

Title	The GWIPS-viz browser
Authors	Kiniry, Stephen J.;Michel, Audrey M.;Baranov, Pavel V.
Publication date	2018-05-16
Original Citation	Kiniry, S. J., Michel, A. M. and Baranov, P. V. (2018) 'The GWIPS-viz browser', Current Protocols in Bioinformatics, 62(1), e50. doi:10.1002/cpbi.50
Type of publication	Other
Link to publisher's version	https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpbi.50 - 10.1002/cpbi.50
Rights	© 2018, John Wiley & Sons Inc. This is the accepted manuscript version of the following article: Kiniry, S. J., Michel, A. M. and Baranov, P. V. (2018) 'The GWIPS-viz browser', Current Protocols in Bioinformatics, 62(1), e50. doi:10.1002/cpbi.50, which has been published in final form at https://doi.org/10.1002/cpbi.50 . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.
Download date	2025-07-03 23:56:38
Item downloaded from	https://hdl.handle.net/10468/7079

**GWIPS-viz**

Journal:	<i>Current Protocols</i>
Manuscript ID	CP-17-0308.R1
Wiley - Manuscript type:	Protocol
Date Submitted by the Author:	15-Jan-2018
Complete List of Authors:	Kiniry, Stephen; University College Cork, School of Biochemistry and Cell Biology Michel, Audrey; University College Cork, School of Biochemistry and Cell Biology Baranov, Pavel; University College Cork, School of Biochemistry and Cell Biology
Keywords:	GWIPS-viz, ribosome profiling, genome browser, translation, mRNA-seq, protein synthesis, gene expression, ribo-seq
Abstract:	GWIPS-viz is a publicly available browser that provides Genome Wide Information on Protein Synthesis through the visualisation of ribosome profiling data. Ribosome profiling (Ribo-seq) is a high throughput technique which isolates fragments of messenger RNA that are protected by the ribosome. The alignment of the ribosome protected fragments or footprint sequences to the corresponding reference genome and their visualisation using GWIPS-viz, allows for unique insights into the genome loci that is expressed as potentially translated RNA. The GWIPS-viz browser hosts both Ribo-seq data and corresponding mRNA-seq data from publicly available studies across a number of genomes avoiding the need for computational processing on the user side. Since its initial publication in 2014, over 1885 tracks have been produced across 24 genomes. This unit describes the navigation of the GWIPS-viz genome browser, the uploading of custom tracks, and the downloading of the Ribo-seq/mRNA-seq alignment data.

The GWIPS-viz browser

Stephen J. Kiniry¹, Audrey M. Michel¹ and Pavel V. Baranov^{1*}.

¹School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland

Email: stephen.kiniry@ucc.ie, A.Michel@ucc.ie, P.Baranov@ucc.ie

Tel: +353 21 420 5419; Fax: +353 21 420 5462;

* To whom correspondence should be addressed.

Significance Statement

Ribosome profiling is a high-throughput technique that captures actively translating ribosomes and isolates the short nucleic acid fragments bound by these ribosomes. Mapping sequences of these short fragments to a reference genome allows for the discovery of novel translated regions or exploration of how translation efficiency is affected under certain conditions. GWIPS-viz is a ribosome profiling dedicated genome browser that allows users to freely explore thousands of ribosome profiling datasets that were uniformly processed to produce alignments to 24 different reference genomes. This protocol describes how to explore and interpret these data within the browser and how to download the relevant data for local use.

Abstract

GWIPS-viz is a publicly available browser that provides **Genome Wide Information on Protein Synthesis** through the visualisation of ribosome profiling data. Ribosome profiling (Ribo-seq) is a high throughput technique which isolates fragments of messenger RNA that are protected by the ribosome. The alignment of the ribosome protected fragments or footprint sequences to the corresponding reference genome and their visualisation using GWIPS-viz, allows for unique insights into the genome loci that is expressed as potentially translated RNA. The GWIPS-viz browser hosts both Ribo-seq data and corresponding mRNA-seq data from publicly available studies across a number of genomes avoiding the need for computational processing on the user side. Since its initial publication in 2014, over 1885 tracks have been produced across 24 genomes. This unit describes the navigation of the GWIPS-viz genome browser, the uploading of custom tracks, and the downloading of the Ribo-seq/mRNA-seq alignment data.

Keywords (3-7): GWIPS-viz, ribosome profiling, genome browser, translation, mRNA-seq, protein synthesis, gene expression

Introduction

Since the inception of ribosome profiling in 2009 (Ingolia, Ghaemmaghami, Newman, & Weissman, 2009), the number of studies that produce ribosome profiling data has been steadily increasing. Processing these data often requires downloading large raw alignment files, removing adapter sequences along with read sequences that map to ribosomal RNA (rRNA), and then aligning the remaining read sequences to a reference genome. These steps can be quite time consuming and computationally expensive. The GWIPS-viz genome browser (Michel et al., 2014) was developed so that users could easily explore publicly-available Ribo-seq and mRNA-seq data for a large number of prokaryotic and eukaryotic genomes without the need to download, pre-process and

1 align raw sequencing data. Created as a customised mirror of the UCSC Genome Browser (Tyner et
2 al., 2017), researchers familiar with using the UCSC Genome Browser will find that using GWIPS-viz
3 is quite similar.
4
5

6 Depending on the approach used to arrest ribosomes, the Ribo-seq technique can capture
7 elongating ribosomes or initiating ribosomes - see the following reviews for a comprehensive
8 overview of the technique, its applications and limitations: (Andreev et al., 2017; Aramayo &
9 Polymenis, 2017; Bartholomaeus, Del Campo, & Ignatova, 2016; Baudin-Baillieu, Hatin, Legendre, &
10 Namy, 2016; Brar & Weissman, 2015; Calviello & Ohler, 2017; Gobet & Naef, 2017; Ingolia, 2014;
11 Michel & Baranov, 2013; Mumtaz & Couso, 2015; Parsons & Myler, 2016; Stern-Ginossar, 2015;
12 Stern-Ginossar & Ingolia, 2015). Data tracks for both initiating and elongating ribosomes are
13 hosted on GWIPS-viz. The total mRNA (or total RNA) datasets that are generated in parallel to
14 Ribo-seq are also provided in GWIPS-viz (referred to as mRNA-seq). For *Saccharomyces cerevisiae*,
15 data tracks for translation complex profile sequencing (TCP-seq) which captures scanning
16 ribosomes (Archer, Shirokikh, Beilharz, & Preiss, 2016) are also available. In addition, for each
17 genome GWIPS-viz provides an aggregate of data from multiple studies which boosts the signal of
18 genuine translation and helps reduce noise.
19
20
21

22 The basic protocol for this unit details how to navigate the GWIPS-viz browser to view and explore
23 the hosted Ribo-seq data and corresponding mRNA-seq data for a given organism along with gene
24 annotations of interest. Support protocol 1 details how a user could upload their own data as a
25 custom track to view it through the browser. Support protocol 2 describes how to use the *Table*
26 *Browser* to access the underlying database tables including gene annotations. Support protocol 3
27 describes how to use the *Downloads* page to download the Ribo-seq/mRNA-seq data alignment
28 files.
29
30
31

32 **Basic Protocol 1** 33 **Using the GWIPS-viz browser** 34 35

36 This protocol describes the use of the GWIPS-viz browser for exploring pre-populated alignment
37 tracks of publicly available Ribo-seq and mRNA-seq data. The GWIPS-viz browser can be accessed
38 at <https://gwips.ucc.ie>.
39
40

41 **Necessary Resources** 42

43 An Internet connected device such as a PC or laptop, capable of running an up-to-date Internet
44 browser that supports JavaScript, such as Firefox, Chrome or Safari.
45

- 46 1. Open the GWIPS-viz home page at <https://gwips.ucc.ie> in a web browser.
47

48 *The GWIPS-viz home page consists of 3 panels in the center of the page (About GWIPS-viz,*
49 *News and Conditions of use), a navigation bar at the top of the page (topbar) and a menu*
50 *to the left of the page. The topbar contains several links: **Genomes** for the main GWIPS-viz*
51 *Genome browser page; **Table Browser** where data can be downloaded from different tracks*
52 *in text format; **Downloads** where Ribo-seq and mRNA-seq alignment files can be*
53 *downloaded in bigWig format; **Sessions** where your current session can be saved and*
54 *shared with others; **Forum** where users can discuss various topics related to GWIPS-viz and*
55 *ribosome profiling in general; **Help** which gives a tutorial on how to use the browser.*
56
57
58
59
60

1
2
3 The menu to the left of the page also gives links to the **GWIPS-viz Genome Browser** page,
4 the **Table Browser** and the **Downloads** page. In addition it provides links to RiboSeq.Org a
5 portal for ribosome profiling (<https://riboseq.org/>), and RiboGalaxy
6 (<https://riboqalaxy.ucc.ie/>) (Michel et al., 2016), a freely-available Galaxy-based platform
7 (Afgan et al., 2016) where researchers can pre-process, align and analyse their own Ribo-
8 seq/mRNA-seq data and with the functionality to explore their data in GWIPS-viz.
9

- 10
11 2. Click on the **Genomes** link in the topbar or the **Genome Browser** link on the sidebar of the
12 home page.
13

14 The **Gateway** page will be displayed (Fig 1.). On the left of this page, the list of species
15 available in GWIPS-viz (currently 24) is displayed with icons of the most popular species
16 displayed at the top. Between these, a text box is provided for the species or common name
17 to be entered. Clicking on either one of the **Popular species** icons, entering a species in the
18 text box or clicking one of the organisms in the **Represented species** list will display the
19 corresponding assembly on the right-hand side of the page (e.g. Human assembly) with a
20 dropdown menu of assembly versions available in GWIPS-viz for that species (e.g. the hg38
21 and hg19 assemblies are available for human).
22
23

- 24 3. Enter genomic co-ordinates or the name of a gene in the **Position/Search term** text box
25 and press the **Go** button.
26
27

28 The text box labelled **Position/Search term** determines which region of the genome will be
29 seen upon hitting the submit button. A position, gene symbol or search term can be used. A
30 genomic position can be entered in the following format:
31

32 <chromosome>:<start_position>-<end_position>
33

34 For example, if you wish to view from position 110,874,784 to position 110,895,712 on
35 chromosome 6 then enter
36
37

38 chr6:110874784-110895712
39

40 in the text box.
41
42

43 A gene name (e.g AMD1) can also be entered in this box in which case a page will be
44 displayed with a list of annotations associated with the gene. Leaving the **Position/Search**
45 **term** text box empty will display either the default genomic location or the most recently
46 cached location if appropriate.
47

48 Upon pressing the **Go** button, the **GWIPS-viz Genome Browser** page will be displayed.
49

- 50
51 4. Explore the **GWIPS-viz Genome Browser** page.
52

53 The **GWIPS-viz Genome Browser** page consists of a main central window pane with a white
54 background in which the data are displayed (see Fig. 2). Data are displayed in the form of
55 one or more tracks which are stacked vertically within the main window. By default the
56 **Global aggregate for Elongating Ribosomes (A-site)** is displayed. This track is coloured red
57
58
59

1 for the majority of genomes except for bacterial genomes where strand orientation is used
2 so that reads aligning to the forward strand are coloured red and reads aligning to the
3 reverse strand are blue. By default the **Global aggregate** for **mRNA-seq** data is also
4 displayed. This track is coloured green (see Fig. 2) for all organisms except bacteria where
5 the mRNA-seq reads are coloured in a strand dependent manner similarly to the Ribo-seq
6 reads. Depending on the selected organism there will be one or more gene annotation
7 tracks displayed by default. In Figure 2, for example, the RefSeq (O'Leary et al., 2016) gene
8 annotations for human are displayed in blue.

11 How to set, use and modify the display of these data tracks along with the other data track
12 groups is explained in step 10.

15 Both above and below the main window, a number of button options exist for navigation
16 and display configuration.

- 19 5. Shift the view left/right and zoom in/out by clicking the buttons at the top of the **GWIPS-**
20 **viz Genome Browser** page. Navigate to a new position by entering co-ordinates in the
21 search box and pressing the **go** button.

