



**Cell Illustrator**  
**Gene Net**  
**User Guide**

© 2002-2007 GNI Ltd. and Human Genome Center, Institute of  
Medical Science, The University of Tokyo

All rights reserved.

## Table of contents:

<b>1 Introduction.....</b>	<b>4</b>
<b>2 Glossary.....</b>	<b>5</b>
<b>3 Installing and Running Cell Illustrator.....</b>	<b>6</b>
3.1.1 Windows.....	6
3.1.2 Linux.....	6
3.1.3 UNIX.....	7
3.1.4 Mac OS X.....	7
3.2 <i>Licensing Information</i> .....	7
3.3 <i>Cell Illustrator Memory Customization</i> .....	7
<b>4 Gene Net Mode.....</b>	<b>8</b>
4.1 <i>Switching to the Gene Net Mode</i> .....	8
4.2 <i>Differences between Gene Net and Biopathway Modeling Modes</i> .....	9
4.3 <i>Documentation for Biopathway Modeling Mode</i> .....	10
<b>5 Importing a Gene Network .....</b>	<b>11</b>
5.1 <i>Import Gene Network from a Text File</i> .....	11
5.2 <i>Gene Network Elements</i> .....	11
5.3 <i>Save the Gene Network to CI Native Format</i> .....	12
<b>6 Viewing the Gene Network.....</b>	<b>14</b>
6.1 <i>Show/Hide Element Properties on the Canvas</i> .....	14
6.2 <i>View the Gene Network in a Tabular View</i> .....	14
6.3 <i>View Detailed Element Properties</i> .....	15
6.4 <i>Optimize the Network Layout</i> .....	15
<b>7 Identification of Subnetworks Regulated by a Specific Gene</b>	<b>17</b>
7.1 <i>Filter Out “Weak” Relations (optional step)</i> .....	17
7.2 <i>Identify a Subnetwork of a Specific Gene</i> .....	18
<b>8 Evaluating a Subnetwork.....</b>	<b>20</b>
8.1 <i>Checking web-based annotations of individual genes</i> .....	20
8.2 <i>Gene Mining</i> .....	21
8.3 <i>Edge Attributes</i> .....	23
8.4 <i>Annotating Edges</i> .....	23
<b>9 Identification of the Regulators of a Target Transcript....</b>	<b>25</b>

<b>10 Identification of Regulatory Hubs.....</b>	<b>26</b>
<b>11 Using a Transcript Known to Play a Role in a Biological Process to Discover the Other Transcripts Involved in this Process</b>	<b>28</b>
11.1 <i>Extract the Subnetwork around the Known Gene.....</i>	28
11.2 <i>Compare the Extracted Subnetwork with Known Facts</i>	28
11.3 <i>Identify Transcripts that Could Be Involved in the Biological Process.....</i>	28
<b>12 Discovering Gene Transcripts Involved in a Biological Process Using Keyword Search.....</b>	<b>29</b>
<b>13 Gene-to-Gene Path Search.....</b>	<b>31</b>

# 1 Introduction

Since the version 3.0, Cell Illustrator (CI) enables the users to view and analyze gene relationships networks. This functionality is available in the special *Gene Net* mode. The graphical user interface of CI offers the user two modes of working:

- (i) The Biopathway Modeling mode dedicated for simulation and modeling of complex biological processes and systems using Petri Net models. This mode is described separately in the **CI User Guide and CI Reference Manual**
- (ii) The new Gene Net mode dedicated for analysis and exploration of gene relationship networks - described in this manual

This guide is an introduction to the analysis and exploration of gene relationship networks and is divided into the following sections:

- installing and running Cell Illustrator
- importing and viewing gene relationship networks
- identification of subnetworks regulated by a specific gene
- evaluating a subnetwork
- identification of regulators of a target gene transcript
- identification of regulatory hubs
- using a transcript known to play a role in a biological process to discover the other transcripts involved in this process
- discovering gene transcripts involved in a biological process using a keyword search
- gene-to-gene pathway search

The documentation for the Gene Net mode includes the manuals listed below:

**CI Gene Net User Guide:** Introduction to CI Gene Net software

**CI Gene Net Reference Manual:** Detailed description functions

## 2 Glossary

**Gene Network or Gene Relationship Network** – a graph that represents the relations between genes. Each graph node represents one gene (or gene transcript), while each edge represents a relationship between the connected genes. For more details see the chapter on Gene Network Elements. Usually, gene networks are inferred by a statistical application, which handles a large amount of micro array data.

