PATHWAY STUDIO™

Finding Major Regulators/Cellular Processes/Diseases in Experimental Data

ELSEVIER

Sub-Network Enrichment Analysis (SNEA) is used to identify and prioritize the targets and regulators most implicated in the experimental dataset. SNEA is based on the Gene Set Enrichment Analysis algorithm. Sub-Networks are user defined networks calculated *de novo* from the information in the database and consist of a seed/regulator and their neighbors (targets) in the database. The seeds of the sub-network whose targets are statistically enriched are implicated as important regulators (or cell processes or diseases) by the experimental data. You can specify the kind of seeds and the kind of relationships for building sub-networks from the SNEA dialog. A seed can be a protein/complex/ functional class, small molecule, cellular process or disease. Once identified, regulators can be further examined to help elucidate cellular processes, mechanisms and pathways impacted in the experiment.

Examples of results from SNEA analysis of experimental data:

- Major gene expression regulators (such as transcription factors) responsible for a differential gene expression profile
- Major miRNA regulators responsible for a differential gene expression profile
- Binding regulatory networks
- Differential gene/protein profiles that are enriched for genes/proteins known to be associated with a particular disease
- Differential gene/protein profiles that are enriched for genes/proteins known to be associated with a particular cellular process

The SNEA tool has easy-to-use preset options for defining the most useful sub-network types, and a customer menu for the experienced user who wants to perform more advanced analysis.

Description of Sub-Network Enrichment Analysis

The Sub-Network Enrichment Analysis (SNEA) algorithm uses existing relationships in the database to build “sub-networks” based on user specified criteria. It then uses these sub-networks with the GSEA algorithm to identify the networks that are significantly enriched. When calculating enrichment, only the targets are considered, but not the seed/regulator.

The user-defined sub-networks consist of a single “regulator” or “seed” and its nearest neighbor network. The type of relationship(s) included in the network and the directionality of the network are user determined.

Outbound (from the seed or regulator) relationships are selected when one wants to identify regulators of targets included in the gene list or experimental data. Inbound (to the seed or regulator) relationships are selected when the seed is a disease or cell process. In this instance the algorithm identifies entities known to be associated with a particular disease or cell process. Many combinations of sub-networks are possible.

Recall: in the mammal database all protein relationships to a disease (or cell process) are inbound to the disease (or cell process). Also, the only type of relationship to disease or cell process is regulation.

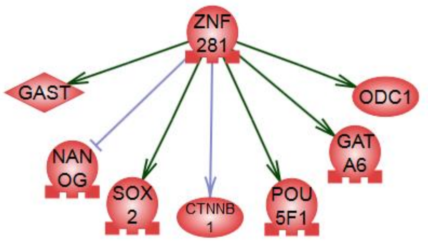
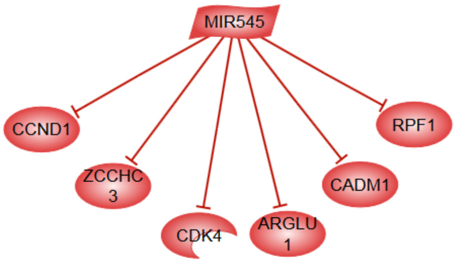
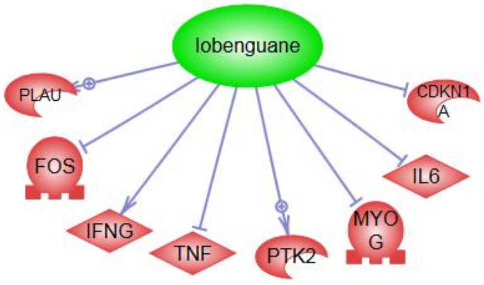
However ChemEffect® and DiseaseFx®, expand on this model. They contain more relation types to diseases and cell processes and some relations such as GeneticChange are outbound from disease. Customer with access to these additional datasets will have additional preset options to allow utilization of the additional relations present in these databases.

