

# TriPOINT User Manual

---

## Table of Contents

<b>Table of Contents</b> .....	<b>2</b>
<b>Getting Started:</b> .....	<b>3</b>
<b>Running TriPOINT</b> .....	<b>4</b>
<b>TriPOINT Options</b> .....	<b>5</b>
<b>Result Files</b> .....	<b>6</b>
<b>Visualizing Pathways in Cytoscape</b> .....	<b>8</b>
<b>Running Time Benchmarks</b> .....	<b>9</b>
<b>How To Provide Input Data</b> .....	<b>10</b>
<b>How To Run With Non-coding Regulators</b> .....	<b>11</b>
<b>How To Configure Scores</b> .....	<b>12</b>
<b>Additional Examples</b> .....	<b>14</b>

## Getting Started:

### 1. JAVA

Download and install the latest version of Java: <https://java.com/en/download/>

### 2. TriPOINT

Download the jar file for TriPOINT from <https://github.uconn.edu/ajt06004/TriPOINT/releases> or download the source code and compile a runnable JAR file with your preferred IDE.

### 3. Setting up R GRAPHITE and Rserve

TriPOINT utilizes GRAPHITE from R and connects to R using Rserve. If not installed already, install R from <https://www.r-project.org/> and install the following packages:

GRAPHITE:

```
source("https://bioconductor.org/biocLite.R")
biocLite("graphite")
```

org.Hs.eg.db

```
biocLite("org.Hs.eg.db")
```

Rserve:

```
install.packages("Rserve")
```

Once Rserve is installed it will need to be started with:

```
library(Rserve)
Rserve()
```

Alternatively if the above method does not start Rserve the following can be used

```
library(Rserve)
Rserve(args='--no-save')
```

Rserve will need to be started for TriPOINT to be executed.

## Running TriPOINT

Use the following command (substituting brackets <> for local file locations and parameter details) to execute TriPOINT:

```
java -jar tripoint.jar <path to your expression file, see example> -db <database e.g. "kegg"> -o <path to your output directory, multiple files are generated>
```

*Example Using Available Data from the data directory which can be found in the release zip folder:*

<https://github.uconn.edu/ajt06004/TriPOINT/releases>

```
java -jar tripoint.jar "<local file path>/combined_log2fold_symbol.txt" -db "kegg" -o <path to your output directory> -rstate 929 -perm 10000 -nct "<local file path>/ENCODE_DnaseMcf7_merged_peaks.bed" -refflat "<local file path>/refFlat.txt" -ci "<local file path>/ENCODE_MCF7_ChIAPET_pooled.txt"
```

*Example Files within the data directory:*

```
combined_log2fold_symbol.txt  
ENCODE_DnaseMcf7_merged_peaks.bed  
ENCODE_MCF7_ChIAPET_pooled.txt  
refFlat.txt
```

## TriPOINT Options

Option	Description
-db	Sets the pathway database to import pathways from GRAPHITE. (REQUIRED)
-o	Sets the output directory for saving result files. (REQUIRED)
-gid	Sets the conversion method for gene identifiers. Refer to 'convertIdentifiers' in GRAPHITE. Default: "SYMBOL"
-up	Sets the up-regulated gene threshold. Default: 0.05
-down	Sets the down-regulated gene threshold. Default: -0.05
-w	Sets the weak activation/inhibition factor when comparing edges. Weak edges are those where the source is inhibited. Default: 0.2
-p	Controls the impact of gene expression when calculating measures. Default: 0.25
-r	Controls the rate of exponential decay to control edge distance in impact score calculation. Default: 0.25
-perm	Sets the number of permutations for p-value calculations. Default: 10000
-rstate	Sets the random state for reproducible results. Default: Current Time (milliseconds)
-refflat	Specifies the UCSC refflat file to use in order to identify TSS locations of genes in pathways. Required for identifying non-coding targets.
-ci	Sets the tab-delimited file (7 column format) for chromatin interaction data to identify non-coding targets of genes.
-nct	Sets the bed for file non-coding targets to be associated with genes.
-prox	Sets the proximity based non-coding target distance which spans the distance both upstream and downstream from the transcription start site.