23 *In order to move your view left or right along the genome you can either place your mouse*
24 *cursor in the center of the main window (where the icon will change to a hand), and click*
25 *and drag left or right. Alternatively you can use the buttons near the top of the page*
26 *labelled <<<, <<, <, >, >>, >>> which will shift the display window along the genome*
27 *upstream or downstream, in steps proportional to the size of your current display window.*

30 *The buttons at top of the page labelled **1.5x**, **3x**, **10x** and **base** are used for zooming in and*
31 ***1.5x**, **3x**, **10x** and **100x** are used for zooming out. Alternatively, if the cursor is placed at the*
32 *top of the main window, clicked and dragged left or right, a pop-up will display asking if the*
33 *user wants to zoom in to (or highlight) this region. A drop down menu within the pop up*
34 *allows you to select the colour of the highlight. Multiple highlights can also be displayed in*
35 *the same window using the **Add Highlight** button.*

38 *Use the **move start** or **move end** buttons located below the main window to increase or*
39 *decrease the breadth of the displayed co-ordinate range, or to shift one or both ends of the*
40 *co-ordinate range to the left or right.*

43 *It is also possible to jump directly to another region on the genome by entering genomic co-*
44 *ordinates or a gene name/transcript ID in the text box labelled **enter position, gene symbol***
45 *or **search term** and pressing the **go** button.*

- 47 6. Configure the parameters for displaying the data tracks on the **GWIPS-viz Genome**
48 **Browser** page.

51 *A number of button options below the main window can be used to configure the displayed*
52 *data (see Fig. 2):*

53 *Use **track search** to search for one or more specific terms in the entire set of track names,*
54 *descriptions, and track groups for the current assembly.*

55 *Click **default tracks** to reset the current display to the default data tracks for the region*
56 *currently displayed.*

1 Click **default order** to restore the order of the tracks to the default order for the current
2 assembly.

3 Click **hide all** to hide all the tracks for the current assembly.

4 Use **add custom tracks** to upload your own Ribo-seq/mRNA-seq data for exploration in the
5 context of published Ribo-seq/mRNA-seq data (see Support Protocol 1).

6 Use **track hubs** to explore tracks hosted by third parties. You can also create your own
7 private track hub.

8 Use **configure** to change the settings related to the GWIPS-viz Genome browser display
9 configuration and the presented tracks (see step 7).

10 Use **multi-region** to slice track viewing (e.g. display exonic regions only -see step 12).

11 Click **reverse** to display the reverse strand of the current location or the forward strand if
12 the reverse is currently displayed.

13 Click **resize** to restore the default window size.

14 Click **refresh** to update the window display if changes have been made to the track settings.

15
16
17
18
19 Note to reset to the overall default settings, including the default genomic locations for all
20 genomes, mouse-over **View** in the topbar in the GWIPS-viz Genome Browser page and click
21 on **Reset All user Settings**. Alternatively, from any page, mouse-over **My Data** in the top
22 bar and click on **My Sessions**. Click on the [Click here to reset](#) link in the **Session**
23 **Management** page.

- 24
25
26 7. Click the **Configure** button below the main window to open an options page for configuring
27 the main window image display and the data tracks display. Click the **Submit** button on the
28 configuration page to apply the changes and return to the **GWIPS-viz Genome Browser**
29 page.

30
31 In the **Configure image** section, options exist to modify the image width, label area width,
32 and text size. A number of checkbox options are also available to control the display
33 characteristics.

34
35
36 The **Configure Tracks on GWIPS-viz** section includes options to show or hide data tracks
37 from the different GWIPS-viz track groups (see step 10).

38
39 Note that in addition to the **Configure** button below the main window in the GWIPS-viz
40 Genome Browser page, the configuration page can be accessed by placing the mouse cursor
41 over the **Genome Browser** tab in the topbar and selecting **Configure**. Another approach is
42 to mouse-over the **View** tab in the topbar and select **Configure Browser**.

- 43
44
45 8. Explore the GWIPS-viz data track groups.

46
47 On the GWIPS-viz Genome Browser page the track groups are denoted by green rectangular
48 banners with the group name in white. Listed below are the currently available GWIPS-viz
49 track groups. Depending on the genome, the available track groups will differ as it depends
50 on the type of ribosome profiling data available for that genome. But at a minimum, all
51 genomes will contain elongating ribosome profiling data along with mRNA-seq.

52
53
54 **Initiating Ribosomes (P-site)** in which treatments are used to capture initiating ribosomes,
55 with an expected peak enrichment at translation initiation codons. The P-site in the group

1 name reflects that each read is given an offset that approximates the P-site of the ribosome
2 at that position. The initiator tRNA is recognized in the P-site during initiation.

3 **Elongating Ribosomes (A-site)** captures elongating ribosomes and an offset is applied to
4 approximate the A-site locations. The recognition of tRNAs during elongation of the nascent
5 peptide occurs in the A-site.

6 **Elongating Ribosomes (Footprints)** is also used for Ribo-seq data but with every position of
7 the read represented. No offset is applied.

8 **mRNA-seq Reads** displays the mRNA-seq data generated in parallel to the Ribo-seq data (if
9 available). Every position in the read is represented.

10 **Annotation Tracks and External Data** is used for displaying gene annotations specific to the
11 assembly such as RefSeq (O'Leary et al., 2016), Ensembl (Aken et al., 2016), Gencode
12 (Harrow et al., 2012), SGD (Cherry, 2015), FlyBase (Gramates et al., 2017), TAIR (Berardini
13 et al., 2015), MaizeGDB (Andorf et al., 2016). This track group also hosts other external
14 data including conservation data such as phyloP (100 way) (Pollard, Hubisz, Rosenbloom, &
15 Siepel, 2010).

16
17
18
19
20 The **Small Ribosome Subunits (Footprints)** track group for scanning ribosome data (Archer
21 et al., 2016) is currently only available for *Saccharomyces cerevisiae*. It is likely, however,
22 that this type of data will soon be generated for other organisms (Andreev et al., 2017) and
23 our intention is to integrate it into GWIPS-viz as it becomes available.

24
25
26 9. Explore the individual Ribo-seq and mRNA-seq data tracks.

27
28 Each Ribo-seq and mRNA-seq data track group provides the names of the published studies
29 from which the Ribo-seq/mRNA-seq data originated. For all genomes, the first author name
30 and publication year are provided for each study. For human, mouse and rat, the cell line or
31 tissue for which the Ribo-seq data was generated is also included in the study label.

32
33
34 Once the study name is clicked a new page showing that study's **track settings** will be
35 displayed with details specific to that study. This consists of the title of the study at the top
36 of the page, followed by the option to view the data quality using RUST plots (see step 13),
37 followed by a number of options related to viewing the tracks (called subtracks) for that
38 study. Below the **list subtracks** radio button there will be one checkbox for each individual
39 data track for the study and one checkbox for the aggregate of **All** tracks within the study.
40 When checked, the track will be displayed on the main **GWIPS-viz Genome browser** page.

41
42
43 In addition, a **Global aggregate** track is provided for each Ribo-seq and mRNA-seq track
44 group. A **Global aggregate** track contains the **All** data tracks from the different studies
45 aggregated into a single display track. The advantage of aggregating all tracks is that
46 inconsistent individual study biases tend to be diluted while consistent genuine translation
47 signals tend to be aggregated into stronger signals. It needs to be noted that such
48 aggregation does not guard against systematic biases pertinent to the majority of the
49 datasets.

50
51
52 Typically the **Global aggregate for Elongating Ribosomes (A-site)** and **Global aggregate**
53 **for mRNA-seq Reads**, along with the gene annotations, are the individual data tracks that
54 are displayed by default for each genome. This helps you to first determine if/where
55 translation occurs in your gene or genomic region of interest before refining your search.

1 You can then determine the study or studies from which the Ribo-seq reads originated by
2 clicking the **Global aggregate** track name which opens a new page called **Global Aggregate**
3 **Track Settings** where you can view the **All** tracks specific to each study. The individual **All**
4 tracks can be viewed by ticking the corresponding checkbox and setting the **Display mode**
5 to **full**.
6
7

8 As the **Global aggregate** tracks are updated when new datasets are added, a list of the
9 studies contributing to the current versions of each Global aggregate track (e.g. version 1) is
10 provided in the **Description** section of the **Global Aggregate Track Settings** page. Also
11 included is a link to the [global aggregate log file](#) which provides the study list of previous
12 versions.
13
14

- 15 10. Select one or more of the individual track display checkboxes. Set the **Display mode**
16 dropdown menu to **full**. Press **submit**.
17

18 *Dropdown menus for setting the track displays are available below each study's entry on*
19 *the GWIPS-viz Genome Browser page as well as for the Global aggregate tracks. These*
20 ***Display mode** menus are also on each study's track settings page beside the subtrack*
21 *checkboxes.*
22
23

24 *For example, to display the individual **All** tracks for each study in the **Global Aggregate***
25 ***track settings** page, select one or more of the checkboxes and set the **Display mode** option*
26 *to **full** in the dropdown menu. Click on the **submit** button (next to the Display mode*
27 *dropdown at the top of the page) to display the **GWIPS-viz Genome Browser** page. New*
28 *tracks should now be visible corresponding to those selected on the Global Aggregate track*
29 *settings page.*
30
31

32 *There are 5 options in the **Display mode** dropdown menu: **hide**, **dense**, **squish**, **pack** and*
33 ***full**. The **dense** mode conveys differences in peak height through shades of gray (darker*
34 *regions correspond to higher peaks); **squish** mode which is the same as **dense** but more*
35 *compact; **pack** mode which is the same as **full** but more compact. Usually the **full** mode is*
36 *used for displaying the Ribo-seq and mRNA-seq data tracks.*
37
38

39 *Note that the above approach is also used for displaying Ribo-seq/mRNA-seq data for each*
40 *study's individual conditions.*
41
42

- 43 11. Configure the annotation track display in the **Annotation Tracks and External Data** track
44 group.
45

46 *Depending on the selected organism there will be one or more gene annotation tracks*
47 *displayed by default. These vary depending on the organism. For example, for human,*
48 *RefSeq (O'Leary et al., 2016) and Gencode (Harrow et al., 2012) annotations are displayed*
49 *by default.*
50
51

52 *Gene annotations are graphically represented by bars of varying thickness: the thickest bars*
53 *represent annotated protein coding regions, the thinner bars represent 5' leaders (5' UTRs)*
54 *and 3' trailers (3' UTRs), while the thinnest bars with arrows represent introns, with the*
55 *arrows indicating the direction of transcription and translation.*
56
57
58
59
60

1 To configure the gene annotations display, click on the gene annotations database name in
2 the **Annotation Tracks and External Data** track group. This will open that particular gene
3 annotations **track settings** page where a number of configuration options will be provided.
4 The main configuration option is the **Display mode** which is set in the same way as the data
5 tracks (see step 10).
6
7

8 In addition to gene annotations, the **Annotation Tracks and External Data** track groups
9 contain additional external data that may help in the interpretation of the Ribo-seq data.
10 The type of external data available depends on the genome. For example, for the human
11 hg38 assembly, phyloP 100-way conservation data (Pollard et al., 2010) is provided.
12
13
14

15 Optional Steps:

- 16
17
18 12. Use the **multi-region** option to hide intronic regions.