**CI** – Cell Illustrator, the software for simulation of Petri Net models of biological processes and pathways.

**Gene Net mode** – a mode of CI, which is dedicated to the handling of gene networks. This mode differs from the normal mode dedicated to the building and simulation of biopathway models. Both modes offer a different user interface.

### 3 Installing and Running Cell Illustrator

In order to run CI correctly, Java version j2sdk 1.5.0 or higher needs to be installed. Java Runtime Environments (JREs) are available for download at: <http://java.sun.com/downloads>. Alternatively, for Windows and Linux platforms, you may execute the Cell Illustrator installation package bundled with a Java installer.

#### 3.1.1 Windows

##### Java-bundled Installation :

1. Uncompress CI3.0\_wj.zip in a temporary directory (You can use WinZip, [www.winzip.com](http://www.winzip.com))
2. Click on CI3.0\_wj.exe to install the program. Java Virtual Machine (JVM) will also be installed.
3. Launch Cell Illustrator from the Start Menu

##### Non Java-bundled Installation:

1. Uncompress CI3.0\_w.zip in a temporary directory
2. Click on CI3.0\_w.exe to install
3. Launch Cell Illustrator from the Start Menu

#### 3.1.2 Linux

##### Java-bundled Installation:

To install Cell Illustrator:

1. Copy CI3.0\_lj.bin to a temporary directory.
2. Execute the command "chmod +x CI3.0\_lj.bin"
3. Execute CI3.0\_lj.bin file

Java Virtual Machine (JVM) will also be installed.

To launch Cell Illustrator:

1. Go to Cell Illustrator folder
2. Execute the command ". /CI"

##### Non Java-bundled Installation:

To install Cell Illustrator:

1. Copy CI3.0\_l.bin to a temporary directory.
2. Execute the command "chmod +x CI3.0\_l.bin"
3. Execute "CI3.0\_l.bin" file

To launch Cell Illustrator:

1. Go to Cell Illustrator directory
2. Execute ". /CI"

### 3.1.3 UNIX

#### Non Java-bundled Installation:

To install Cell Illustrator:

1. Copy CI3.0\_u.jar to a temporary directory.
2. Execute the command "java -jar CI3.0\_u.jar"

To launch Cell Illustrator:

1. Go to Cell Illustrator folder
2. Execute ". /CI"

### 3.1.4 Mac OS X

#### Non Java-bundled Installation:

To install Cell Illustrator:

1. Copy CI3.0\_m.tgz to a temporary directory.
2. Double-click on CI3.0\_m.tgz
3. Double-click on InstallApp
4. A Cell Illustrator icon should appear on the desktop, double click to launch Cell Illustrator.

CI3.0 requires JRE version 1.5.0 or higher. The jre1.5.0 for Mac OS X Tiger can be downloaded from:

<http://www.apple.com/support/downloads/java2se50release1.html>

## 3.2 Licensing Information

To run Cell Illustrator, you need to obtain a license file from GNI or your local distributor of Cell Illustrator. Copy the license file to a directory where the Cell Illustrator is installed. For example, on Windows systems, the Cell Illustrator is installed to "C:\Program Files\Cell Illustrator 3.0" if the default installation settings are used. The license can be bound to your machine and/or time-limited. To obtain a hardware-bound license, you need to send your machine hardware identifier (Hardware ID) which can be revealed by clicking on the About Cell Illustrator option in the Help menu. In the About Cell Illustrator dialog, the Hardware ID is displayed as the License Details tab. The Hardware ID is also displayed in the Invalid License dialog which appears at the application startup if Cell Illustrator cannot locate a valid license. Press the Copy All button to copy the error message and the Hardware ID into the clipboard. You can later paste this text into your message to the Cell Illustrator technical support team.

## 3.3 Cell Illustrator Memory Customization

Using the auxiliary application *Cell Illustrator Memory Configuration*, you can set the maximum memory for the Cell Illustrator application.

## 4 Gene Net Mode

This guide describes only the functionality of the Gene Net mode. In this chapter it will be described:

- how to switch to the Gene Net mode
- what are the differences between the Gene Net and Biopathway Modeling Modes
- where to find the documentation for the Biopathway mode

### 4.1 Switching to the Gene Net Mode

The menu command View | Gene Net Mode switches between the biopathway modeling and Gene Net modes. If this menu item is checked then you are in the Gene Net mode, otherwise you are in the Biopathway modeling mode.

Any model that has been opened in the Cell Illustrator workspace (or imported) can be viewed in each of both Cell Illustrator modes: Biopathway Modeling and Gene Net mode. However there are the following limitations:

- Do not view large gene networks in the Biopathway Modeling mode

Displaying large networks with hundreds of edges in the biopathway modeling mode will cause performance problems. Therefore they should be displayed in the Gene Net mode only.

