Table of Contents

Introduction: page 1
Disclaimers: page 1
Quickstart tutorial: page 2
Advanced considerations: page 4
An illustrated example (Histones in the RNAseq dataset): page 5

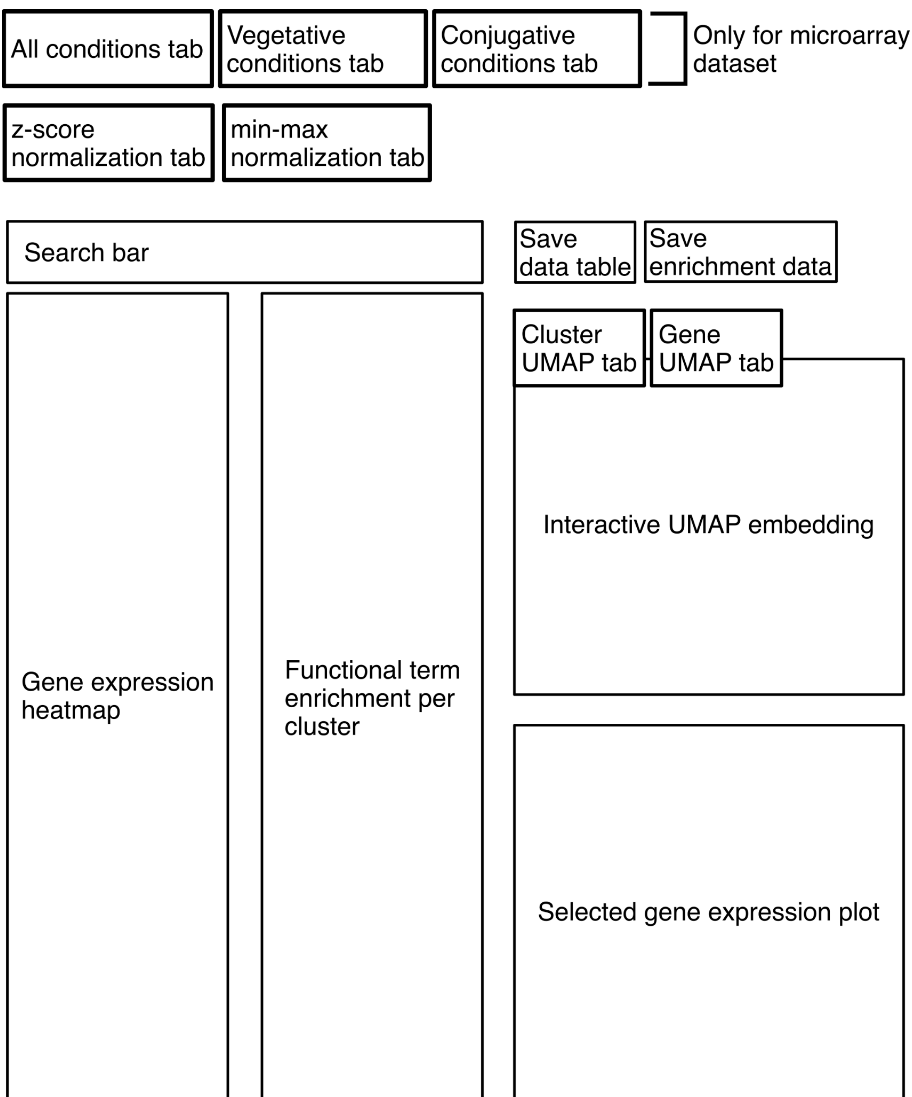
Introduction

The Tetrahymena Gene Network Explorer (TGNE) is a tool for interactively exploring the microarray data across bulk physiological conditions from Xiong et al. (2011) and cell-synchronized RNAseq data across 1.5 doublings from Zhang et al. (2023). We corrected the raw data from these studies using the newest macronuclear genome model from Ye et al. (2024) and computational quality control checks. We applied extensive new clustering algorithms to explore the co-expression of genes, which we will describe in detail in our forthcoming publication. For the time being, please keep in mind that this is a "beta" version of the tool, and it is subject to change. Feedback on the tool or user guide is welcome!