19
20 *Many genes, particularly in higher eukaryotes, can contain numerous intronic regions or*
21 *extremely long intronic regions which makes visual interpretation of the Ribo-seq data*
22 *difficult. To alleviate this problem the **multi-region** option can be used. This option can be*
23 *accessed by clicking on the **multi-region button** below the main window in the GWIPS-viz*
24 *Genome Browser page or mouse-over the **View** tab in the topbar and select **Multi-region**.*
25

26
27 *When selected, this option will show a pop-up with 4 radio buttons, the first of which is*
28 *simply **Exit multi-region mode**, the others are as follows:*
29

30 **Show exons using RefSeq Genes.** *Note the gene annotation database proposed will vary*
31 *according to the default database for each genome. When the option is selected, intronic*
32 *regions will be hidden (see Fig. 7 in the section guidelines for understanding results). The*
33 *text box to the right allows you to enter a padding between exons. So if left at the default*
34 *value of 6 there will be a gap of 6 nucleotides between all exons. We recommend putting a*
35 *value of 0 here. We wish to mention that this option works best for gene loci that have few*
36 *transcript isoform predictions. The display can, however, be improved further by entering*
37 *custom regions (see below).*
38
39

40 **Show genes using RefSeq Genes.** *When this option is selected, intergenic regions will be*
41 *hidden meaning neighbouring genes will appear as if directly next to each other on the*
42 *chromosome with only the specified padding between them.*
43
44

45 **Enter Custom regions as BED, or a URL to them.** *This section allows you to enter multiple*
46 *genomic regions in the provided text box. Only the regions supplied by you will be displayed.*
47 *This option can be particularly useful when trying to view the Ribo-seq data for a specific*
48 *transcript isoform when a gene locus has many isoforms. The genomic regions can be*
49 *entered in the format*
50

51
52 `<chromsome_name> <start_position> <end_position>`
53

54 *with a space separating each. If multiple regions are required each should be given its own*
55 *line in the text box:*
56
57
58
59
60

1 <chromsome_name1> <start_position1> <end_position1>
2 <chromsome_name2> <start_position2> <end_position2>
3 <chromsome_name3> <start_position3> <end_position3>
4
5
6
7

8 *Beneath the 4 radio buttons lies a checkbox labelled **Highlight alternating regions in multi-***
9 ***region view** which if checked will colour every second region in blue, allowing for easy*
10 *visual interpretation of exon junctions. When you have made your selection press the*
11 ***submit** button to view the relevant changes.*
12
13

14. Explore the RUST profiles for individual tracks.

15
16 *This is an optional step to explore the quality of individual Ribo-seq datasets by viewing the*
17 *profiles that have been generated using the **Riboseq Unit Step Transformation** (or RUST) tool*
18 *(O'Connor, Andreev, & Baranov, 2016). RUST profiles are available for the majority of recent*
19 *Ribo-seq datasets generated in eukaryotes and can be viewed by clicking on the label for a*
20 *particular study in the **Elongating Ribosome (A-site)** track group in the **GWIPS-viz Genome***
21 ***Browser** page. This will open the study's **track settings** page near the top of which will be a*
22 *button labelled **View Data Quality (RUST plots)**. Clicking this button will display a pdf file*
23 *where each Ribo-seq file from that study will have its own page as in Figure 3. Each page*
24 *consists of the following plots:*
25
26

27
28 **Metafootprint profile with Kullback-Leibler divergence** (top panel). The RUST
29 metafootprint profiles in GWIPS-viz show the differences between the observed and
30 expected codon RUST statistics (reflecting frequencies of codons) in the region of the
31 ribosome footprint for each Ribo-seq dataset. The Kullback-Leibler (K-L) divergence of RUST
32 statistics ratios of all 61 codons at a single codon position is represented by the blue line.
33 The green line shows the K-L divergence at adjacent codons. Typically the highest K-L values
34 (reflecting variation in codon frequencies) is expected to occur at the ribosome decoding
35 center (A- and/or P- site). The position of the A-site is indicated with the red vertical line,
36 see Figure 3. If a high variation occurs near the ends of the footprints, it likely represents
37 sequencing biases (Fig. 3). In this way, the metafootprint profiles can be used to assess
38 biases in individual datasets. In the top-right of this panel, the total number of mapped
39 reads for the individual dataset is also displayed.
40
41

42
43 **Relative inferred A-site dwell time** (bottom left panel). The y-axis shows the ratio of
44 ribosome occupancy (relative to the minimum occupancy) i.e. the amount of time that the
45 ribosome spends at that codon (dwell time). The x-axis groups the codons according to their
46 corresponding amino acids.
47

48
49 **Triplet Periodicity** (bottom right panel). This plot shows the Ribo-seq counts for each of the
50 three possible reading frames, here labelled zero, one and two. These are broken down by
51 read length which is displayed on the x-axis. Ideally the triplet signal of the decoding
52 ribosome should be strong in Ribo-seq data (Bazzini et al., 2014; Ingolia et al., 2009; Michel
53 et al., 2012). The more random the distribution of footprints across the sub-codon positions,
54 the weaker the periodicity signal.
55

14. Use the **PDF/PS** option to generate images of GWIPS-viz data tracks.

1
2
3 *This is an optional step to generate images of GWIPS-viz data tracks for use in manuscripts,*
4 *posters, etc. In the **GWIPS-viz Genome Browser** page, mouse-over the **View** tab and click on*
5 *the **PDF/PS** option.*

6
7 *In the **PDF Output** page, click on the link to download **the current browser graphic in PDF***
8 *format. Another option is to download the graphic in EPS format.*
9

10 11 12 **Support Protocol 1: Uploading a custom track**

13
14 The **custom track** functionality allows users to upload their own Ribo-seq and/or mRNA-seq data
15 to GWIPS-viz and explore their data in the context of published data. Only the user that uploads a
16 custom track can view it (unless the user shares it with collaborators). In other words, the custom
17 track is not accessible to anybody but the person who uploads it. Typically the uploaded data will
18 be Ribo-seq data, mRNA-seq data or gene annotation data. These tracks are only temporary, being
19 automatically removed after 7 days.
20

21
22 The following support protocol will cover the steps needed to upload a custom track.
23

24 ***Necessary Resources***

25
26 An Internet connected device such as a PC or laptop, capable of running an up-to-date Internet
27 browser that supports JavaScript, such as Firefox, Chrome or Safari.
28

- 29
30
31 1. Before uploading your alignment data as a custom track, convert your alignment file to
32 either the bedGraph format or the compressed bigWig format if your file is large.
33

34
35 *As the custom track feature is inherited from the UCSC Genome Browser (Tyner et al.,*
36 *2017), a number of data formats are supported. Alignment files in SAM format, or the*
37 *compressed BAM format, can be displayed as custom tracks. Their display, however, will be*
38 *different to the public Ribo-seq and mRNA-seq tracks in GWIPS-viz. For this reason we*
39 *recommend converting alignment data to the bigWig format as it is suitable for dense,*
40 *continuous data that will be displayed as a graph in the browser. It is also the format used*
41 *for all public alignment tracks in GWIPS-viz. The bedGraph format can also be used if the*
42 *alignment file is not too large. However, converting to the compressed bigWig format will*
43 *decrease the time taken to upload the file to GWIPS-viz.*
44

45
46 *One way to generate a bigWig file is to use RiboGalaxy (<http://ribogalaxy.ucc.ie>) (Michel et*
47 *al., 2016) which is a Galaxy-based platform (Afgan et al., 2016) for pre-processing, mapping*
48 *and analysing Ribo-seq and corresponding mRNA-seq data through a browser. The ‘Convert*
49 *formats’ suite of tools in RiboGalaxy contains utilities to convert a SAM alignment file to*
50 *BAM format and then convert the BAM file to bigWig. There are also a number of published*
51 *workflows in RiboGalaxy that are tailored for the most commonly used genomes in GWIPS-*
52 *viz and automatically output the alignment files in bigWig format with a direct link for*
53 *exploration as a GWIPS-viz custom track.*
54
55
56
57
58
59
60

Alternatively command line utilities exist for converting the BAM format to bed format and then from bed to bigWig (e.g. <http://hgdownload.soe.ucsc.edu/admin/exe/>).

2. Click on the **add custom tracks** button either on the **GWIPS-viz Genome Browser** page or mouse-over the **My Data** tab in the topbar and select **Custom Tracks**.

There are multiple ways to upload a custom track (Fig. 4). Here we will focus on the standard approaches for uploading bedGraph and bigWig file formats, the most suitable for exploring Ribo-seq data. For other file formats, we recommend checking the documentation on the GWIPS-viz custom track help page <https://qwips.ucc.ie/goldenPath/help/customTrack.html> and also <https://genome.ucsc.edu/goldenpath/help/customTrack.html>, since the custom track functionality is inherited from the UCSC Genome Browser.

First select the **clade**, **genome** and **assembly** from the dropdown menus at the top of the Custom Track Management page.

- When uploading a bedGraph file, use the **Choose file** button to navigate to the file on your computer and click the **Submit** button.
- A bigWig file can only be uploaded via a URL. The file (e.g bigWigExample.bw) needs to be copied to a web-accessible http, https, or ftp location. Paste the URL into the custom track text box and click **Submit**. Alternatively, construct a track line in the text box that references the bigWigExample.bw file:

```
track type=bigWig name="Example One" description="A bigwig file"
bigDataUrl=https://qwips.ucc.ie/goldenPath/help/examples/bigWigExample.bw
```

You can customize this track line further and define certain parameters on how your custom data should display (see <https://qwips.ucc.ie/goldenPath/help/customTrack.html#TRACK>).

3. Then explore the Ribo-seq and mRNA-seq custom track data as per the instructions outlined in the **Basic Protocol: Using the GWIPS-viz browser**.

Custom tracks can be shared with other users using a custom track URL. Note that the functionality to share a custom track as a Session is not available on GWIPS-viz. An alternative option is to create a track hub (see <https://qwips.ucc.ie/goldenPath/help/hqTrackHubHelp.html>) as a track hub may be shared privately with colleagues or shared publicly.

Support Protocol 2: Using the Table Browser

The GWIPS-viz Table Browser can be accessed at <https://qwips.ucc.ie/cgi-bin/hqTables>.

The GWIPS-viz Table Browser is a graphical user interface which provides the ability to download tracks from the GWIPS-viz browser. This includes Ribo-seq and mRNA-seq data that has already been processed and aligned to a reference genome, as well as annotation data. It also provides

1 the ability to apply filters to this data such as returning specific chromosomes. There are a number
2 of output formats available depending on the type of data being retrieved.
3

4 **Necessary Resources**

5
6 An Internet connected device such as a PC or laptop, capable of running an up-to-date Internet
7 browser that supports JavaScript, such as Firefox, Chrome or Safari.
8
9

- 10
11
12 1. Click on the **Table Browser** link in the topbar or in the navigation bar on the left hand side of the
13 GWIPS-viz homepage. Alternatively on the GWIPS-viz Genome Browser page, mouse-over the
14 **Tools** tab in the topbar and click on **Table Browser**.
15

16 *This will bring you to a new page which consists of two parts. The top half of the page is the*
17 **Table Browser** itself and is made up of multiple inputs allowing you to customise the output
18 to suit your needs (Fig. 5). The bottom half of the page provides explanations of each the
19 **Table Browser** parameters.
20

- 21
22 2. Select the **clade**, **genome**, and **assembly**.
23

24 *The **Table Browser** settings will be set to the same settings last used on the **Gateway** page.*
25 *Use the **clade**, **genome** and **assembly** dropdown menus, if required, to select the organism*
26 *of interest.*
27

- 28
29 3. Select the **group**, **track**, and **table** of interest using the dropdown menus.
30

31 *The **group** is the list of track groups as they appear on the main **GWIPS-viz Genome***
32 ***Browser** page. So, for example, if you wanted to access data for elongating ribosome*
33 *profiles, you would set the **group** to **Elongating ribosomes (A-site)**. The track is the name as*
34 *it appears within the **track group** on the **GWIPS-viz Genome Browser** page. For the Ribo-*
35 *seq and mRNA-seq groups, the list of tracks will be the names of hosted studies.*
36
37

38 *The **table** is the name of the table file you wish to download. If you had selected a study in*
39 ***track** dropdown menu, then the table dropdown menu would be a list of individual tracks*
40 *associated with that study, as well as an **All** table which is the aggregation of all Ribo-seq or*
41 *mRNA-seq tracks (depending on which group is selected) for that study.*
42
43

- 44 4. Specify the **query** region by selecting the **genome** or **position** radio buttons.
45

46 *There are two radio buttons in this field, one labelled **genome** the other labelled **position**. If*
47 *the **genome** radio button is selected any queries submitted will return data from all regions*
48 *of the genome for the currently selected organism. Alternatively if set to **position** any*
49 *submitted queries will only return data from the region specified in the text box adjacent to*
50 *the **position** radio button. Positions should be entered in the form*
51

52
53 *<chromosome name>:<start_position>-<end_position>*
54

55 *for example:*
56
57
58
59
60

1 *chrX:151283001-151290000*

2
3
4 *If the exact genomic co-ordinates are not known a gene name or a transcript ID can be*
5 *entered in the **position** text box. Pressing the **lookup** button to the right of this box will*
6 *display another page with a list of genomic co-ordinates. Clicking on one of these will go*
7 *back to the **Table Browser** with the relevant co-ordinates automatically entered into the*
8 ***position** text box.*

9
10
11 *The **define regions** button allows the user to enter multiple regions (one per line, to a limit*
12 *of 1000) either by using a text box or by uploading a text file.*

13
14 5. Choose an **output format** from the dropdown menu and enter a filename in the **output file**
15 field.