- A typical gene network cannot be simulated.

Originally a gene network does not contain any data useful for the simulation. After switching to the Biopathway Modeling mode you will see that all simulation parameters for a gene network have default values. Simulating such a net will not give any reasonable results. However such a net can be used as the starting point for creating a simulatable model.

- A biopathway model might not be a correct gene network model

This is the case if one process has more than one input or output entities, which is normal for Petri net models of biopathways, but forbidden for gene networks. In the case of such a conflict some connection between entities will be ignored. In the Gene Net view one process is represented by one edge connecting the input and output entity of this process. For each process with multiple input or output entities, the edge connects only one input and output entity of the process.



## 4.2 Differences between Gene Net and Biopathway Modeling Modes

The differences between the Biopathway modeling mode and the Gene Net mode are listed in the table below.

	<b>Biopathway Modeling Mode</b>	<b>Gene Net Mode</b>
<i>Model</i>	<p><i>Includes 3 base elements: entity, process and connector.</i></p> <p>Models are created by the user within CI workspace.</p> <p>Biopathway models are smaller but more precisely defined than the Gene Network models.</p> <p>Biopathway models include quantitative simulation properties for studying the kinetics of biological processes.</p>	<p>Includes 2 basic elements: gene and edge.</p> <p>Gene networks are automatically generated by a special external software.</p> <p>Originally gene networks are large models without precise simulation data.</p>
<i>Visualization</i>	<p><i>More advanced graphics: images for each graph elements, polygon and curved edges for connectors, visualization and animation of simulation progress, etc.</i></p> <p><i>Visualization is slower but allows for displaying more details.</i></p>	<p>Simple graphics: default view styles for each gene and edge.</p> <p>Visualization is faster but includes less details. Thus it is more suitable for large networks.</p>
<i>Editing</i>	<p><i>Many commands enable building biopathway models from the scratch and defining the simulation properties (e.g. kinetic styles)</i></p>	<p>Extraction and Merging of subnetworks. Other edit commands are used in seldom cases.</p>
<i>Simulation</i>	<p><i>Simulation using Petri Net algorithm in</i></p>	<p>no Simulation capabilities</p>

	<i>interactive or remote mode.</i>	
<i>Analysis</i>	<i>Analysis of simulation results: charts</i>	Analysis of gene networks: Gene Mining, Pathway and Keyword Search, tools for extraction of subnets

### 4.3 Documentation for Biopathway Modeling Mode

Please refer to the **Cell Illustrator User Guide** and **Reference Manual** to learn more about the biopathway modeling and simulation capabilities of Cell Illustrator.

## 5 Importing a Gene Network

This chapter describes how a gene network file is imported into Cell Illustrator workspace. You will see how to:

- import Gene Networks from a Text File
- save the gene network to Cell Illustrator native format
- open a native file

### 5.1 Import Gene Network from a Text File

To import text files you can use File | Import | GN Text File. This function allows you to import gene networks from a normal text file with tab separated values. Each row in this text file represents an edge. The edge is described by 4 tab separated values as follows:

<parent\_gene\_name> <child\_gene\_name> <edge\_score> <relation\_type>

Please note that the text file must not contain a header row.

### 5.2 Gene Network Elements

In Cell Illustrator, the gene relationship network is represented as a graph containing nodes connected with edges. Each node represents one gene, while each edge represents a relationship between the connected genes. Genes (nodes) may possess the properties described below.

Gene	
Property	Meaning
Gene Name	The gene name as it is known in public databases. It can be either the official or alternate gene name
ProbeID	The ID for the given gene assigned during micro array experiment.
Accession	The AccessionNumber. Normally it will be the same as gene name.
Extension	Additional comments, e.g. GO annotations, etc.
Comments	Gene Comments

Edges describe relations between genes. Two genes are related if one gene influences (regulates) another gene. Each edge (relation) has a direction, which defines the parent-child hierarchy. The parent node is the regulator gene that regulates

the child node (gene). The regulator gene is also called the parent gene, while the regulated gene is called the child gene.

