OCSANA_Plus Documentation

Release 0.0.1

Lauren Marazzi

Jan 15, 2021

OCSANA+ is a Cytoscape app for identifying driver nodes that control non-linear systems' long-term dynamics, prioritizing combinations of interventions in large scale complex networks, and estimating the effects of node perturbations in signaling networks, all based on the analysis of the network's structure. OCSANA+ includes an update to the previously introduced OCSANA (optimal combinations of interventions from network analysis) software tool with cutting-edge and rigorously tested algorithms, together with recently-developed structure-based control algorithms for non-linear systems and an algorithm for estimating the signal flow. All these algorithms are based on the network's topology.

OCSANA+ features:

Optimal Combinations of Interventions OCSANA (Optimal Combinations of Interventions from Network Analysis), originally introduced in Vera-Licona et al., 2013, identifies and prioritizes optimal minimal combinations of interventions (CIs) that disrupt the elementary paths from selected source nodes to the specified target nodes. When indicated by the user, OCSANA seeks to additionally minimize the side-effects that CIs can cause on specified off-target nodes.

Feedback Vertex Set Control Feedback Vertex Set Control (FC) is an attractor based control method specifically developed for networks with non-linear dynamics that uses a component of the network's topology, the Feedback Vertex Set (FVS) (Mochizuki et al., 2013)The FVS of a network is defined as the minimal set of nodes whose removal would leave a graph without cycles. Later, an extended FC control version adding the the network's source nodes was proposed by Zanudo et al., 2017.

Signal Flow Analysis The SFA algorithm (Lee and Cho, 2018) estimates the steady state activity of a node by a linear difference equation that considers (i) The activity of a node at the previous time step, (ii) The effect (activating or inhibiting) and influence of incoming edges to a node, and (iii) the initial activities of the node

In this manual we include instructions for intallation, CI discovery, FC identification, SFA analysis, and a walkthrough tutorial of the features of OCSANA+.

1.1 Installing OCSANA+

The OCSANA+ jar file can be downloaded from the OCSANA+ github repository

To install apps within Cytoscape, go to the menu bar and choose **Apps** → **App Manager...**. At the top of the **App Manager** window, make sure you have the **Install Apps** tab selected.
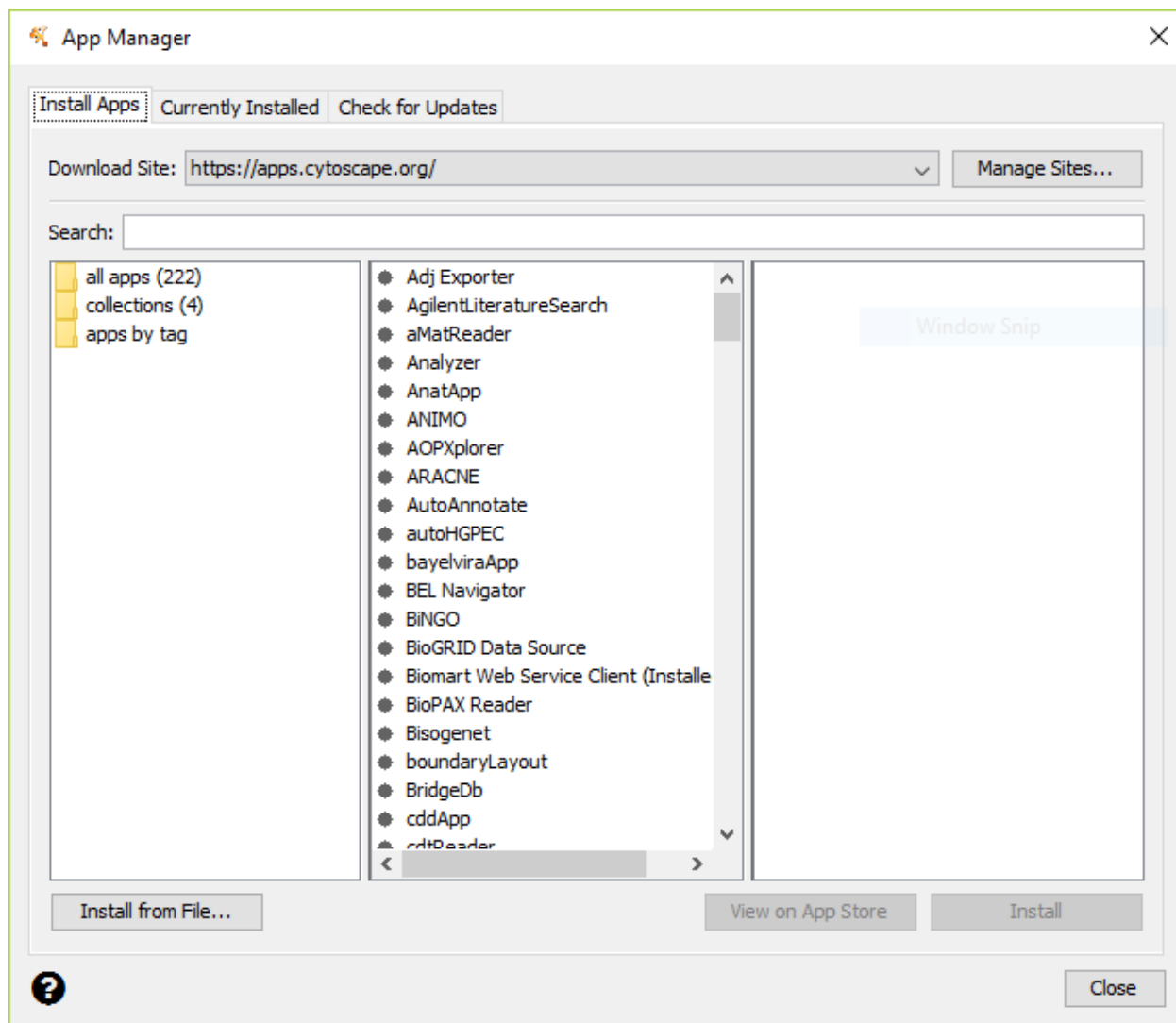
You can install OCSANA+ by clicking the "Install from File" button on the bottom-left, and navigating to the downloaded folder.

1.2 Computing Optimal Combinations of Intervention with OCSANA

OCSANA (Optimal Combinations of Interventions from Network Analysis), originally introduced in Vera-Licona et al., 2013, identifies and prioritizes optimal minimal combinations of interventions (CIs) that disrupt the elementary paths from selected source nodes to the specified target nodes. When indicated by the user, OCSANA seeks to additionally minimize the side-effects that CIs can cause on specified off-target nodes.

1.2.1 Running OCSANA in Cytoscape

After installation of OCSANA+ and loading of a network, the OCSANA configuration menu will appear as a tab in the Cytoscape Control Panel (left panel).
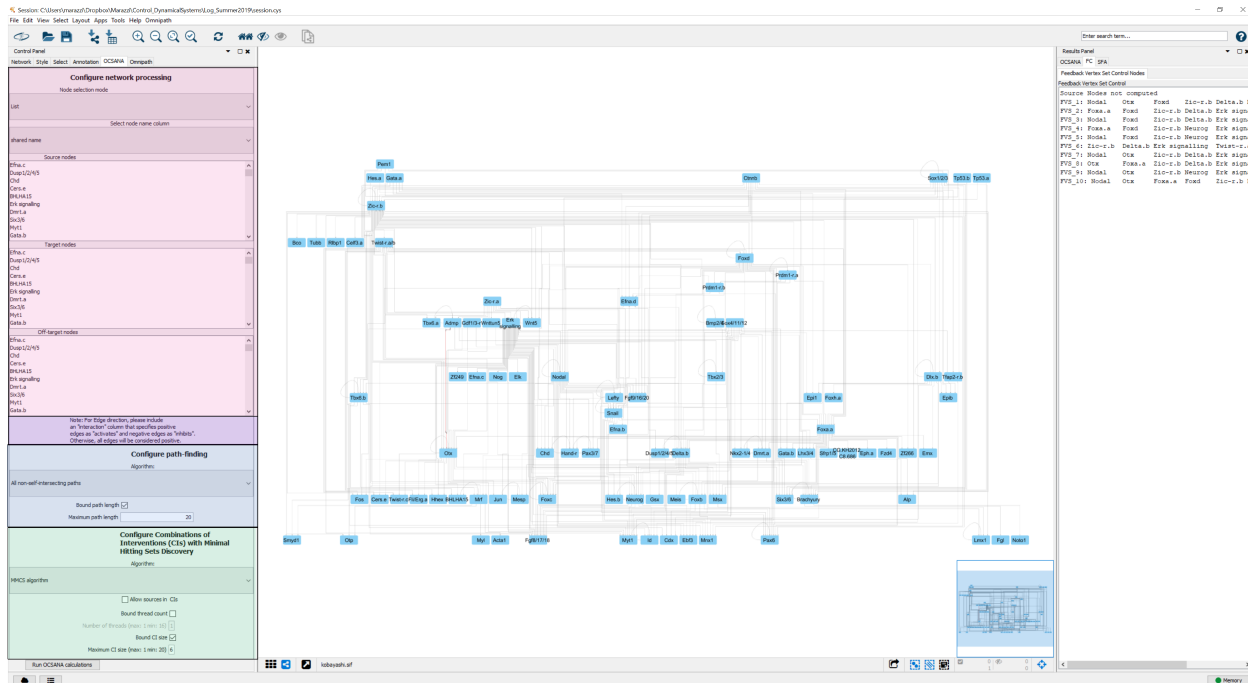
Configuring your OCSANA Run

Configure Network Processing (pink panel)

This subpanel allows users to select how OCSANA processes your network (which columns to read for node names) and which nodes to build paths from (source nodes to targets, source nodes to off-targets).

Node Selection Mode: Select whether you will enter a string of source, target, and off target nodes, or will click-to-select the nodes.

Select Node Name Column: This option allows users to select which column of the node attribute table will be used to identify network nodes.

Source Nodes: The Source nodes section allows the user to specify the source nodes (i.e. the nodes from which the signals' cascades will start). The nodes can be selected manually from the list of nodes appearing in this section or by copying and pasting the list of desired nodes in the dialog window after clicking on the set source nodes button.

Target Nodes: The Target nodes section allows the user to select the target nodes (i.e. the nodes intended to be blocked). The nodes can be selected manually from the list of nodes appearing in this section or by copying and pasting a list of nodes in the dialog window after clicking on the set of target nodes button.

Off-target Nodes: The off-target nodes section is an optional column for the user that allows to select offtarget (side-effect) nodes, that is, nodes that are preferred to be avoided when constructing CIs. A CI that contains a node from such given list, will then receive a penalty. The nodes can be selected manually from the list of nodes appearing in this column or by copying and pasting the list of nodes in the dialog window after clicking on the set side-effects nodes button.

Configuring Edge Sign (purple panel)

To configure edge signs, please include an "interaction" column in your network table that specifies positive edges as "activates" and negative edges as "inhibits." Otherwise, all edges will be considered positive.

Configure Path Finding (blue panel)

This panel allows users to select which algorithm should be used for finding paths between source nodes and target/off-target nodes.

Shortest Paths: searches for the paths with the shortest length among all the elementary paths connecting source nodes to target nodes and source nodes to side-effect nodes (via Dijkstra's algorithm).

All Non-Self-intersecting paths: considers all the paths, between specified nodes, that do not contain loops (self-intersections).

Configure Combinations of Interventions (CIs) with Minimal Hitting Sets Discovery (green panel)

This panel allows you to choose which MHS solving algorithm you would like to use to discover CIs. A CI is as a set of nodes such that each elementary path (a path from user-selected source to target nodes) contains at least one node from this set. This CI set indicates the nodes to be intervened to disrupt all the identified elementary paths. Optimality of CIs is defined in terms of a heuristic scoring. The scoring of a node is based on (i) the lengths of the paths from the node of interest to the targets, (ii) the type of effect on target nodes (e.g. activation/inhibition effect), (iii) side effects with respect to off-target nodes, (iv) the number of elementary paths in which the node participates and (v) the number of targets that such node can reach simultaneously.

