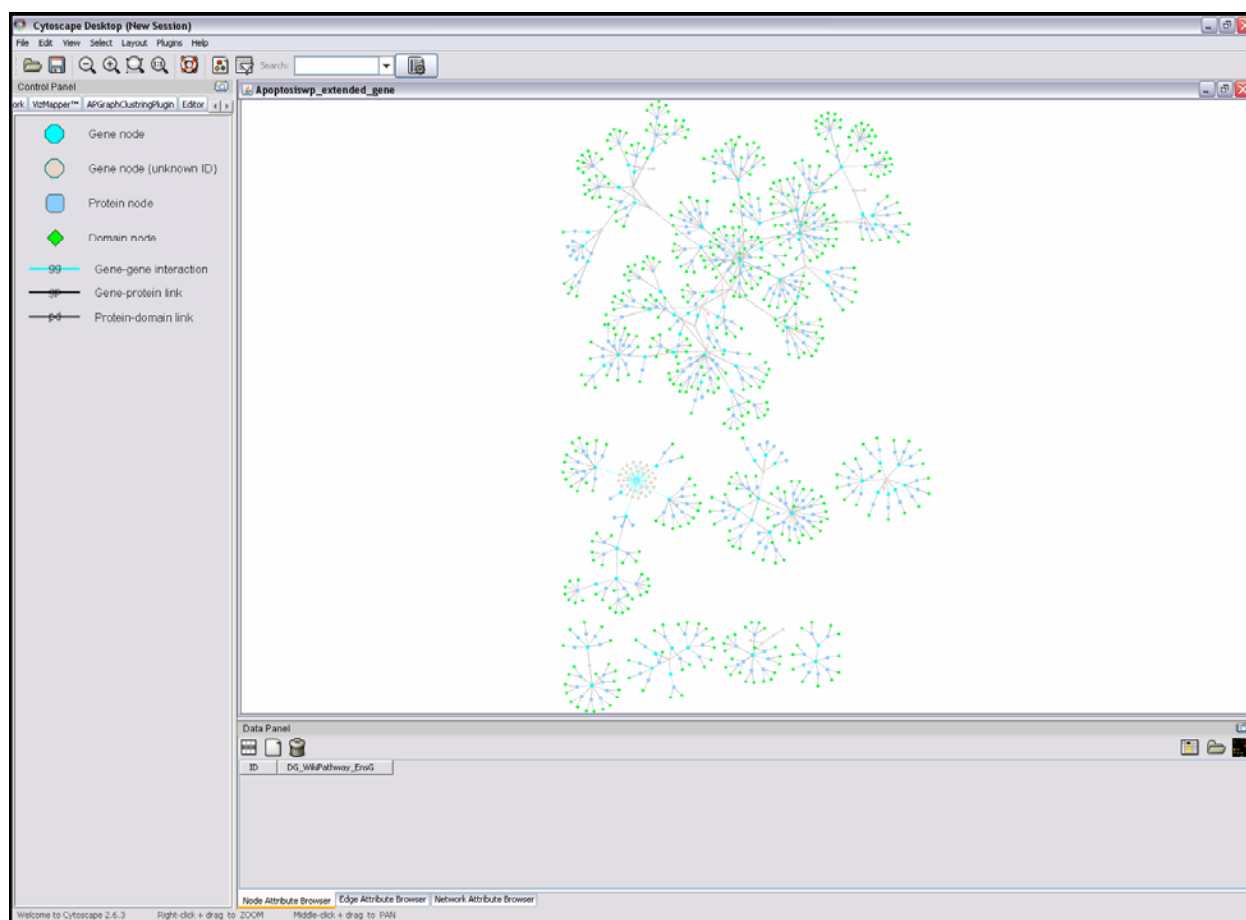


- Pathway nodes associated with differentially expressed probeset are highlighted.
- The table in the results panel shows the node ids, gene ids and gene names for regulated nodes.
- Please note that you can import this pathway into DomainGraph if you want to further analyze it. The WikiPathway nodes are automatically annotated with an attribute called 'DG_WikiPathway_EnsG'. This attribute is used for making this pathway a Domain Graph.

