CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

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	2026-11-07 19:04:41
Item downloaded from	https://hdl.handle.net/10468/7079

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh



Current Protocols

GWIPS-viz

Journal:	*Current Protocols*
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.

SCHOLARONE™
Manuscripts

The GWIPS-viz browser

Stephen J. Kiniry[1], Audrey M. Michel[1] and Pavel V. Baranov[1*].
[1]School 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 **G**enome **W**ide **I**nformation on **P**rotein **S**ynthesis 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

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

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

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

Basic Protocol 1
Using the GWIPS-viz browser

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

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. Open the GWIPS-viz home page at https://gwips.ucc.ie in a web browser.

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

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

2. Click on the **Genomes** link in the topbar or the **Genome Browser** link on the sidebar of the home page.

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

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

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

   <chromosome>:<start_position>-<end_position>

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

   chr6:110874784-110895712

   in the text box.

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

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

4. Explore the **GWIPS-viz Genome Browser** page.

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

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

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

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

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

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

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

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

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

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

A number of button options below the main window can be used to configure the displayed data (see Fig. 2):
*Use **track search** to search for one or more specific terms in the entire set of track names, descriptions, and track groups for the current assembly.*
*Click **default tracks** to reset the current display to the default data tracks for the region currently displayed.*

*Click **default order** to restore the order of the tracks to the default order for the current assembly.*
*Click **hide all** to hide all the tracks for the current assembly.*
*Use **add custom tracks** to upload your own Ribo-seq/mRNA-seq data for exploration in the context of published Ribo-seq/mRNA-seq data (see Support Protocol 1).*
*Use **track hubs** to explore tracks hosted by third parties. You can also create your own private track hub.*
*Use **configure** to change the settings related to the GWIPS-viz Genome browser display configuration and the presented tracks (see step 7).*
*Use **multi-region** to slice track viewing (e.g. display exonic regions only -see step 12).*
*Click **reverse** to display the reverse strand of the current location or the forward strand if the reverse is currently displayed.*
*Click **resize** to restore the default window size.*
*Click **refresh** to update the window display if changes have been made to the track settings.*

*Note to reset to the overall default settings, including the default genomic locations for all genomes, mouse-over **View** in the topbar in the GWIPS-viz Genome Browser page and click on **Reset All user Settings**. Alternatively, from any page, mouse-over **My Data** in the top bar and click on **My Sessions**. Click on the **Click here to reset** link in the **Session Management** page.*

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

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

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

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

8. Explore the GWIPS-viz data track groups.

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

***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. The initiator tRNA is recognized in the P-site during initiation.
Elongating Ribosomes (A-site) *captures elongating ribosomes and an offset is applied to approximate the A-site locations. The recognition of tRNAs during elongation of the nascent peptide occurs in the A-site.*
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 *displays the mRNA-seq data generated in parallel to the Ribo-seq data (if available). Every position in the read is represented.*
Annotation Tracks and External Data *is used for displaying gene annotations specific to the assembly such as RefSeq (O'Leary et al., 2016), Ensembl (Aken et al., 2016), Gencode (Harrow et al., 2012), SGD (Cherry, 2015) , FlyBase (Gramates et al., 2017), TAIR (Berardini et al., 2015), MaizeGDB (Andorf et al., 2016). This track group also hosts other external data including conservation data such as phyloP (100 way) (Pollard, Hubisz, Rosenbloom, & Siepel, 2010).*

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

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

Each Ribo-seq and mRNA-seq data track group 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 was generated is also included in the study label.

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

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

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

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

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

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

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

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

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

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

11. Configure the annotation track display in the **Annotation Tracks and External Data** track group.

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

    Gene annotations are graphically represented by bars of varying thickness: the thickest bars represent annotated protein coding regions, the thinner bars represent 5' leaders (5' UTRs) and 3' trailers (3' UTRs), while the thinnest bars with arrows represent introns, with the arrows indicating the direction of transcription and translation.

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

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

Optional Steps:

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

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

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

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

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

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

<chromsome_name> <start_position> <end_position>

with a space separating each. If multiple regions are required each should be given its own line in the text box:

<chromsome_name1> <start_position1> <end_position1>
<chromsome_name2> <start_position2> <end_position2>
<chromsome_name3> <start_position3> <end_position3>

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

13. Explore the RUST profiles for individual tracks.

*This is an optional step to explore the quality of individual Ribo-seq datasets by viewing the profiles that have been generated using the **R**iboseq **U**nit **S**tep **T**ranformation (or RUST) tool (O'Connor, Andreev, & Baranov, 2016). RUST profiles are available for the majority of recent Ribo-seq datasets generated in eukaryotes and can be viewed by clicking on the label for a particular study in the **Elongating Ribosome (A-site)** track group in the **GWIPS-viz Genome Browser** page. This will open the study's **track settings** page near the top of which will be a button labelled **View Data Quality (RUST plots).** Clicking this button will display a pdf file where each Ribo-seq file from that study will have its own page as in Figure 3. Each page consists of the following plots:*

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

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

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

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

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

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

Support Protocol 1: Uploading a custom track

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

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

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. Before uploading your alignment data as a custom track, convert your alignment file to either the bedGraph format or the compressed bigWig format if your file is large.