DISCLAIMERS

1. The dashboards require browsers with WebGL support to work quickly. These include Google Chrome, Microsoft Edge, Mozilla Firefox, Apple Safari, and Opera One. To check if your browser has WebGL enabled, visit this page.
2. The dashboard HTML files are very large (100-300 MB), so they may load slowly depending on your internet connection and computer memory. Once loaded, with WebGL enabled, they will work quickly.
3. Some of the TTHERM IDs based on Ye et al. (2024) are not yet available on the TGD. Clicking on links in the annotation tables will normally take you to the corresponding TGD page. If it says "Feature Not Found!" the information will become available when the TGD is updated with the new gene annotations. In the meantime, the peptide sequences in our annotation tables reflect the newest gene models.

Quickstart tutorial



There are two TGNE dashboard interfaces: one based on the microarray data from Xiong et al. (2011) and one based on the RNAseq data from Zhang et al. (2023). Each interface spans whole-transcriptome and single-gene views of expression data, and every plot provides interactive tools for exploring the data, such as zooming, click/click-and-drag to select, and information when hovering the mouse cursor over a data point. The clickable tabs constrain the different data displays in hierarchical order from top to bottom.

For users who are familiar with the co-expression resources that are available through the TFGD, the most comparable dashboard is the full microarray analysis with min-max normalization. A more detailed discussion of the different normalization frameworks and data subsets is available after the tutorial.

Tutorial (a programmed genome rearrangement gene cluster):

1. Start with the full microarray dataset using min-max normalization.
2. Type in TTHERM_00125280 or PDD1 into the search bar at the top of the dashboard and press the enter/return key on your keyboard.
    a. The data for all other genes in the plots will be deselected: only one row of the heatmap will be selected; the enrichment data will be blank; one point on the UMAP will be selected; one gene expression profile will be shown; the data table under the dashboard will show information for just TTHERM_00125280.
3. To explore genes with very similar expression profiles to the query, look at the module ID (m232) associated with the gene in the annotation table below the dashboard. Type it into the search bar and press the enter/return key on your keyboard.
    a. All the genes in that module are now selected.
    b. Scroll sideways through the annotation table to explore all the available information for the genes.
4. Click between the tabs above the UMAP plot to compare the displays.
    a. The "UMAP of clusters" tab has a single point highlighted, representing the selected module relative to all other modules. The "UMAP of genes" tab shows every gene in the selected module highlighted relative to all the genes in the dataset.
5. On the right-hand side of the UMAP plot, use the box-zoom tool ( ) to inspect the region next to the selected points. Pan the mouse over the selected and nearby transparent points to see information about the clusters or genes.
    a. Nearby points that are not selected have related, but less similar expression profiles to the selected gene(s).
6. Use the reset tool ( ) to return the plot to its original state
7. On the right hand side of the functional term enrichment plot, click the scroll-zoom tool ( ) and use your mouse wheel or trackpad to zoom in on the highlighted cluster corresponding to the PDD1 module.
8. Hover your cursor over the enriched functional terms.
    a. The hover data for each point reports the identity of the term, additional information about it, its fold-change enrichment, and its Bonferroni-corrected p-value.
9. Press the "💾Data Table" button in the top right corner of the dashboard to download the contents of the table below the dashboard.
10. Press the "💾Enrichment" button in the top right corner of the dashboard to download the functional term enrichment information for all modules the dataset (specific to the tab(s) selected in the top left corner of the page).

Advanced considerations

There are four general approaches to using the tool:

1) Search for a specific gene using the search bar (TTHERM_ID or common name) and explore the genes co-expressed with it;
2) Search for a functional term using the search bar (GO, PFAM, KEGG orthology, Interpro) and explore genes and expression clusters (modules) associated with them;
3) Find genes that have the desired expression profiles in the heatmap and then search for them;
4) Explore cluster or gene UMAP embedding and identify interesting regions in the gene expression space. Genes or clusters that are close in the UMAP embedding tend to have similar expression profiles.

We provide two versions of the tool:

1) The RNAseq version (cell cycle synchronized dataset over one and a half cell cycles)
2) The microarray version (the culturing conditions as on the TFGD and in Xiong et al. (2011), minus conditions that did not pass our quality controls)
    a) This is subdivided into full, vegetative-only, and conjugative-only profiles that are selectable via tabs in the top left corner of the dashboard. This subdivision is relevant in cases such as when a gene plays different biological roles during vegetative or conjugative functions, and considering co-expression across both life stages may be uninformative.