Transforming a WikiPathway pathway into a gene interaction Domain Graph

- Go to the plugins menu, select DomainGraph -> Start DomainGraph.
- Select 'Create domain graph for gene or protein interaction network'.
- Click the Start button.
- Select 'Visualize gene interaction network'.
- Click on the drop-down menu to select/import gene interactions.
- Select 'Apoptosis'. (This is the WikiPathway that you just loaded).
- Leave the network view on 'extended' and identifier on 'Ens Gene'.
- Click the Submit button.
- A window pops up asking which node attribute contains the gene ids. Select 'DG_WikiPathway_EnsG' and click OK.

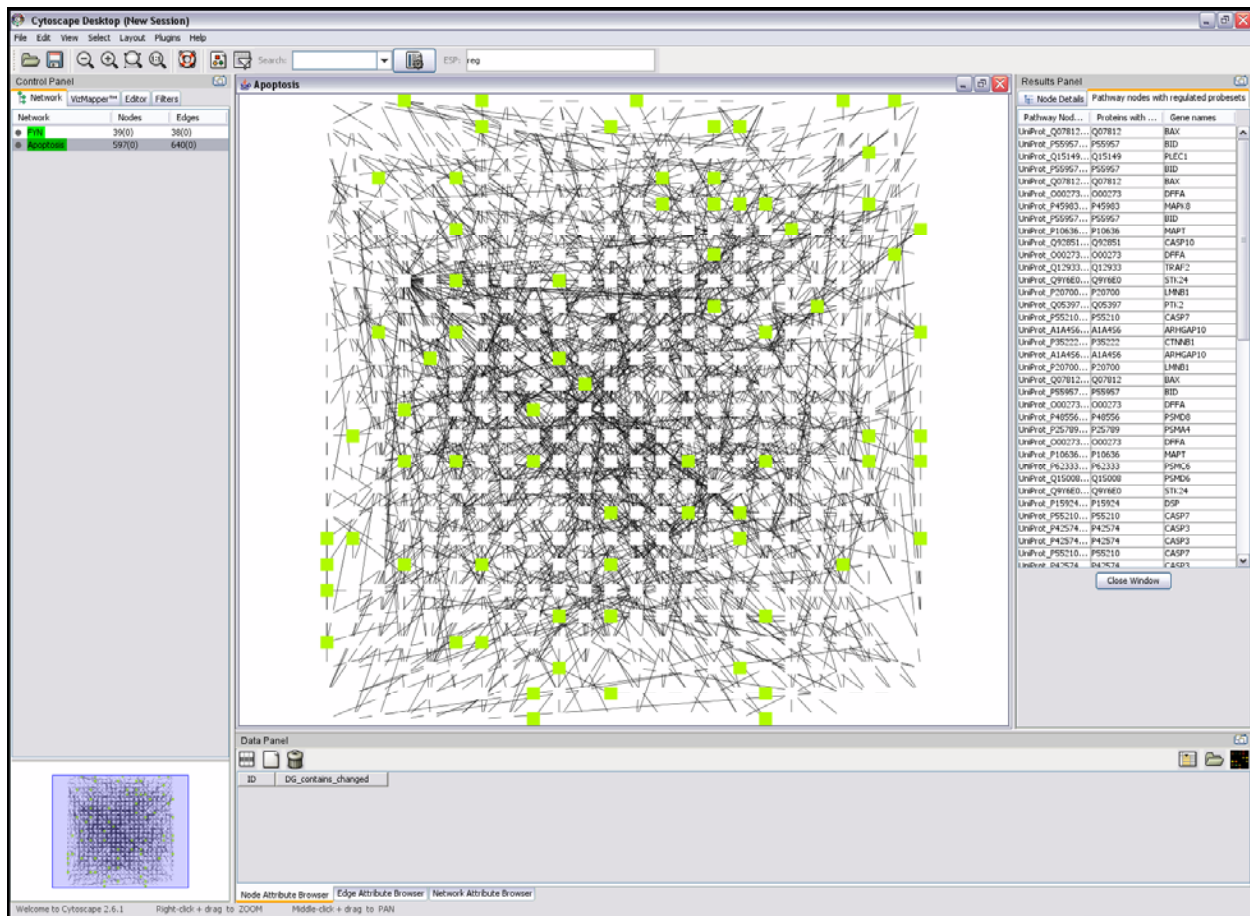
- The resulting domain graph consists of the input genes and gene interactions and is enriched by the protein isoforms encoded by the genes together with their domain compositions.
- Your resulting network should look similar to this **screenshot**:



- If you want to overlay the network with your AltAnalyze results and further analyze your data, continue to Workflow 3 below.

