aura.science.unitn.it

AURA 2 TUTORIAL

1. OVERVIEW

This tutorial provides a brief overview of AURA 2 interface and use. It is organized in the following topics, encompassing three sections:

- individual search modes and data-mining features description (part 2 and 3)
- detailed overview of the annotations available for trans-factors and UTRs (part 4)

2. WAYS TO QUERY AURA

AURA can be queried in multiple ways:

- The "**target locus**" mode provides the annotated UTRs associated with your query gene (here named target locus).
- The "**trans-factor**" mode provides the annotated UTRs that are regulated by your query post-transcriptional factor (here named trans-factor), on a genome-wide scale.
- The "**co-regulation**" mode provides the annotated UTRs that are co-regulated by all the trans-factors in your query.
- The "**sequence**" mode provides the annotated UTRs that contain at least a match to your query sequence.
- The "**batch**" mode provides ways to display, and investigate with data mining tools, the annotated UTRs associated with the multiple genes in your query.
- The "**bioMart**" mode allows querying and retrieving all data contained in AURA through a *BioMart* interface.

2.1. Querying AURA by "target locus"

In this worked example we will show how to obtain information about UTRs belonging to a single gene. The gene that will be investigated is **FOS** ("*FBJ murine osteosarcoma viral oncogene homolog*"), from the *Homo sapiens* specie.

2.1.1. Searching for FOS in AURA

- Click on [*target locus*] on the navigation bar
- Type "FOS" in the text box and Click on [submit]. This will take to the [*UTR Searcher*] page.
- If multiple gene names or synonyms contain the "FOS" substring, select the proper "FOS" gene card on the page with [*UTR Searcher*] results.

2.1.2. Searching for FOS UTRs

- Drag and drop your UTRs of interest in the [*Selected UTRs*] side menu. Click on [*Add all*] to add all UTRs to the side menu. Please note that:
    - A composite view of the exon/intron structures of the 5′ and 3′ UTRs, arranged by UCSC transcript ids, help identifying the UTRs of interest.
    - UTRs marked by gold stars belong to protein-coding transcripts present in the Consensus CDS (*CCDS*) database.
- Click on [*Explore UTRs*] to view UTRs. [*Reset*] resets the selections made so far, while [*Back*] takes again to the initial page.

2.1.3. Browsing FOS UTRs

UTRs are automatically sorted by their UCSC transcript identifier, and individually shown in separate panels on the same page. Take the following steps:

- Select an UTR of choice
- Scroll the UTR image window to view the annotation feature tracks
- Click on [*Show gene expression of FOS*] to view total RNA expression profiles according to "*The Genotype-Tissue Expression project (GTEx)*" (http://www.broadinstitute.org/gtex/). Please note that this information is currently available for human genes.

- Click on [*Show UTR secondary structure*] to view annotations within the UTR secondary structure.
- Click on [*Download UTR card*] to get an alternative text-based access to the complete information displayed in the annotation tracks window.
- Click on [*Download FASTA sequence*] to download the intronless version of the selected UTR.

Annotation feature tracks, gene expression measurements and the UTR secondary structure are also briefly presented by appropriate reference cards in the corresponding UTR genome-browser, the gene expression window, and the UTR reference card.

2.2. Querying AURA by "trans-factor"

In this worked example we will show how to obtain information about UTRs regulated by a single post-transcriptional factor on a genome-wide scale. The worked example will investigate the **ELAVL2** ("*ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B)*") gene of the *Homo sapiens* specie.

2.2.1. Setting search preferences

Query results may be filtered by UTR type (5', 3') and by the technique supporting the results. Eventually, the expression filter allows to choose, through its selection box, one of the 48 available gene expression profiles: this will cause the display of a subset of all regulatory interactions, namely the ones for which both the trans-factor and the target genes are expressed in that particular tissue or cell line. Results may be ordered by Gene Ontology classification or by chromosome localization.