For each of these, there are two variants: one has each gene's expression normalized to its respective minimum and maximum of expression; the other has each gene's expression normalized to its respective mean and standard deviation (z-score) of expression. These normalization modes are selected via tabs in the top left corner of the dashboard interface.

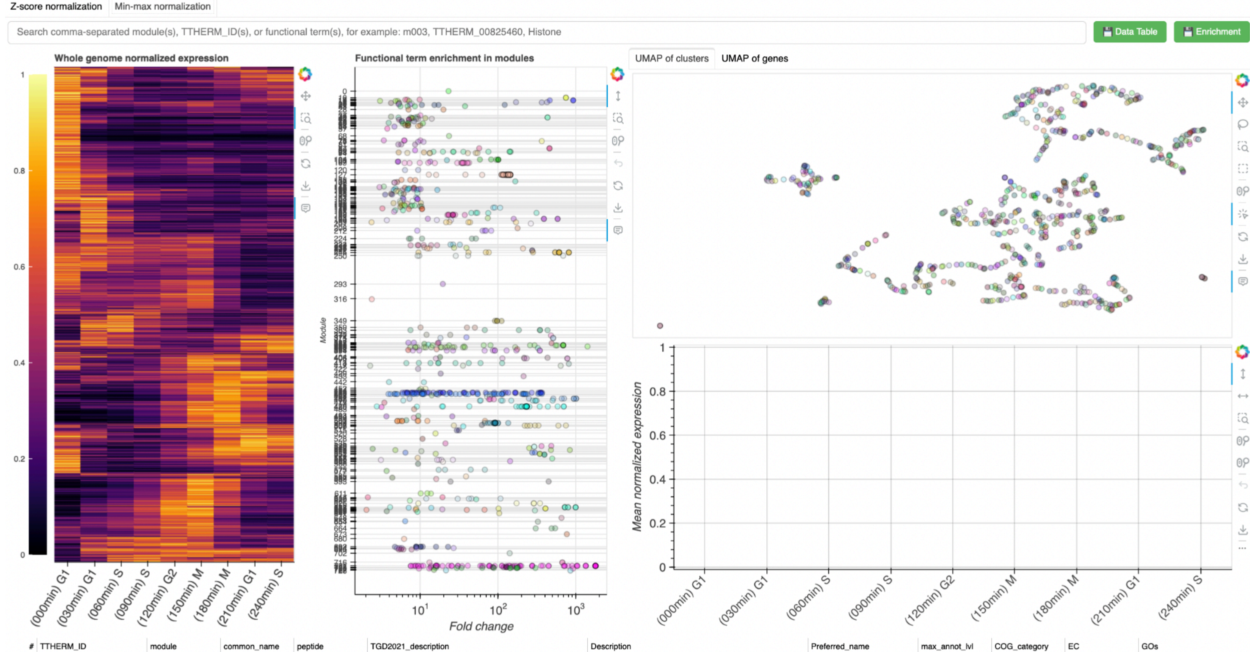
The minimum-maximum normalization is most useful for identifying co-expressed genes solely by the shape of their expression profile. This is most similar to the information on co-expressed genes as provided on the TFGD. However, it is important to note that our co-expression clusters are mutually exclusive (every gene is assigned to a unique cluster), unlike the similarity network reported in the TFGD. Consequently, the lists of genes that are co-expressed with a given query tend to be smaller.

The z-score normalization is most useful for identifying co-expressed genes by both the shape of the expression profile ***and*** expression intensity. In general this results in more stringent clustering, separating genes that may have similar expression patterns but different degrees of expression variance over the analyzed conditions. This is useful when the relative regulation of expression is biologically relevant.