Analysis of a Reactome pathway

- Sort the table according to Reactome pathways by clicking on the 'Reactome pathway(s)' header.
- Scroll to an entry containing 'Apoptosis' in the Reactome column.
- Click on 'Apoptosis'.
- A new window opens displaying all pathways associated with the probeset's gene. Click on 'Apoptosis' in this window and then on the 'Submit' button.
- Wait for the Reactome pathway to be loaded. Please note that you need to have an internet connection to load pathways!
- The resulting pathway should look similar to this **screenshot**:

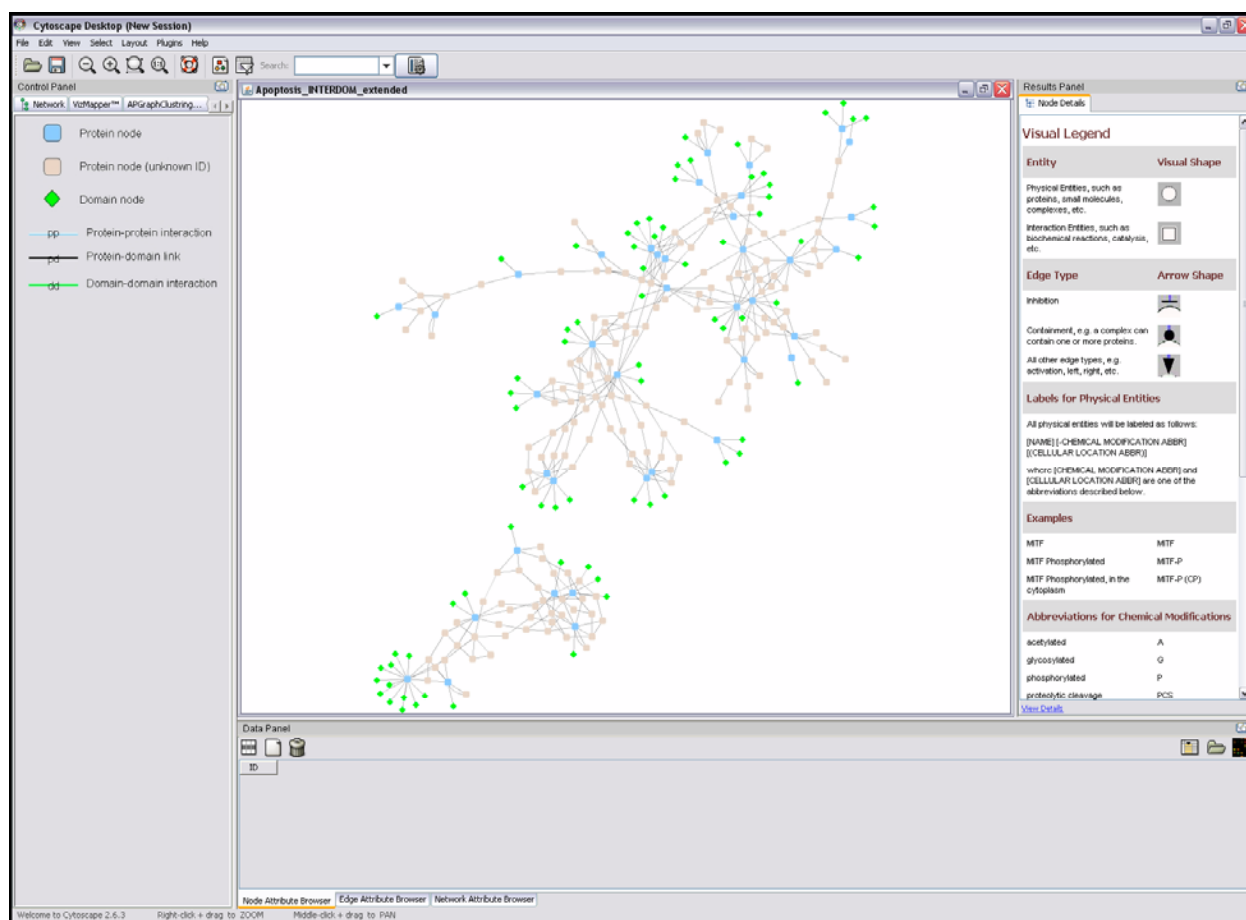


- Pathway nodes associated with differentially expressed probeset are highlighted.
- The table in the results panel shows the node ids, UniProt proteins and gene names for regulated nodes.
- Please note that you can import this pathway into DomainGraph if you want to further analyze it (see workflows 2 and 3).