2.2.2. Searching for ELAVL2 in AURA

You may want to use the auto-complement enabled in the search box. The query takes to the page listing all search results for your trans-factor in AURA

2.2.3. Searching for ELAVL2 trans-regulated UTRs

- Drag and drop your UTRs of interest in the [*Selected UTRs*] side menu.
- Click on [*Explore UTRs*] to view UTRs. [*Reset*] resets the selection of UTRs, [*Back*] takes to the initial page.

2.2.4. Browsing ELAVL2 trans-regulated UTRs

UTRs are automatically sorted by transcript UCSC identifier, and individually shown in separate views.
- Select an UTR of choice
- Scroll the UTR image window to view the annotation feature tracks
- Click on [*Show gene expression of ELAVL2*] to view total RNA expression profiles according to "*The Genotype-Tissue Expression project (GTEx)*" (http://www.broadinstitute.org/gtex/). Please note that this information is currently available only for human genes.
- Click on [*Show UTR secondary structure*] to view annotations within the UTR secondary structure.
- Click on [*Download UTR card*] to get an alternative text-based access to the full information displayed in the annotation tracks window.
- Click on [*Download FASTA sequence*] to download the intronless version of the selected UTR.

Note that the annotation feature tracks, the gene expression heat map and the UTR secondary structure are briefly presented by appropriate reference cards in the corresponding UTR genome-browser styled reference card, the gene expression reference card, and the UTR reference card in secondary.

2.3. Querying AURA by "co-regulation"

This query mode allows executing the "**trans-factor**" query on multiple query trans-factors and to obtain annotated UTRs which are co-regulated by all the trans-factors in your query. See also section 2.2 for further details

2.4. Querying AURA by "sequence"

- Click on [*sequence*] on the navigation bar
- Type the sequence in the text box and Click on [*submit*]. Optionally, the user can set two parameters: the e-value threshold and the minimum percent sequence identity to be used by BLAST. When not set, default values will be used (e-value 10, minimum identity 0%).
This will take to the [*sequence search*] page. You may scroll the 5′ and 3′ UTRs matching your query sequence. Furthermore, you may choose the following options:

- "**Browse all matching transcript UTRs**" allows selecting the UTRs matching your query sequence which are of your interest and to access to the UTR annotations. The *blastn-short* set of parameters is employed to obtain these results.
  See section 2.1 for further details.
- "**Download results**" allows downloading, in a textual format, the matching results data, including position in UTRs, BLAST score and significance value.

2.5. Querying AURA by "batch mode"

This query mode provides multiple options:

- The "**Browser display**" option allows executing the "**target locus**" query on multiple query genes simultaneously.
    - The "**Post-transcriptional regulatory network**" option provides a network view of direct trans-factor-mRNA binding interactions which involve the UTRs associated with query genes. The expression filter allows to select one of the 48 available gene expression profiles: this will cause the display of a subset of all regulatory interactions, namely the ones for which both the regulator and the target gene are expressed in that particular tissue or cell line. The network is displayed by an interactive plugin and can also be downloaded in PDF or GraphML formats. Node size increases with the number of genes regulated by the factor represented by the node. On the other hand, network edges are weighted by the number of binding sites of the regulator (source node) on the regulated gene UTRs (destination node): this is displayed by increasing edge thickness.
- The "**Regulatory element enrichment**" option provide: (A) trans-factors whose binding sites are significantly overrepresented in your query gene list, (B) regulatory elements (e.g. ARE) which are significantly overrepresented in your query gene list.
  The expression filter allows to select one of the 48 available gene expression profiles: this will cause the computation of the enrichment p-value for only a subset of all potential regulators, namely the ones expressed in that particular tissue or cell line.
  This option displays the number and proportion of query genes which are regulated by each trans-factor or regulatory element along with the enrichment p-value adjusted for multiple testing by the *Benjamini-Hochberg* method.
  Note that, in order to allow visualization and download of results which may not be supported by statistically significant P-values, all P-values are shown but P-values above 0.05 are highlighted in red.