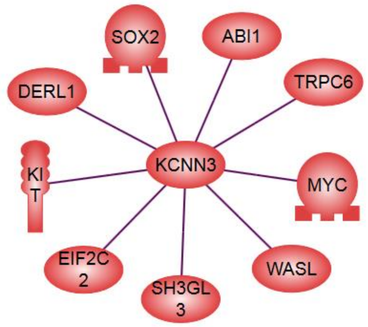
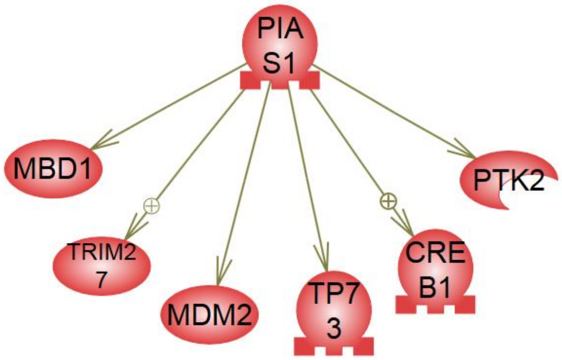
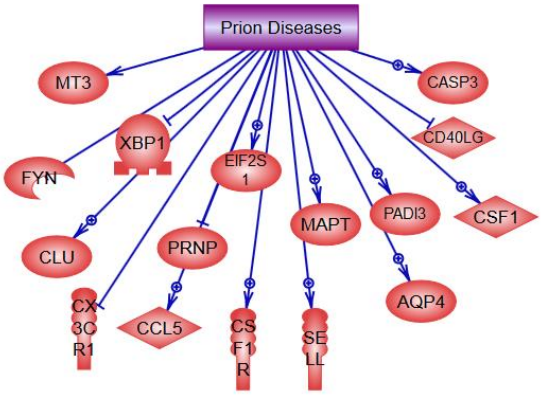
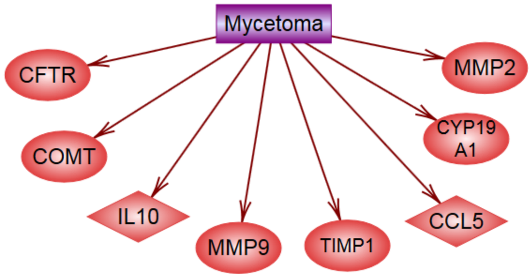
Sub-networks that have as the seed a disease or cell process are useful in defining a collection of proteins known to be associated with these entities, without any implied function by the designated directionality of the relationship.

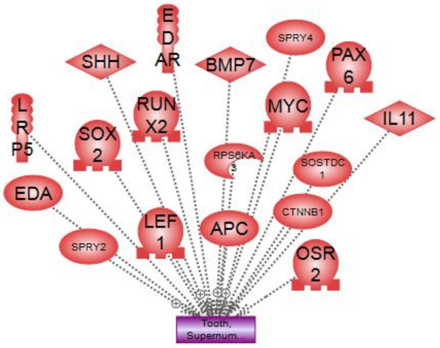
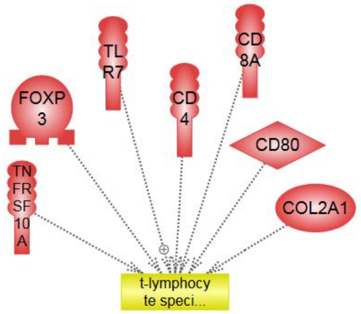
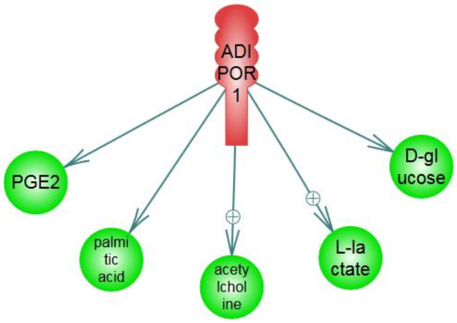
Defining the Sub-Networks

Selecting user-defined sub-networks involves first defining the "regulator" or "seed" type" and the nearest "neighbor" network by selecting specific relationship types and directionality.