The algorithms included are:

- Minimal-to-Maximal Conversion Search (MMCS)
- Reverse Search (RS)
- Berge's algorithm
- Greedy Heuristic Algorithm

If you are interested in descriptions of these algorithms, please see the *OCSANA+ supplementary file* or Gainer-Dewar and Vera-Licona, 2017.

Allow sources in CIs: Select this option if you would like source nodes to be considered in CIs

Bound thread count: Select this option if you would like to bind the number of threads used to CI Discovery if selected, enter the number of threads to be used in the below box (min:1,max:16)

Bound CI size: Select this option if you would like to bind the number of nodes that can be in a CI if selected, enter the max CI size in the below box (min:1,max:20)

Additional configuration option of Greedy Heuristic Algorithm

For the Greedy Heuristic Algorithm, you may also set a bound on the number of candidates (the number of potential CIs). If selected, enter the number below. Note this is converted to millions (if you input "1", OCSANA will generate 1,000,000 candidates)

After configuring your OCSANA Run, click 'Run OCSANA Calculations'

1.2.2 OCSANA Results

After the OCSANA run has completed, the results will be displayed in the OCSANA subpanel of the Cytoscape Results Panel (right panel)

CIs

The CIs results tab displays the discovered CIs for your OCSANA run. The information produced includes:

- **CI**: an optimal Combination of Intervention

- **Size**: the number of nodes in the CI

- **OCSANA Score**: The calculated OCSANA score for the CI

- **Target Score**: The CI score for targets (component of OCSANA Score)

- **Side Effect Score**: The CI score for off-targets (component of OCSANA Score), if off-target nodes were included in the search parameters

Paths to Targets

This tab of the OCSANA output panel displays the paths from source nodes to each target node. The paths displayed will be in accordance with your run configuration and path algorithm selection.

Saving Results

By clicking 'Save Report', you can save a text file that contains both your OCSANA configuration settings and results from your OCSANA run.

1.3 Identifying Minimal Feedback Vertex Set Control (FC) Sets

Feedback Vertex Set Control (FC) is an attractor based control method specifically developed for networks with non-linear dynamics that uses a component of the network's topology, the Feedback Vertex Set (FVS) (Mochizuki et al., 2013)

1.3.1 Running FC in Cytoscape

After installation of OCSANA+ and loading of a network, you can select Feedback Vertex Set Control from the main toolbar>Apps>OCSANA
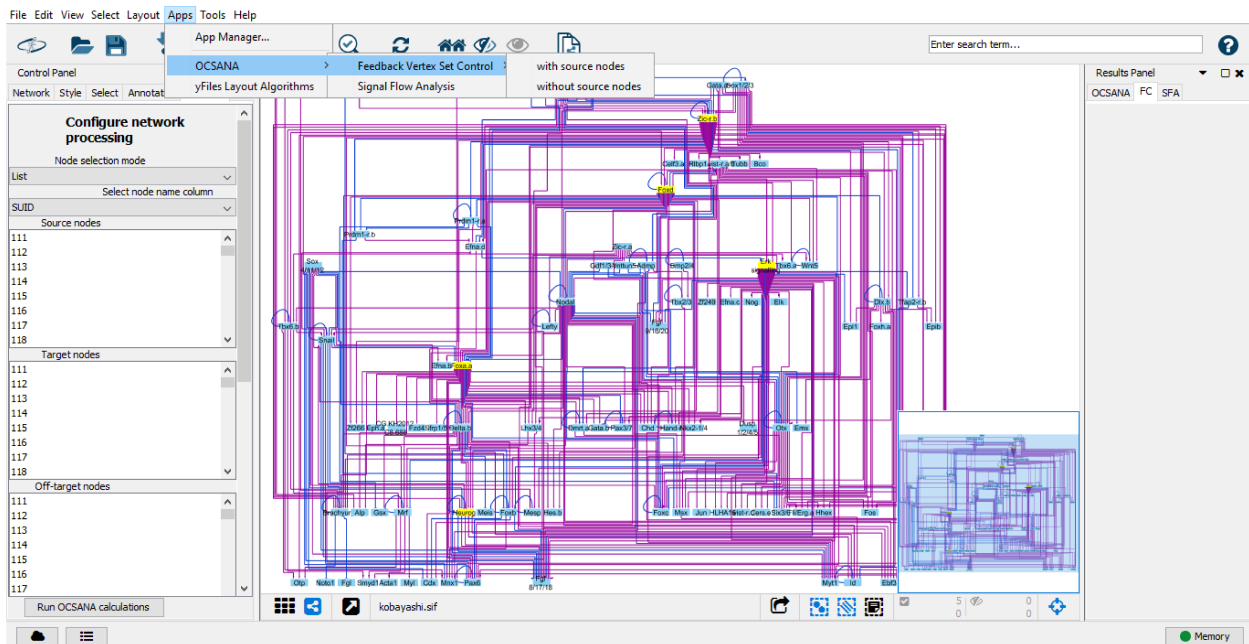
Configuring your FC Run

Two FC algorithms are implemented in OCSANA+. Mochizuki et al., 2013 mathematically identified the FC set to be comprised of the network's minimal FVS. The FVS of a network is defined as the minimal set of nodes whose removal would leave a graph without cycles. Later, an extended FC control version adding the the network's source nodes was proposed by Zanudo et al., 2017.

You can choose either version to run in OCSANA+.

After selecting an FC option, a new pop-up menu will appear.
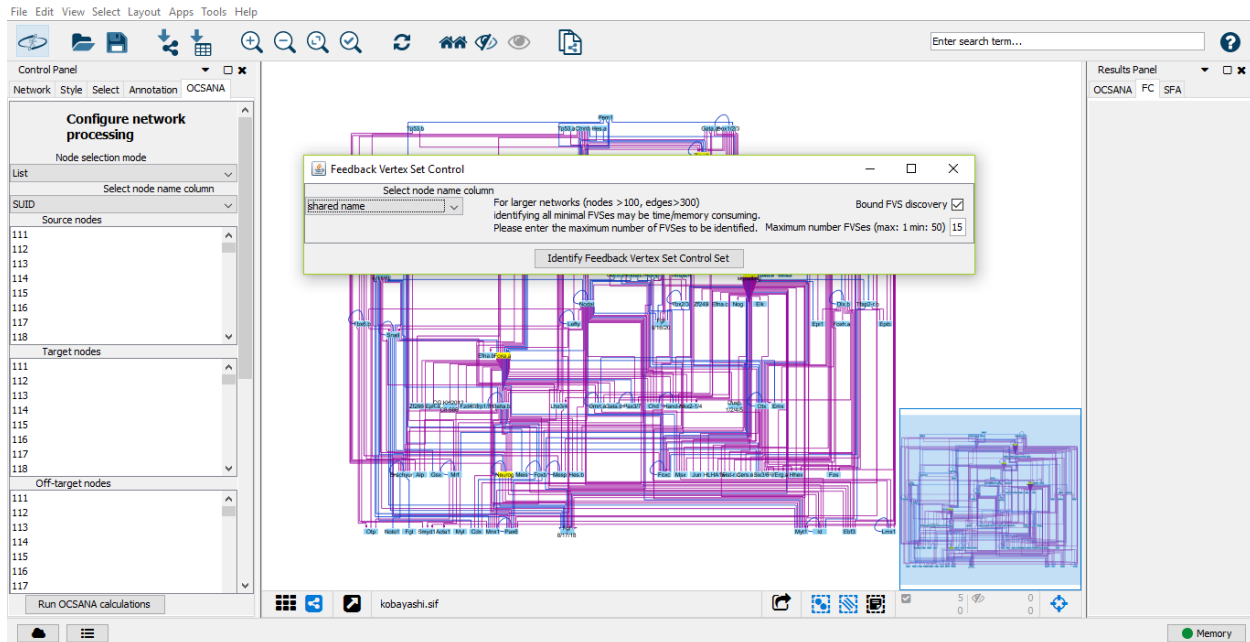
Select Node Name Column

This option allows users to select which column of the node attribute table will be used to identify network nodes.

Bound FVS discovery

For larger, more complex networks (those with many cycles), FVS discovery can be time/memory consuming. We suggest limiting the number of FVSs runs, with a maximum of 50. Please note that the number of FVSes returned may be less than the set boundary, as there may be fewer FVSes in the network.

After configuring your run, select "Identify Feedback Vertex Set Control Set"

1.3.2 FC Results

After the OCSANA run has completed, the results will be displayed in the FC subpanel of the Cytoscape Results Panel (right panel).
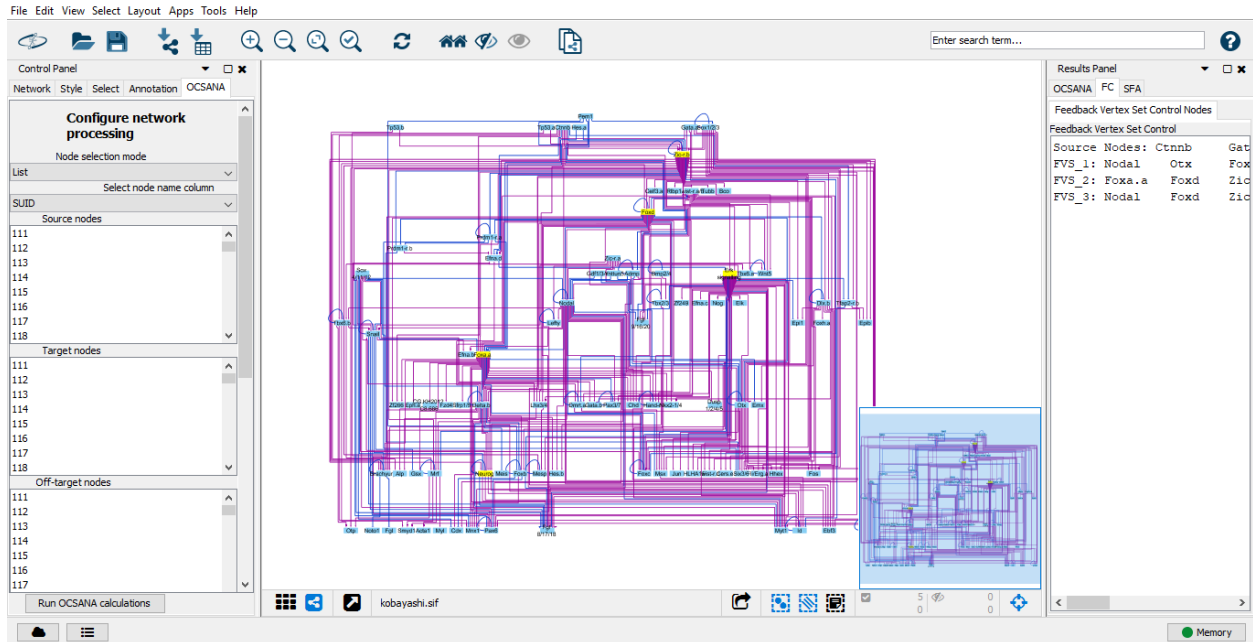
The results will include:

- Source Nodes: either the list of source nodes in the network, or "Source nodes not computed" depending on your run configuration

- FVS_i: results for each discovered FVS.