Transforming a Reactome pathway into a protein interaction Domain Graph

- Go to the plugins menu, select DomainGraph -> Start DomainGraph.
- Select 'Create domain graph for gene or protein interaction network'.
- Click the Start button.
- Select 'Visualize protein interaction network'.
- Click on the drop-down menu to select/import protein interactions.
- Select 'Apoptosis'. (This is the Reactome pathway that you just loaded).
- Click on the second drop-down menu to select domain interactions. Choose 'InterDom' as domain interaction dataset.
- Leave the network view on 'extended' and identifier on 'UniProt'.
- Click the Submit button.
- A window pops up asking which node attribute contains the gene ids. Select 'biopax.xref.UniProt' and click OK.

- The resulting domain graph consists of the input proteins and protein interactions and is enriched by the domain compositions and domain interactions occurring according to InterDom.
- Your resulting network should look similar to this **screenshot**:



- If you want to overlay the network with your AltAnalyze results and further analyze your data, continue to Workflow3 below.

Workflow 2: Analyzing a protein interaction network

Installing DomainGraph

- Download and install Cytoscape 2.6
 - Go to the plugins manager in the Cytoscape menu, 'Network inference' and select DomainGraph for install.
 - Start DomainGraph via the plugins menu.
 - Upon first usage you need to register and accept the DomainGraph license.
-

Installing the DomainGraph database (WITH internet connection)

- Go to the plugins menu, select DomainGraph -> Manage DomainGraph database -> Import data for selected species into database.
 - Choose a local folder to install the database. Make sure to have write permission for this folder.
 - Select 'Homo sapiens'.
 - Select database version ENS_52.
 - Click on the import button.
 - Wait until a window pops up telling you the database is complete.
-

Installing the DomainGraph database (on a machine WITHOUT internet connection)

- You first have to install the database on a computer that is connected to the internet, as described above.
 - Once you have installed the database, you can copy it to any other machine as follows:
 - Copy the database directory (the folder that you chose for the database) to the non-connected machine.
 - Example: You installed the database to "C:\Cytoscape\Database\" on the internet computer.
 - You copy the "Database" folder anywhere to the non-internet computer.
 - Go to the plugins menu, select DomainGraph -> Manage DomainGraph database -> Change path to database.
 - Choose the "Database" directory as new path.
 - The database can now be used on your non-internet computer.
-