Examples of commonly used sub-networks available through presets:

Preset Name: Expression Targets seed/regulator = protein/complex/functional class; relations = promoter binding, expression neighbors/targets = proteins input data set = gene expression (most common), miRNA array results: identifies major expression regulators active in experiment	 ZNF 281 GAST ODC1 NAN OG GAT A6 SOX 2 CTNNB 1 POU 5F1
Preset Name: miRNA Targets seed/regulator = miRNA relations = miRNAEffect neighbors/targets = proteins input data set = gene expression (most common), miRNA array results: identifies major miRNAs regulating gene expression in experiment	 MIR545 CCND1 RPF1 ZCCHC 3 CADM1 CDK4 ARGLU 1
Preset Name: Chemical Expression Targets seed/regulator = small molecules relations = expression neighbors/targets = proteins input data set = gene expression results: identifies small molecules/drugs that regulate expression of proteins	 Iobenguane PLAU CDKN1 A FOS IL6 IFNG TNF PTK2 MYO G

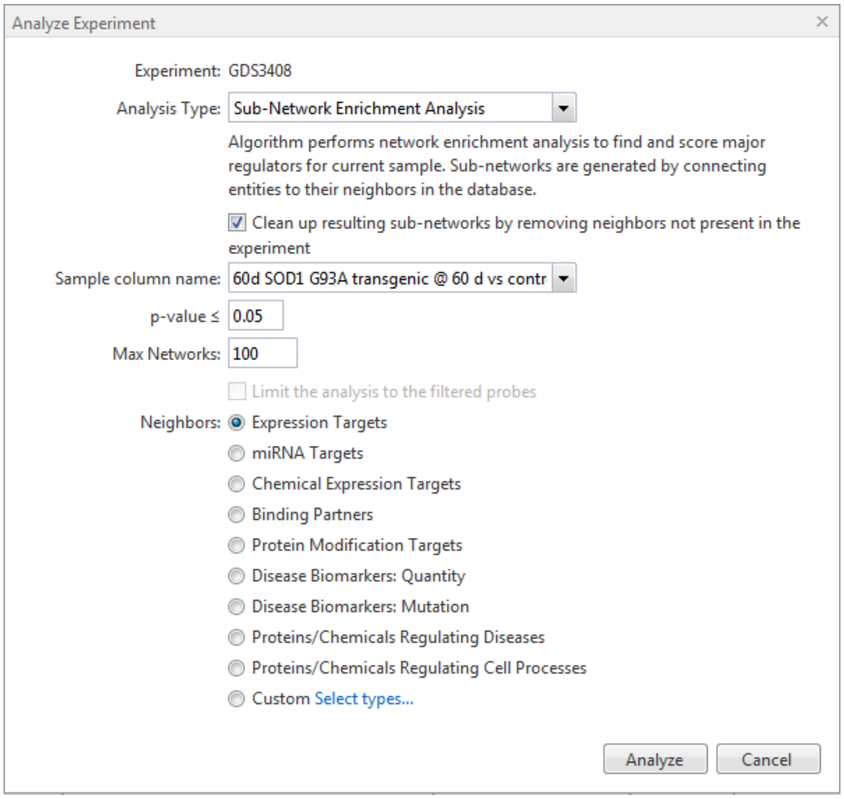
Preset Name: Binding Proteins seed/regulator = protein relations = binding neighbors/targets = protein input data set = proteomics results: identifies enriched binding sub-networks mediated by an individual protein	 SOX2 ABI1 DERL1 TRPC6 KIT KCNN3 MYC EIF2C2 SH3GL3 WASL
Preset Name: Protein Modification Targets seed/regulator = proteins relations = protein modification neighbors/targets = proteins input data set = proteomics (most common) results: identifies binding protein modification sub-networks through activities such as: acetylation, cleavage, deacetylation, demethylation, dephosphorylation, direct interaction, methylation, phosphorylation,posttrascriptional inhibition, proteolysis, ubiquitination	
Preset Name: Disease Biomarkers: Quantity seed/regulator = disease relations = quantitative change neighbors/targets = proteins input data set = gene expression (most common), proteomics results: identifies enrichment of proteins that are associated with specific diseases through changes in the proteins abundance, expression or activity	 Prion Diseases MT3 CASP3 CD40LG XBP1 EIF2S1 FYN MAPT PADI3 CSF1 CLU PRNP AQP4 CX3CR1 CCL5 CSF1R SELL
Preset Name: Disease Biomarkers: Mutation seed/regulator = disease relations = genetic change neighbors/targets = proteins input data set = gene expression (most common), proteomics results: identifies enrichment of proteins that are associated with a specific disease through genetic changes in the genes such as: gene amplification, epigenic methylation	 Mycetoma CFTR MMP2 COMT CYP19A1 IL10 MMP9 TIMP1 CCL5