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

   One way to generate a bigWig file is to use RiboGalaxy (http://ribogalaxy.ucc.ie) (Michel et al., 2016) which is a Galaxy-based platform (Afgan et al., 2016) for pre-processing, mapping and analysing Ribo-seq and corresponding mRNA-seq data through a browser. The 'Convert formats' suite of tools in RiboGalaxy contains utilities to convert a SAM alignment file to BAM format and then convert the BAM file to bigWig. There are also a number of published workflows in RiboGalaxy that are tailored for the most commonly used genomes in GWIPS-viz and automatically output the alignment files in bigWig format with a direct link for exploration as a GWIPS-viz custom track.

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://gwips.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://gwips.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://gwips.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://gwips.ucc.ie/goldenPath/help/hgTrackHubHelp.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://gwips.ucc.ie/cgi-bin/hgTables.

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

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

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 **Table Browser** link in the topbar or in the navigation bar on the left hand side of the GWIPS-viz homepage. Alternatively on the GWIPS-viz Genome Browser page, mouse-over the **Tools** tab in the topbar and click on **Table Browser**.

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

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

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

3. Select the **group**, **track**, and **table** of interest using the dropdown menus.

> *The **group** is the list of track groups as they appear on the main **GWIPS-viz Genome Browser** page. So, for example, if you wanted to access data for elongating ribosome profiles, you would set the **group** to **Elongating ribosomes (A-site)**. The track is the name as it appears within the **track group** on the **GWIPS-viz Genome Browser page**. For the Ribo-seq and mRNA-seq groups, the list of tracks will be the names of hosted studies.*
>
> *The **table** is the name of the table file you wish to download. If you had selected a study in **track** dropdown menu, then the table dropdown menu would be a list of individual tracks associated with that study, as well as an **All** table which is the aggregation of all Ribo-seq or mRNA-seq tracks (depending on which group is selected) for that study.*

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

> *There are two radio buttons in this field, one labelled **genome** the other labelled **position**. If the **genome** radio button is selected any queries submitted will return data from all regions of the genome for the currently selected organism. Alternatively if set to **position** any submitted queries will only return data from the region specified in the text box adjacent to the **position** radio button. Positions should be entered in the form*
>
> *<chromosome name>:<start_position>-<end_position>*
>
> *for example:*

chrX:151283001-151290000

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

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

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

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

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

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

6. Hit the **get output** button.

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

Optional steps:

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

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

8. Press the **Create** button next to **subtrack merge**.

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

9. Press the **Create** button next to **intersection**.

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

10. Press the **Create** button next to **correlation**.

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

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

Support Protocol 3: Using the Downloads page

The GWIPS-viz Downloads page can be accessed at https://gwips.ucc.ie/downloads/index.html.

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

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.

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

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

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

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

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

GUIDELINES FOR UNDERSTANDING RESULTS

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

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

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

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

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

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

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

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

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

COMMENTARY

Development of the GWIPS-viz Genome Browser

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

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

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

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

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

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

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

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

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

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

CRITICAL PARAMETERS AND TROUBLE SHOOTING

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

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

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

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

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

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

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

REFERENCES

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., . . . Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res, 44*(W1), W3-W10. doi:10.1093/nar/gkw343

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., . . . Searle, S. M. (2016). The Ensembl gene annotation system. *Database (Oxford), 2016.* doi:10.1093/database/baw093

Andorf, C. M., Cannon, E. K., Portwood, J. L., 2nd, Gardiner, J. M., Harper, L. C., Schaeffer, M. L., . . . Lawrence-Dill, C. J. (2016). MaizeGDB update: new tools, data and interface for the maize model organism database. *Nucleic Acids Res, 44*(D1), D1195-1201. doi:10.1093/nar/gkv1007

Andreev, D. E., O'Connor, P. B., Fahey, C., Kenny, E. M., Terenin, I. M., Dmitriev, S. E., . . . Baranov, P. V. (2015). Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife, 4*, e03971. doi:10.7554/eLife.03971