Edge	
Property	Meaning
Parent	Parent Gene
Child	Child Gene
Type	Relation Type, can be one of the following: Type 0: up-regulation Type 1: down-regulation Type 2: Conflict Type 3: Unknown
Edge Score	Edge Score is a statistical value that describes how strong a relation is. The higher the edge score the stronger is the relation between genes.
Comments	Process Comment

### 5.3 Save the Gene Network to CI Native Format

The CSML (Cell Systems Modeling Language) file format is the native format of Cell Illustrator. It is recommended to save each imported gene network in this file format. Working with CSML files in Cell Illustrator is faster and easier than working with text files. To save the imported text in the CSML format:

1. Do one of the following:
  - a. Choose File | Save As menu command or
  - b. Press the File Save As button on the left hand toolbar

The File Save dialog appears with the file name, which has been generated on the base of the imported text file name. The file extension of the native files is “GON”, which stands for Genomic Object Net.

2. In the File Save dialog box, enter the desired file location and name
3. Click OK to save the file.
4. Click File | Close to close the canvas.

After saving a native file it is added to the recent files list, which is available in the File | Open Recent File menu. Use this menu to open the recently saved native file. To open another native file:

1. Do one of the following:

- a. Choose File | Open menu command or
  - b. Press the File Open button on the left hand toolbar
2. In the File Open dialog box, select the file location and name
3. Click OK to open the file.

## 6 Viewing the Gene Network



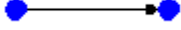

After importing the GN text file all the information from the file has been stored in the CI model. This chapter describes the most common operations to view a gene network in the CI workspace. These basic operations are useful for the analysis of large networks and for small subnetworks as well. You'll learn how to:

- show/hide element properties on the canvas
- view the gene network in a tabular view
- view detailed element properties
- optimize the network layout

### 6.1 Show/Hide Element Properties on the Canvas

The main node or edge properties, such as node name, edge score, node/edge comments can be displayed directly on the canvas as property labels around the node or edge. In Cell Illustrator you can:

- display or hide labels for all elements in the canvas with View | Show Edge Score, View | Show Name, View | Show Comments. Or
- show or hide labels for selected nodes by using the commands: Element | Show Name, Element | Show Edge Score, Element | Show Comments.
- Color edges by their relation type. For this, set a checkmark by the View | Color by Relation Type menu item. Please refer to the table below to see how edge types are colored and displayed on the canvas:

Edge Type	Description	View Style
Type 0	<i>up-regulation</i>	
Type 1	down-regulation	
Type 2	Conflict	
Type 3	Unknown	

### 6.2 View the Gene Network in a Tabular View

A gene network consists of two lists: the gene list and the relation list. You can view this information in the tabular view in the Element List dialog box:

1. Do one of the following
  - a. Choose Dialog | Element List or
  - b. Select the Element List button from the right hand toolbar.

2. Click the column header to sort the gene/edge list by the given property.
3. With the Element List dialog open click the Path tab to switch to the Edge property table

### 6.3 View Detailed Element Properties

Properties for a single node/edge can be displayed and/or edited in the Element Settings dialog box. To view/edit properties for a single node or edge:

1. Do one of the following
  - a. Double click the selected node or edge.
  - b. Select a node and then click on the Element Settings button on the right hand side toolbar.
  - c. Right click the node and select Details from the context menu
2. Select another node on the canvas without closing the dialog to view its attributes.

**Hint:** You can also view the node details as a fly-over tool tip. To do this hold the mouse pointer over the desired node/edge. A tool tip with node details will popup after a short while.

### 6.4 Optimize the Network Layout

You can improve the layout manually by dragging nodes on the canvas or automatically using the Graph Layout tool.

To move a single node:

1. Click on it
2. Drag it to the desired location without releasing the mouse button.

You can also move multiple nodes:

1. Make a rectangular selection by dragging the mouse on the canvas. This should select multiple nodes.
2. To move selected nodes, click on one of them and drag it to a desired location without releasing the mouse button.

To apply an automatic network layout:

1. Choose Dialog | Graph Layout or click the Graph Layout button on the right toolbar.

2. In the Graph Layout dialog, press the Import button to display the current network in the dialog.
3. Choose one of the layout algorithms.
4. Choose the Options button to select additional layout settings. This is an optional step.
5. Finally, click on the Export button to apply the layout to the network in the canvas

The most often used layouts are Sugiyama and CCL Sugiyama. A short description of each layout is given in the table below

<b>Layout Name</b>	<i>Description</i>
Sugiyama	<i>Nodes are organized in levels according to their parent child relation as in a company organization chart. Nodes without parents are at the top. A child node is always below all its parent nodes.</i>
CCL Sugiyama	<i>CCL stands for Child Count Layout. Nodes are organized in levels according to their child count. Nodes with the most children are at the top. Nodes with the same number of children are in one row. In this layout a child node can be above its parent.</i>
Circle	<i>Nodes are positioned on a circle</i>
Tree	<i>Requires the selection of one node inside the Graph Layout window. This node becomes the root (starting point) of the tree. The tree is built by placing connected nodes below the current node. The edge direction (parent-child or child parent) does not matter.</i>
Moen	<i>Improved tree layout</i>
Radial Tree	<i>A tree layout converted to a circle. The tree root is in the middle of the circle.</i>

Both tree layouts, called Tree and Moen, require the selection of one node inside the Graph Layout window as the starting point for the tree.