16
17
18 *The **output format** field consists of a dropdown menu to specify the format that the output*
19 *will take and two radio buttons to specify where to send the file if appropriate. The type of*
20 *formats available in the dropdown menu will vary depending on the selected table. For*
21 *example, if a Ribo-seq or mRNA-seq track is chosen, the type of output formats available*
22 *will be **data points**, **bed format** or **custom track**. **Bed format** will output a tab separated file*
23 *of chromosomal positions where reads are found while **data points** will output the same*
24 *but with the number of reads at each position also included. The **custom track** option*
25 *allows you to save the output as a custom track (See support protocol 1: **Uploading a***
26 ***custom track**).*

27
28
29 *Choosing a table in the annotation group will give more options for the output format (e.g.*
30 *FASTA format or GTF format). By choosing **selected fields from primary and related tables***
31 *the exact columns in the output file can be specified. The two radio buttons labelled **Galaxy***
32 *and **Great** allow you to send the output directly to a Galaxy instance that you are logged*
33 *into e.g. RiboGalaxy (Michel et al., 2016) or to the Genomic Regions Enrichment of*
34 *Annotations Tool (Great) web server (McLean et al., 2010) respectively for further*
35 *processing.*

36
37
38 *The filename of the output when downloaded will be named according to the text box in*
39 *the **output file** field. This field can also be left blank, in which case the output will appear in*
40 *the browser. In the **file type returned** field choose either **plain text** or **gzip compressed** to*
41 *specify if the output file should be compressed or not. For large files it will be quicker to*
42 *download the compressed version of the file and then decompress once downloaded.*

43
44
45 6. Hit the **get output** button.

46
47
48 *In some cases the file will start downloading (or displaying in the browser) once you press*
49 *this button, however in certain cases you will be brought to another page. For example,*
50 ***selected fields from primary and related table** will bring you to another page to select the*
51 *desired fields, clicking the **get output** button on this page will then start the download.*

52
53 **Optional steps:**

54
55 7. Press the **Create** button next to **filter**.

1 Like the region **position** filter in step 4, this can be used to restrict the information returned
2 by a query. The type of filters available depends on the type of table selected, for example if
3 downloading Ribo-seq data you can return only counts greater than 100, or if downloading
4 annotation data only download genes with multiple exons.
5
6

7 8. Press the **Create** button next to **subtract merge**.

9 The **subtract merge** option allows the user to concatenate data from a number of different
10 tables within a particular track. This is done by using the checkboxes to select all files
11 needed and then choosing a merge operation which defines how the values will be
12 combined. For instance, if the **average** operation is chosen the final output at each position
13 will be an average of all selected tracks. The user can also apply a filter at this stage, for
14 instance discarding any positions with scores less than a minimum threshold after the
15 merge operation has been applied.
16
17

18 19. Press the **Create** button next to **intersection**.

20 The **intersection** option allows the user to specify conditions that multiple tracks must fulfil
21 and then return only the positions that match both criteria (like a Venn diagram).
22
23

24 25. Press the **Create** button next to **correlation**.

26 The purpose of this step is to determine the correlation between two tracks. Once clicked
27 you will be brought to another page where you will pick another table using the **group**,
28 **track** and **table** dropdown menus. Then select how many data points to correlate using the
29 dropdown menu after **Limit total data points in result**. The maximum possible value here is
30 300,000,000 data points due to memory constraints. You can also set the window size (in
31 bases) using the text box in **Window data to: x bases**. Where x is a user defined value of at
32 least one and no upper limit, if set to a value of one the correlation will be based on data
33 points from every base pair. If set to a value higher than 1 correlations will be based on
34 average data points from every window of size x.
35
36
37

38 Upon hitting **submit** a table will display where the rows are the chromosomes of the
39 selected organism. The columns are various values related to the correlation between the
40 two files in this chromosome, for example, r^2 , correlation coefficient, mean, variance. The
41 final row in this table will be the overall values for all chromosomes. Four plots will also be
42 produced beneath the table. The first is a scatter plot of the counts between the two files.
43 The second is a plot of the residuals vs. fitted values. And the final two are histograms of
44 both files showing the counts of each.
45
46
47
48
49
50

51 **Support Protocol 3: Using the Downloads page**

52 The GWIPS-viz Downloads page can be accessed at <https://qwips.ucc.ie/downloads/index.html>.
53

54 The GWIPS-viz **Downloads** page is a graphical user interface which provides the ability to
55 download the GWIPS-viz Ribo-seq and mRNA-seq alignment files in bigWig format.
56
57
58
59

Necessary Resources

An Internet connected device such as a PC or laptop, capable of running an up-to-date Internet browser that supports JavaScript, such as Firefox, Chrome or Safari.

1. Click on the **Downloads** link in the topbar or in the navigation bar on the left hand side of the GWIPS-viz homepage.

*This will bring you to the **Download alignment files** page (Fig. 6). This page contains multiple dropdown menus so that you can specify the alignment file(s) you want to download.*

2. Select the **clade**, **genome**, and **assembly**.

*Use the **clade**, **genome** and **assembly** dropdown menus to select the genome assembly of interest.*

3. Select the **group** for the type of Ribo-seq or mRNA-seq data that you wish to download.

*Depending on the genome selected, Ribo-seq and mRNA-seq alignment files will be available in one or more of the following track **groups**:*

***Initiating Ribosomes (P-site)** in which treatments are used to capture initiating ribosomes, with an expected peak enrichment at translation initiation codons. The P-site in the group name reflects that each read is given an offset that approximates the P-site of the ribosome at that position.*

***Elongating Ribosomes (A-site)** captures elongating ribosomes and an offset is applied to approximate the A-site locations.*

***Elongating Ribosomes (Footprints)** is also used for Ribo-seq data but with every position of the read represented. No offset is applied.*

***mRNA-seq Reads** contain the mRNA-seq data generated in parallel to the Ribo-seq data (if available). Every position in the read is represented.*

***Small Ribosome Subunits (Footprints)** contains data generated for scanning ribosomes. Every position in the read is represented. This track group is currently available for *Saccharomyces cerevisiae* only.*

4. Select the **track** for the study data that you wish to download.

*The **track** dropdown menu provides the names of the published studies from which the Ribo-seq/mRNA-seq data originated. For all genomes, the first author name and publication year are provided for each study. For human, mouse and rat, the cell line or tissue for which the Ribo-seq data were generated is also included in the track study label.*

5. Select the **table** that represents the alignment file you wish to download.

1 The **table** dropdown menu provides the names of the individual Ribo-seq/mRNA-seq tracks
2 hosted on GWIPS-viz. Each table name corresponds to a single track and hence a single
3 alignment file.
4

5
6 An alignment file called **All** that contains the aggregation of all the individual track
7 alignment files will also be available for each study. This is not to be confused with the **All**
8 **tables** entry in the same dropdown menu which provides the option to download all the
9 individual alignment files in a single click operation.
10

11
12 In addition, the alignment files for presenting the **Global aggregate** tracks in the **GWIPS-viz**
13 **Genome Browser** page are available for download. A **Global aggregate** track contains the
14 **All** data tracks from the different studies aggregated into a single display track (see **Basic**
15 **Protocol Using the GWIPS-viz browser**).
16

- 17
18 6. Click on the **Download** button to start downloading the selected alignment file(s).
19

20 The alignment files are in bigWig format. This is a compressed format that is useful for
21 displaying dense, continuous data like Ribo-seq and mRNA-seq data in a web browser
22 (<https://qwips.ucc.ie/goldenPath/help/bigWig.html>). A number of tools exist for
23 downstream analysis of bigWig files including bwtool (Pohl & Beato, 2014) and deeptools2
24 (Ramirez et al., 2016).
25
26
27

28 GUIDELINES FOR UNDERSTANDING RESULTS 29

30 In general if viewing an area of the genome with high Ribo-seq density, it is likely that this region is
31 undergoing translation. However as with all high throughput sequencing data, ribosome profiling
32 is susceptible to noise. We try to address this by aggregating the available Ribo-seq data into
33 **Global aggregate** tracks. We recommend using the Global aggregate tracks as a first option to
34 determine if translation occurs in the region of interest. Then you can turn on/off individual study
35 Ribo-seq tracks to refine your exploration.
36
37

38 To illustrate with an example, we provide the ribosome profile from GWIPS-viz of the human
39 mitochondrial elongation factor 1 gene (*MIEF1*) (Fig. 7) which encodes the mitochondrial dynamics
40 protein MID51. Panel A of Figure 7 shows the Ribo-seq profile (red) of the human *MIEF1* gene
41 taken from the **Global aggregate** track in the **Elongating Ribosomes (A-site)** track group and the
42 mRNA-seq (green) taken from the **Global Aggregate** of the **mRNA-seq Reads** track group in the
43 human hg38 assembly. Also shown are the **phyloP** track (blue for positive scores and brown for
44 negative scores) and the **RefSeq** gene annotation track (blue), both in the **Annotation Tracks and**
45 **External Data** track group. The direction of the arrows in the RefSeq annotated intronic regions for
46 this locus shows the direction of transcription and also translation and so the 5' end of the
47 transcript is on the left-hand side and the 3' end of the transcript is on the right-hand side in
48 Figure 7.
49
50

51
52 Exploring the Ribo-seq profile for the human *MIEF1* gene in panel A shows a much higher Ribo-seq
53 density in the annotated coding regions (thickest blue bars) compared to the 3' trailer of this gene.
54 This correlates well with data from the phyloP track which also has positive conservation peaks
55 which are relatively higher in the coding region than in the 3' trailer.
56
57
58
59
60

1 The 5' leader region of the *MIEF1* gene, however, reveals that the Ribo-seq density is even more
2 prominent than in the coding region, even though none of the RefSeq annotated transcript
3 isoforms for the human *MIEF1* gene show this as a coding region. Interestingly, the phyloP track
4 also shows that this region is conserved, providing additional supporting evidence of translation in
5 this region.
6
7

8 Due to the introns within the 5' leader, interpretation of the Ribo-seq data in GWIPS-viz is
9 somewhat difficult. To improve this the multi-region view can be used. As the *MIEF1* locus has
10 multiple isoform predictions, not all of the intronic regions can be collapsed with this functionality.
11 Nevertheless, the interpretation can be made easier: Panel B of Figure 7 shows the human *MIEF1*
12 gene but with the view specifically adapted for the exons of one of the *MIEF1* transcript isoforms,
13 i.e. the third isoform from the top (see **Enter Custom regions as BED in the multi-region section of**
14 **the Basic Protocol**).
15
16

17 Turning on the **All** tracks in the Global aggregate (Fig. 8) and setting the display mode to **Pack**,
18 shows that the higher density of Ribo-seq in the 5' leader region of *MIEF1* is systematic across all
19 of the studies included in the GWIPS-viz hg38 assembly. This shows how GWIPS-viz can be used to
20 compare data across studies. GWIPS-viz can also be utilised for cross-species exploration. For
21 example, the *Mief1* gene in Mouse also clearly shows translation in the 5' leader (Fig. 9, Panel A)
22 as does the *Mief1* gene in rat (Fig. 9, Panel B).
23
24

25 Exploring the human *MIEF1* gene further, keeping the introns collapsed with the multi-region view
26 and zooming in to the 5' leader of *MIEF1* reveals the open reading frame architecture at the top of
27 the window, where AUG start codons are coloured green and stop codons are coloured red (Fig
28 10.). In this figure, the highlighting feature of GWIPS-viz has been utilised to colour in three
29 distinct parts of the 5' leader which appear to be undergoing translation. The light blue highlight
30 shows a relatively short upstream open reading frame (uORF), 19 codons in length. The phyloP
31 track for this particular region shows some, albeit not strong, conservation.
32
33

34 The orange highlight shows a long translated uORF of length 70 codons; the phyloP scores show
35 that this uORF is highly conserved. Phylogenetic analysis of this uORF suggested that it codes for a
36 functional protein (Andreev, O'Connor, Fahey, et al., 2015). Indeed, the product of this uORF
37 (LOR8F8) has been found to play a role in mitoribosome biogenesis (Brown et al., 2017).
38
39

40 Finally the region in yellow also has a high Ribo-seq density but there is no AUG start codon at the
41 5' end of that region. There is a short AUG-initiated ORF downstream in this region, but this would
42 not account for the high Ribo-seq density across the entire yellow region. The phyloP scores are
43 lower here than for the preceding two uORFs. While it is difficult to elucidate the exact ORF origin
44 of these footprints using GWIPS-viz, it may be that these footprints originate from a non-AUG
45 initiated uORF.
46
47
48
49

50 COMMENTARY

51 Development of the GWIPS-viz Genome Browser

52 GWIPS-viz was launched in September 2012 in response to the rapidly increasing interest in
53 ribosome profiling. At the time there was no freely available resource for exploring Ribo-seq and
54 corresponding mRNA-seq data. GWIPS-viz was developed to provide researchers with an easy and
55
56
57
58
59
60