1.4 Performing Topological Estimation of Signal Flow with Signal Flow Analysis (SFA)

The SFA algorithm (Lee and Cho, 2018) estimates the steady state activity of a node by a linear difference equation that considers:

1. The activity of a node at the previous time step

2. The effect (activating or inhibiting) and influence of incoming edges to a node

3. The initial activities of the node

1.4.1 Running SFA in Cytoscape

After installation of OCSANA+ and loading of a network, you can select Signal Flow Analysis from the main toolbar>Apps>OCSANA
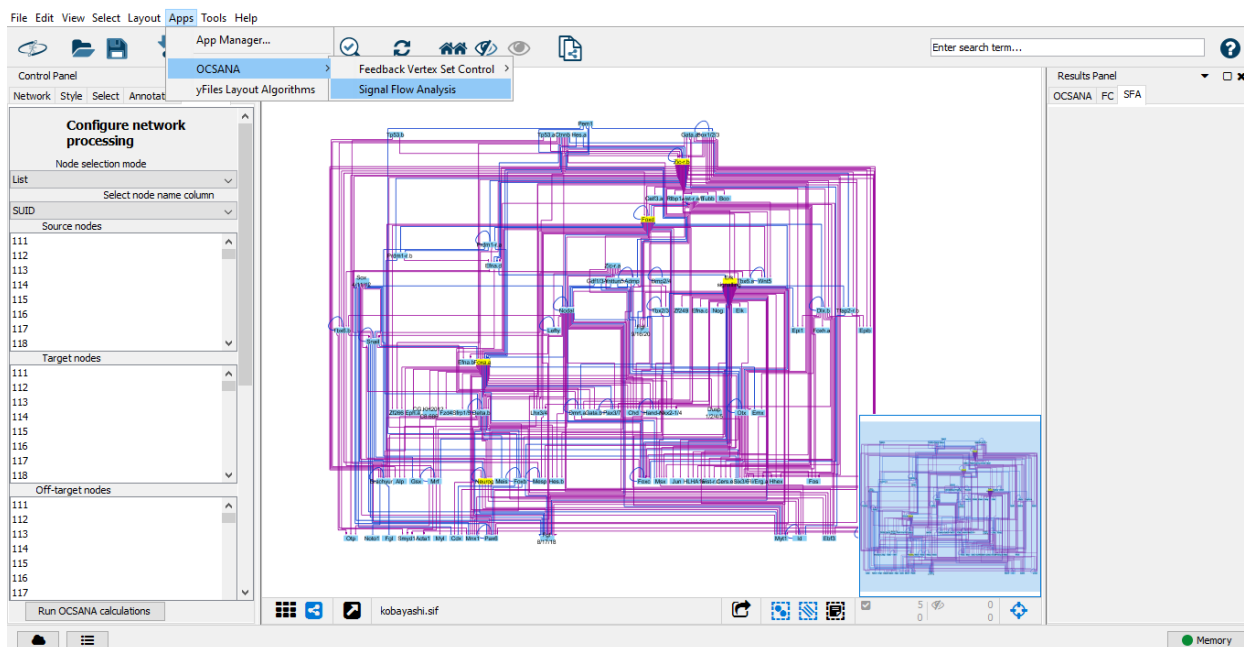
Configuring your SFA Run

After selecting an SFA option, a new pop-up menu will appear.
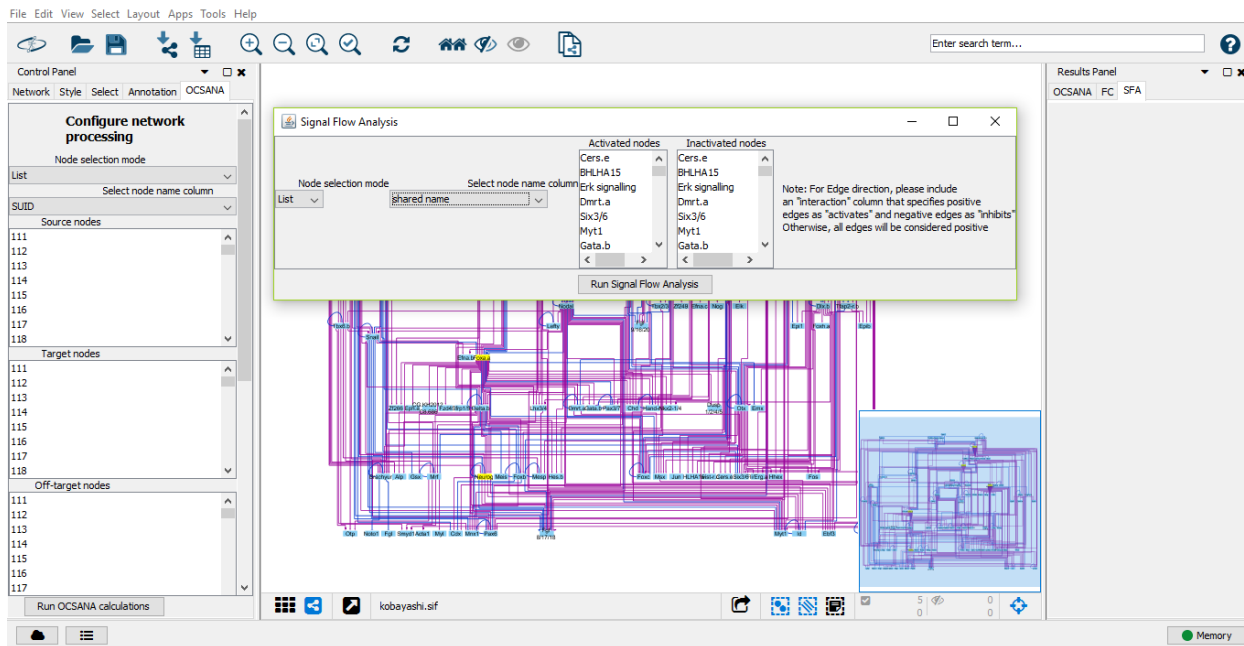
Node Selection Mode

Select whether you will enter a string of activated and inhibited nodes, or will click-to-select the nodes.

Select Node Name Column

This option allows you to select which column of the node attribute table will be used to identify network nodes.

Activated Nodes

The Activated nodes section allows you to specify which nodes will have a postive basal activity (activated state). The nodes can be selected manually from the list of nodes appearing in this section or by copying and pasting the list of desired nodes in the dialog window after clicking on the set source nodes button.

Inhibited Nodes

The Inhibited nodes section allows you to specify which nodes will have a negative basal activity (inhibited state). The nodes can be selected manually from the list of nodes appearing in this section or by copying and pasting a list of nodes in the dialog window after clicking on the set of target nodes button.

Edge Sign Configuration

To configure edge signs, please include an "interaction" column in the network table that specifies positive edges as "activates" and negative edges as "inhibits." Otherwise, all edges will be considered positive.

After you have configured your run, click "Run Signal Flow Analysis"

1.4.2 SFA Results

After the SFA run has completed, the results will be displayed in the SFA subpanel of the Cytoscape Results Panel (right panel)

SFA Config

This subpanel displays the configuration set for this run (activated nodes and inhibited nodes)
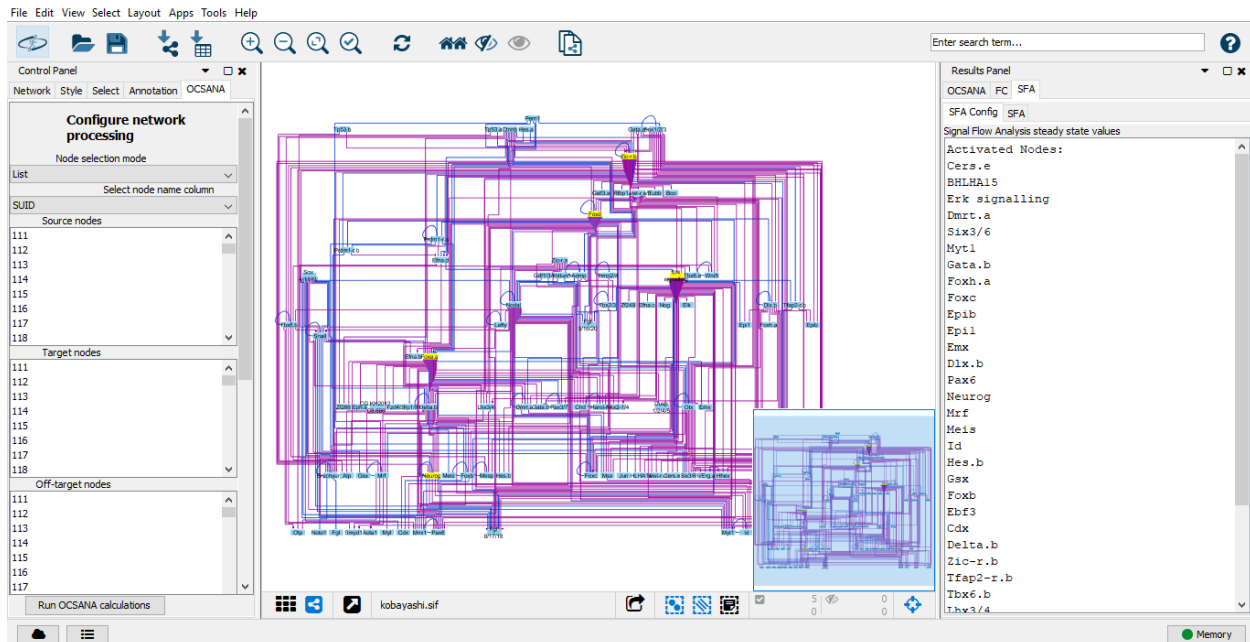
SFA

This subpanel contains the steady state log values for all network nodes.
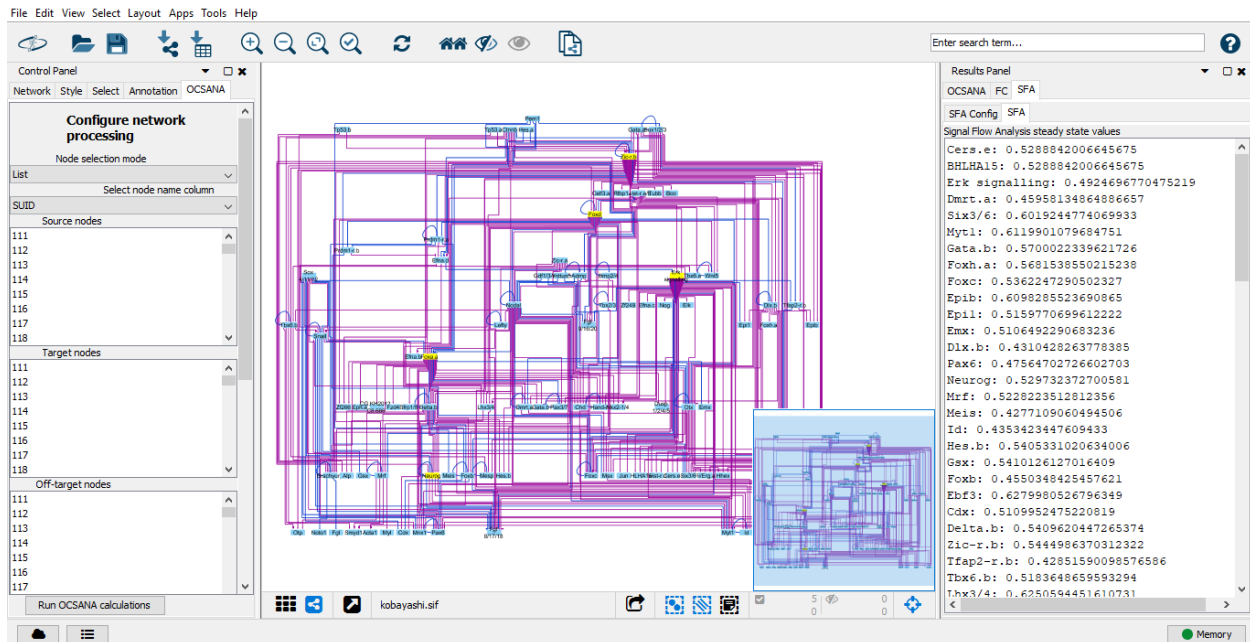
1.5 Walk-Through Example: Cell Fate Specification in Ascidian Embryo

In Kobayashi et al., 2018, the authors performed FC analysis from Mochizuki et al., 2013 on a gene regulatory network (GRN) of seven tissues that specifies cell fates in embryos of the ascidian Ciona intestinalis (type A; also called Ciona

robusta). After performing FC analysis, the authors identified five key molecules. By controlling the activities of these key molecules, the specific gene expression of six of seven tissues observed in the embryo was successfully reproduced.