See also section 2.1 for further details.

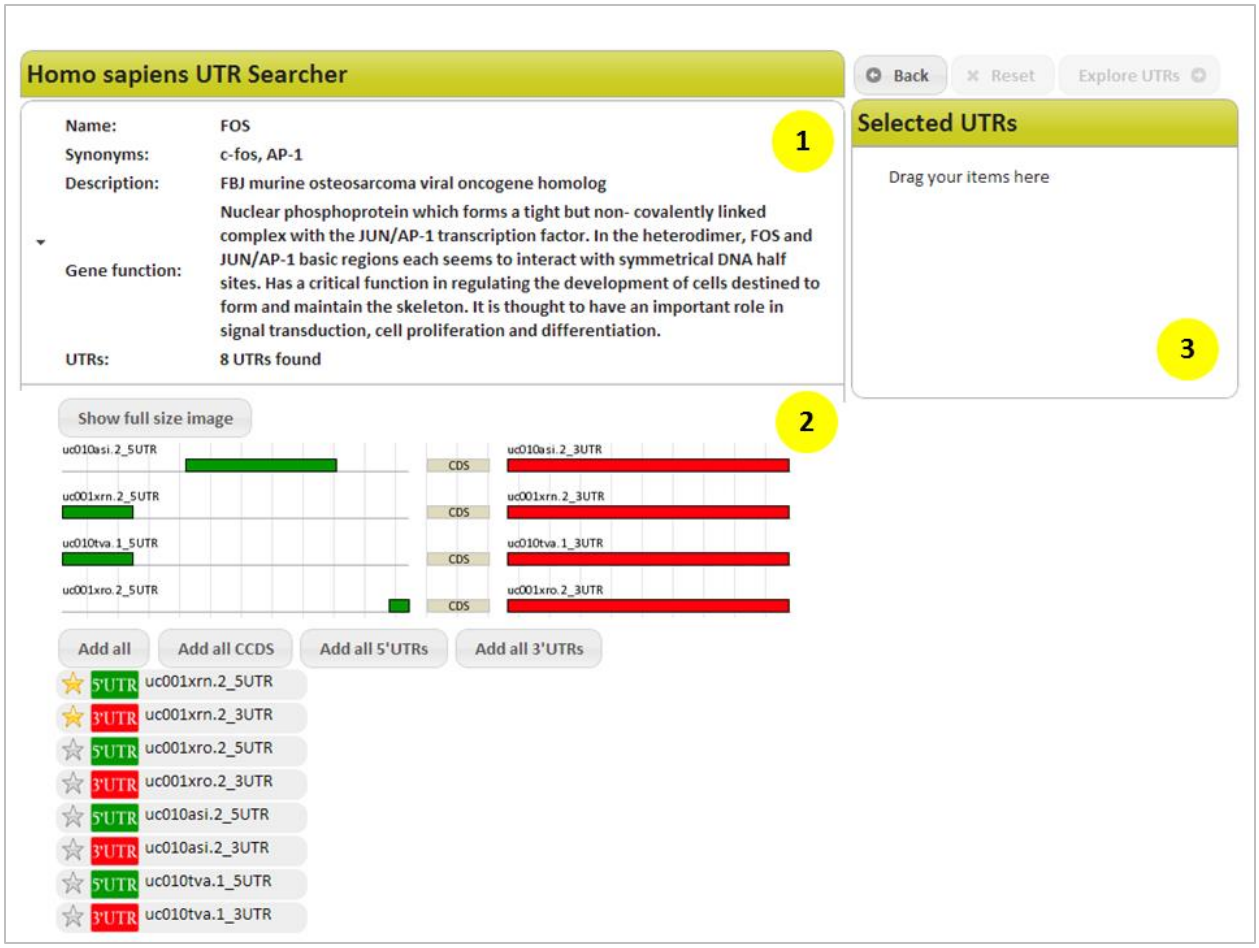
2.6. Querying AURA by "bioMart"