Importing a protein interaction network into DomainGraph

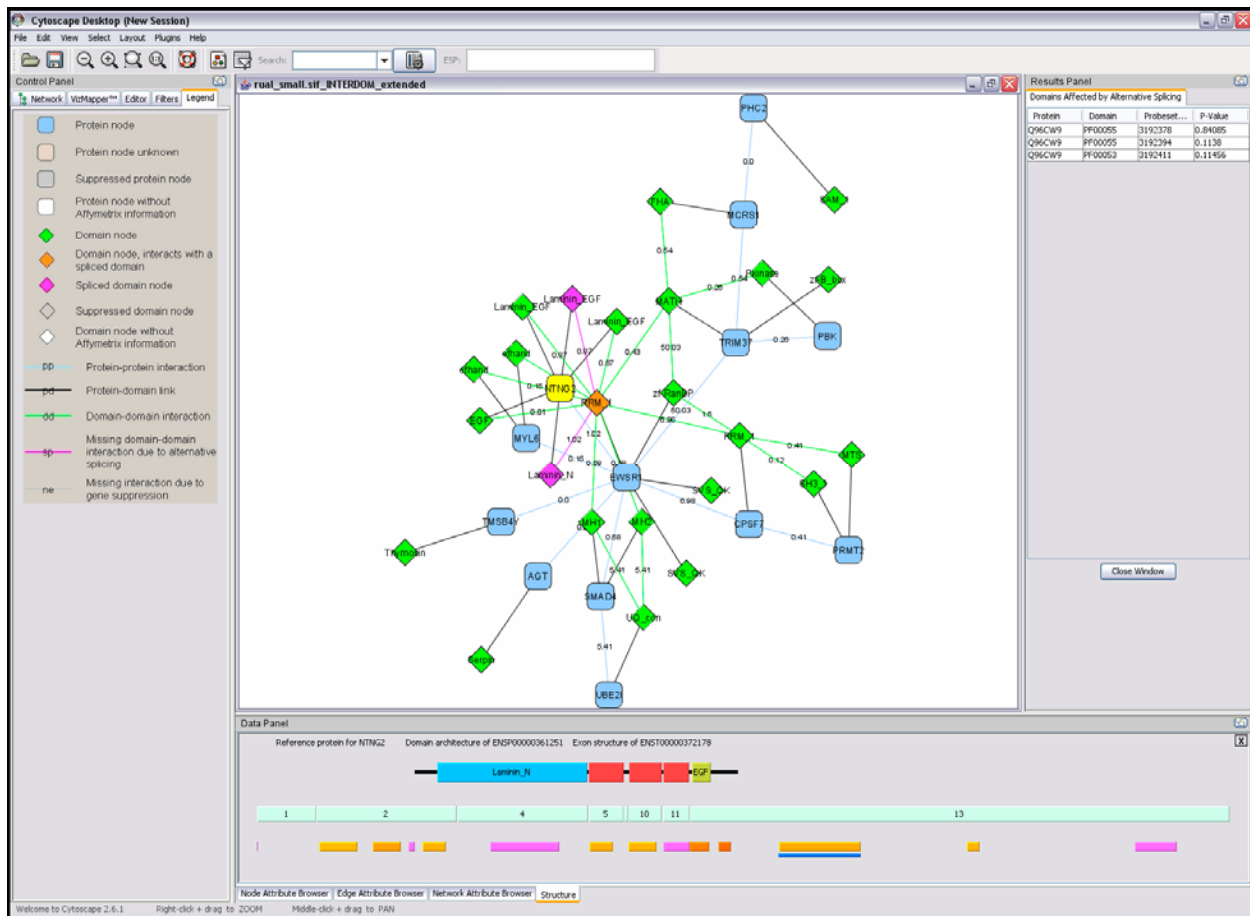
This step requires a protein interaction network, given with Ensembl or UniProt ids. You can download a toy protein interaction network containing human proteins via right-click this link: http://domaingraph.bioinf.mpi-inf.mpg.de/sampledata/Toy_network_PROTEIN_UniProtIDs.sif. Save the file to your disk.

- Go to the plugins menu, select DomainGraph -> Start DomainGraph.
- Select 'Create domain graph for gene or protein interaction network'.
- Click the Start button.
- Select 'Visualize protein interaction network'.

- Click the OK button and wait for the data integration.
- In the resulting domain graph, domains partly or completely not expressed (=spliced out) are highlighted in pink. Orange domains are those being normally expressed but would form interactions with a spliced-out domain.
- Your resulting network should look similar to this **screenshot**:

- In the results panel on the right you can see the proteins, domains, and probesets that are not expressed in the sample according to their DABG p-values.
- Click on a row in the results panel to select the respective node in the network.

- Double-click on the protein node labelled 'NTNG2' (you might have to zoom in to be able to read the node labels).
- The graphics is displayed in the data panel and should look similar to this **screenshot**:



- The top row of an entry shows the protein with its contained domains (colored rectangles).
- Below is the exon structure including 5' and 3' UTRs.
- Below are the probesets. Pink coloring indicates that the probeset is not expressed. A color gradient from yellow to red is applied for expressed probesets representing the expression strength.
- Please note the absent probesets (colored in pink) covering most of the sequence of the Laminin-N domain and the 3rd Laminin-EGF domain.
- The blue rectangle below one probeset indicates a microRNA binding site.
- Move the mouse over the domains, exons, probesets, and miRNAs. Tooltips providing additional information are shown.
- Move the mouse over the pink probesets. The tooltips display the p-values computed by DABG.
- Right-click somewhere in the data panel. A menu will pop up for customizing and saving the graphics.
- Click on an exon. This will open the corresponding Ensembl website for further information.