In this example, we show how OCSANA+ can reproduce the FC analysis in addition to the results from control experiments in Ascidian embryos. Additionally, we explore how the extended Feedback Vertex Set Control (FC) with source nodes (Zanudo et al., 2017) and OCSANA functionalities could increase the accuracy of simulated results.

For the full results of this example, please see **'OCSANA+supplementary file'_**

You can download the .sif file used in this tutorial here

1.5.1 Step 1: Initial Cytoscape Setup

Upon launching Cytoscape, a new session will be opened. You can load the ascidian embryo network by selecting "Import Network from file System" icon in the toolbar
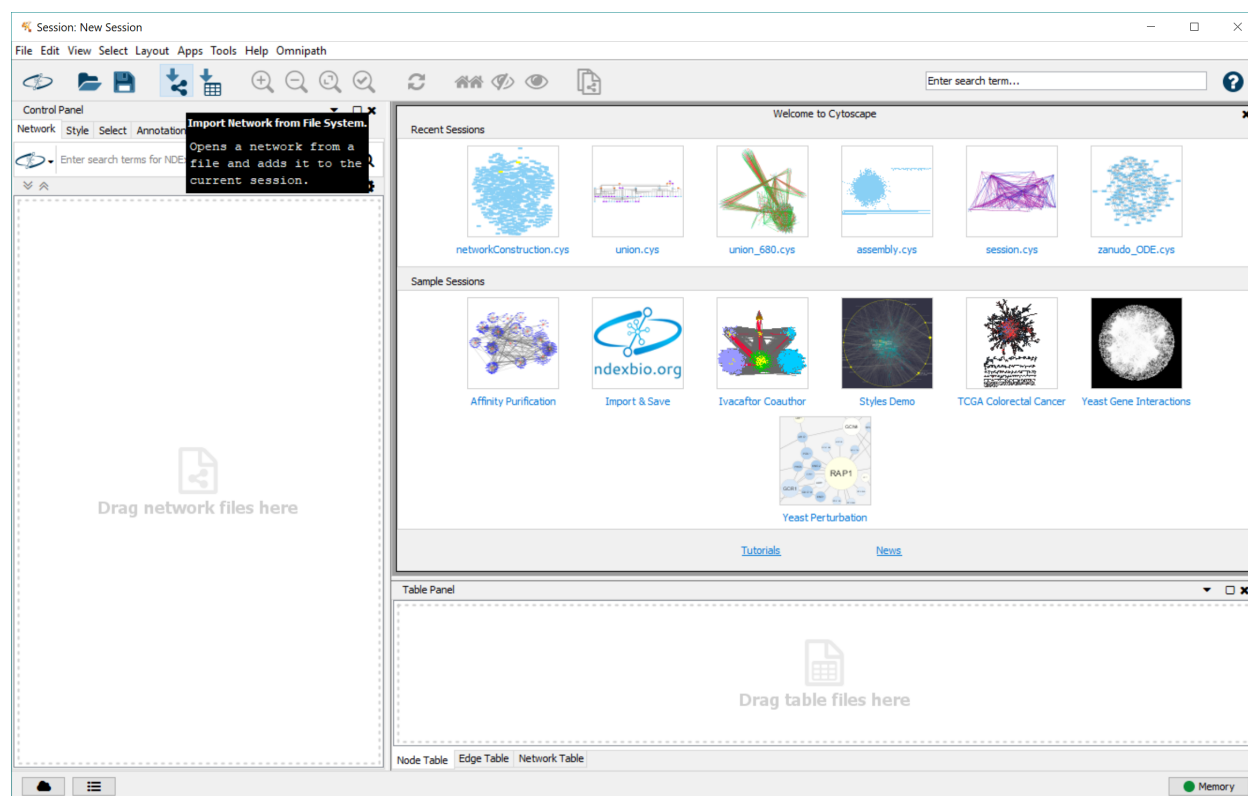
In the pop-up menu, navigate to the folder where you stored the network, and select the file.
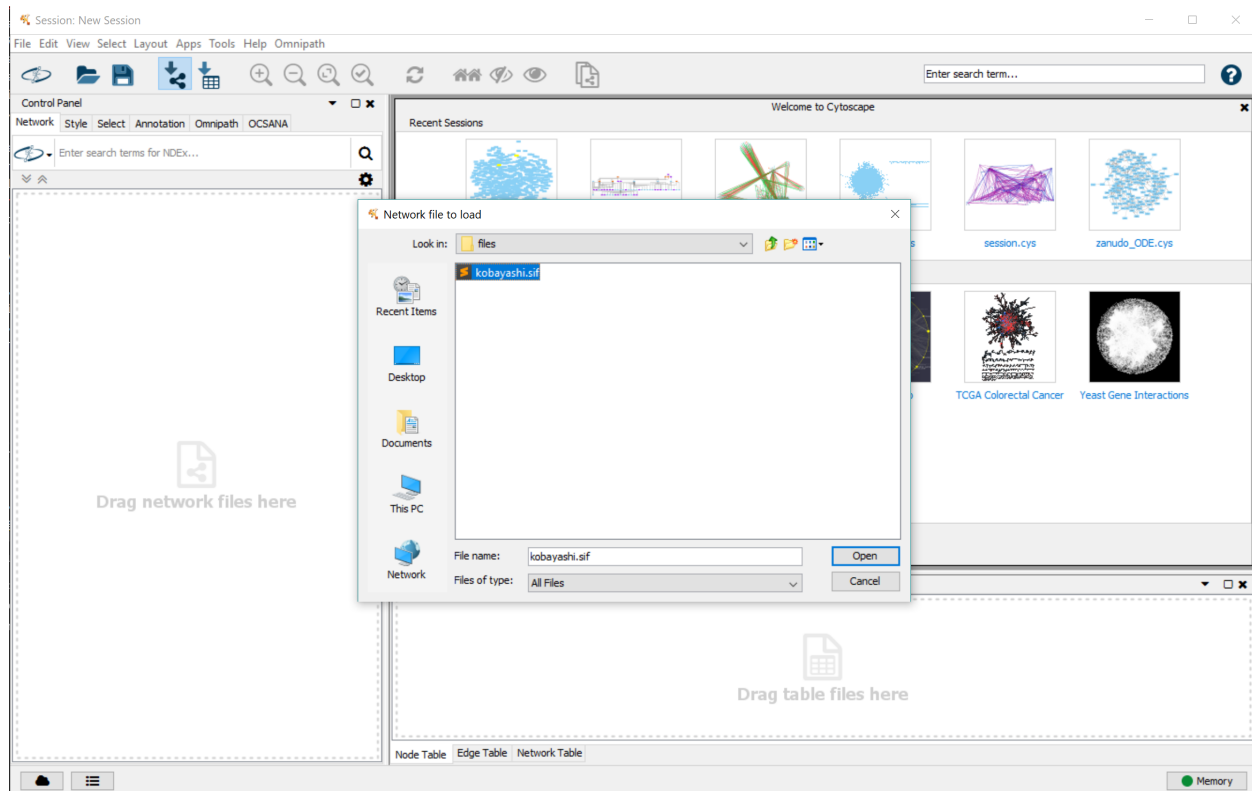
1.5.2 Step 2: Identifying Feedback Vertex Set Control Sets (FC)

To first reproduce the results of Kobayashi et al., 2018, we perform Feedback Vertex Set Control (FC) Discovery.

From the Apps dropdown in the toolbar navigate to "OCSANA>Feedback Vertex Set Control. OCSANA+ provides two algorithms for identifying FC sets: with or without source nodes (You can read more about these two algorithms in <'OCSANA+ paper supplement'_, Mochizuki et al., 2013, or Zanudo et al., 2017.).

To replicate the results of Kobayashi et al., 2018, select "without source nodes"
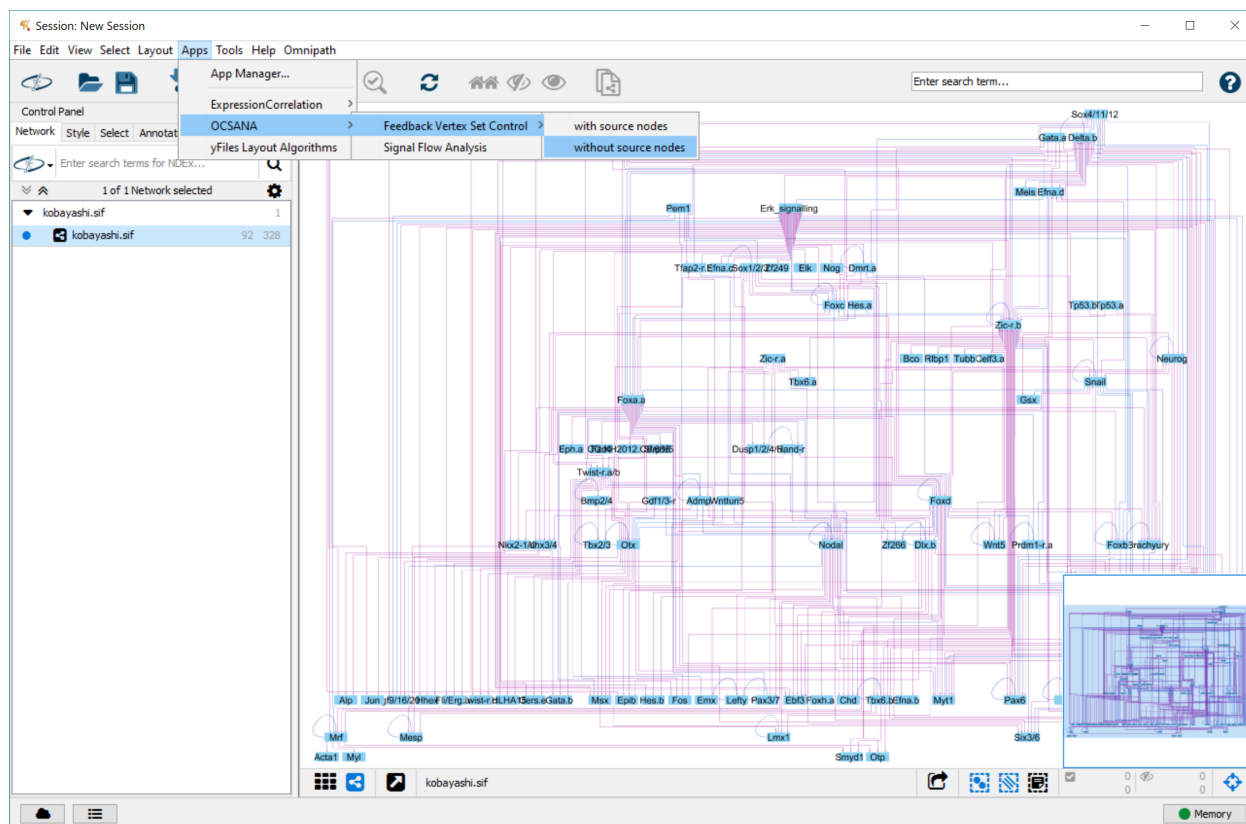
In the pop up menu:

1. Set "select node name column" to "shared name"

2. For this tutorial example, we will do an unbound FVS Discovery search. This will run the FVS algorithm 50 times to identify FVSes. Note that there may not be 50 FVSes in the network, so the number of FVSes returned may be fewer than 50.
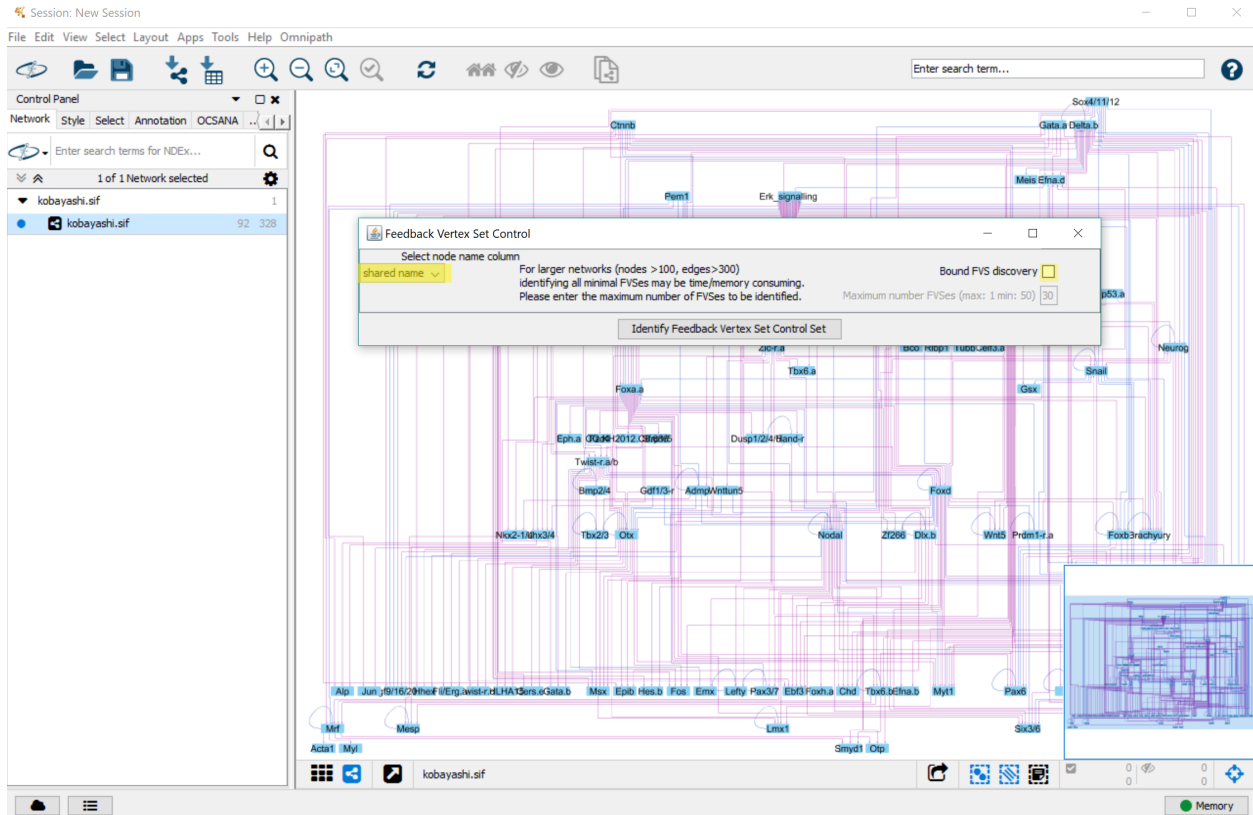
Select "Identify Feedback Vertex Set Control Set"
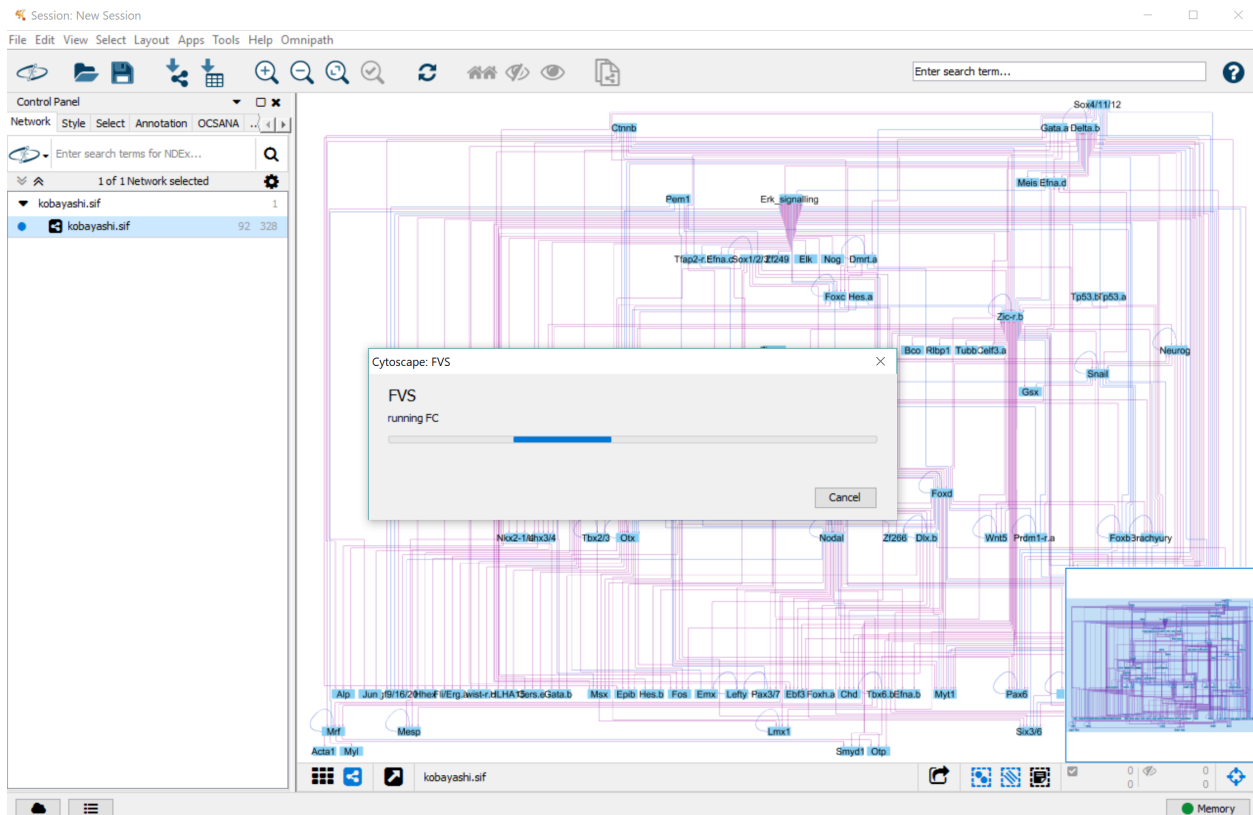
The results will be displayed in the FC subpanel of the Results Panel. Minimal FC sets are highlighted
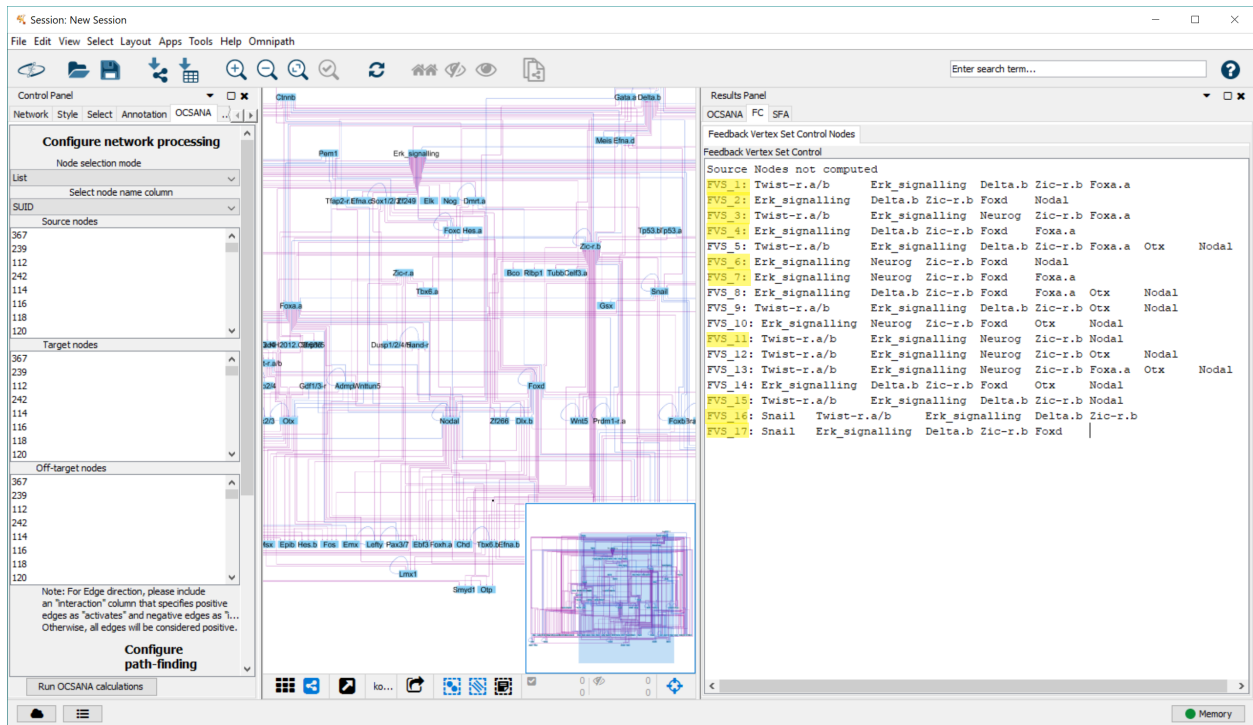
1.5.3 Step 3: Simulating FC node perturbations using SFA

We will focus on the FC set used in Kobayashi et al., 2018 in vivo perturbations for our simulation with SFA. The FC set consists of the following 5 nodes:

> Foxa.a, Foxd, Neurog, Zic-r.b, and Erk signaling

To simulate the experimental results in Kobayashi et al., 2018, we will use Signal Flow Analysis (SFA) to estimate the log steady state values of network nodes (Lee and Cho, 2018)

First, we simulate the "unperturbed" state (no up-regulation or down-regulation of FC nodes), by activation of Gata.a and Zic-r.a which are noted to initiate the zygotic developmental program.