## Result Files

### *\*\_triangulation.txt*

Contains measures for each pathway gene to determine genes either most perturbed or most supporting.

#### Columns

<b>Column</b>	<b>Description</b>
Gene ID	Gene Identifier
Pathway ID	Pathway identifier (Pathway Database:Pathway Name)
Expression Value	Expression value provided as input
Inconsistency Score	Raw inconsistency score
Inconsistency p-value	Permuted p-value of inconsistency scores
Inconsistency q-value	Benjamini Hochberg adjusted p-value of inconsistency p-values
Support Score	Raw support score
Support p-value	Permuted p-value of support scores
Support q-value	Benjamini Hochberg adjusted p-value of support p-values
Consistency Score	Raw consistency score
Consistency p-value	Permuted p-value of the consistency score
Consistency q-value	Benjamini Hochberg adjusted p-value of consistency p-values
Impact Score	Raw impact score
Impact p-value	Permuted p-value of the impact score
Impact q-value	Benjamini Hochberg adjusted p-value of impact p-values
Number of NonCoding Regulators	Number of non-coding regulators targeting the gene
NonCoding Poisson p-value	P-value derived from the poisson distribution representing the one tail significance (under or over the average) of the number of non-coding regulators targeting the gene.
Triangulation Score	Triangulation score combining consistency, impact, and number of non-coding regulators (if provided)
Triangulation p-value	Permuted p-value of the triangulation score
Triangulation q-value	Benjamini Hochberg adjusted p-value of triangulation p-values

**\*\_noncoding.txt**

Tab delimited file containing a list of genes and their non-coding targets.

*Format*

<b>Gene</b>	<b>Chr</b>	<b>Start</b>	<b>End</b>

**\*.json**

JSON file format to visualize the pathway in Cytoscape. Each JSON file encodes the pathway used in calculating measures in TriPOINT as well as providing links to non-coding targets.

***TriPOINT\_Parameters.txt***

Lists the parameters used to obtain the output in the current directory.

## Visualizing Pathways in Cytoscape

To visualize pathways, open the latest version of Cytoscape and import the pathway of interest:

File -> Import -> Network -> File...

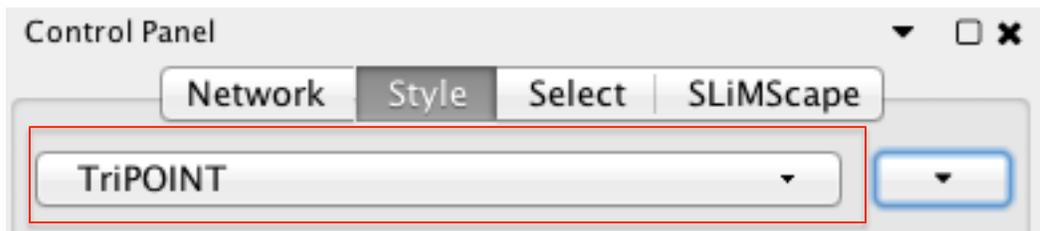
Reposition the nodes in the network using your favorite layout pattern. For example:

Layout -> Hierarchical Layout

To import the TriPOINT style which colors nodes by their expression representation and labels node/edge types, first download the XML file from: <https://github.uconn.edu/raw/ajt06004/TriPOINT/master/tripoint.xml>. Once downloaded, import the style file using:

File -> Import -> Styles...

Finally, load the style by selecting the Style tab in the control panel



and selecting the TriPOINT style from the drop down menu.

## Running Time Benchmarks

<b>Operating System</b>	<b>Processor</b>	<b>Memory</b>	<b># of Permut.</b>	<b>Time (minutes)</b>
Window 10	2.3 GHz Intel Core i7	8 GB	1000	5.27
Window 10	2.3 GHz Intel Core i7	8 GB	10000	20.05
OS X 10.9.5	2.6 GHz Intel Core i7	16 GB	10000	13.93
Ubuntu	4.2 GHz Intel Core i7	16GB	10000	7.63

## How To Provide Input Data

### How to format Gene Expression/Score Files

To format gene expression/score files for usage with TriPOINT, provide a two-column tab-delimited text file where the first column contains the official gene symbol and the second column contains the scores.