1 quickly accessible way to explore this data. To this end, a customised version of the UCSC Genome
2 Browser was developed as it provided the high performance necessary for rendering large
3 alignment files, as well as being an already well-established resource which many researchers
4 would be familiar with navigating. In addition the team behind the UCSC genome browser have a
5 proven track record of updating and maintaining the resource as well as consistently responding
6 to user enquires. All of these points made the UCSC genome browser an excellent starting point
7 for GWIPS-viz development. Prior to GWIPS-viz, a researcher wishing to explore publicly available
8 Ribo-seq data would have to download, process and align the data, then upload to the UCSC
9 Genome Browser or use an alternative genome browser. This is a time consuming and
10 computationally expensive process.
11
12
13

14 The GWIPS-viz pipeline involves downloading publicly available Ribo-seq and mRNA-seq data
15 (usually via NCBI's sequence read archive (Leinonen, Sugawara, Shumway, & International
16 Nucleotide Sequence Database, 2011)), removing any adapter sequence using the Cutadapt tool
17 (Martin, 2011), removing rRNA contamination by aligning the reads to known rRNA sequences and
18 finally using the short read alignment tool bowtie (Langmead, Trapnell, Pop, & Salzberg, 2009) to
19 align unambiguous reads to a reference genome.
20
21

22 To date there are 1885 tracks of both Ribo-seq and mRNA-seq data across 24 genomes (Michel,
23 Kiniry, O'Connor, Mullan, & Baranov, 2017). One of the advantages of exploring these datasets in
24 GWIPS-viz is that all datasets are processed in the same manner making it easier for both cross-
25 study comparisons as well as cross-species comparisons. According to Google Analytics, over
26 15,000 unique IP addresses have been used to access GWIPS-viz across the globe (Fig. 11).
27
28

29 More recently other online resources have become available for analysing and exploring Ribo-seq
30 data. RPFdb (Xie et al., 2016) is a freely available database that provides mapping statistics for
31 Ribo-seq datasets generated for 8 model organisms and also provides the functionality to explore
32 Ribo-seq data for individual genes within a genome browser. TranslatomeDB is an online platform
33 that provides the functionality to analyse published Ribo-seq and mRNA-seq data including
34 differential gene expression analysis (Liu, Xiang, Zheng, Jin, & Zhang, 2017). Riboviz (Carja, Xing,
35 Plotkin, & Shah, 2017) is another online visualization tool for Ribo-seq data which also provides
36 analytical functionalities including metagenomic analysis and triplet periodicity analysis. sORFs.org
37 is a repository of small ORFs identified by Ribo-seq (Olexiouk et al., 2016) while the uORFdb
38 database provides information on upstream open reading frames that have been reported in the
39 literature along with Ribo-seq support for the translation of the uORF if available (Wethmar,
40 Barbosa-Silva, Andrade-Navarro, & Leutz, 2014). SmProt also uses ribosome profiling in
41 conjunction with proteomics and literature mining to identify small open reading frames and
42 provide a way to visualise their location in the form of a UCSC mirror (Hao et al., 2017).
43
44
45

46 These online resources are similar to GWIPS-viz in that they host processed Ribo-seq data and
47 related information which can save the researcher much time when searching for evidence of
48 translation in different regions of genes.
49
50

51 Of interest also are the tools that help the researcher to analyse their own Ribo-seq data. As
52 mentioned, RiboGalaxy (<http://ribogalaxy.ucc.ie/>) is a Galaxy platform specifically tailored for
53 Ribo-seq data analysis (Michel et al., 2016). Like GWIPS-viz, RiboGalaxy can also be used through
54 most web browsers making it easy for the user as no software needs to be downloaded.
55 RiboGalaxy hosts a number of tools so that the user can upload their raw Ribo-seq/mRNA-seq data
56 in FASTQ format, remove adapter sequences, remove sequences that map to rRNA and map the
57
58
59
60

1 remaining sequences to a reference genome/transcriptome. As well as the option to generate
2 profiles that can be viewed directly in GWIPS-viz as custom tracks, the user can use the RiboPlot
3 suite in RiboGalaxy to generate sub-codon profiles (Michel et al., 2012) in both graphical format
4 and spreadsheet format with the individual counts for each sub-codon position. There are also a
5 number of tool suites for analysing Ribo-seq data available on RiboGalaxy such as the RUST suite
6 (O'Connor et al., 2016) mentioned in the basic protocol, along with RiboTools (Legendre, Baudin-
7 Baillieu, Hatin, & Namy, 2015) and the riboSeqR (Chung et al., 2015) tools suite which includes
8 baySeq for differential expression analysis of Ribo-seq/mRNA-seq data.
9

10
11
12 Other web-based tools include PausePred (<http://pausepred.ucc.ie/>) (Kumari, Michel, & Baranov,
13 2017) for the detection of potential pause sites using Ribo-seq data and Rfeet
14 (<http://pausepred.ucc.ie/rfeet2.html>) (Kumari et al., 2017) a web-based tool for plotting ribosome
15 profiles. Command-line versions of both of these tools are also available for download.
16

17
18 Indeed, quite a number of command-line based Ribo-seq analysis tools have become available
19 over the last number of years. While these usually offer more flexibility than online tools, they do
20 require the user to download and install the software along with its dependencies. The currently
21 available software resources for exploring Ribo-seq data has been reviewed in detail in (Calviello &
22 Ohler, 2017; Wang, Wang, & Xie, 2017).
23

24 25 **CRITICAL PARAMETERS AND TROUBLE SHOOTING**

26
27
28 As mentioned, caution should be exercised when interpreting the Ribo-seq data tracks in GWIPS-
29 viz as there may be cases where the occurrence of footprints does not represent genuine
30 translation. There could be several sources of artefactual footprints. First is very low level noise
31 that is present in almost all studies. This is readily apparent when viewing the **Global aggregate** on
32 hg38 where it can be seen that almost all 5' leaders and 3' trailers have at least some Ribo-seq
33 peaks associated with them. These peaks tend to be much lower and sparser than the peaks in
34 annotated protein coding regions, likely representing random fragments of naked RNA. This type
35 of noise can be mitigated by viewing the global aggregate where the signal from genuine
36 translation is quite strong compared to the noise.
37
38

39
40 The second case is due to non-ribosome protected fragments. There are other RNP complexes that
41 could have a sedimentation density similar to that of ribosomes. Proteins as well as strong RNA
42 secondary structures in them may protect fragments from their digestion. While such protected
43 fragments have a different length distribution from that of genuine footprints (Ingolia et al., 2014),
44 the information on read length distribution for a given transcript is not readily available in GWIPS-
45 viz, hence the presence of such peaks could mislead the user.
46

47
48 The third case is due to misalignment, where footprints are aligned to a location different from its
49 origin. While only uniquely mapped reads are displayed in GWIPS-viz, the transcriptome
50 sequences of samples in which the ribosome profiling was carried out differs from the reference
51 sequence of the human genome. Due to polymorphisms a footprint originated in a specific locus of
52 the human genome may produce a better scoring alignment with a different location. The
53 problem could be particularly acute due to structural polymorphisms and copy number variations
54 which are particularly pertinent to cancer or immortalized cell lines that are often used for
55 ribosome profiling experiments.
56
57
58
59
60

1 The opposite scenario can also take place, when a translated locus may have little or no footprints.
2 The main source of this problem is ambiguous alignments. GWIPS-viz discards reads that could be
3 aligned to multiple locations. This increases the mapping confidence of the relatively short
4 ribosome protected fragment read sequence and decreases false positives. This approach,
5 however, negatively affects the density profiles of genes whose sequences share regions of
6 similarity with other genes or pseudogenes. For the same reason, this also affects highly repetitive
7 regions of the genome.
8
9

10 Another limitation is the reduction in read density across coding exon-exon boundaries. The short
11 read alignment tool bowtie (Langmead et al., 2009) is used to map reads to the genome but it is
12 not splice aware, although the mapping of short footprints of ~30 nucleotides across splice
13 junctions is challenging even for splice aware alignment tools. This means that Ribo-seq or mRNA-
14 seq reads which span exon-exon junctions cannot be mapped back to the genome. This can be
15 particularly apparent when using the multi-region mode. Distorted footprint densities are also
16 inherited in the datasets themselves due to technical biases caused by drug effects (Gerashchenko
17 & Gladyshev, 2014; Hussmann, Patchett, Johnson, Sawyer, & Press, 2015) and the substrate
18 sequence specificities of the enzymes used in the protocol (Gerashchenko & Gladyshev, 2017).
19
20
21

22 One other current limitation of GWIPS-viz is that a static offset is used to infer the ribosome
23 decoding centers. Irrespective of the variation in read lengths across Ribo-seq datasets, 12
24 nucleotides from the 5' end of the reads are used to infer the P-site while 15 nucleotides are used
25 to infer the A-site for Ribo-seq data generated in eukaryotes. In bacteria, interactions of the
26 ribosome with Shine-Dalgarno result in a higher variability of the protected length 5' of the
27 decoding center (O'Connor, Li, Weissman, Atkins, & Baranov, 2013). Therefore, in bacteria an
28 offset of 12 nucleotides from the 3' end of the read is used to infer the A-site (Martens, Taylor, &
29 Hilser, 2015; Woolstenhulme, Guydosh, Green, & Buskirk, 2015). A number of publicly available
30 tools now exist for determining a refined offset for individual read lengths such as Plastid (Dunn &
31 Weissman, 2016), RiboWaltz, (Lauria et al., 2017) RP-BP (Malone et al., 2017), and RiboProfiling
32 (Popa et al., 2016) . We hope to soon incorporate a similar approach into the GWIPS-viz
33 computational pipeline that assigns offset values for individual read lengths.
34
35
36

37 If you encounter issues or have questions when using GWIPS-viz, the GWIPS-viz forum
38 (<https://gwips.ucc.ie/Forum/>) is a useful tool to ask for help with queries related to GWIPS-viz (Fig.
39 12). You can also browse questions and solutions posted by others, and not just for GWIPS-viz but
40 related topics such as ribosome profiling in general. The GWIPS-viz forum can be accessed via the
41 link in the topbar on all of the GWIPS-viz browser pages.
42
43
44

45 REFERENCES

- 46
47 Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., . . . Goecks, J.
48 (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical
49 analyses: 2016 update. *Nucleic Acids Res*, *44*(W1), W3-W10. doi:10.1093/nar/gkw343
50 Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., . . . Searle, S. M. (2016).
51 The Ensembl gene annotation system. *Database (Oxford)*, *2016*.
52 doi:10.1093/database/baw093
53 Andorf, C. M., Cannon, E. K., Portwood, J. L., 2nd, Gardiner, J. M., Harper, L. C., Schaeffer, M.
54 L., . . . Lawrence-Dill, C. J. (2016). MaizeGDB update: new tools, data and interface for the
55 maize model organism database. *Nucleic Acids Res*, *44*(D1), D1195-1201.
56 doi:10.1093/nar/gkv1007
57
58
59
60