This query mode allows obtaining all the data contained in AURA by exploiting the power of the BioMart query system. The selectable datasets are:

- "**Gene**" provides Gene Ontology, Uncoupling and other annotations of the query genes in addition to gene-associated UTRs positional information, sequence conservation and structural information.
- "**UTR**" allows obtaining the identity of regulatory factors, regulated UTRs and the positional information of the binding sites within the UTRs. This datasets also provides UTR-associated gene level information, UTR sequence conservation and structural information, variation data (single nucleotide polymorphisms). The user may construct his query by, for instance, specifying the type and identity of the regulatory element and by specifying the technique supporting the interactions that will be returned.

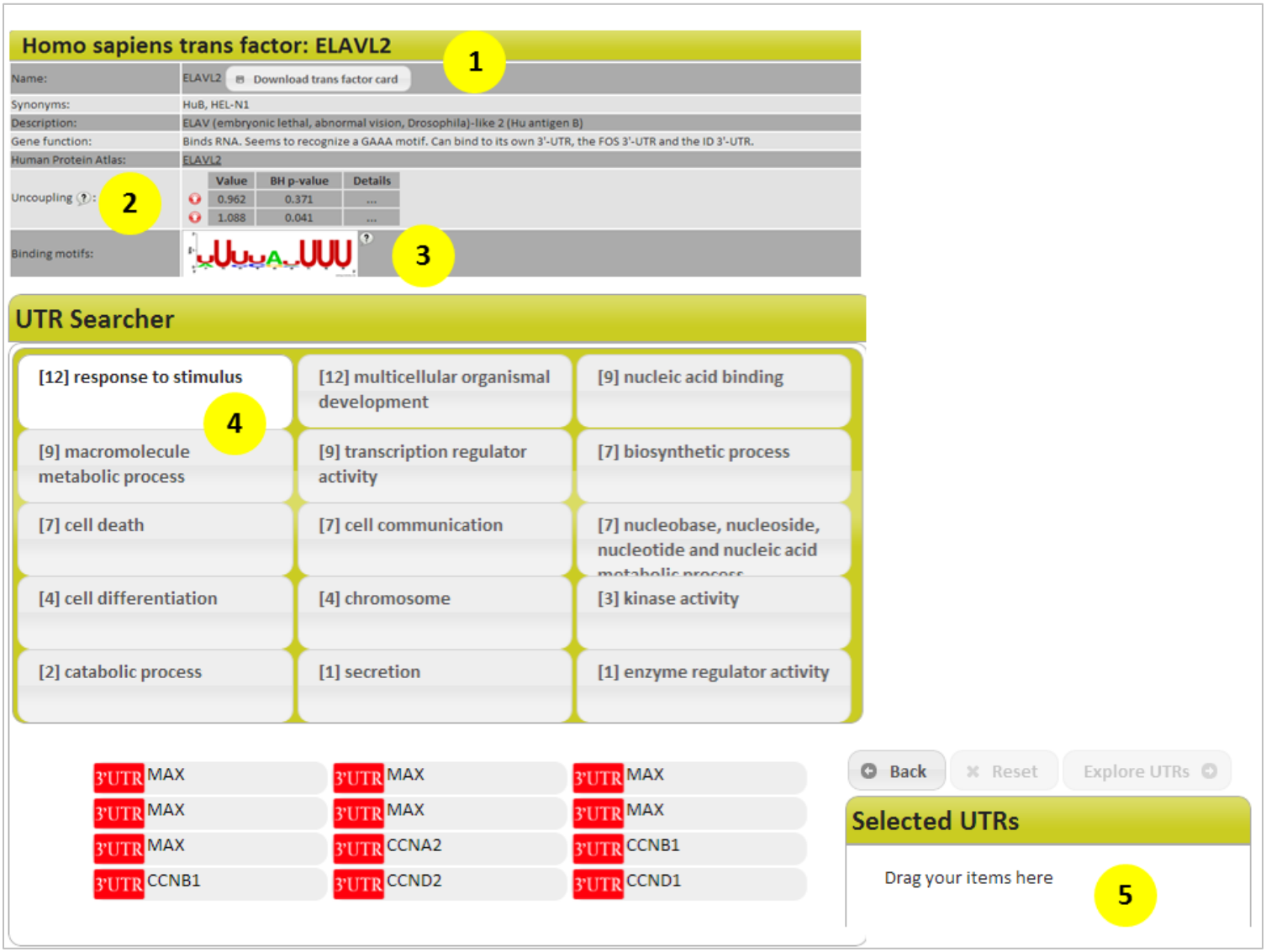
3. REFERENCE CARDS

3.1. UTR search



1. The panel contains gene-level information for the target locus that the user inputs in the query.