Workflow 3: Analyzing a gene interaction network

Installing DomainGraph

- Download and install Cytoscape 2.6
 - Go to the plugins manager in the Cytoscape menu, 'Network inference' and select DomainGraph for install.
 - Start DomainGraph via the plugins menu.
 - Upon first usage you need to register and accept the DomainGraph license.
-

Installing the DomainGraph database (WITH internet connection)

- Go to the plugins menu, select DomainGraph -> Manage DomainGraph database -> Import data for selected species into database.
 - Choose a local folder to install the database. Make sure to have write permission for this folder.
 - Select 'Homo sapiens'.
 - Select database version ENS_52.
 - Click on the import button.
 - Wait until a window pops up telling you the database is complete.
-

Installing the DomainGraph database (on a machine WITHOUT internet connection)

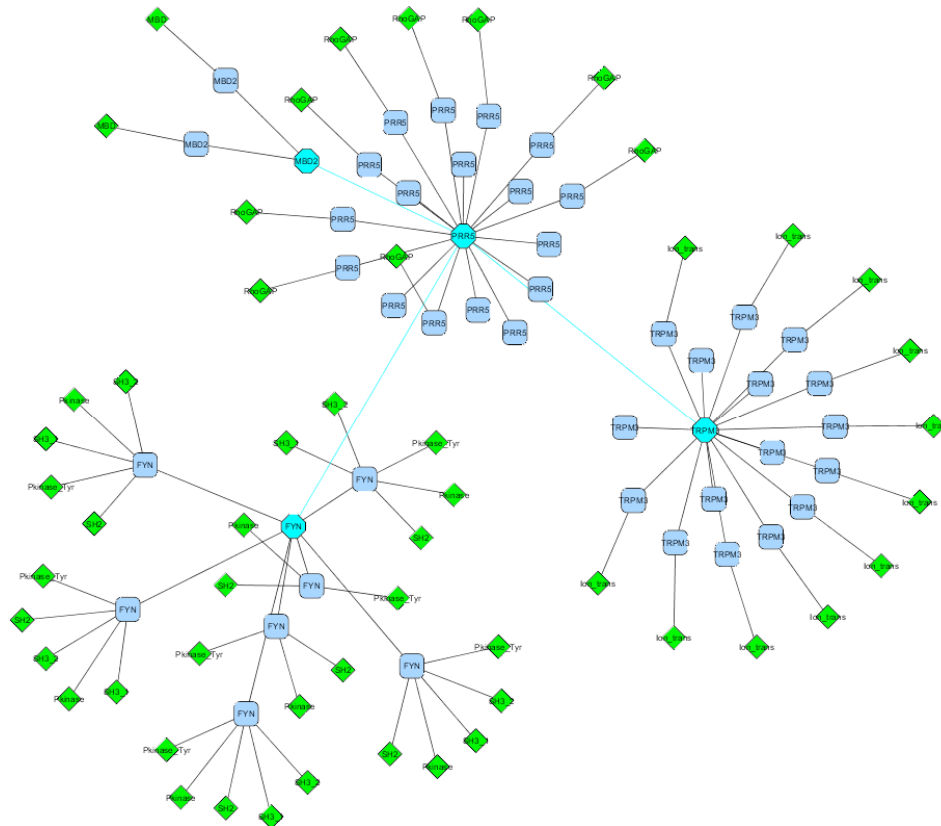
- You first have to install the database on a computer that is connected to the internet (as described above).
 - Once you have installed the database, you can copy it to any other machine as follows:
 - Copy the database directory (the folder that you chose for the database) to the non-connected machine.
 - Example: You installed the database to "C:\Cytoscape\Database" on the internet computer.
 - You copy the "Database" folder anywhere to the non-internet computer.
 - Go to the plugins menu, select DomainGraph -> Manage DomainGraph database -> Change path to database.
 - Choose the "Database" directory as the new path
 - The database can now be used on your non-internet computer.
-

Importing a gene interaction network into DomainGraph

This step requires a gene interaction network, given with Ensembl or Entrez gene ids. You can download a toy gene interaction network containing human genes via right-click on this link: http://domaingraph.bioinf.mpi-inf.mpg.de/sampledata/Toy_network_GENE_EnsemblIDs.sif. Save the file to your disk.

- Go to the plugins menu, select DomainGraph -> Start DomainGraph.
- Select 'Create domain graph for gene or protein interaction network'.
- Click the Start button.
- Select 'Visualize gene interaction network'.