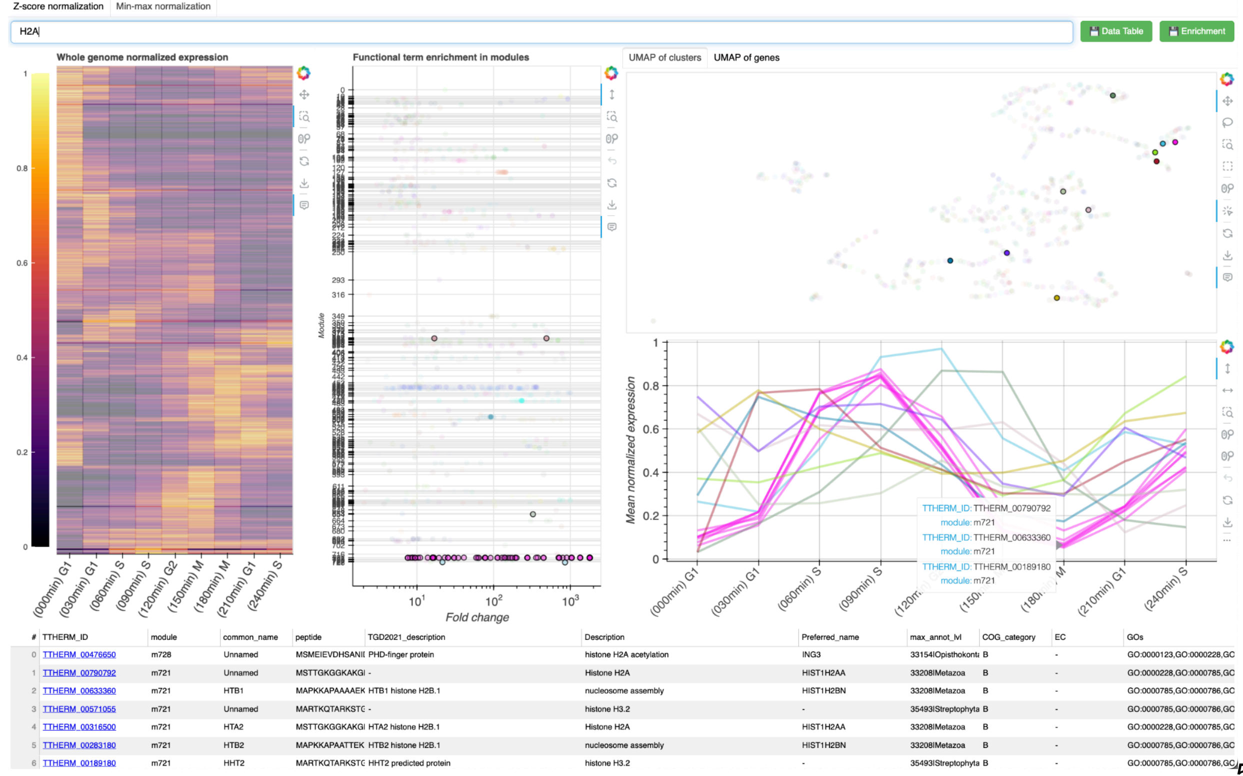
We discuss the normalization approaches and their merits in more detail in our forthcoming manuscript.

An illustrated example (Histones in the RNAseq dataset)

1. Before searching for anything, the min-max normalized RNAseq dataset looks as follows. The default view shows the UMAP embedding where each point represents an entire cluster (clicking on a point will select all the genes in that cluster).



2. Searching for “H2A” selects all genes that include this phrase in their annotation data and highlights their clusters. There is one dominant cluster that is apparent from the expression profiles: m721, indicated by hovering the cursor over the expression profiles. The annotation table is truncated in this view, but can be scrolled down further.