Andreev, D. E., O'Connor, P. B., Loughran, G., Dmitriev, S. E., Baranov, P. V., & Shatsky, I. N. (2017). Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res, 45*(2), 513-526. doi:10.1093/nar/gkw1190

Andreev, D. E., O'Connor, P. B., Zhdanov, A. V., Dmitriev, R. I., Shatsky, I. N., Papkovsky, D. B., & Baranov, P. V. (2015). Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol, 16*, 90. doi:10.1186/s13059-015-0651-z

Aramayo, R., & Polymenis, M. (2017). Ribosome profiling the cell cycle: lessons and challenges. *Curr Genet*. doi:10.1007/s00294-017-0698-3

Archer, S. K., Shirokikh, N. E., Beilharz, T. H., & Preiss, T. (2016). Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature, 535*(7613), 570-574. doi:10.1038/nature18647

Bartholomaus, A., Del Campo, C., & Ignatova, Z. (2016). Mapping the non-standardized biases of ribosome profiling. *Biol Chem, 397*(1), 23-35. doi:10.1515/hsz-2015-0197

Baudin-Baillieu, A., Hatin, I., Legendre, R., & Namy, O. (2016). Translation Analysis at the Genome Scale by Ribosome Profiling. *Methods Mol Biol, 1361*, 105-124. doi:10.1007/978-1-4939-3079-1_7

Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., . . . Giraldez, A. J. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J, 33*(9), 981-993. doi:10.1002/embj.201488411

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis, 53*(8), 474-485. doi:10.1002/dvg.22877

Brar, G. A., & Weissman, J. S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol, 16*(11), 651-664. doi:10.1038/nrm4069

Brown, A., Rathore, S., Kimanius, D., Aibara, S., Bai, X. C., Rorbach, J., . . . Ramakrishnan, V. (2017). Structures of the human mitochondrial ribosome in native states of assembly. *Nat Struct Mol Biol*. doi:10.1038/nsmb.3464

Calviello, L., & Ohler, U. (2017). Beyond Read-Counts: Ribo-seq Data Analysis to Understand the Functions of the Transcriptome. *Trends Genet*. doi:10.1016/j.tig.2017.08.003

Carja, O., Xing, T., Plotkin, J. B., & Shah, P. (2017). riboviz: analysis and visualization of ribosome profiling datasets. *bioRxiv*.

Cherry, J. M. (2015). The Saccharomyces Genome Database: A Tool for Discovery. *Cold Spring Harb Protoc, 2015*(12), pdb top083840. doi:10.1101/pdb.top083840

Chung, B. Y., Hardcastle, T. J., Jones, J. D., Irigoyen, N., Firth, A. E., Baulcombe, D. C., & Brierley, I. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA, 21*(10), 1731-1745. doi:10.1261/rna.052548.115

Dunn, J. G., & Weissman, J. S. (2016). Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics, 17*(1), 958. doi:10.1186/s12864-016-3278-x

Gerashchenko, M. V., & Gladyshev, V. N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res, 42*(17), e134. doi:10.1093/nar/gku671

Gerashchenko, M. V., & Gladyshev, V. N. (2017). Ribonuclease selection for ribosome profiling. *Nucleic Acids Res, 45*(2), e6. doi:10.1093/nar/gkw822

Gobet, C., & Naef, F. (2017). Ribosome profiling and dynamic regulation of translation in mammals. *Curr Opin Genet Dev, 43*, 120-127. doi:10.1016/j.gde.2017.03.005

Gramates, L. S., Marygold, S. J., Santos, G. D., Urbano, J. M., Antonazzo, G., Matthews, B. B., . . . the FlyBase, C. (2017). FlyBase at 25: looking to the future. *Nucleic Acids Res, 45*(D1), D663-D671. doi:10.1093/nar/gkw1016

Hao, Y., Zhang, L., Niu, Y., Cai, T., Luo, J., He, S., . . . Chen, R. (2017). SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief Bioinform*. doi:10.1093/bib/bbx005

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., . . . Hubbard, T. J. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res, 22*(9), 1760-1774. doi:10.1101/gr.135350.111

Hussmann, J. A., Patchett, S., Johnson, A., Sawyer, S., & Press, W. H. (2015). Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast. *PLoS Genet, 11*(12), e1005732. doi:10.1371/journal.pgen.1005732

Ingolia, N. T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet, 15*(3), 205-213. doi:10.1038/nrg3645

Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J., Jackson, S. E., . . . Weissman, J. S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep, 8*(5), 1365-1379. doi:10.1016/j.celrep.2014.07.045

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science, 324*(5924), 218-223. doi:10.1126/science.1168978

Kumari, R., Michel, A. M., & Baranov, P. V. (2017). PausePred and Rfeet: Webtools for inferring ribosome pauses and visualizing footprint density from ribosome profiling data. Manuscript submitted for publication.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol, 10*(3), R25. doi:10.1186/gb-2009-10-3-r25

Lauria, F., Tebaldi, T., Bernabo, P., Groen, E. J., Gillingwater, T. H., & Viero, G. (2017). riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. *bioRxiv*, 169862.