- Click on the drop-down menu to select/import gene interactions.
- Select 'Import own network'.
- Choose the location of your gene interaction network.
- Leave the network view on 'extended' and identifier on 'Ens Gene'.
- Click the Submit button.
- The resulting domain graph consists of the input genes and gene interactions and is enriched by the protein isoforms encoded by the genes together with their domain compositions.
- Your resulting network should look similar to this **screenshot**:

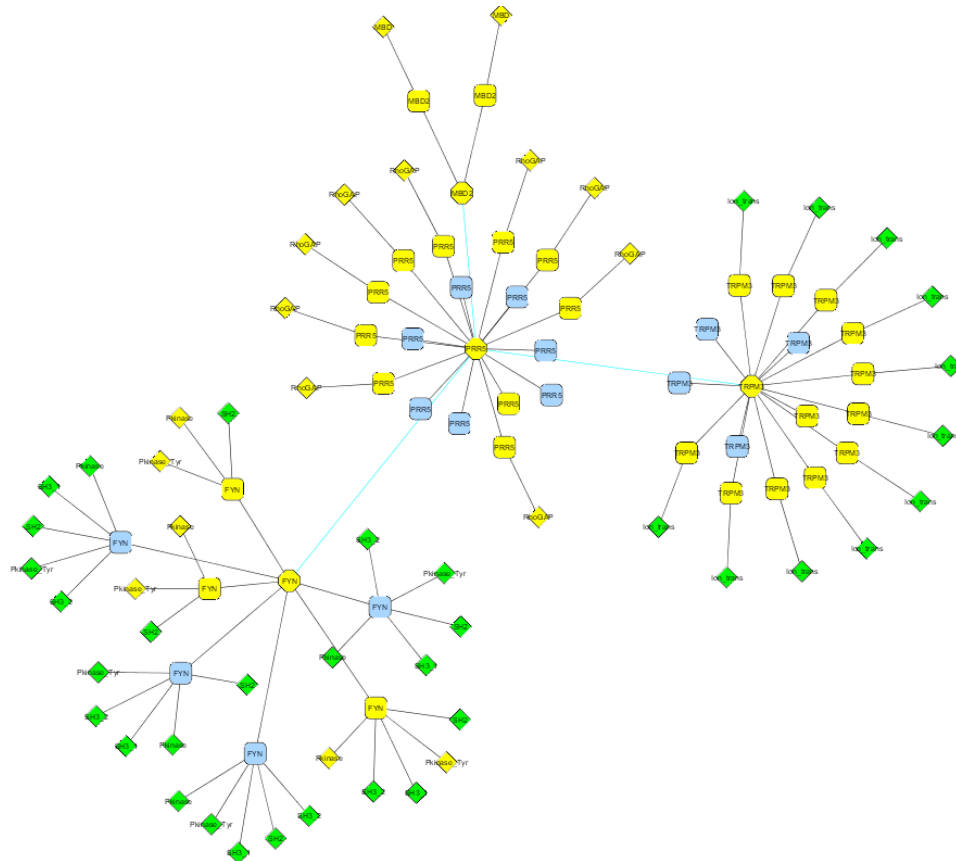


Integrating Affymetrix Exon Array data produced by AltAnalyze for analyzing differential exon expression

This step requires a results file produced by AltAnalyze (in the results folder of AltAnalyze, called "YOUR_DATASET_NAME-DomainGraph.txt"). You can download a sample results file via right-click on this link: http://domaingraph.bioinf.mpi-inf.mpg.de/sampledData/AltAnalyze-results.Hs_Exon_CP_vs_hESC-splicing-index-DomainGraph.txt. This sample file compares human embryonic stem cells (control group) to cardiac progenitor cells (experimental group). Save the file to your disk.

- Go to the plugins menu, select DomainGraph -> Integrate Affymetrix exon expression data.
- Select the 'AltAnalyze data' tab.
- Click the Import button.