- 1 Andreev, D. E., O'Connor, P. B., Fahey, C., Kenny, E. M., Terenin, I. M., Dmitriev, S. E., . . .
2 Baranov, P. V. (2015). Translation of 5' leaders is pervasive in genes resistant to eIF2
3 repression. *Elife*, *4*, e03971. doi:10.7554/eLife.03971
- 4 Andreev, D. E., O'Connor, P. B., Loughran, G., Dmitriev, S. E., Baranov, P. V., & Shatsky, I. N.
5 (2017). Insights into the mechanisms of eukaryotic translation gained with ribosome
6 profiling. *Nucleic Acids Res*, *45*(2), 513-526. doi:10.1093/nar/gkw1190
- 7 Andreev, D. E., O'Connor, P. B., Zhdanov, A. V., Dmitriev, R. I., Shatsky, I. N., Papkovsky, D. B.,
8 & Baranov, P. V. (2015). Oxygen and glucose deprivation induces widespread alterations in
9 mRNA translation within 20 minutes. *Genome Biol*, *16*, 90. doi:10.1186/s13059-015-0651-z
- 10 Aramayo, R., & Polymenis, M. (2017). Ribosome profiling the cell cycle: lessons and challenges.
11 *Curr Genet*. doi:10.1007/s00294-017-0698-3
- 12 Archer, S. K., Shirokikh, N. E., Beilharz, T. H., & Preiss, T. (2016). Dynamics of ribosome
13 scanning and recycling revealed by translation complex profiling. *Nature*, *535*(7613), 570-
14 574. doi:10.1038/nature18647
- 15 Bartholomaeus, A., Del Campo, C., & Ignatova, Z. (2016). Mapping the non-standardized biases of
16 ribosome profiling. *Biol Chem*, *397*(1), 23-35. doi:10.1515/hsz-2015-0197
- 17 Baudin-Baillieu, A., Hatin, I., Legendre, R., & Namy, O. (2016). Translation Analysis at the
18 Genome Scale by Ribosome Profiling. *Methods Mol Biol*, *1361*, 105-124. doi:10.1007/978-
19 1-4939-3079-1_7
- 20 Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., .
21 . . . Giraldez, A. J. (2014). Identification of small ORFs in vertebrates using ribosome
22 footprinting and evolutionary conservation. *EMBO J*, *33*(9), 981-993.
23 doi:10.1002/embj.201488411
- 24 Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The
25 Arabidopsis information resource: Making and mining the "gold standard" annotated
26 reference plant genome. *Genesis*, *53*(8), 474-485. doi:10.1002/dvg.22877
- 27 Brar, G. A., & Weissman, J. S. (2015). Ribosome profiling reveals the what, when, where and how
28 of protein synthesis. *Nat Rev Mol Cell Biol*, *16*(11), 651-664. doi:10.1038/nrm4069
- 29 Brown, A., Rathore, S., Kimanius, D., Aibara, S., Bai, X. C., Rorbach, J., . . . Ramakrishnan, V.
30 (2017). Structures of the human mitochondrial ribosome in native states of assembly. *Nat*
31 *Struct Mol Biol*. doi:10.1038/nsmb.3464
- 32 Calviello, L., & Ohler, U. (2017). Beyond Read-Counts: Ribo-seq Data Analysis to Understand the
33 Functions of the Transcriptome. *Trends Genet*. doi:10.1016/j.tig.2017.08.003
- 34 Carja, O., Xing, T., Plotkin, J. B., & Shah, P. (2017). riboviz: analysis and visualization of ribosome
35 profiling datasets. *bioRxiv*.
- 36 Cherry, J. M. (2015). The *Saccharomyces* Genome Database: A Tool for Discovery. *Cold Spring*
37 *Harb Protoc*, *2015*(12), pdb top083840. doi:10.1101/pdb.top083840
- 38 Chung, B. Y., Hardcastle, T. J., Jones, J. D., Irigoyen, N., Firth, A. E., Baulcombe, D. C., &
39 Brierley, I. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-
40 friendly software package for Ribo-seq data analysis. *RNA*, *21*(10), 1731-1745.
41 doi:10.1261/rna.052548.115
- 42 Dunn, J. G., & Weissman, J. S. (2016). Plastid: nucleotide-resolution analysis of next-generation
43 sequencing and genomics data. *BMC Genomics*, *17*(1), 958. doi:10.1186/s12864-016-3278-
44 x
- 45 Gerashchenko, M. V., & Gladyshev, V. N. (2014). Translation inhibitors cause abnormalities in
46 ribosome profiling experiments. *Nucleic Acids Res*, *42*(17), e134. doi:10.1093/nar/gku671
- 47 Gerashchenko, M. V., & Gladyshev, V. N. (2017). Ribonuclease selection for ribosome profiling.
48 *Nucleic Acids Res*, *45*(2), e6. doi:10.1093/nar/gkw822
- 49 Gobet, C., & Naef, F. (2017). Ribosome profiling and dynamic regulation of translation in
50 mammals. *Curr Opin Genet Dev*, *43*, 120-127. doi:10.1016/j.gde.2017.03.005
- 51
52
53
54
55
56
57
58
59
60

- 1 Gramates, L. S., Marygold, S. J., Santos, G. D., Urbano, J. M., Antonazzo, G., Matthews, B. B., . . .
2 the FlyBase, C. (2017). FlyBase at 25: looking to the future. *Nucleic Acids Res*, *45*(D1),
3 D663-D671. doi:10.1093/nar/gkw1016
- 4 Hao, Y., Zhang, L., Niu, Y., Cai, T., Luo, J., He, S., . . . Chen, R. (2017). SmProt: a database of
5 small proteins encoded by annotated coding and non-coding RNA loci. *Brief Bioinform*.
6 doi:10.1093/bib/bbx005
- 7 Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., . . . Hubbard,
8 T. J. (2012). GENCODE: the reference human genome annotation for The ENCODE
9 Project. *Genome Res*, *22*(9), 1760-1774. doi:10.1101/gr.135350.111
- 10 Hussmann, J. A., Patchett, S., Johnson, A., Sawyer, S., & Press, W. H. (2015). Understanding
11 Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in
12 Yeast. *PLoS Genet*, *11*(12), e1005732. doi:10.1371/journal.pgen.1005732
- 13 Ingolia, N. T. (2014). Ribosome profiling: new views of translation, from single codons to genome
14 scale. *Nat Rev Genet*, *15*(3), 205-213. doi:10.1038/nrg3645
- 15 Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J., Jackson, S. E., . . .
16 Weissman, J. S. (2014). Ribosome profiling reveals pervasive translation outside of
17 annotated protein-coding genes. *Cell Rep*, *8*(5), 1365-1379.
18 doi:10.1016/j.celrep.2014.07.045
- 19 Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009). Genome-wide
20 analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*,
21 *324*(5924), 218-223. doi:10.1126/science.1168978
- 22 Kumari, R., Michel, A. M., & Baranov, P. V. (2017). PausePred and Rfeet: Webtools for inferring
23 ribosome pauses and visualizing footprint density from ribosome profiling data. Manuscript
24 submitted for publication.
- 25 Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient
26 alignment of short DNA sequences to the human genome. *Genome Biol*, *10*(3), R25.
27 doi:10.1186/gb-2009-10-3-r25
- 28 Lauria, F., Tebaldi, T., Bernabo, P., Groen, E. J., Gillingwater, T. H., & Viero, G. (2017).
29 riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. *bioRxiv*,
30 169862.
- 31 Legendre, R., Baudin-Baillieu, A., Hatin, I., & Namy, O. (2015). RiboTools: a Galaxy toolbox for
32 qualitative ribosome profiling analysis. *Bioinformatics*, *31*(15), 2586-2588.
33 doi:10.1093/bioinformatics/btv174
- 34 Leinonen, R., Sugawara, H., Shumway, M., & International Nucleotide Sequence Database, C.
35 (2011). The sequence read archive. *Nucleic Acids Res*, *39*(Database issue), D19-21.
36 doi:10.1093/nar/gkq1019
- 37 Liu, W., Xiang, L., Zheng, T., Jin, J., & Zhang, G. (2017). TranslatomeDB: a comprehensive
38 database and cloud-based analysis platform for translatome sequencing data. *Nucleic Acids*
39 *Research*, *46*(D1). doi:10.1093/nar/gkx1034
- 40 Malone, B., Atanassov, I., Aeschmann, F., Li, X., Grosshans, H., & Dieterich, C. (2017). Bayesian
41 prediction of RNA translation from ribosome profiling. *Nucleic Acids Res*, *45*(6), 2960-
42 2972. doi:10.1093/nar/gkw1350
- 43 Martens, A. T., Taylor, J., & Hilser, V. J. (2015). Ribosome A and P sites revealed by length
44 analysis of ribosome profiling data. *Nucleic Acids Res*, *43*(7), 3680-3687.
45 doi:10.1093/nar/gkv200
- 46 Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing
47 Reads. *EMBnetjournal*, *17*.
- 48 McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., . . . Bejerano, G.
49 (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat*
50 *Biotechnol*, *28*(5), 495-501. doi:10.1038/nbt.1630
- 51
52
53
54
55
56
57
58
59
60

- 1 Michel, A. M., & Baranov, P. V. (2013). Ribosome profiling: a Hi-Def monitor for protein
2 synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA*, 4(5), 473-490.
3 doi:10.1002/wrna.1172
- 4 Michel, A. M., Choudhury, K. R., Firth, A. E., Ingolia, N. T., Atkins, J. F., & Baranov, P. V.
5 (2012). Observation of dually decoded regions of the human genome using ribosome
6 profiling data. *Genome Res*, 22(11), 2219-2229. doi:10.1101/gr.133249.111
- 7 Michel, A. M., Fox, G., A. M. K., De Bo, C., O'Connor, P. B., Heaphy, S. M., . . . Baranov, P. V.
8 (2014). GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res*,
9 42(Database issue), D859-864. doi:10.1093/nar/gkt1035
- 10 Michel, A. M., Kiniry, S. J., O'Connor, Patrick B. F., Mullan, J. P., & Baranov, P. V. (2017).
11 GWIPS-viz: 2018 update. *Nucleic Acids Research*. doi:10.1093/nar/gkx790
- 12 Michel, A. M., Mullan, J. P., Velayudhan, V., O'Connor, P. B., Donohue, C. A., & Baranov, P. V.
13 (2016). RiboGalaxy: A browser based platform for the alignment, analysis and visualization
14 of ribosome profiling data. *RNA Biol*, 13(3), 316-319. doi:10.1080/15476286.2016.1141862
- 15 Mumtaz, M. A., & Couso, J. P. (2015). Ribosomal profiling adds new coding sequences to the
16 proteome. *Biochem Soc Trans*, 43(6), 1271-1276. doi:10.1042/BST20150170
- 17 O'Connor, P. B., Andreev, D. E., & Baranov, P. V. (2016). Comparative survey of the relative
18 impact of mRNA features on local ribosome profiling read density. *Nat Commun*, 7, 12915.
19 doi:10.1038/ncomms12915
- 20 O'Leary, N. A., Wright, M. W., Brister, J. R., Ciuffo, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D.
21 (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic
22 expansion, and functional annotation. *Nucleic Acids Res*, 44(D1), D733-745.
23 doi:10.1093/nar/gkv1189
- 24 Olexiouk, V., Crappe, J., Verbruggen, S., Verhegen, K., Martens, L., & Menschaert, G. (2016).
25 sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res*,
26 44(D1), D324-329. doi:10.1093/nar/gkv1175
- 27 Park, J. E., Yi, H., Kim, Y., Chang, H., & Kim, V. N. (2016). Regulation of Poly(A) Tail and
28 Translation during the Somatic Cell Cycle. *Mol Cell*, 62(3), 462-471.
29 doi:10.1016/j.molcel.2016.04.007
- 30 Parsons, M., & Myler, P. J. (2016). Illuminating Parasite Protein Production by Ribosome Profiling.
31 *Trends Parasitol*, 32(6), 446-457. doi:10.1016/j.pt.2016.03.005
- 32 Pohl, A., & Beato, M. (2014). bwtool: a tool for bigWig files. *Bioinformatics*, 30(11), 1618-1619.
33 doi:10.1093/bioinformatics/btu056
- 34 Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral
35 substitution rates on mammalian phylogenies. *Genome Res*, 20(1), 110-121.
36 doi:10.1101/gr.097857.109
- 37 Popa, A., Lebrigand, K., Paquet, A., Nottet, N., Robbe-Sermesant, K., Waldmann, R., & Barbry, P.
38 (2016). RiboProfiling: a Bioconductor package for standard Ribo-seq pipeline processing.
39 *F1000Res*, 5, 1309. doi:10.12688/f1000research.8964.1
- 40 Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., . . . Manke, T.
41 (2016). deepTools2: a next generation web server for deep-sequencing data analysis.
42 *Nucleic Acids Res*, 44(W1), W160-165. doi:10.1093/nar/gkw257
- 43 Sendoel, A., Dunn, J. G., Rodriguez, E. H., Naik, S., Gomez, N. C., Hurwitz, B., . . . Fuchs, E.
44 (2017). Translation from unconventional 5' start sites drives tumour initiation. *Nature*,
45 541(7638), 494-499. doi:10.1038/nature21036
- 46 Stern-Ginossar, N. (2015). Decoding viral infection by ribosome profiling. *J Virol*, 89(12), 6164-
47 6166. doi:10.1128/JVI.02528-14
- 48 Stern-Ginossar, N., & Ingolia, N. T. (2015). Ribosome Profiling as a Tool to Decipher Viral
49 Complexity. *Annu Rev Virol*, 2(1), 335-349. doi:10.1146/annurev-virology-100114-054854
- 50 Tyner, C., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Eisenhart, C., . . . Kent, W. J.
51 (2017). The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res*, 45(D1),
52 D626-D634. doi:10.1093/nar/gkx1134