From the Apps dropdown in the toolbar navigate to "OCSANA>Signal Flow Analysis".
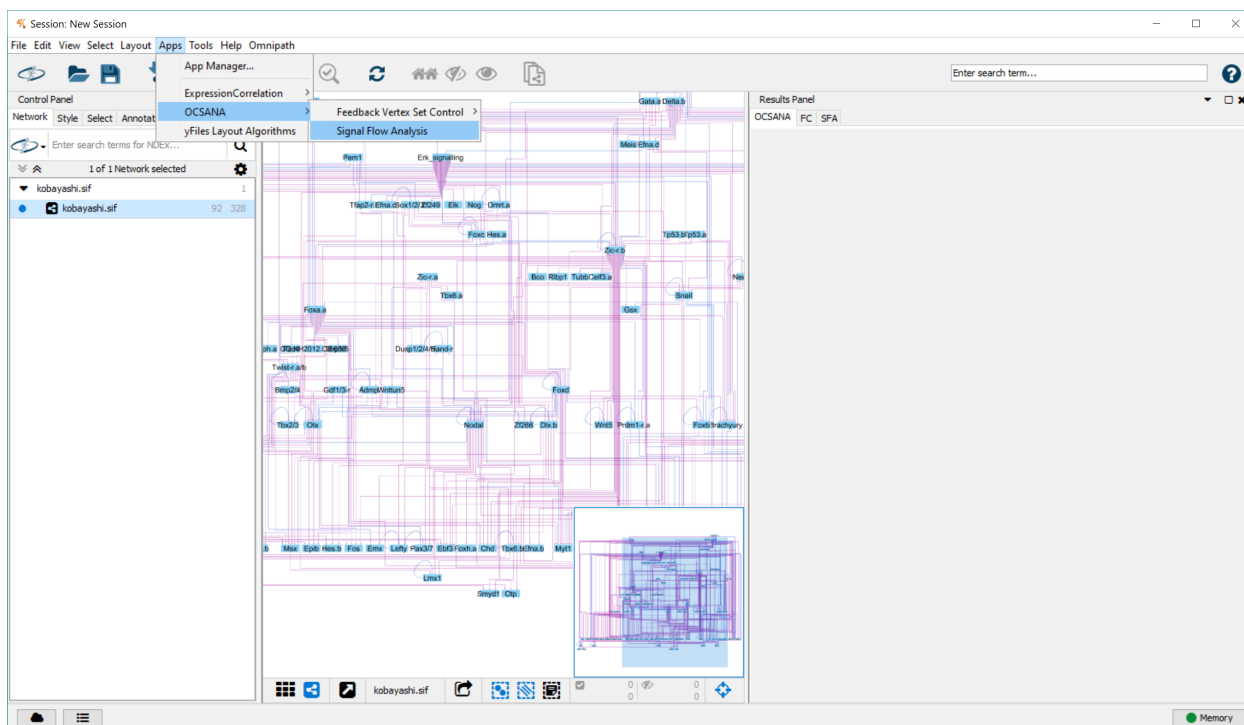
In the pop-up menu:

1. select either "List" or "String" for Node Selection Mode

2. select "shared name" for Select Node Name column 3a. If "List" is selected, scroll in the "Activated Nodes" box and ctrl+click Gata.a and Zic-r.a
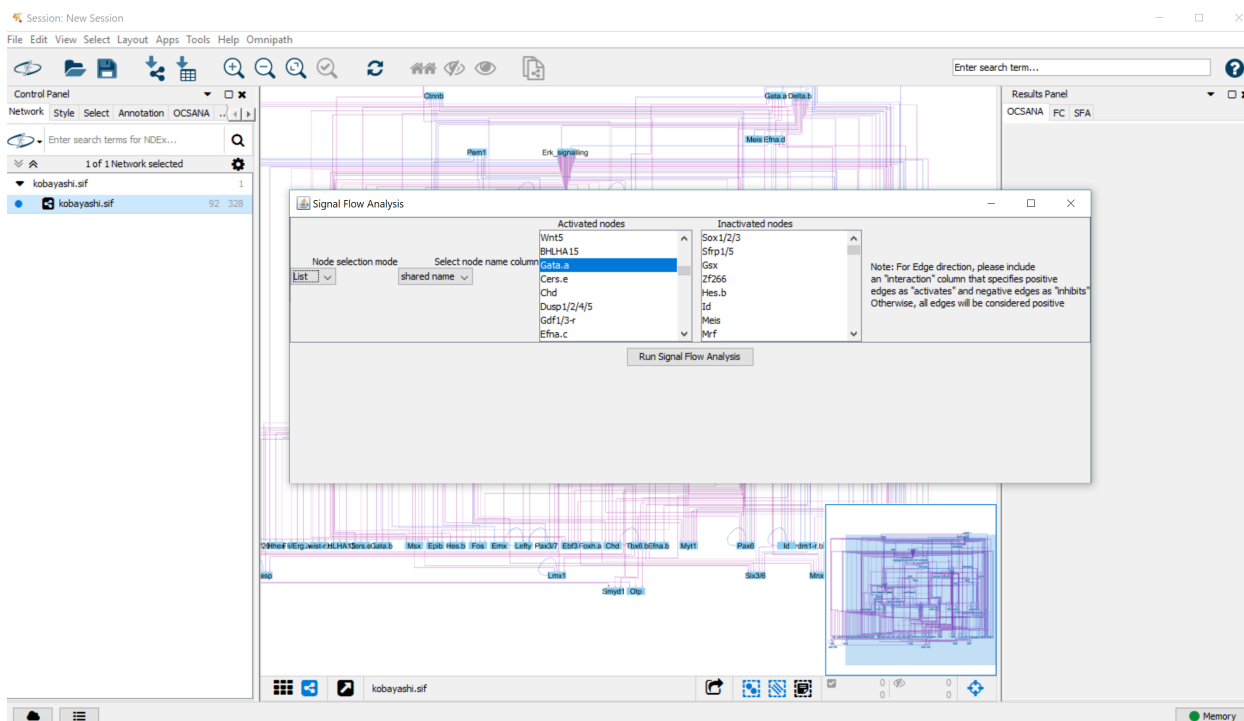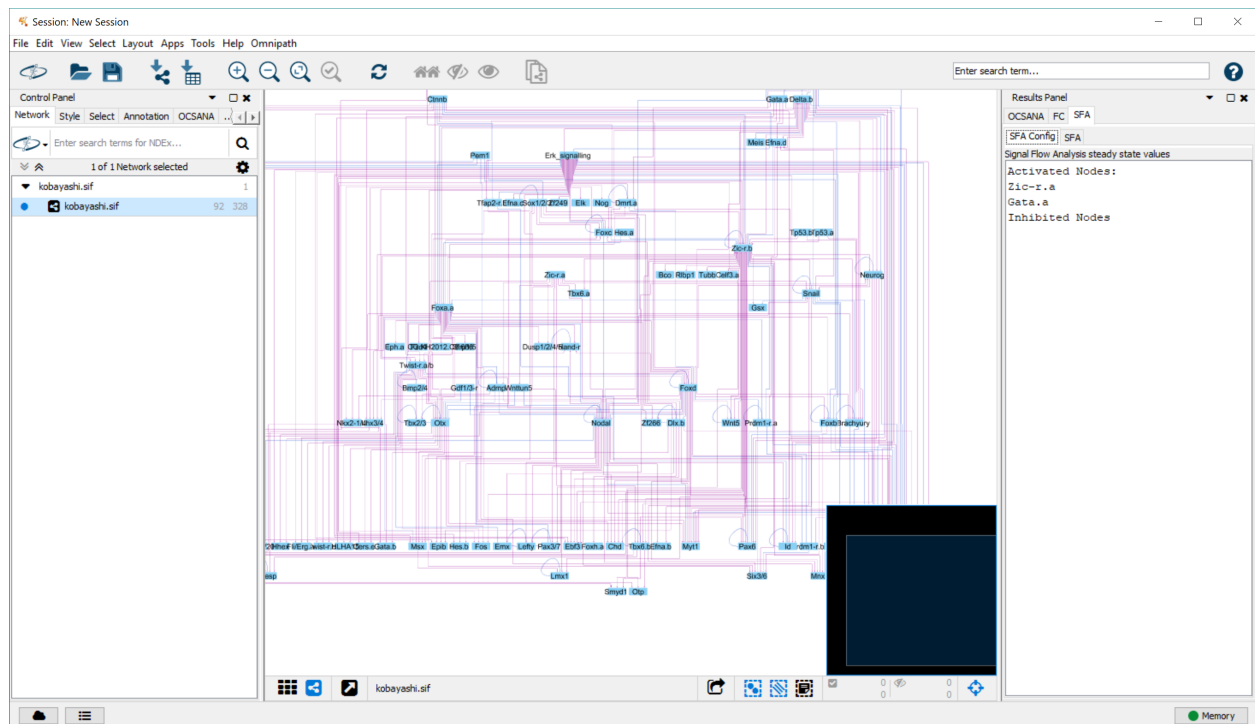

3b. If "String" is selected, type Gata.a and Zic-r.a into the "Activated Nodes" box (either comma delimited, or newline delimited)
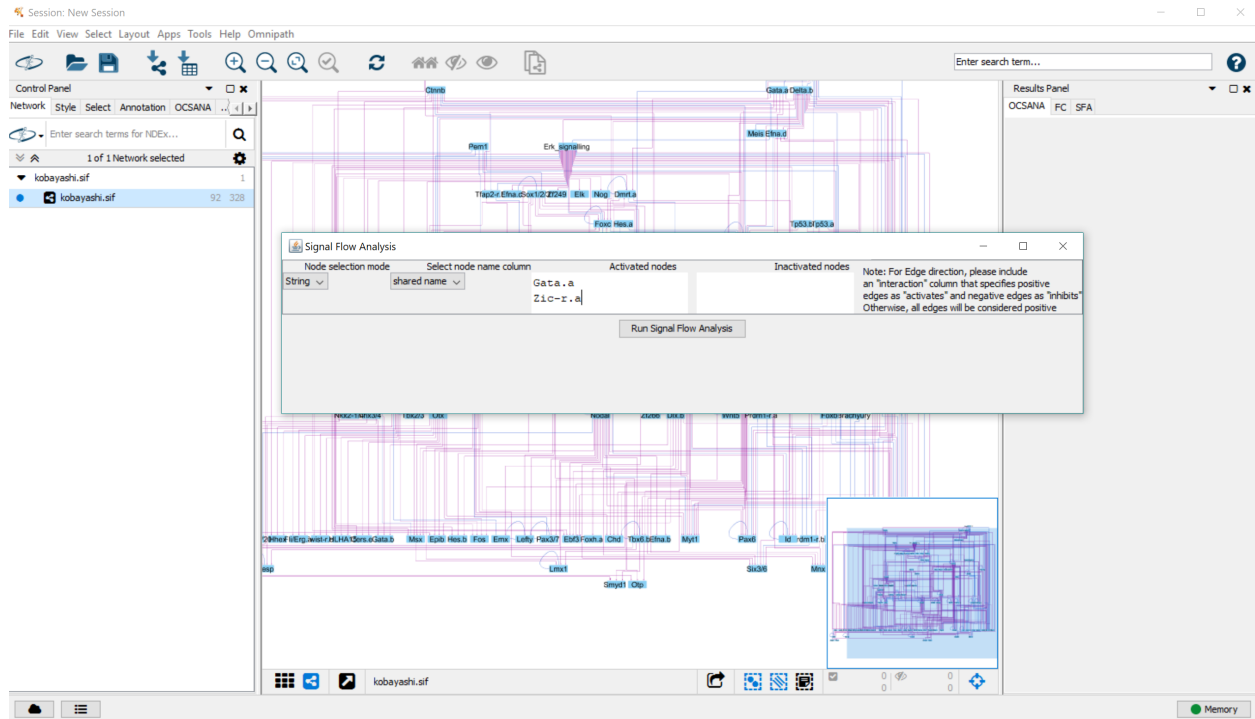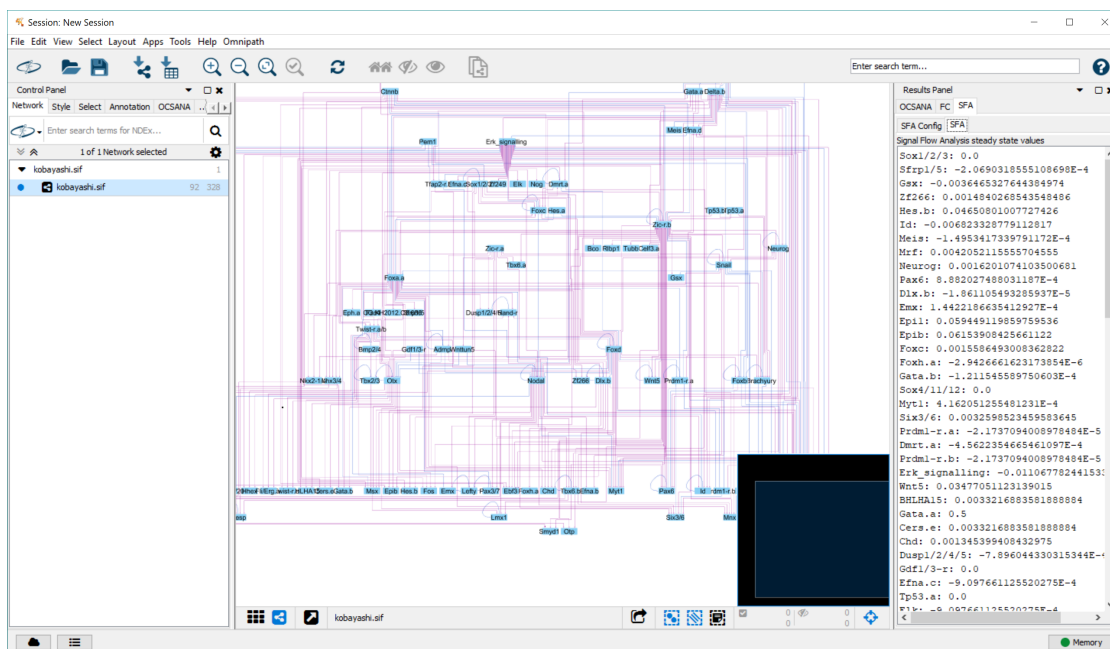
To start SFA, click "Run Signal Flow Analysis"

After completion of SFA, the results will appear in the SFA tab of the Results Panel.