Legendre, R., Baudin-Baillieu, A., Hatin, I., & Namy, O. (2015). RiboTools: a Galaxy toolbox for qualitative ribosome profiling analysis. *Bioinformatics, 31*(15), 2586-2588. doi:10.1093/bioinformatics/btv174

Leinonen, R., Sugawara, H., Shumway, M., & International Nucleotide Sequence Database, C. (2011). The sequence read archive. *Nucleic Acids Res, 39*(Database issue), D19-21. doi:10.1093/nar/gkq1019

Liu, W., Xiang, L., Zheng, T., Jin, J., & Zhang, G. (2017). TranslatomeDB: a comprehensive database and cloud-based analysis platform for translatome sequencing data. *Nucleic Acids Research, 46*(D1). doi:10.1093/nar/gkx1034

Malone, B., Atanassov, I., Aeschimann, F., Li, X., Grosshans, H., & Dieterich, C. (2017). Bayesian prediction of RNA translation from ribosome profiling. *Nucleic Acids Res, 45*(6), 2960-2972. doi:10.1093/nar/gkw1350

Martens, A. T., Taylor, J., & Hilser, V. J. (2015). Ribosome A and P sites revealed by length analysis of ribosome profiling data. *Nucleic Acids Res, 43*(7), 3680-3687. doi:10.1093/nar/gkv200

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnetjournal, 17*.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., . . . Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol, 28*(5), 495-501. doi:10.1038/nbt.1630

Michel, A. M., & Baranov, P. V. (2013). Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA, 4*(5), 473-490. doi:10.1002/wrna.1172

Michel, A. M., Choudhury, K. R., Firth, A. E., Ingolia, N. T., Atkins, J. F., & Baranov, P. V. (2012). Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res, 22*(11), 2219-2229. doi:10.1101/gr.133249.111

Michel, A. M., Fox, G., A, M. K., De Bo, C., O'Connor, P. B., Heaphy, S. M., . . . Baranov, P. V. (2014). GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res, 42*(Database issue), D859-864. doi:10.1093/nar/gkt1035

Michel, A. M., Kiniry, S. J., O'Connor, Patrick B. F., Mullan, J. P., & Baranov, P. V. (2017). GWIPS-viz: 2018 update. *Nucleic Acids Research*. doi:10.1093/nar/gkx790

Michel, A. M., Mullan, J. P., Velayudhan, V., O'Connor, P. B., Donohue, C. A., & Baranov, P. V. (2016). RiboGalaxy: A browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biol, 13*(3), 316-319. doi:10.1080/15476286.2016.1141862

Mumtaz, M. A., & Couso, J. P. (2015). Ribosomal profiling adds new coding sequences to the proteome. *Biochem Soc Trans, 43*(6), 1271-1276. doi:10.1042/BST20150170

O'Connor, P. B., Andreev, D. E., & Baranov, P. V. (2016). Comparative survey of the relative impact of mRNA features on local ribosome profiling read density. *Nat Commun, 7*, 12915. doi:10.1038/ncomms12915

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res, 44*(D1), D733-745. doi:10.1093/nar/gkv1189

Olexiouk, V., Crappe, J., Verbruggen, S., Verhegen, K., Martens, L., & Menschaert, G. (2016). sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res, 44*(D1), D324-329. doi:10.1093/nar/gkv1175

Park, J. E., Yi, H., Kim, Y., Chang, H., & Kim, V. N. (2016). Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Mol Cell, 62*(3), 462-471. doi:10.1016/j.molcel.2016.04.007

Parsons, M., & Myler, P. J. (2016). Illuminating Parasite Protein Production by Ribosome Profiling. *Trends Parasitol, 32*(6), 446-457. doi:10.1016/j.pt.2016.03.005

Pohl, A., & Beato, M. (2014). bwtool: a tool for bigWig files. *Bioinformatics, 30*(11), 1618-1619. doi:10.1093/bioinformatics/btu056

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res, 20*(1), 110-121. doi:10.1101/gr.097857.109

Popa, A., Lebrigand, K., Paquet, A., Nottet, N., Robbe-Sermesant, K., Waldmann, R., & Barbry, P. (2016). RiboProfiling: a Bioconductor package for standard Ribo-seq pipeline processing. *F1000Res, 5*, 1309. doi:10.12688/f1000research.8964.1

Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., . . . Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res, 44*(W1), W160-165. doi:10.1093/nar/gkw257

Sendoel, A., Dunn, J. G., Rodriguez, E. H., Naik, S., Gomez, N. C., Hurwitz, B., . . . Fuchs, E. (2017). Translation from unconventional 5' start sites drives tumour initiation. *Nature, 541*(7638), 494-499. doi:10.1038/nature21036

Stern-Ginossar, N. (2015). Decoding viral infection by ribosome profiling. *J Virol, 89*(12), 6164-6166. doi:10.1128/JVI.02528-14

Stern-Ginossar, N., & Ingolia, N. T. (2015). Ribosome Profiling as a Tool to Decipher Viral Complexity. *Annu Rev Virol, 2*(1), 335-349. doi:10.1146/annurev-virology-100114-054854

Tyner, C., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Eisenhart, C., . . . Kent, W. J. (2017). The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res, 45*(D1), D626-D634. doi:10.1093/nar/gkw1134

Yordanova, M. M., Loughran, G., Zhdanov, A. V., Mariotti, M., Kiniry, S. J., O'Connor, P. B., . . . Baranov, P. V. (2018). AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation. *Nature*. doi:10.1038/nature25174

Wang, H., Wang, Y., & Xie, Z. (2017). Computational resources for ribosome profiling: from database to Web server and software. *Brief Bioinform*. doi:10.1093/bib/bbx093

Wethmar, K., Barbosa-Silva, A., Andrade-Navarro, M. A., & Leutz, A. (2014). uORFdb--a comprehensive literature database on eukaryotic uORF biology. *Nucleic Acids Res, 42*(Database issue), D60-67. doi:10.1093/nar/gkt952

Woolstenhulme, C. J., Guydosh, N. R., Green, R., & Buskirk, A. R. (2015). High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep, 11*(1), 13-21. doi:10.1016/j.celrep.2015.03.014

Xie, S. Q., Nie, P., Wang, Y., Wang, H., Li, H., Yang, Z., . . . Xie, Z. (2016). RPFdb: a database for genome wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res, 44*(D1), D254-258. doi:10.1093/nar/gkv972

Figure legends

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

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

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

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

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

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

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

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

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

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

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

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

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

Figure 12. The GWIPS-viz forum. Queries about using GWIPS-viz can be posted to the GWIPS-viz forum (https://gwips.ucc.ie/Forum/). Other related topics of discussion include RiboGalaxy, ribosome profiling, translation/protein synthesis in general, sequencing and Bioinformatics.

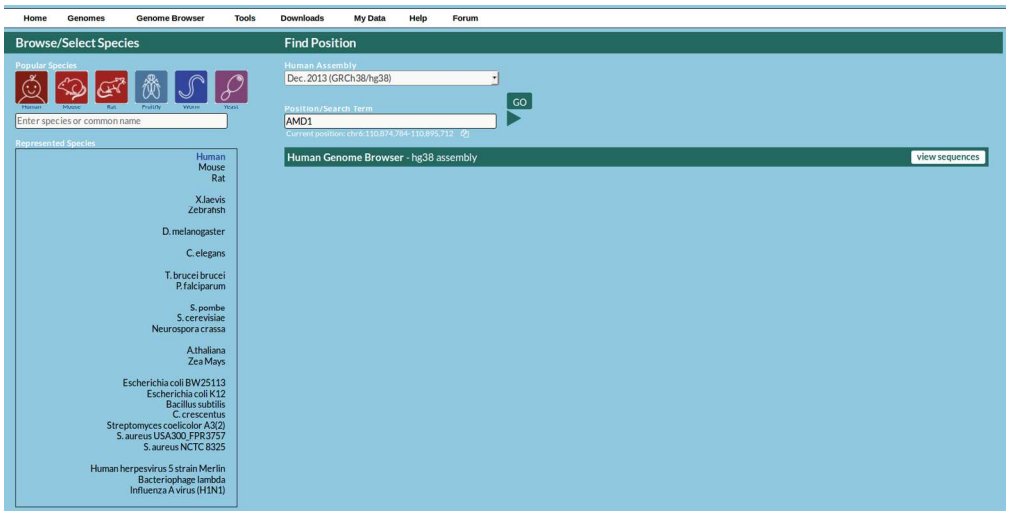


Figure 1

644x325mm (72 x 72 DPI)



Figure 2

575x136mm (72 x 72 DPI)