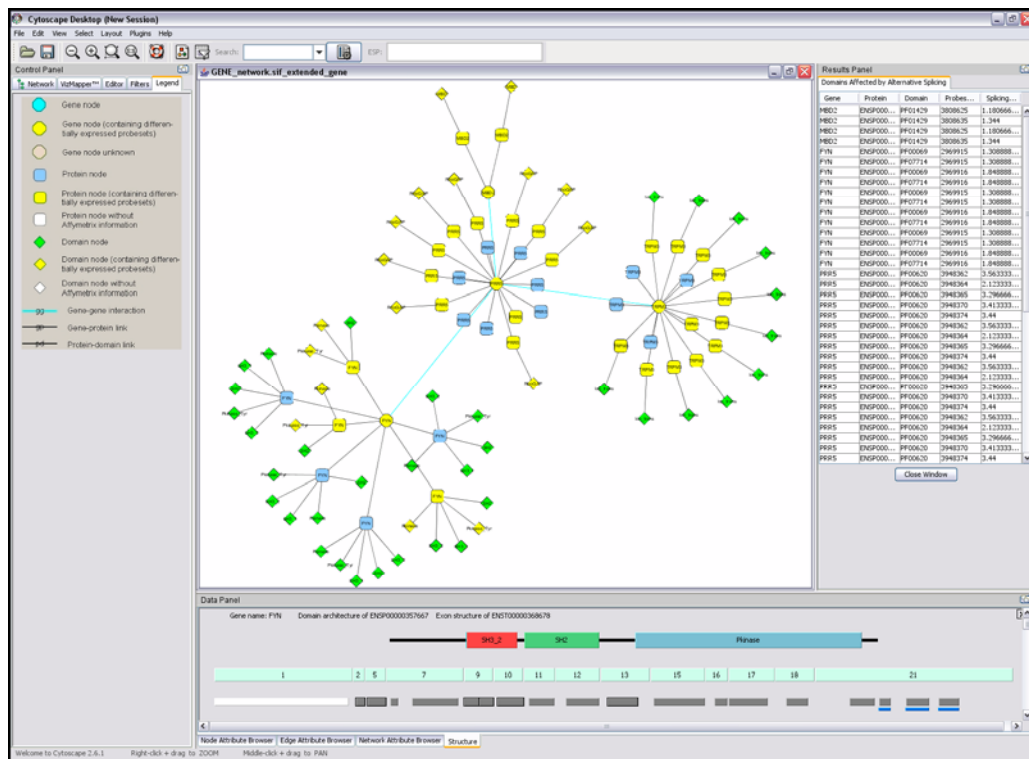
- Select the sample data for import.
- Type in a name for this dataset and click on Import.
- Once the data is imported, it is shown in the drop-down menu.
- Click the OK button and wait for the data integration.
- In the resulting domain graph, all genes, proteins and domains containing differentially expressed probeset are highlighted in yellow.
- Your resulting network should look similar to this **screenshot**:



- In the results panel on the right you can see the genes, proteins, domains, and probesets that are differentially expressed in the two samples together with their splicing indices.
- Click on a row in the results panel to select the respective node in the network.

Viewing proteins, domains, exons, probesets, and miRNA binding sites as graphics

- Double-click on the gene node labelled 'FYN' (you might have to zoom in to be able to read the node labels).
- The graphics is displayed in the data panel and should look similar to this **screenshot**:



- The top row of an entry shows the protein with its contained domains (colored rectangles).
- Below is the exon structure including 5' and 3' UTRs. Identical exon numbers in different transcripts indicate identical exons.
- Below are the probesets. White coloring means the probeset was not statistically analyzed by AltAnalyze due to pre-filtering steps. Gray probesets are expressed at similar levels in both groups of samples. Differentially expressed probesets are shown in red (up-regulated in the experimental group) and green (down-regulated in the experimental group).
- Scroll down the graphics. Please note the down-regulated (highlighted by green-colored probesets) exon no 14. This result indicates that protein isoforms containing exon no 14 are down-regulated in cardiac progenitor cells.
- Blue rectangles below some probesets indicate microRNA binding sites.
- Move the mouse over the domains, exons, probesets, and miRNAs. Tooltips providing additional information are shown.
- Move the mouse over the green-colored probesets annotated to exon no 14. The tooltips display that exon no 14 is known to be a cassette exon, supporting the potential alternative splicing event. The Cross-hybridization information shows that both these probesets are 'unique'. This means, the probesets were specifically designed for this genomic locus and (at design-time of the microarray) did not align to any other transcript.
- Right-click somewhere in the data panel. A menu will pop up for customizing and saving the graphics.
- Click on an exon. This will open the corresponding Ensembl website for further information.