The "SFAConfig" tab displays which nodes were activated or inhibited in the initial user configuration

The "SFA" tab displays the steady state log values for all network nodes.

We repeat this analysis performing a perturbation of the FC nodes. For example, as described in Kobayashi et al., 2018, down-regulation of Foxa.a, Foxd, and Neurog and up-regulation of Zic-r.b, and Erk signaling would lead to the mesenchymal tissue cell fate in the ascidian embryo.

We can simulate this experiment using SFA by selecting Zic-r.b and Erk signaling as activated nodes, and Foxa.a, Foxd, and Neurog as inactivated nodes
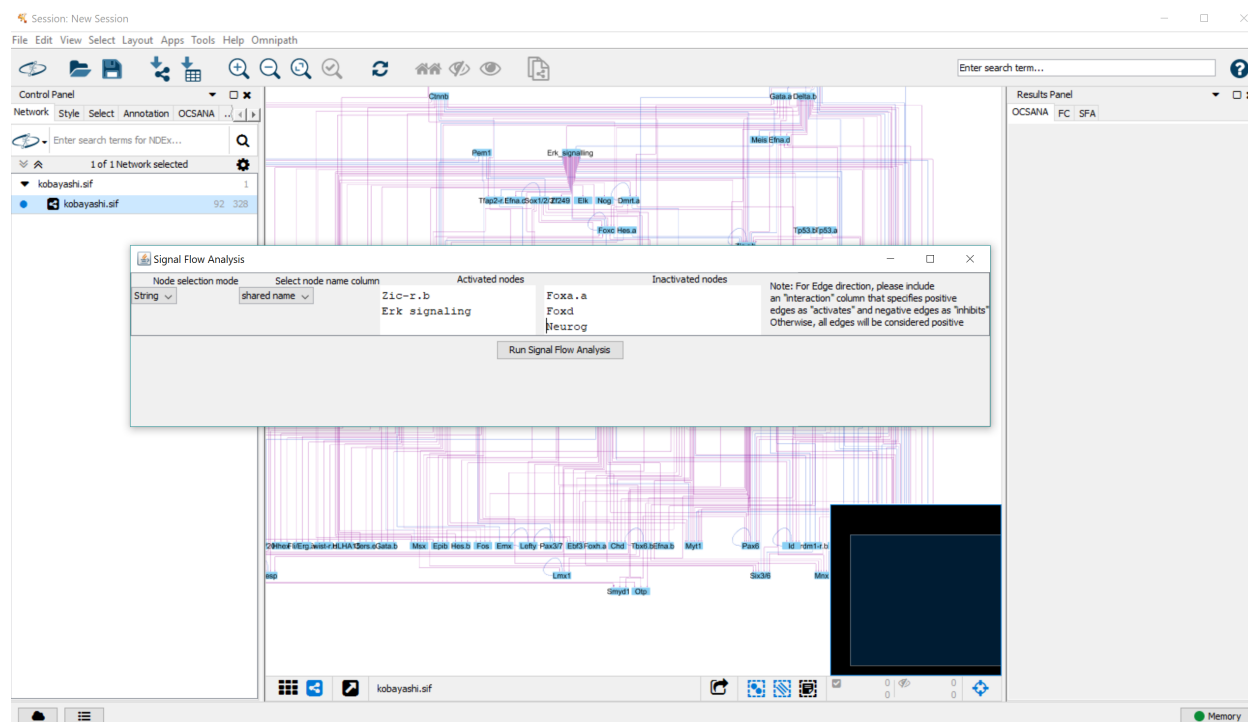
Again, the configuration and results appear in the Results Panel.

Using the steady state log values produced by SFA, we can calculate a direction of activity change between the perturbed and unperturbed state, synonymous with a logFC value.

For example, the log steady state value produced of mesenchymal marker Fli/Erg.a in the unperturbed state is 0.00332. The value in the above perturbation is 0.02310. by computing the difference of perturbed vs unperturbed (0.02310-0.00332=0.01978), and computing a positive result, we interpret this as Fli/Erg.a upregulated in the perturbed steady state vs unperturbed steady state (to read more about interpreting the output of SFA, see Lee and Cho, 2018).

Let's check another cell fate marker node. Alp is a marker gene for endoderm specification. The steady state log value of Alp in the unperturbed state is 3.2139E-5. The value steady state log value of Alp in the perturbed state is -0.02999. When we compute the direction of activity change, (-0.02999-3.2139E-5=-0.03002) and receive a negative value, we interpret this as Alp being downregulated under this perturbation.

1.5.4 Identifying Feedback Vertex Set Control Sets (FC) with source nodes

Considering that the GRN studied contains source nodes, which in principle can affect the dynamical attractors in the system, we applied the extended FC approach from Zanudo et al., 2017.

We can use OCSANA+ to compute FC sets with source nodes.

From the Apps dropdown in the toolbar navigate to "OCSANA>Feedback Vertex Set Control>with source nodes"

The pop-up menu is identical to that of step 2.

After setting your configurations and running FC discovery, the FC results subpanel of the Results Panel will display both the FVSes and source nodes
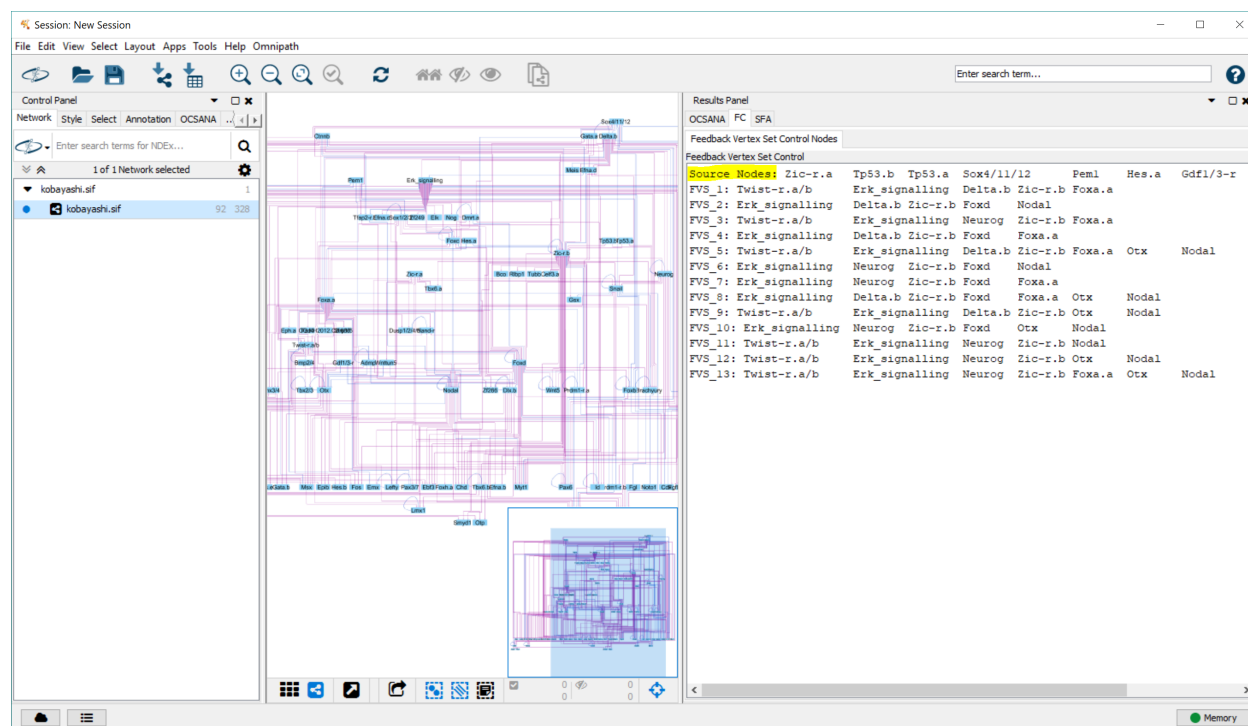
We now have identified the nine source nodes in the network: Ctnnb, Gata.a, Gdf1/3-r, Hes.a, Pem1, Sox4/11/12, Tp53.a,Tp53.b, Zic-r.a

1.5.5 Identifying CIs with OCSANA

We can use OCSANA in OCSANA+ to canalize the signal from the network source nodes to a specified cell fate. For example, if we want to predict additional nodes that may control the signal to epidermal specification, we identify combinations of interventions that can be used to intervene in paths from the nine source nodes to epidermis marker Epi1.

In the OCSANA panel in the Cytoscape Control Panel:

1. select either "List" or "String" for Node Selection Mode

2. select "shared name" for Select Node Name column

3. Enter the nine source nodes in the source nodes box

4. Enter Epi1 in the target nodes box

5. We have set to discover all non-self intersecting paths with a length limit of 20 nodes from source to target. This setting can be changed to suit your network needs and size

6. To configure CI discovery we have chosen the RS algorithm. We will check "allow sources in CIs" so that source nodes can be considered in CIs. We will not bind the number of threads. We will bind CI size at 6 nodes.