Gene Symbol	Score

### Non-Coding Regulators

Non-coding regulators can be defined with a 3-column tab-delimited text file or bed file with the following columns:

1. chromosome
2. start
3. end

chr	start	end

### Chromatin Interaction Files

To link non-coding regulators to promoters, a chromatin interaction file can be provided. A chromatin interaction is a 7-column tab-delimited file with the following columns:

1. Chromosome of the first interaction anchor
2. Start of the first interaction anchor
3. End of the first interaction anchor
4. Chromosome of the second interaction anchor
5. Start of the second interaction anchor
6. End of the second interaction anchor
7. Paired-end tag score/interaction count (integer value)

chr1	start1	end1	chr2	start2	end2	Interaction count

### Refflat Files

Refflat files are used to define transcription start sites of genes in order to identify gene targets of non-coding regulators. Refflat files can be found at the UCSC genome browser. For example, the hg19 refflat file can be found here:

<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/refFlat.txt.gz>

This file must be unzipped to use with TriPOINT.

## How To Run With Non-coding Regulators

There are currently two options for identifying targets of non-coding regulators. In both cases, please provide the filepath to the non-coding regulator locations (-nct <filepath> option when using terminal) and provide the path to the refflat file for TSS definitions to inform the algorithm which sites are promoters of genes (-refflat <filepath> option when using terminal).

### **1. Chromatin Interactions**

Chromatin interaction networks are built using QuIN (<https://quin.jax.org>). Provide the chromatin interactions file (-ci <filepath> option when using terminal), to use this option. In this case all non-coding regulators targeting genes will be based on evidence from chromatin interactions.

### **2. Proximity Based**

If chromatin interactions are not provided, TriPOINT also allows for assigning gene targets based on proximity. Provide the distance (in bp, e.g., 50000) using the proximal distance option (-prox <distance in bp> when using terminal).

## How To Configure Scores

In total, there are 5 different parameters for configuring scoring using TriPOINT:

### Active (Upregulated) Gene Threshold

The active gene threshold sets at what values a gene will be active (or upregulated in the pathway). Using log2 fold change of gene expression, a threshold of 0.05 has been selected in our tool, however this can be changed using the `-up` option.

If using other scoring metrics other than log2 fold change of gene expression, be sure to set the threshold appropriate for the score.

### Inactive (Downregulated) Gene Threshold

The inactive gene threshold sets at what values a gene will be inactive (or downregulated in the pathway). Using log2 fold change of gene expression, a threshold of -0.05 has been selected in our tool, however this can be changed using the `-down` option.

If using other scoring metrics other than log2 fold change of gene expression, be sure to set the threshold appropriate for the score.

### Weak Pathway Association Factor (w)

Weak relationships are defined in TriPOINT as associations in pathways where the source gene is inactive (downregulated). The assumption here is that if the gene is inactive, it doesn't necessarily imply that there would be an effect on its downstream genes in the pathway.

$$\text{Support}(g_i, p_m) = \sum_{g_k \in U^s(g_i, p_m)} s_s |\Delta \text{Exp}(g_k)|^p + w \sum_{g_k \in U^s(g_i, p_m)} s_w |\Delta \text{Exp}(g_k)|^p$$

$$s_s = \begin{cases} 1 & \text{if } E_{\text{type}}(g_k, g_i, p_m) = \text{ACTIVATION } \Delta \text{Exp}(g_k) > T_u \quad \Delta \text{Exp}(g_i) > T_u \\ 1 & \text{if } E_{\text{type}}(g_k, g_i, p_m) = \text{INHIBITION } \Delta \text{Exp}(g_k) > T_u \quad \Delta \text{Exp}(g_i) < T_d \\ 0 & \text{otherwise} \end{cases}$$