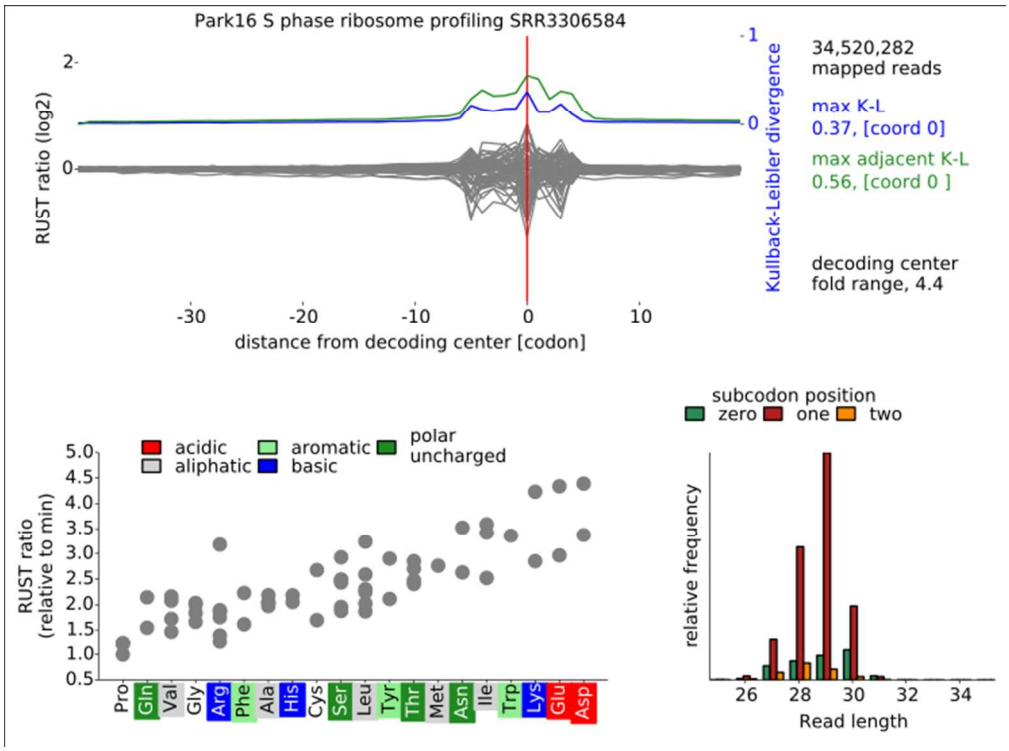


Figure 3

196x145mm (96 x 96 DPI)

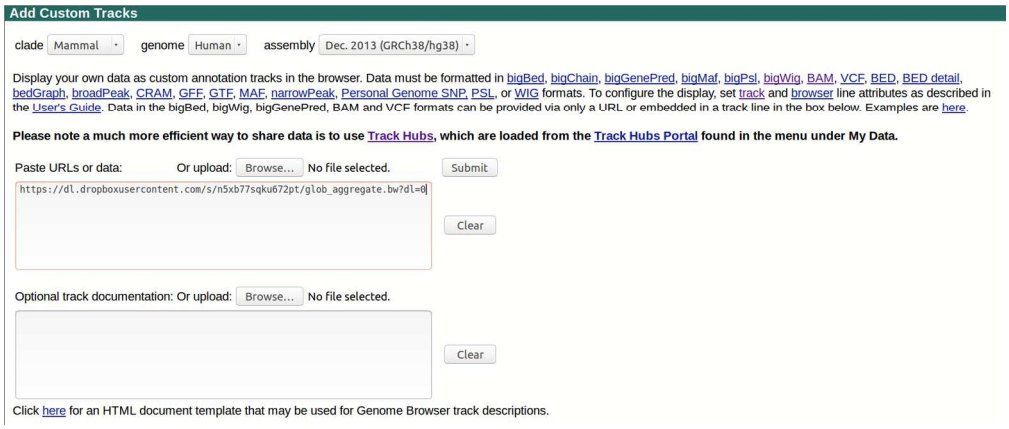


Figure 4

483x202mm (96 x 96 DPI)