## 7 Identification of Subnetworks Regulated by a Specific Gene

Finding genes that depend on the specific regulator gene is a very important task. It can be relatively easy accomplished using the Pathway Search tool. You will see how to find and extract subnetworks of genes regulated (knocked-down) by a specific gene. In this example you will learn how to:

- filter out “weak” relations from the gene network
- identify the subnetwork regulated by a specific gene
- extract the subnetwork

To start this example:

1. Open the network comprising several hundred
2. Optimize the network layout as described in the chapter: Optimize the Network Layout (optional step)

### 7.1 Filter Out “Weak” Relations (optional step)

This optional step is recommended when you want to filter out relations with a small edge score value (called “weak” relations). This step reduces the network complexity and allows you to focus on the most important relations. To cut out edges with low edge scores:

1. Open the Element List dialog by doing one of the following:
  - a. Choose Dialog | Element List or
  - b. Press the Element List button on the right hand toolbar.
2. Select the Path tab
3. Click the Edge Score column header to sort by the edge score property in descending order. The first click will sort the edges in ascending order - from the lowest to the highest value. The second click will sort in the opposite order.
4. Use the scrollbar to locate the threshold value. E.g. if you want to cut out 50% of edges, scroll to the middle of the bar.
5. Click the edge to select it. This will be the “threshold” edge – the edge with the highest edge score that will be removed.
6. Scroll down to the bottom most edge. This edge should have the lowest edge score.

7. SHIFT+click this edge. All edges between the “threshold” edge and the bottom edge should be selected.
8. Right click (or CTRL+click on Mac) on the selection to open the context menu.
9. Select the Delete command from the context menu. The selected edges are deleted.
10. Choose File | Save or File | Save As to save your modifications.
11. Close the Element List dialog.

## 7.2 Identify a Subnetwork of a Specific Gene

The Pathway Search dialog box is a powerful tool used to identify subnetworks related to a selected gene. Using it, you can identify:

- the subnetwork of parent nodes. This will include all the genes that regulate the specified gene, or
- the subnetwork of child nodes. This will include all the genes which regulated by the specified gene, or
- the subnetwork of all related nodes. This will include both the child and parent nodes.

To find a subnetwork related to one gene:

1. Select the desired knock-down gene on the canvas
2. Do one of the following:
  - a. Choose Analyze | Pathway Search or
  - b. Press the Pathway Search button on the right hand toolbar.

The Pathway Search dialog box is opened and the selected gene appears in the Search From field.

3. If you want to select another gene, select it from the Search From drop down list. By default this list contains gene names only. However it might happen that two genes with the same name appear on the canvas. To solve the ambiguity of gene names, you can:
4. Check the Show Probe ID option to see the probe IDs. The Search From and Search To combo boxes will show Probe ID next to the gene name.
5. Uncheck the Show Probe ID to see gene names only.

6. Select the number of edges to be searched for. A recommended value is 3 to 5.
7. Set the Search Direction.

Select Forward, if you want to find the subnetwork that is regulated by the specified gene. Select Backward, if you want to find the subnetwork that regulates the specified gene. Select both, if you want to find the subnetwork that contains all related genes.

In this specific example, we want to identify only the genes which are regulated (knocked-down) by the selected gene. Therefore choose the option Forward.

8. Press the Execute button to start the search.

The list of found pathways is displayed in the Search Results dialog box. You are interested in the subnetwork that contains all the found pathways. To create this subnetwork:

9. Select the first pathway in the Search Results list. The selected pathway is added to the subnetwork. It is marked in red on the canvas.
10. Scroll down to the end of the list
11. Select the bottom most pathway Search Results list. All found pathways are added to the subnet.
12. Choose Analyze | Extract Subnet from the menu. This should open the subnet regulated by the knock-down gene in a new canvas.
13. Choose File | Save As to save the subnet.

You have seen how a subnetwork with genes regulated by a knock-down gene can be extracted. After extracting the subnetwork you are ready to go to the final step of the analysis - Evaluating a Subnetwork. Since CI offers many methods of accomplishing this, it is described in a separate section.