#	TTHERM_ID	module	common_name	peptide	TGD2021_description	Description	Preferred_name	max_annot_lvl	COG_category	EC	GOs
0	TTHERM_00476650	m728	Unnamed	MSMEIEVDHSANII	PHD-finger protein	histone H2A acetylation	ING3	33154|Opisthokonta	B	-	GO:0000123,GO:0000228,GO
1	TTHERM_00790792	m721	Unnamed	MSTTGKGGKAKGI	-	Histone H2A	HIST1H2AA	33208|Metazoa	B	-	GO:0000228,GO:0000785,GO
2	TTHERM_00633360	m721	HTB1	MAPKKAPAAAAEK	HTB1 histone H2B.1	nucleosome assembly	HIST1H2BN	33208|Metazoa	B	-	GO:0000785,GO:0000786,GO
3	TTHERM_00571055	m721	Unnamed	MARTKQTARKSTG	-	histone H3.2	-	35493|Streptophyta	B	-	GO:0000785,GO:0000786,GO
4	TTHERM_00316500	m721	HTA2	MSTTGKGGKAKGI	HTA2 histone H2B.1	Histone H2A	HIST1H2AA	33208|Metazoa	B	-	GO:0000228,GO:0000785,GO
5	TTHERM_00283180	m721	HTB2	MAPKKAPAATTEK	HTB2 histone H2B.1	nucleosome assembly	HIST1H2BN	33208|Metazoa	B	-	GO:0000785,GO:0000786,GO
6	TTHERM_00189180	m721	HHT2	MARTKQTARKSTG	HHT2 predicted protein	histone H3.2	-	35493|Streptophyta	B	-	GO:0000785,GO:0000786,GO

3. Searching for m721 selects all 14 genes in the cluster. Zooming in on the functional enrichment plot and hovering over the points corresponding to m721 reveals the significantly enriched functional terms. Clicking on a row in the annotation table will highlight that gene's expression profile in the plot.

#	TTHERM_ID	module	common_name	peptide	TGD2021_description	Description	Preferred_name	max_annot_lvl	COG_category	EC	GOs
0	TTHERM_00823720	m721	HHO1	MAPRSSTSKSATR	HHO1 Tonsoku-like protein	-	-	-	-	-	-
1	TTHERM_00790792	m721	Unnamed	MSTTGKGGKAKGI	-	Histone H2A	HIST1H2AA	33208IMetazoa	B	-	GO:0000228,GO:0000785,GC
2	TTHERM_00660180	m721	HMGB1	MSKAASQYATLED	HMGB high mobility group (HMG) box protein	high mobility group	-	5878ICiliophora	K	-	-
3	TTHERM_00633360	m721	HTB1	MAPKKAPAAAAEK	HTB1 histone H2B.1	nucleosome assembly	HIST1H2BN	33208IMetazoa	B	-	GO:0000785,GO:0000786,GC
4	TTHERM_00571055	m721	Unnamed	MARTKQTARKSTC	-	histone H3.2	-	35493IStreptophyta	B	-	GO:0000785,GO:0000786,GC
5	TTHERM_00529532	m721	Unnamed	MGSTKKSATKSAS	-	-	-	-	-	-	-
6	TTHERM_00516300	m721	Unnamed	MEQNHPDFPQLR	hypothetical protein	-	-	5878ICiliophora	-	-	-
7	TTHERM_00498190	m721	HHF1	MAGGKGGKGMGI	HHF1 predicted protein	Core component of nucleosome. Nucleosomes wrap and cor	-	35493IStreptophyta	B	-	-
8	TTHERM_00476610	m721	Unnamed	MGKTRSVSVSQK	Uncharacterized oxidoreductase C2F3.05c	-	-	-	-	-	-
9	TTHERM_00316500	m721	HTA2	MSTTGKGGKAKGI	HTA2 histone H2B.1	Histone H2A	HIST1H2AA	33208IMetazoa	B	-	GO:0000228,GO:0000785,GC
10	TTHERM_00283180	m721	HTB2	MAPKKAPAATTEK	HTB2 histone H2B.1	nucleosome assembly	HIST1H2BN	33208IMetazoa	B	-	GO:0000785,GO:0000786,GC
11	TTHERM_00257230	m721	HMGB2	MAKSKDDSKPAPF	HMGC high mobility group (HMG) box protein	high mobility group	-	2759IEukaryota	B	-	-
12	TTHERM_00189180	m721	HHT2	MARTKQTARKSTC	HHT2 predicted protein	histone H3.2	-	35493IStreptophyta	B	-	GO:0000785,GO:0000786,GC
13	TTHERM_00189170	m721	HHF2	MAGGKGGKGMGI	HHF2 histone H4	Core component of nucleosome. Nucleosomes wrap and cor	-	35493IStreptophyta	B	-	-
14	TTHERM_00143660	m721	HTA3	MAGGKGGKGGKC	HTA3 histone H2A	Belongs to the histone H2A family	-	5878ICiliophora	B	-	GO:0000228,GO:0000785,GC