Preset Name: Proteins/Chemicals Regulating Diseases seed/regulator = disease relations = regulation neighbors/targets = proteins (shown) or small molecules input data set = gene expression (most common), proteomics or metabolomics results: identifies proteins (gene expression or proteomics data) or small molecules (metabolomics data) enriched for a specific disease	 E D AR SHH BMP7 SPRY4 PAX 6 L R P5 RUN X2 MYC IL11 SOX 2 RPS6KA 3 SOSTDC 1 EDA LEF 1 APC CTNNB1 SPRY2 OSR 2 Tooth, Supernum...
Preset Name: Proteins/Chemicals Regulating Cell Processes seed/regulator = disease relations = regulation neighbors/targets = proteins (shown) or small molecules input data set = gene expression (most common), proteomics or metabolomics results: identifies proteins (gene expression or proteomics data) or small molecules (metabolomics data) enriched for a specific cellular process	 TL R7 CD 8A FOXP 3 CD 4 CD80 TN FR SF 10 A COL2A1 t-lymphocy te speci...
Preset Name: Metabolomics Targets seed/regulator = protein relations = molsynthesis neighbors/targets = small molecules input data set = metabolomics results: identifies enrichment of small molecules where changes in abundance regulated by a specific protein	 ADI POR 1 PGE2 D-gl ucose palmi tic acid acety lchol ine L-la ctate

Running Sub-Network Enrichment Analysis in Pathway Studio

With an experiment open, go to the experiment view Tools menu and select “Analyze Experiment.” From the Analyze Experiment window select “Sub-Network Enrichment Analysis” in the Analysis Type drop down menu.

The box “clean up resulting sub-networks by removing neighbors not present in the experiment” is checked by default. This will limit the sub-networks to only entities included in the experiment being analyzed.



Experiment: GDS3408

Analysis Type: Sub-Network Enrichment Analysis

Algorithm performs network enrichment analysis to find and score major regulators for current sample. Sub-networks are generated by connecting entities to their neighbors in the database.

☑ Clean up resulting sub-networks by removing neighbors not present in the experiment

Sample column name: 60d SOD1 G93A transgenic @ 60 d vs contr

p-value ≤ 0.05

Max Networks: 100

□ Limit the analysis to the filtered probes

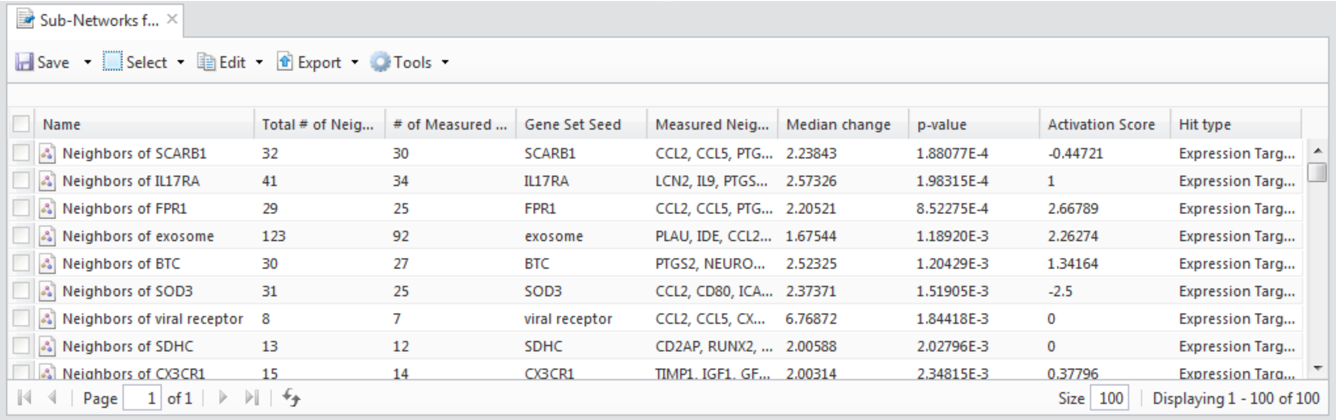
Neighbors:
- ● Expression Targets
- ○ miRNA Targets
- ○ Chemical Expression Targets
- ○ Binding Partners
- ○ Protein Modification Targets
- ○ Disease Biomarkers: Quantity
- ○ Disease Biomarkers: Mutation
- ○ Proteins/Chemicals Regulating Diseases
- ○ Proteins/Chemicals Regulating Cell Processes
- ○ Custom Select types...