- 1 Yordanova, M. M., Loughran, G., Zhdanov, A. V., Mariotti, M., Kiniry, S. J., O'Connor, P. B., . . .
2 Baranov, P. V. (2018). AMD1 mRNA employs ribosome stalling as a mechanism for
3 molecular memory formation. *Nature*. doi:10.1038/nature25174
4 Wang, H., Wang, Y., & Xie, Z. (2017). Computational resources for ribosome profiling: from
5 database to Web server and software. *Brief Bioinform*. doi:10.1093/bib/bbx093
6 Wethmar, K., Barbosa-Silva, A., Andrade-Navarro, M. A., & Leutz, A. (2014). uORFdb--a
7 comprehensive literature database on eukaryotic uORF biology. *Nucleic Acids Res*,
8 42(Database issue), D60-67. doi:10.1093/nar/gkt952
9 Woolstenhulme, C. J., Guydosh, N. R., Green, R., & Buskirk, A. R. (2015). High-precision analysis
10 of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep*, 11(1), 13-
11 21. doi:10.1016/j.celrep.2015.03.014
12 Xie, S. Q., Nie, P., Wang, Y., Wang, H., Li, H., Yang, Z., . . . Xie, Z. (2016). RPFdb: a database for
13 genome wide information of translated mRNA generated from ribosome profiling. *Nucleic*
14 *Acids Res*, 44(D1), D254-258. doi:10.1093/nar/gkv972
15
16
17
18
19
20
21

22 Figure legends

23
24 **Figure 1.** The GWIPS-viz Gateway page set up to display the human *AMD1* gene locus. The hg38
25 human genome assembly was set by clicking on Human in the Represented Species list and
26 selecting Dec. 2013 (GRCh38/hg38) in the assembly dropdown menu. The text box labelled
27 Position/Search term determines which region of the genome will be seen upon hitting submit. A
28 position, gene symbol (e.g. *AMD1*) or search term can be used. A genomic position can be entered
29 in the form <chromosome>:<start_position>-<end_position>. For example, if you wish to view
30 from position 110,874,784 to 110,895,712 on chromosome 6 then enter chr6:110874784-
31 110895712 in the text box.
32
33

34
35 **Figure 2.** The GWIPS-viz Genome Browser page displaying the default data tracks for the human
36 *AMD1* gene locus on chromosome 6 (chr6:110,874,784-110,895,712) in the Dec 2013 assembly
37 (GRCh38/hg38). The navigation and configuration buttons are visible above and below the main
38 window image. The red rectangle in the ideogram directly above the main image indicates the
39 location of the currently displayed region of the chromosome. The default tracks are shown in the
40 main window: the Global Aggregate for Elongating Ribosomes (A-site) (red), the Global aggregate
41 for mRNA-seq reads (green) and the RefSeq gene annotations (blue). The direction of the arrows
42 in the intronic regions of the gene annotations for the *AMD1* gene indicates the direction of
43 transcription and translation. The recently discovered ribosome stalling site downstream of the
44 *AMD1* annotated stop codon (Yordanova et al. 2018) could easily be identified as a prominent
45 peak of footprint density in the region annotated as the 3' trailer (3' UTR).
46
47

48
49 **Figure 3.** A RUST metafootprint profile (O'Connor et al., 2016) taken from GWIPS-viz for the (Park,
50 Yi, Kim, Chang, & Kim, 2016) study. The impact of mRNA codons on the relative read density in the
51 region of the ribosome is shown in grey in the top panel. The Kullback-Leibler divergence (blue for
52 a single codon, green for adjacent codons) indicates the influence of each mRNA location on the
53 frequency of footprint occurrence in the library. This is an example of a typical dataset with likely
54 sequencing biases at both the 5' and 3' end of the ribosome but where the A-site codon influence
55 is the highest. The lower left panel shows the RUST estimates of relative codon dwell times for
56
57
58
59

1 each amino acid. The lower right panel shows the triplet periodicity signal for individual read
2 lengths.
3

4
5 **Figure 4.** The Custom Track page where users can upload their own Ribo-seq/mRNA-seq data
6 tracks and explore them in the context of existing public tracks. An example is shown in the text
7 box of a URL pointing to a bigWig file. When submitted, the bigWig file will be displayed as a
8 separate track in the human hg38 GWIPS-viz Genome Browser page. This custom track, however,
9 will only be visible to the user who uploaded it.
10

11
12 **Figure 5.** The Table Browser page where the database tables underlying the GWIPS-viz tracks can
13 be accessed. In this example, the Table Browser is set up to provide access to one of the
14 Elongating Ribosome (A-site) tracks in the mouse mm10 assembly for the (Sendoel et al., 2017)
15 study. The bed format has been selected as the output format and when submitted, the output
16 will be displayed on screen for the genomic region specified in the position text box.
17

18
19 **Figure 6.** The Downloads page where the alignment files for the GWIPS-viz Ribo-seq and mRNA-
20 seq tracks can be downloaded. In this example, when the Download button is clicked the
21 alignment file (in bigWig format) of the Ribo-seq data generated under one of the conditions in
22 the (Andreev, O'Connor, Zhdanov, et al., 2015) study and mapped to the rat rn6 genome assembly
23 will be downloaded.
24

25
26 **Figure 7.** The human *MIEF1* gene locus in GWIPS-viz. Panel A shows the elongating ribosome
27 profile (A-sites) (red) along with the mRNA-seq reads (green) for the human *MIEF1* gene locus. The
28 phyloP 100-way track provides conservation scores (blue for positive scores and brown for
29 negative scores) for the alignment of 100 vertebrate genomic sequences (Pollard et al., 2010). The
30 RefSeq gene annotations for this locus are in blue. Panel B shows the same tracks for the same
31 gene locus but the intronic regions have been collapsed, where feasible, using the multi-region
32 view functionality.
33

34
35 **Figure 8.** Cross-study comparison using the *All* tracks for each study in the Global Aggregate
36 Elongating Ribosomes (A-site) track for the human *MIEF1* gene locus in GWIPS-viz. The 30
37 individual study tracks currently composing the Global Aggregate track for the human hg38
38 assembly show that the high Ribo-seq density in the 5' leader region relative to the annotated
39 protein coding region of the *MIEF1* gene is consistent across the majority of studies.
40

41
42 **Figure 9.** Cross-species comparison of Ribo-seq data using the mouse *Mief1* gene and the rat
43 *Mief1* gene in GWIPS-viz. Both the elongating ribosome (A-site) Global Aggregate track for the
44 *Mief1* gene in mouse mm10 assembly (panel A) and the *Mief1* gene in rat rn6 assembly (panel B)
45 show the high Ribo-seq density in the 5' leader region relative to the annotated protein coding
46 region. The intronic regions in the RefSeq gene annotations have been collapsed using the multi-
47 region view functionality.
48

49
50 **Figure 10.** A zoomed-in view of the open reading frame (ORF) organisation of the 5' leader region
51 of the human *MIEF1* gene in GWIPS-viz. The upstream ORFs (uORFs) that are likely translated are
52 highlighted light-blue for the first AUG- initiated uORF and orange for the second (longer) AUG-
53 initiated uORF. The third region highlighted in yellow shows ribosome footprints that span across a
54 region than contains a short AUG-initiated uORF. The footprints just upstream of the third AUG-
55 initiated uORF may originate from a non-AUG initiated uORF in an alternative frame, but this
56 cannot be directly determined using GWIPS-viz. These uORFs may play a functional role in stress
57
58
59
60

1 resistance (Andreev, O'Connor, Fahey, et al., 2015). The phyloP conservation scores are
2 consistently high for the second uORF (orange) indicating that this uORF likely produces a
3 functional peptide (Andreev, O'Connor, Fahey, et al., 2015) which has been recently characterised
4 by (Brown et al., 2017) .
5
6

7 **Figure 11.** The usage statistics of GWIPS-viz across the world. The map shows the user session
8 statistics according to Google Analytics for the period January 1st, 2013 to December 31st, 2017.
9

10 **Figure 12.** The GWIPS-viz forum. Queries about using GWIPS-viz can be posted to the GWIPS-viz
11 forum (<https://gwips.ucc.ie/Forum/>). Other related topics of discussion include RiboGalaxy,
12 ribosome profiling, translation/protein synthesis in general, sequencing and Bioinformatics.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

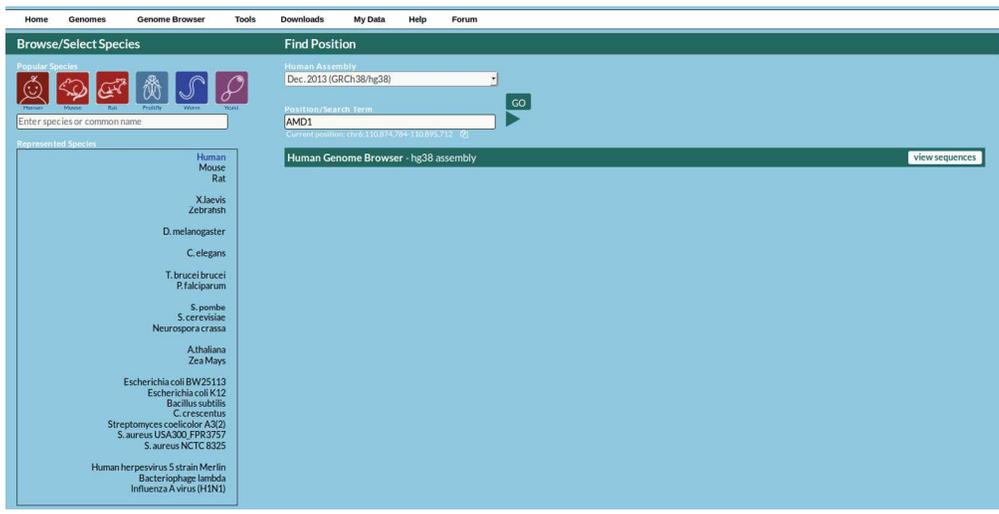


Figure 1
644x325mm (72 x 72 DPI)

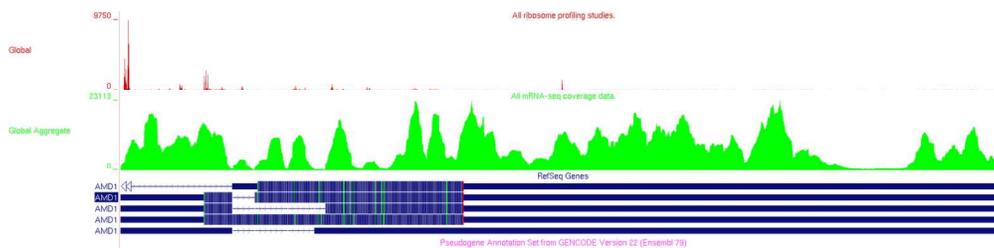


Figure 2

575x136mm (72 x 72 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

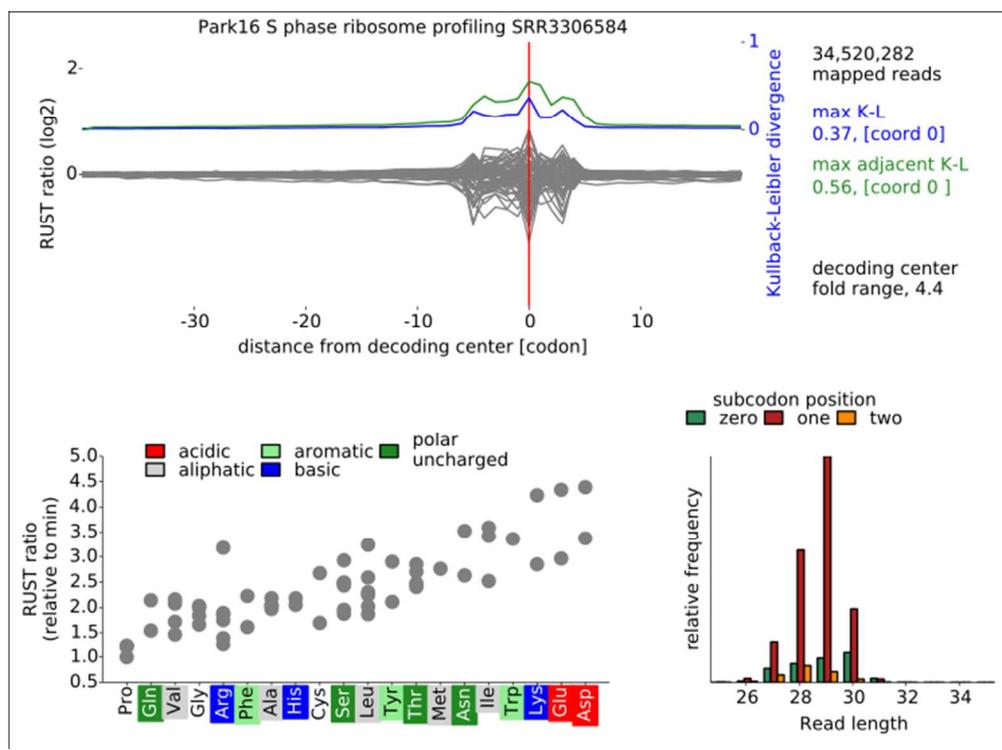


Figure 3
196x145mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Add Custom Tracks

clade genome assembly

Display your own data as custom annotation tracks in the browser. Data must be formatted in [bigBed](#), [bigChain](#), [bigGenePred](#), [bigMaf](#), [bigPsi](#), [bigWig](#), [BAM](#), [VCF](#), [BED](#), [BED_detail](#), [bedGraph](#), [broadPeak](#), [CRAM](#), [GFF](#), [GTF](#), [MAF](#), [narrowPeak](#), [Personal Genome SNP](#), [PSL](#), or [WIG](#) formats. To configure the display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#). Data in the bigBed, bigWig, bigGenePred, BAM and VCF formats can be provided via only a URL or embedded in a track line in the box below. Examples are [here](#).