## 8 Evaluating a Subnetwork

Analyzing gene networks containing thousands of atoms is best accomplished by analyzing their subnetworks. In this way you can focus only on that relatively small portion of the gene network that is interesting to you. Cell Illustrator offers various methods for finding and extracting subnetworks, which are described in the other sections of this document. Having such a subnetwork, you can make the assessment of the quality of predictions encoded as edges in the subnetwork. One of the idea of evaluating a subnetwork is comparing known regulation facts with what can be concluded from the experimental subnetwork data. Another one is to apply more precise statistical analysis. This is done by:

- checking web-based annotations of individual genes
- searching for Pubmed, KEGG or MIM articles that concord with the edges – gene mining
- annotating edges

### 8.1 Checking web-based annotations of individual genes

You can query public databases for details about a selected gene in the following way:

1. Click the selection tool on the top toolbar to make sure that the selection mode is active.
2. Right click on a selected gene.
3. Choose URL Templates from the context menu. The available URLs should pop up on the right hand.
4. Select one item from the list of URLs. This should open the selected web page in your default web browser.

By default one URL template is defined. You may add other URL templates in the Cell Illustrator Preferences dialog. These templates will be stored persistently. If you want to add a new URL template:

1. Open the Preferences dialog box by doing one of the following:
  - a. Choose Dialog | Preferences or
  - a. Select the Preferences button from the right hand toolbar.
2. Select the Hyperlink Templates | Entity section in the left panel of the Preferences dialog box. You should see the list of currently defined templates on the right panel.
3. Press the Add button on the right panel.

4. Specify the new URL Template. The template should consist of the URL address e.g.  
<http://www.ncbi.nih.gov/IEB/Research/Acembly/av.cgi?db=human&l=>  
and the gene name symbol, i.e. <%%GENE\_NAME%%>.

## 8.2 Gene Mining

Searching for information and articles in public databases is a very important evaluation step. Using this method you can:

- obtain new details about individual genes
- prove that the gene network concords with facts published in scientific articles.

The Cell Illustrator Gene Mining dialog enables you search the public databases Pubmed, KEGG and MIM for articles about genes in the network. With this tool you can:

- display Pubmed, KEGG or MIM articles concerning individual genes
- open KEGG and MIM links describing the selected gene
- see article Pubmed ID
- find Pubmed, KEGG or MIM articles that include several genes.  
Finding such articles that link two or more genes is useful to evaluate edges.

To start the gene mining:

1. Open a subnetwork and activate it in the Cell Illustrator workspace

In this example we apply the Gene Mining to a small subnetwork. But it can be also applied to large networks in the same way as described below

2. Define the genes to be incorporated into the analysis.

Gene Mining searches for articles about a specified set of genes. Therefore you need to define the set of genes by making a selection on the canvas. The set of genes that will be incorporated into Gene Mining depends on the canvas selection in the following way

- a. if a gene subnetwork exists on the canvas, then all genes from this subnetwork are loaded onto the gene list
- b. Otherwise, if the normal selection is not empty then the genes from the normal selection are loaded to the gene list
- c. Otherwise all genes are loaded to the list

By default it is recommended the user should perform the gene mining on all genes. For this you need to make sure that nothing is selected on the canvas. For an advanced analysis, if you want to narrow the analysis to a specific subset of genes, create a subnetwork containing these genes.

3. Open the Gene Mining dialog box by doing one of the following:
  - a. Choose Analyze | Gene Mining or
  - b. Select the Gene Mining button from the right hand toolbar.

You should see the Processing Database dialog after which the Gene Mining dialog is displayed with the search results. On the left hand you will see the list of genes that have been used in the article search. Initially the first gene in the list is selected and detailed information is displayed in the Gene Info tab.

4. Select another gene in the Gene List.

The Gene Info tab is updated with details of the selected gene.

5. View the information about the selected gene.

The information about each gene is divided into 6 panels described below:

- a. Gene Info panel

Review the main information about the gene: gene name, official name, aliases and number of PubMed articles about this gene. Get the gene ontology information about gene function, process and component. If a MIM or KEGG Pathway information is present for the selected gene, you may click on the hyperlink to open it in the web browser.

- b. Gene Pubmed panel.

In this panel you see the Pubmed article list. Each article includes the following details: authors, title, journal and publication date.

- c. View the Link Summary panel.

For a selected gene, this tab lists all related genes in the analyzed set as well as the number of PubMed, KEGG and MIM identifiers which the gene has in common with other genes in the analyze set.

- d. View the PubMed Links panel.

For a selected gene, this tab lists Entrez PubMed entries that link this gene to one or more genes in the analyzed set. The list of related genes is displayed next to each link.