2. The image depicts a composite view of the exon/intron structures for the 5′ and 3′ UTRs corresponding to distinct splice variants of the target locus. UTRs marked by gold stars belong to protein-coding transcripts in the CCDS database.
3. Drag the selected UTRs from the UTR searcher panel and drop them on the selected UTRs panel which will allow access to the UTR views.

3.2. Trans-factor search



1. Click the ***download trans-factor card*** to download a copy of gene level information on the trans-factor as well as of the binding activities.
2. The image allows visualizing the quantification of uncoupling between total RNA and polysomal RNA variations.
3. The sequence logo visually depicts the sequence binding preferences of the trans-factor.
4. The UTR Searcher panel groups the UTRs regulated by the trans-factor by the criterion previously selected by the user. Click a Gene Ontology category or chromosome radio button to get the corresponding UTRs.

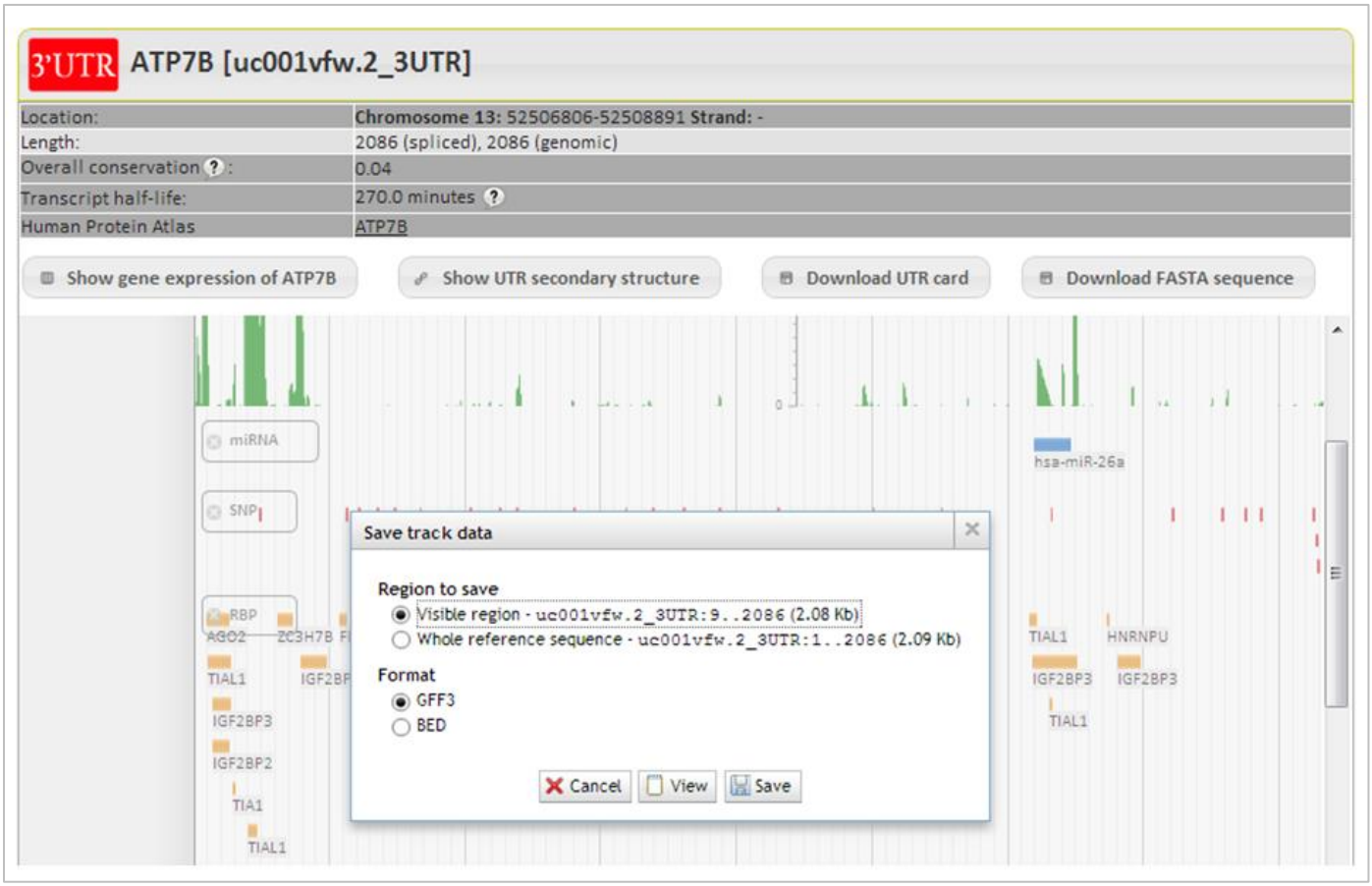
5. Drag the selected UTRs from the ***UTR searcher*** panel and drop them on the ***selected UTRs*** panel which will allow access to the UTR views.