If the experiment has multiple comparisons, select the desired comparison for analysis. The default *p*-value is < 0.05 and the maximum number of networks is set to 100. These values can be changed at the user's discretion.

The list of preset options is dependent on the type of experiment. For example, a metabolomics data set will have a different list of preset options than will a gene expression experiment. For most users' needs, the list of preset options will suffice. However, there is an option to define custom sub-networks for more advanced users.

For the example below the experiment type is gene expression and the preset selected is "Expression Targets." This analysis will identify the top gene expression regulators for this experiment.

The results of the analysis are displayed in the list pane at the bottom of the screen.

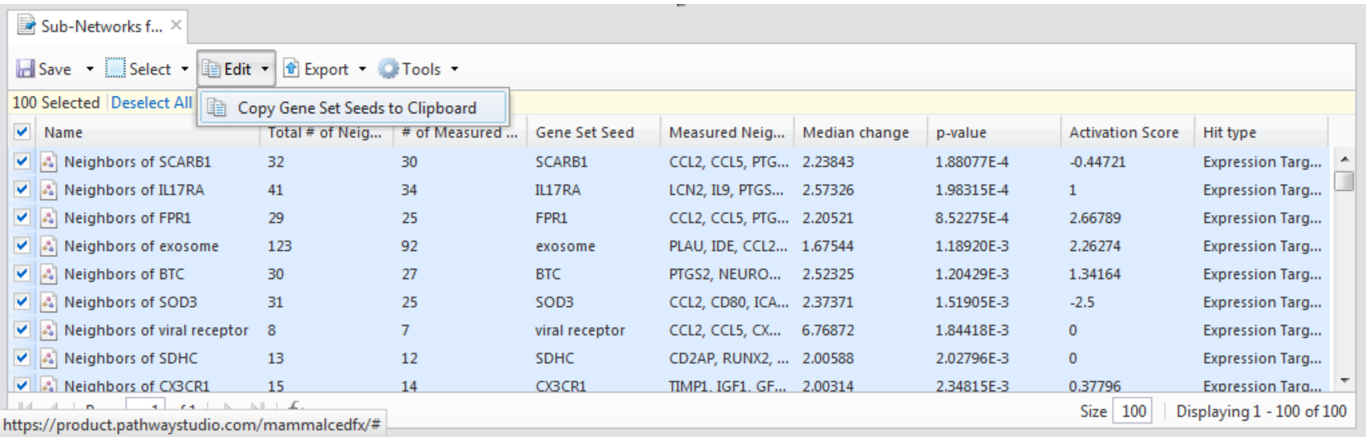


□	Name	Total # of Neig...	# of Measured ...	Gene Set Seed	Measured Neig...	Median change	p-value	Activation Score	Hit type
□	Neighbors of SCARB1	32	30	SCARB1	CCL2, CCL5, PTG...	2.23843	1.88077E-4	-0.44721	Expression Targ...
□	Neighbors of IL17RA	41	34	IL17RA	LCN2, IL9, PTGS...	2.57326	1.98315E-4	1	Expression Targ...
□	Neighbors of FPR1	29	25	FPR1	CCL2, CCL5, PTG...	2.20521	8.52275E-4	2.66789	Expression Targ...
□	Neighbors of exosome	123	92	exosome	PLAU, IDE, CCL2...	1.67544	1.18920E-3	2.26274	Expression Targ...
□	Neighbors of BTC	30	27	BTC	PTGS2, NEURO...	2.52325	1.20429E-3	1.34164	Expression Targ...
□	Neighbors of SOD3	31	25	SOD3	CCL2, CD80, ICA...	2.37371	1.51905E-3	-2.5	Expression Targ...
□	Neighbors of viral receptor	8	7	viral receptor	CCL2, CCL5, CX...	6.76872	1.84418E-3	0	Expression Targ...
□	Neighbors of SDHC	13	12	SDHC	CD2AP, RUNX2, ...	2.00588	2.02796E-3	0	Expression Targ...
□	Neighbors of CX3CR1	15	14	CX3CR1	TIMP1, IGF1, GF...	2.00314	2.34815E-3	0.37796	Expression Targ...

Page 1 of 1 Size 100 Displaying 1 - 100 of 100

Each *de novo* calculated sub-network is named based on its seed/regulator. The total number of neighbors in the sub-network may be higher than the # of measured neighbors. If the box was checked to “clean up” the resulting sub-networks, then only the # of measured neighbors will be displayed in the graph view when selected.