- e. View the KEGG Link panel.

For a selected gene, this tab lists KEGG pathways that link this gene to one or more genes in the analyzed set. The list of related genes is displayed next to each link. Click on the http link to access the WWW site with information on a particular MIM identifier.

- f. View the MIM Links panel.

For a selected gene, this tab lists the MIM identifiers that link this gene to one or more genes in the analyzed set. The list of related genes is displayed next to each link. Click on the http link to access the WWW site with information on a particular MIM identifier.

Note: You may use the shortcut keys Tab and SHIFT+Tab to navigate between the panels. The Tab key switches you one panel forward (to the right), while SHIFT+Tab one panel backward (to the left).

6. Repeat steps 4 and 5 for other genes.



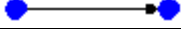

If you want to change the set of genes which are to be analyzed, do the following:

7. Define a new subnet on the canvas.
8. Press the Refresh button in the Gene Mining dialog

This restarts the gene mining analysis using the set of genes currently selected in the active canvas window as the input list. In this way, you can switch back and forth between canvas and Gene Mining dialog while decreasing or expanding the set of genes of interest.

### 8.3 Edge Attributes

Edges represent relations between genes. The edge score, describes how “strong” or “weak” a relation is. The edge type describes the relation and is visualized on the canvas as illustrated below.

Edge Type	Description	View Style
Type 0	<i>up-regulation</i>	
Type 1	down-regulation	
Type 2	Conflict	
Type 3	Unknown	

To see the edge types on the canvas use the View | Color by Edge type menu command. To view/edit the edge score and type use the standard methods described in the Viewing the Gene Network section.

### 8.4 Annotating Edges

The results of subnetwork investigation can be saved as edge comments in the following way:

1. Select the edge to be annotated on the canvas
2. Do one of the following:
  - a. Choose Element | Comments from the menu

- b. Right click on the edge and select Comments from the context menu
3. Enter the comment in the Comments dialog box.
4. Press OK to close the dialog.
5. Save the file.

**Hint:** You can view the edge comments as a fly-over tool tip. To do this hold the mouse pointer over the desired edge. A tool tip with edge details will popup after a short while.



## 9 Identification of the Regulators of a Target Transcript

With Cell Illustrator you can identify genes and pathways that regulate a potential therapeutic target. This can be achieved using the Pathway Search tool in the following way:

1. Open the network comprising several hundred nodes
2. Optimize the network layout as described in the chapter: Optimize the Network Layout (optional step)
3. Filter out “weak” relations as described in the chapter: Filter Out “Weak” Relations (optional step)
4. Identify a subnetwork of regulators of the target transcript.

Use the Pathway Search dialog to identify a subnetwork that regulates the target gene by executing a search of 3 to 5 steps BACKWARD for the target transcript. For details on how to perform pathway searches see the chapter: Identify a Subnetwork of a Specific Gene

As the result you should get the subnetwork with “regulators” (or “parents”) of the selected target gene. After extracting the subnetwork you are ready to go to the final step of the analysis - see Evaluating a Subnetwork.

## 10 Identification of Regulatory Hubs

Hubs are genes that regulate many other genes. In the graph hubs are nodes connected by many edges to child genes. The idea of finding hubs is:

- to identify transcripts that directly regulate many other transcripts
- to evaluate the network by comparing hubs discovered in the network against known data
- to identify transcripts that regulate the hubs

The steps below can be repeated to investigate more than one hub in a network. To find a hub:

1. Open a network comprising several hundred nodes
2. Optimize the network layout as described in the chapter Optimize the Network Layout (optional step)
3. Open the Element List dialog by doing one of the following:
  - a. Choose Dialog | Element List or
  - b. Select the Element List button from the right hand toolbar.
4. Select the Gene tab
5. Click the Children column header to sort by the child count property in descending order. The first click will sort the edges in ascending order - from the lowest to the highest value. The second click in the opposite order.
6. Select one of the genes with the highest child count

The hub gene is selected on the canvas. Now you are ready to extract the subnetwork consisting of children and parents of the hub using the Pathway Search dialog.

7. For the selected gene (hub) open the Pathway Search dialog box and execute a pathway search with 1 step in BOTH directions. See the chapter Identify a Subnetwork of a Specific Gene for instructions on how this can be done. As the result you should get a subnetwork with all genes that are related to the hub gene.
8. Improve the network layout to place the hub in between its parents and children

Open the Graph Layout dialog and import the subnetwork. Apply the Sugiyama layout and export the new layout back to the canvas. For more details – see chapter Optimize the Network Layout.