3.3. UTR Reference Card

1. The upper right corner shows the main navigation bar that provides access to other pages. Click the ***home*** menu to return to this home page at any time and the ***help*** menu to access to detailed supporting information.
2. Click the ***reset*** menu to return to the UTR searcher page at any time.
3. The header panel displays summary-level information on UTR location, average conservation, protein expression heat map and post-transcriptional regulatory

characteristics such as expressed in terms of transcript half-life and transcriptome versus translatome uncoupling.

4. The radio button panel lists the annotations you can download a copy of. Click the ***show gene expression*** menu to get gene expression heat maps across tissues, diseases and cell lines. Click the ***show UTR secondary structure*** menu to get per base annotations within the representative UTR secondary structure. Click the download UTR card menu to download a copy of the summary-level as well as of the per base-level annotations corresponding to the UTR.
5. Click the header of any track shown in the central panel to (1) save the track data in the visible region or in the whole UTR sequence in BED or GFF3 format or (2) to simplify the UTR view by disabling the tag labels along the track (*shown in the picture below*).
6. Click on a specific annotation of a feature track to get the corresponding coordinates and direct links to supporting references.



4. DESCRIPTION OF ANNOTATION FEATURES

4.1. Displaying the general UTR annotations

The initial panel in the UTR view shows:

- UTR average conservation score
- Transcript half-life estimated in multiple cellular contexts which are described in the cited papers
- Protein expression profile by direct link to the Human Protein Atlas
- Uncoupling is displayed if the gene has been been involved in genome-wide studies profiling and comparing dynamic changes in the transcriptome and the translatome mRNA levels. Translatome levels are profiled using polysomal sucrose gradient sedimentation, followed by microarray quantification of transcripts contained in polysomal fractions. Transcriptome levels are profiled by microarray quantification of transcripts from the whole RNA cellular extract.
  Each row of the displayed table corresponds to a different study. Details on the original experiments and publications can be accessed by clicking the "Details" button. All data were extracted from public repositories of high-throughput data. For a more detailed explanation of the data, see also (PMID: 22672192).