The Gene Set Seed column provides the list of identified regulators. In this example these are the most important expression regulators for this experiment. You can copy this column by selecting the entire table and then going to Edit > Copy Gene Seeds to Clipboard.


Sub-Networks f...
Save
Select
Edit
Export
Tools
100 Selected
Deselect All
Copy Gene Set Seeds to Clipboard
Size 100
Displaying 1 - 100 of 100
https://product.pathwaystudio.com/mammalcedfx/#

☑	Name	Total # of Neig...	# of Measured ...	Gene Set Seed	Measured Neig...	Median change	p-value	Activation Score	Hit type
☑	Neighbors of SCARB1	32	30	SCARB1	CCL2, CCL5, PTG...	2.23843	1.88077E-4	-0.44721	Expression Targ...
☑	Neighbors of IL17RA	41	34	IL17RA	LCN2, IL9, PTGS...	2.57326	1.98315E-4	1	Expression Targ...
☑	Neighbors of FPR1	29	25	FPR1	CCL2, CCL5, PTG...	2.20521	8.52275E-4	2.66789	Expression Targ...
☑	Neighbors of exosome	123	92	exosome	PLAU, IDE, CCL2...	1.67544	1.18920E-3	2.26274	Expression Targ...
☑	Neighbors of BTC	30	27	BTC	PTGS2, NEURO...	2.52325	1.20429E-3	1.34164	Expression Targ...
☑	Neighbors of SOD3	31	25	SOD3	CCL2, CD80, ICA...	2.37371	1.51905E-3	-2.5	Expression Targ...
☑	Neighbors of viral receptor	8	7	viral receptor	CCL2, CCL5, CX...	6.76872	1.84418E-3	0	Expression Targ...
☑	Neighbors of SDHC	13	12	SDHC	CD2AP, RUNX2, ...	2.00588	2.02796E-3	0	Expression Targ...
☑	Neighbors of CX3CR1	15	14	CX3CR1	TIMP1, IGF1, GF...	2.00314	2.34815E-3	0.37796	Expression Targ...

The Median Change value indicates the median expression value for all the targets of the sub-network.

The Activation Score indicates how closely the changes in the expression of the targets closely match the predicted effect (positive or negative) that the regulator has on the target, where a positive number indicates concordance and a negative number indicated discordance.

The Activation Score is a measure of whether the regulator is “active” or “repressed” in the experimental conditions.

Regulator Effect on Target	**Target Experimental Results**	**Concordant/ Disconcordant**	***Implied* activation/repression of seed/regulator***
Positive	Down regulated	Disconcordant	Repression
Positive	Up regulated	Concordant	Activated
Negative	Down regulated	Concordant	Activated
Negative	Up regulated	Disconcordant	Repressed

* The user must examine if the actual state of the seed/regulator matches or does not match the implied state of activation or repression.

Activation Score calculation considerations:

- Any genes with absolute log change below 0.5 are disregarded from calculation
- Any genes that are connected with relationships without effect are disregarded from calculation

Thus, differential genes (that may or may not been filtered by a user) with log changes above 0.5 threshold AND connected with signed relationships to the seed are analyzed. The ones that differentially change consistent with the sign of the relationship are considered “concordant.”
The activation score = (N_concordant – N_disconcordant)/sqrt(N_total)

Further insight into the experiment may be gained by examining the list of top regulators individually or as a group and examining the concordance/disconconrdance of the sub-networks.

If you have any questions about Pathway Studio please contact Customer Care:

USA, Canada and Latin America:

(8am-8pm CET - St.Louis)
Tel: US toll-free: +1 (888) 615 4500
Tel: Non toll-free: +1 (314) 523 4900
Email: usinfo@elsevier.com
Email Brazil: brinfo@elsevier.com

Europe, Middle East and Africa:

(9am-6pm GMT+1, Amsterdam office)
Tel: +31 20 485 3767
Email: nlinfo@elsevier.com

Japan:

(9,30am-5,30pm JST, Tokyo office)
Tel: +81 (3) 5561 5035
Email: jpinfo@elsevier.com
Website: japan.elsevier.com

Asia and Australasia:

(9am-6pm SST, Singapore office)
Tel: +65 6349 0222
Email: sginfo@elsevier.com