After configuration, click "Run OCSANA analysis"
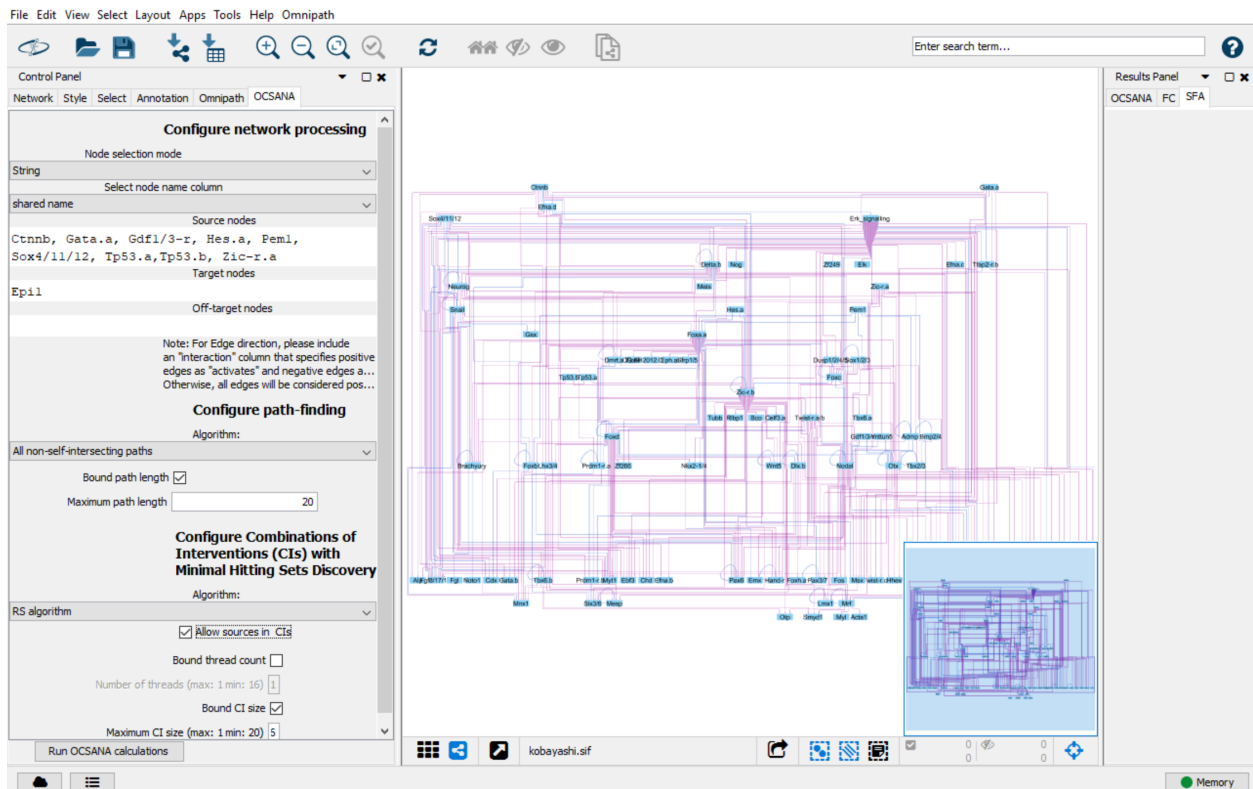
Once the OCSANA Run has completed, the results will appear in the OCSANA Subpanel of the results panel.

The CI tab displays the minimal CIs discovered under user settings
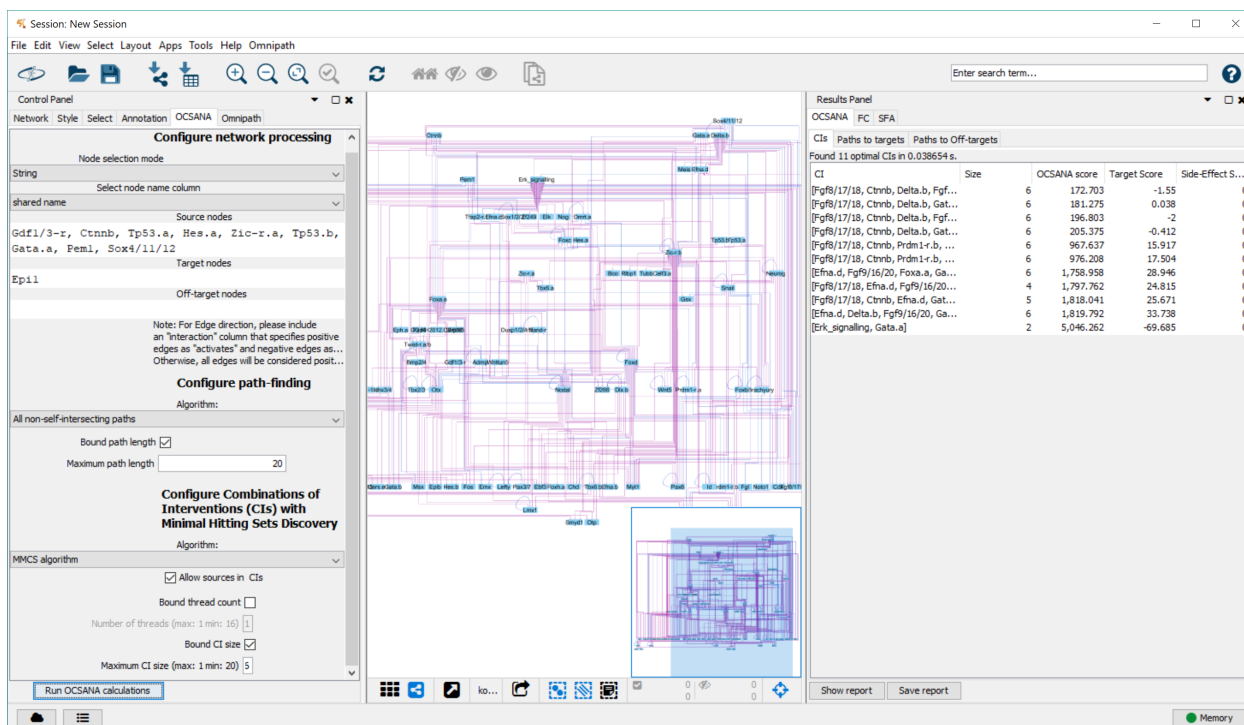
The paths to targets tab displays the paths from the nine source nodes to Epi1
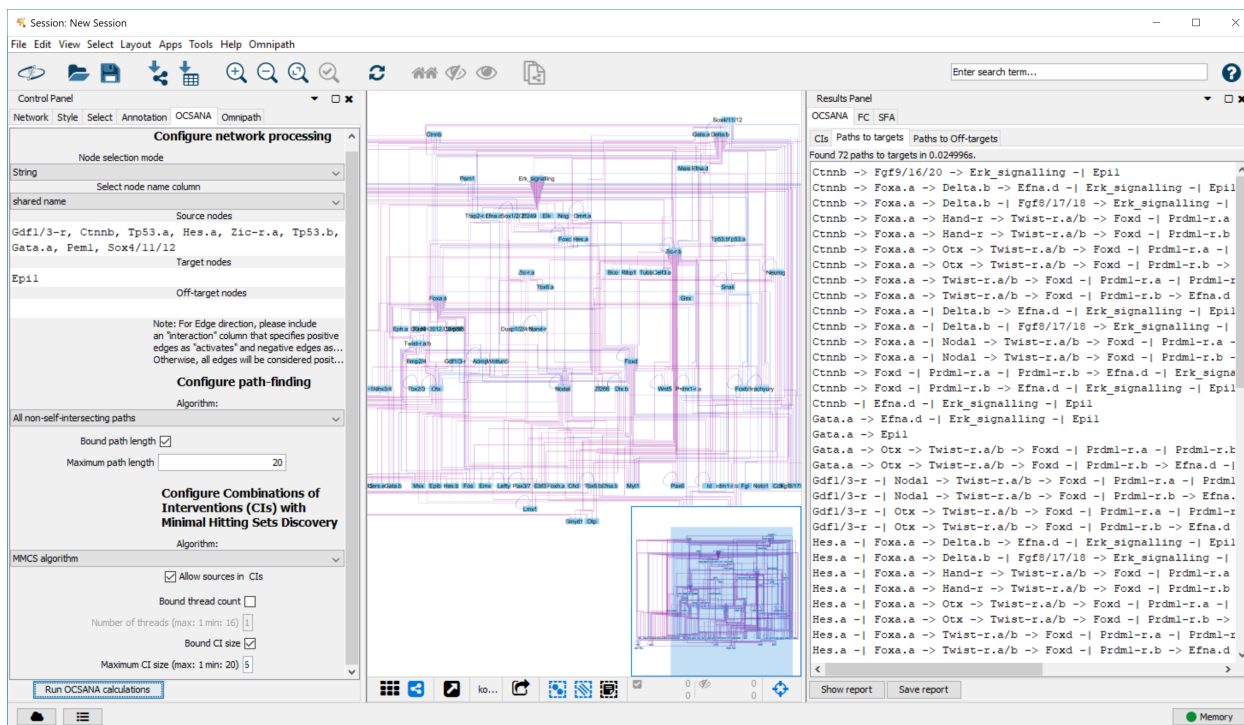
1.5.6 Simulating FC and CI node perturbations using SFA

We will choose the smallest CI of 2 nodes (Gata.a and Erk signalling) to use in combination with the FC perturbation specified in Kobayashi et al., 2018 (down regulation of FC nodes) to simulate the results for epidermal tissue

specification. Do note that Erk signalling is a node within the FC set.

Using SFA, we will set the activated nodes to Gata.a and inactivated nodes as Foxa.a, Foxd, Neurog, Zic-r.b, and Erk signalling

After running SFA, we see the results in the results panel
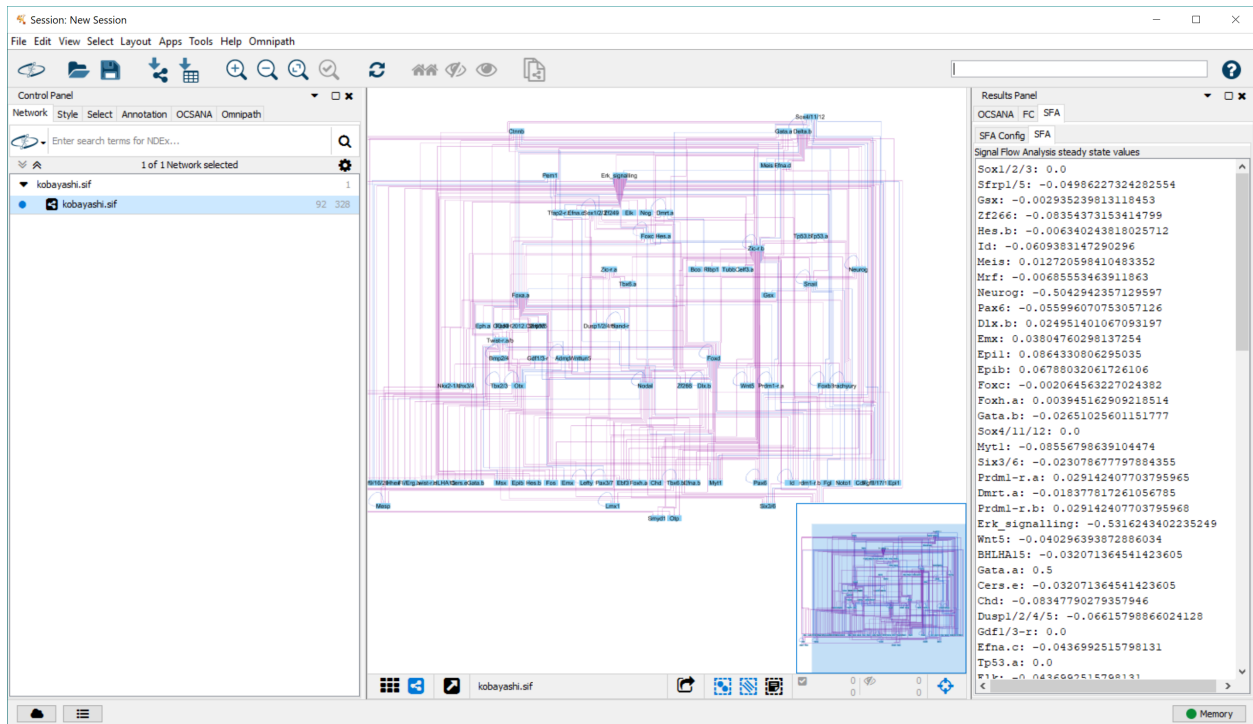
Again, we can calculate the logFC between the above perturbation and the unperturbed steady state for Epi1 (0.08643-0.05944=0.02699). We see that the value is positive; therefore Epi1 is upregulated in the perturbed steady state when compared to the unperturbed steady state.