9. Evaluate the subnetwork - see section Evaluating a Subnetwork

## **11 Using a Transcript Known to Play a Role in a Biological Process to Discover the Other Transcripts Involved in this Process**

Let us assume you know a gene that plays a role in a biological process. Having such information and a gene network, you can check whether this gene network is in agreement with the known fact. If the evaluation is successful, you can find other transcripts that are involved in this biological process.

Discovering of other transcripts that are involved in a known process is completed by 3 main steps:

- extract the subnetwork around the known gene
- compare the extracted subnetwork with what is known about the biological process
- identify transcripts that could be involved in the biological process

As a result you should get another gene that can be further investigated by making a new experiment in the lab or by extracting and analyzing a network around it.

### **11.1 Extract the Subnetwork around the Known Gene**

To extract a subnetwork around a gene known to play a role in the process, execute pathway search with typically **3 steps** in BOTH directions – see chapter Identify a Subnetwork of a Specific Gene. Depending on the number of found items select the whole list in the Pathway Search Results or a manageable portion of it. Use the SHIFT+click to select a range or CTRL+click (Apple+click on Mac) to select/deselect single rows. Create and save a sub-network.

### **11.2 Compare the Extracted Subnetwork with Known Facts**

To compare the extracted subnetwork with what is known about the biological process use the standard procedure described in the section: Evaluating a Subnetwork.

### **11.3 Identify Transcripts that Could Be Involved in the Biological Process**

During the evaluation step, you will be able to identify other transcripts that may influence (regulate) the biological process of interest. A transcript identified in this way can be further investigated by performing a new experiment in the lab or by extracting and analyzing a network around it.

## 12 Discovering Gene Transcripts Involved in a Biological Process Using Keyword Search

The gene network file may include additional annotations added automatically by a program or manually by the user. These annotations can be very useful to divide genes into groups or categories. E.g. in the Entrez Pubmed XML files each gene is described by gene ontology information which assigns the gene to one or more functions, processes and components.

The goal of the Keyword Search is to evaluate the network by comparing an extracted sub-network with known facts about the biological process. This can be achieved with Cell Illustrator using the Keyword Search dialog to extract a sub-network composed of genes annotated with specific information related to a biological process.

To extract a subnetwork annotated with specific information:

1. Open a network comprising several hundred nodes
2. Open the Keyword Search dialog by doing one of the following:
  - a. Choose Analyze | Keyword Search from the menu
  - b. Click on the Keyword Search button on the right hand side toolbar
3. In the Keyword Search dialog box, type in a keyword in the Search For entry.
4. Specify the advanced search options Case Sensitive or Regular Expression. This is an optional step. By default you can leave both fields unchecked.
5. Choose the options in the Search In panel.

The Search In panel allows you to specify the search area. Select those gene/edge properties that should be taken into account during the search. The unchecked properties will be ignored in the search. To have a quick preview of what is stored in the gene/edge properties open the Element List dialog box as described in the chapter: View the Gene Network in a Tabular View.

6. Press the Find button to start the search

The search results appear in the form of a gene/edge list in the same dialog box.

7. Inspect the search results.

Depending on the outcome:

- a. repeat the search from the beginning with a different keyword
- b. add the search results to the current subnet – see below

8. Select several genes/edges in the Search Results list.

Use the standard SHIFT+click or CTRL+click (Apple+click on Mac) operations. Corresponding genes/edges are selected on the canvas.

9. Right click (or CTRL+click on Mac) on the selection to open the context menu
10. Choose Add to Subnet from the context menu to create a subnet with the desired genes/edges

You may extend the found subnet. If you want to do this start the search from the beginning with a different keyword. Alternatively, you may narrow down the search. If you want to do this, extract the subnetwork and start the Keyword Search again.

## 13 Gene-to-Gene Path Search

Let us assume you know a regulator gene and a target gene. In such cases you may perform a gene-to-gene search to find a therapeutic target in the path between the regulator transcript and the target transcript.

To perform a path-to-path search:

1. Open a gene network
2. Use the Pathway Search dialog to extract a sub-network composed of genes on a path between two genes A and B: In the Pathway Search dialog, specify the two gene names:
  - a. Select the regulator gene in the Search From entry
  - b. Select the target gene in the Search To entry

Execute 3-5 steps search in the FORWARD direction.

3. Highlight the edge list in the Search Results dialog to make a selection in the network. Create and save sub-network
4. Evaluate the sub-network, i.e., make the assessment of the quality of predictions and identify the regulator genes transcripts – see the section Evaluating a Subnetwork.