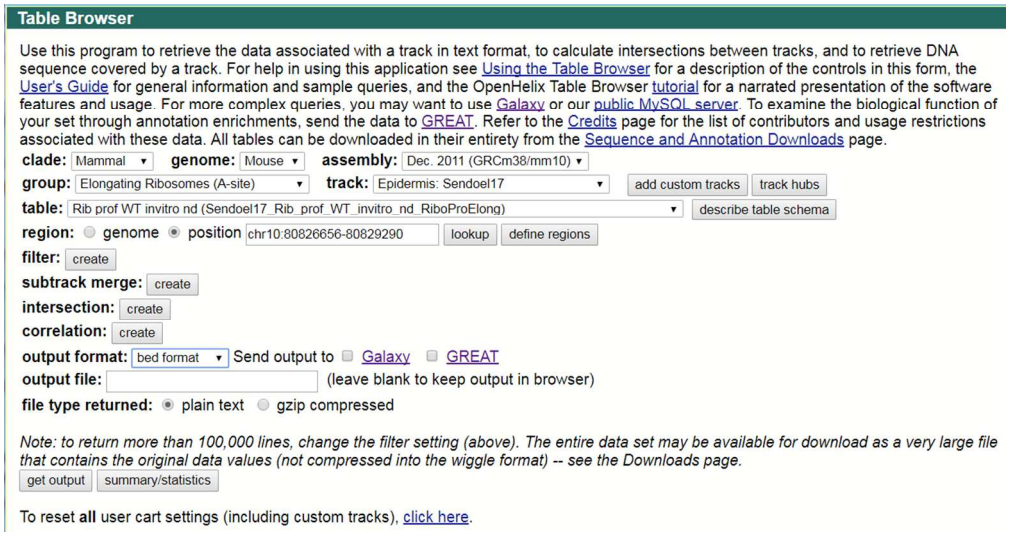


Figure 5

341x180mm (96 x 96 DPI)

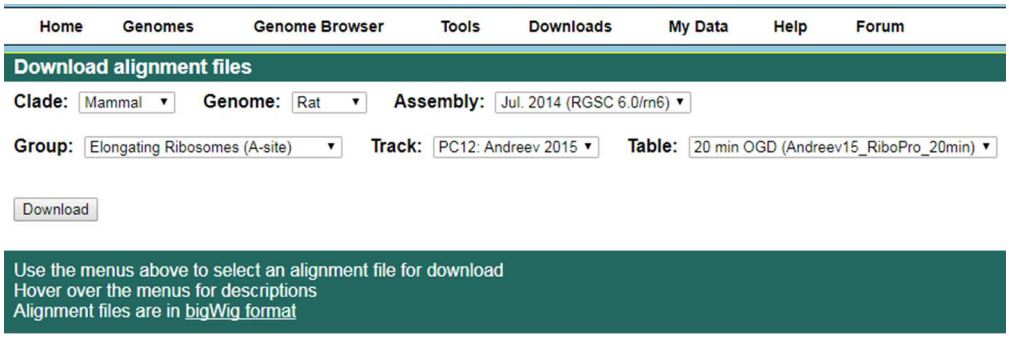


Figure 6

236x77mm (96 x 96 DPI)

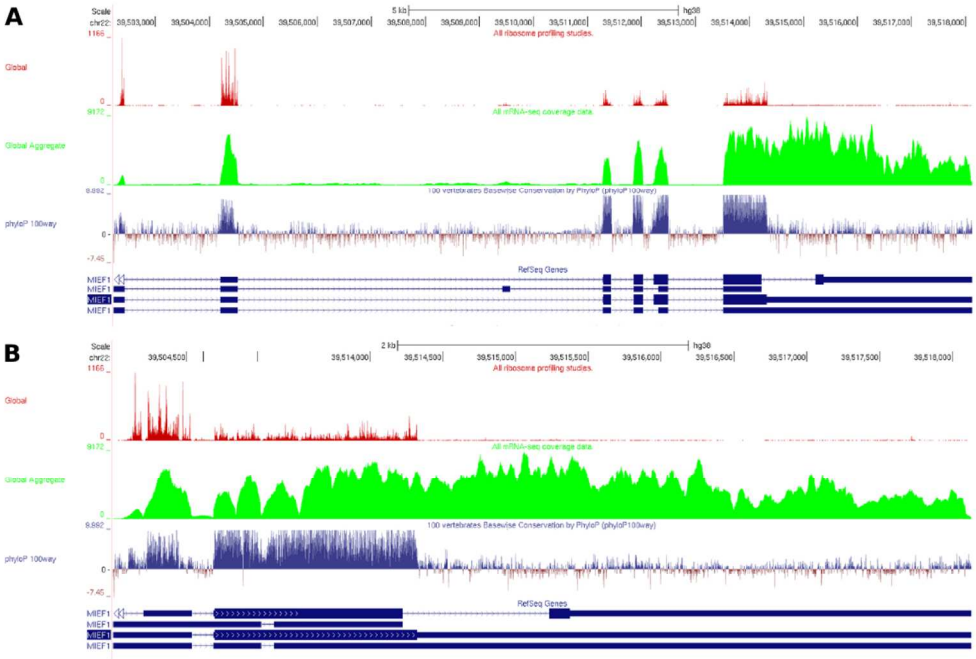


Figure 7

417x280mm (72 x 72 DPI)