Please note a much more efficient way to share data is to use [Track Hubs](#), which are loaded from the [Track Hubs Portal](#) found in the menu under My Data.

Paste URLs or data: Or upload: No file selected.

Optional track documentation: Or upload: No file selected.

Click [here](#) for an HTML document template that may be used for Genome Browser track descriptions.

Figure 4

483x202mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: genome: assembly:

group: track:

table:

region: genome position

filter:

subtrack merge:

intersection:

correlation:

output format: Send output to Galaxy GREAT

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

Note: to return more than 100,000 lines, change the filter setting (above). The entire data set may be available for download as a very large file that contains the original data values (not compressed into the wiggle format) -- see the Downloads page.

To reset all user cart settings (including custom tracks), [click here](#).

Figure 5

341x180mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Home Genomes Genome Browser Tools Downloads My Data Help Forum

Download alignment files

Clade: Genome: Assembly:

Group: Track: Table:

Use the menus above to select an alignment file for download
Hover over the menus for descriptions
Alignment files are in [bigWig format](#)

Figure 6

236x77mm (96 x 96 DPI)

18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

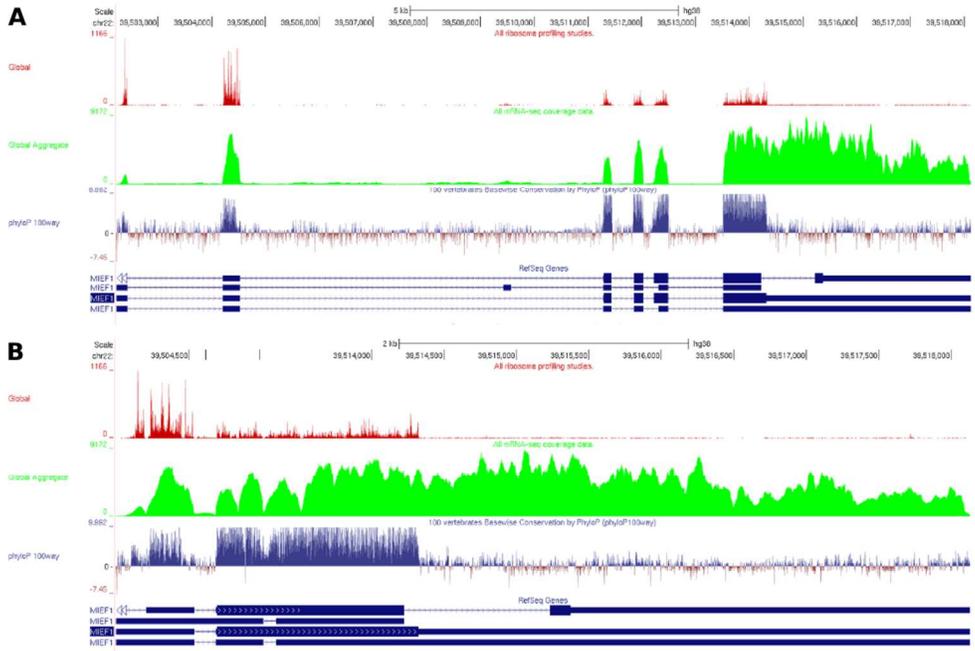


Figure 7

417x280mm (72 x 72 DPI)

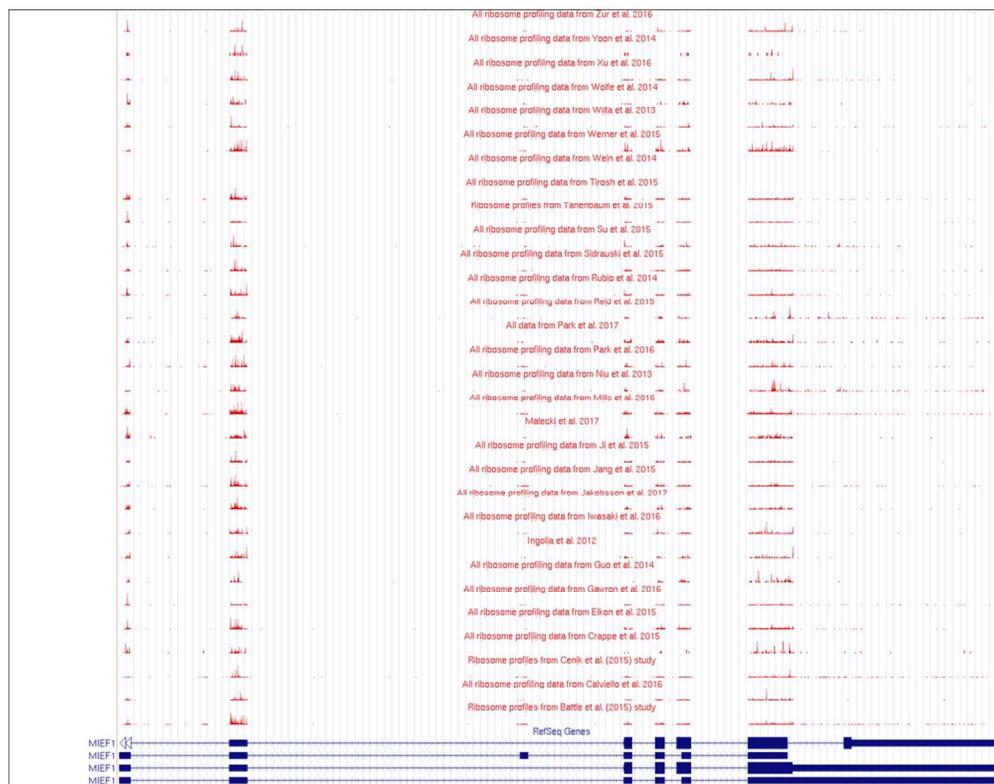


Figure 8

299x233mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

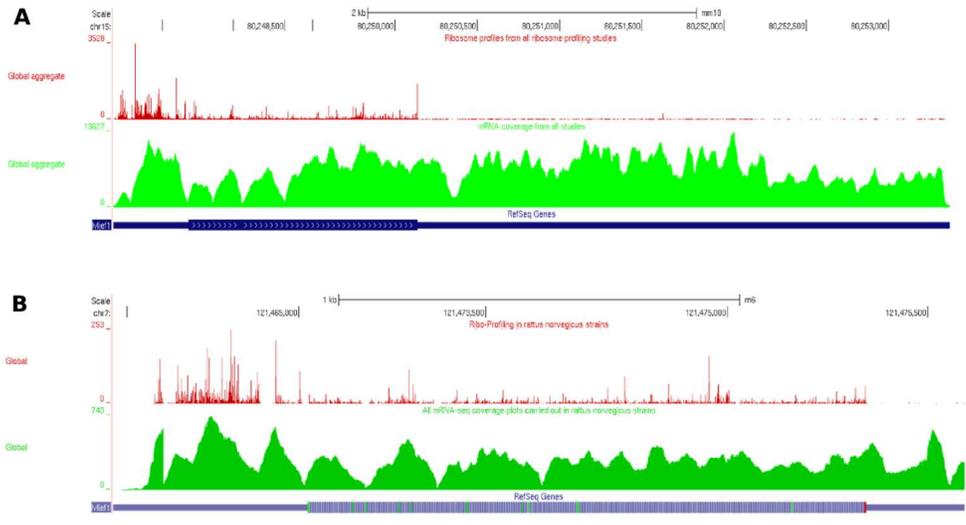


Figure 9

449x249mm (72 x 72 DPI)

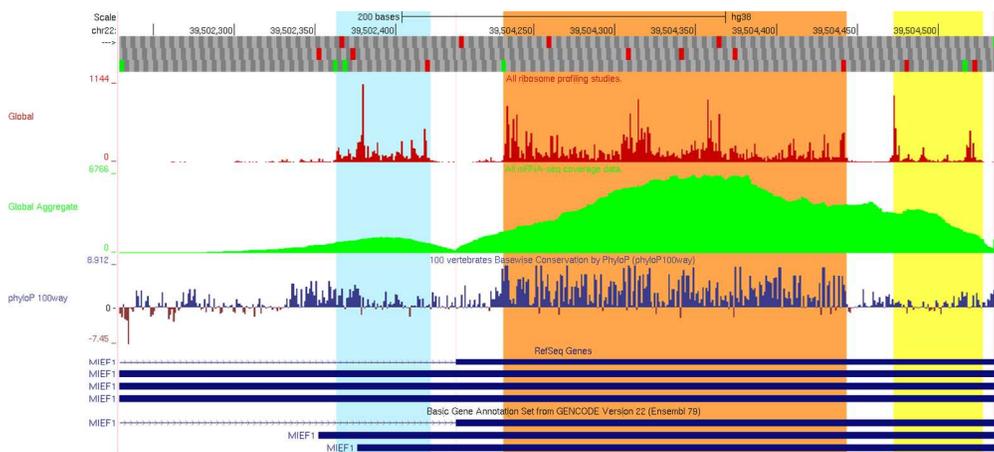


Figure 10

427x189mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

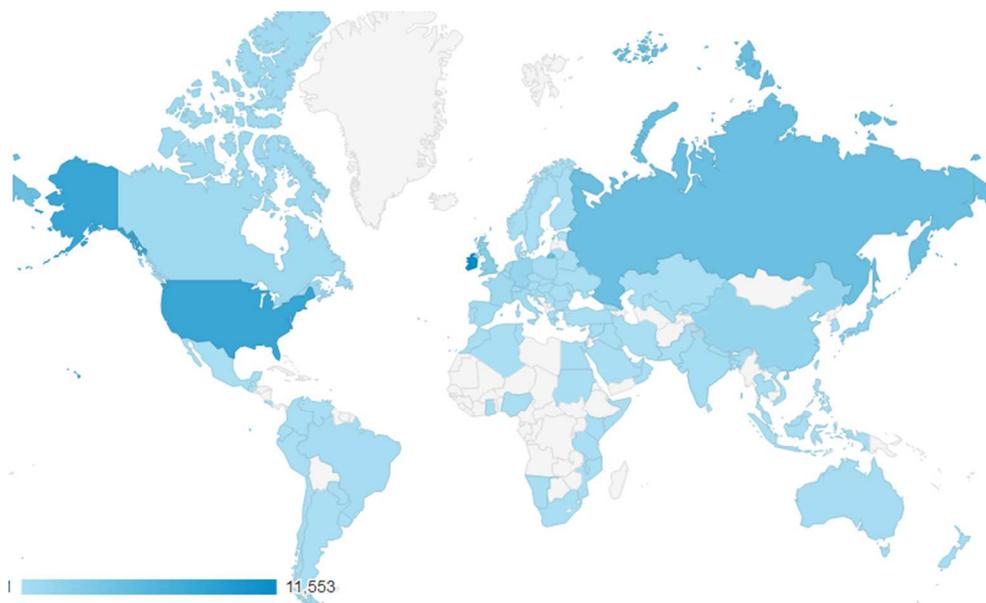


Figure 11
300x181mm (72 x 72 DPI)



Figure 12

221x140mm (96 x 96 DPI)