$$s_w = \begin{cases} 1 & \text{if } E_{\text{type}}(g_k, g_i, p_m) = \text{ACTIVATION } \Delta \text{Exp}(g_k) < T_d \quad \Delta \text{Exp}(g_i) < T_d \\ 1 & \text{if } E_{\text{type}}(g_k, g_i, p_m) = \text{INHIBITION } \Delta \text{Exp}(g_k) < T_d \quad \Delta \text{Exp}(g_i) > T_u \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Inconsistency}(g_i, p_m) = \sum_{g_j \in U^s(g_i, p_m)} i_s |\Delta \text{Exp}(g_j)|^p + w \sum_{g_j \in U^s(g_i, p_m)} i_w |\Delta \text{Exp}(g_j)|^p$$

$$i_s = \begin{cases} 1 & \text{if } E_{\text{type}}(g_j, g_i, p_m) = \text{ACTIVATION } \Delta \text{Exp}(g_j) > T_u \quad \Delta \text{Exp}(g_i) < T_d \\ 1 & \text{if } E_{\text{type}}(g_j, g_i, p_m) = \text{INHIBITION } \Delta \text{Exp}(g_j) > T_u \quad \Delta \text{Exp}(g_i) > T_u \\ 0 & \text{otherwise} \end{cases}$$

$$i_w = \begin{cases} 1 & \text{if } E_{\text{type}}(g_j, g_i, p_m) = \text{ACTIVATION } \Delta \text{Exp}(g_j) < T_d \quad \Delta \text{Exp}(g_i) > T_u \\ 1 & \text{if } E_{\text{type}}(g_j, g_i, p_m) = \text{INHIBITION } \Delta \text{Exp}(g_j) < T_d \quad \Delta \text{Exp}(g_i) < T_d \\ 0 & \text{otherwise} \end{cases}$$

To adjust for this use the `-w` parameter in the terminal. By default TriPOINT weights these interactions by 0.2, giving strong associations more impact than weak ones. This weighting can be adjusted to having equal weights (i.e., `-w 1`) or having more weight for weak associations (i.e., `-w 1.5`)

### Exponential Decay Rate (r)

To calculate impact scores, all downstream paths in a pathway are calculated where each downstream gene is added to the cumulative sum for the impact score:

$$\text{Impact}(g_i, p_m) = \sum_{g_m \in Ds(g_i, p_m)} \frac{|\Delta \text{Exp}(g_m)|^p}{e^{r * d(g_i, g_m)}}$$

The exponential decay rate adjusts to weight genes that are further downstream away from the source gene less than those that are more immediate.

### Gene Score Influence (p)

Gene score influence (adjusted using `-p`), adjusts the impact of the gene score. Values should range between 0 and 1 where a value of 0 will remove all gene score information other than whether or not it's active (upregulated) or inactive (downregulated) based on the prior thresholds set and a value of 1 will be adjusted the most from the score associated with the gene. See the following formula:

$$|\Delta \text{Exp}(g_m)|^p$$

A default value of 0.25 is used to value whether a gene is active (upregulated) or inactive (downregulated) more than its total gene expression.

## Additional Examples

### Running without Non-Coding Regulators

```
java -jar <path to jar>/tripoint.jar "<path to example files>/examplefiles/combined_log2fold_symbol.txt" -db "kegg" -o <path to output directory> -rstate 929 -perm 1000
```

### Running with Non-Coding Regulators (using chromatin interactions)

```
java -jar <path to jar>/tripoint.jar "<path to example files>/examplefiles/combined_log2fold_symbol.txt" -db "kegg" -o <path to output directory> -rstate 929 -perm 1000 -nct <path to example files>/examplefiles/ENCODE_DnaseMcf7_merged_peaks.bed" -refflat "<path to example files>/refFlat.txt" -ci "/Users/athib/Desktop/Pathway/testrun/examplefiles/ENCODE_MCF7_ChIAPET_pooled.txt"
```

### Running with Non-Coding Regulators (by proximity)

```
java -jar <path to jar>/tripoint.jar "<path to example files>/examplefiles/combined_log2fold_symbol.txt" -db "kegg" -o <path to output directory> -rstate 929 -perm 1000 -nct <path to example files>/examplefiles/ENCODE_DnaseMcf7_merged_peaks.bed" -refflat "<path to example files>/refFlat.txt" -prox 50000
```