  Uncoupling is defined as the difference in the dynamic variation of a gene in the translatome with respect to the corresponding variation in the transcriptome. It is numerically measured as the difference between the translatome log2 fold change and the transcriptome log2 fold change of the gene in the experiment. An uncoupling value corresponding to zero indicates that the gene shows exactly the same transcriptional and translational variations in the experiment. A positive uncoupling value indicates that the gene is translated more than what it would be expected looking at the variation in the transcriptome levels (the gene shows an increase in the translational efficiency, measured as the ratio of the translatome and the transcriptome levels), while a negative value suggests that the gene is translated less than what it would be expected looking at the variation in the transcriptome levels (the gene shows a decrease in the translational efficiency, measured as the ratio of the translatome and the transcriptome levels). Each uncoupling value is associated to a moderated t-test statistics p-value, calculated with the *Limma Bioconductor* package, measuring the statistical significance of the difference between the gene variations in the translatome and in the transcriptome. A *Benjamini-Hochberg* correction has been applied to the raw p-value in order to correct for multiple testing.

  A high uncoupling between variations in the translatome and the transcriptome suggests that the gene could be targeted by translational regulation mechanisms,

ultimately altering its loading on polysomes, the active translational component of the cell. Therefore the uncoupling measure becomes a powerful information when integrated with other post-transcriptional properties of the gene, for example the presence of several binding sites for proteins or miRNAs in the UTRs..

4.2. Displaying the UTR annotation tracks

The annotation tracks displayed in the image use a common set of display conventions:

Display mode of an individual annotation track: The track is displayed with each annotation feature shown separately and labelled, but not necessarily displayed into a single line. When the number of features within the requested UTR range is large, the track automatically defaults to a more tightly-packed display, where the features are shown unlabelled. In this case, you can restore the track display to full mode by narrowing the UTR range displayed.

Annotation track display control: All annotation tracks pertaining to the selected UTR are automatically displayed. To hide an annotation track, drag the side label to the left of the undesired annotation track to the left of the image window.

Zooming and scrolling the tracks display: To adjust the display, click the zoom in and zoom out buttons at the top of the window: this will change the zoom level over the annotation tracks window.
To scroll the annotation tracks sideways to the left or right of the displayed range, click the corresponding move arrow at the top of the window. To scroll the annotation tracks upwards or downwards, move the vertical side bar on the right of the image window.

Changing the order of the displayed tracks: At times you may want to vertically reposition a track in the annotation track window, click-and-hold the mouse button on the side label to the left of the track, then drag the track up or down within the image. Release the mouse button when the track is in the desired position.

4.3. Interpreting the UTR annotation tracks

There are potentially up to 9 annotation features shown in an UTR view:

- RNA-binding protein (RBP) binding site
- MicroRNA (miR) binding site
- Cis-regulatory regions without known regulatory trans-factors (either RBPs or miRNAs)
- Single nucleotide polymorphisms

- Per base conservation scores
- Alternative polyadenylation site
- Post-transcriptional RNA modifications: N6-methyladenosine (m^6A), 5-methylcytosine (m^5C), deamination of adenosine to inosine (A-to-I editing)

Display mode of an annotation feature: A feature may have two display modes that reflect the feature mapping quality allowed by the primary source of information:

- **Tagging:** The annotation features are displayed by tags along the annotation track, defining the start and end coordinates of the annotation features along the selected UTR.

- **Lining:** Unmapped annotation features are individually displayed on separate lines, spanning the selected UTR. When an annotation feature is unmapped, the track automatically defaults to this mode. The continuous fill effect shows up the cases where the annotation feature (trans-factor regulatory activity) can only be assigned to the transcript; in this situation AURA displays the feature in the transcript's 5′ and 3′ UTRs. The striped fill effect shows up the cases where the annotation feature (trans-factor regulatory activity) can only be assigned to the UTR, without finer positional details.

Enabling feature popup window: Each feature within an annotation track has an associated details window that can be displayed by clicking on its label or item. The details contained in the popup window vary by annotation track, but may include basic position information about the feature, and references to primary source of information.