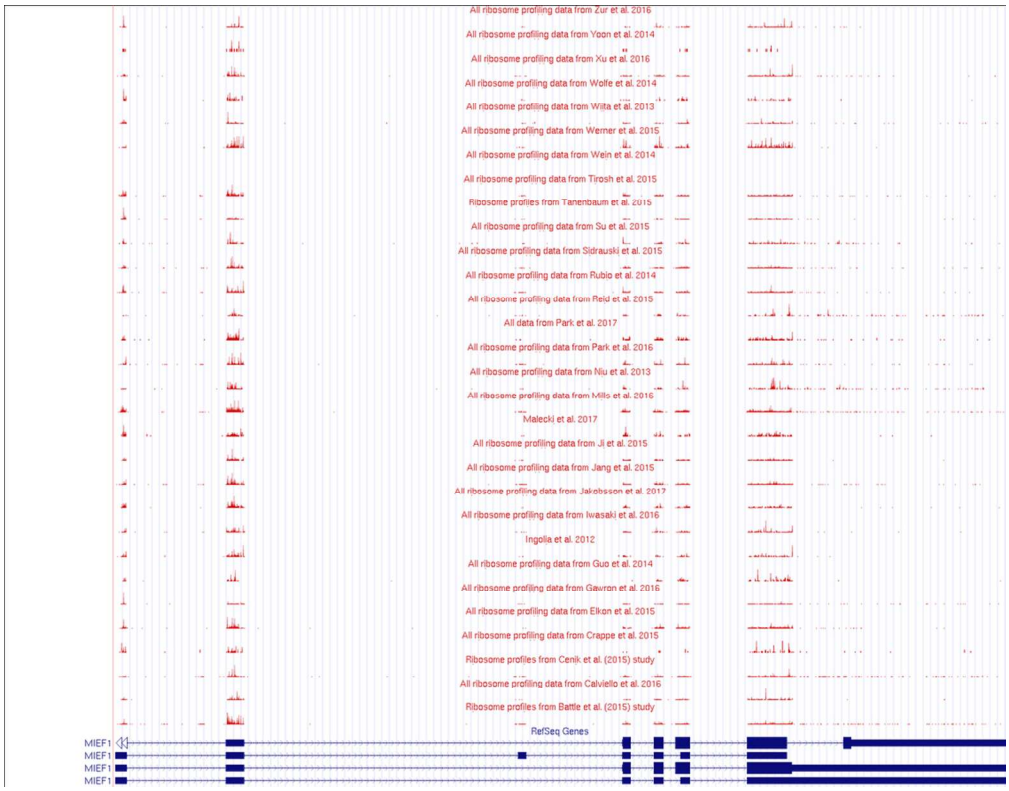


Figure 8

299x233mm (96 x 96 DPI)

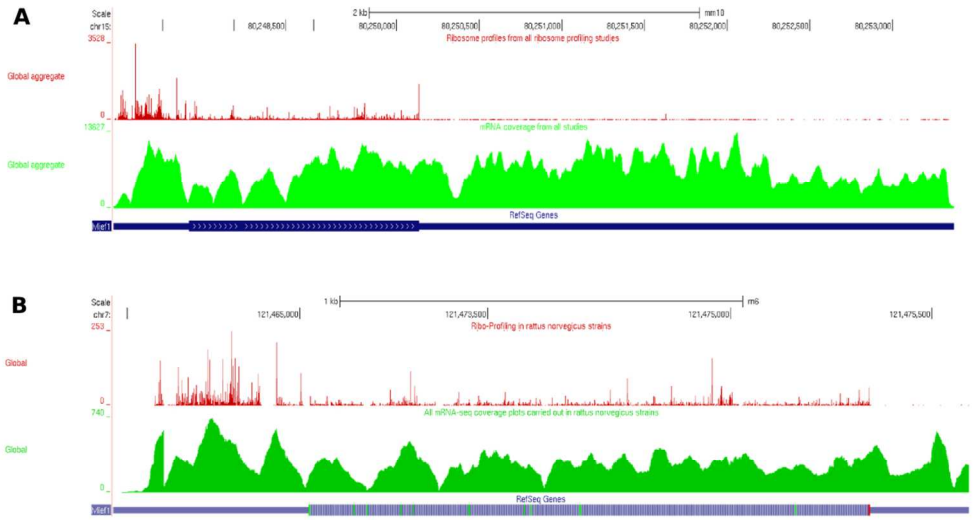


Figure 9

449x249mm (72 x 72 DPI)

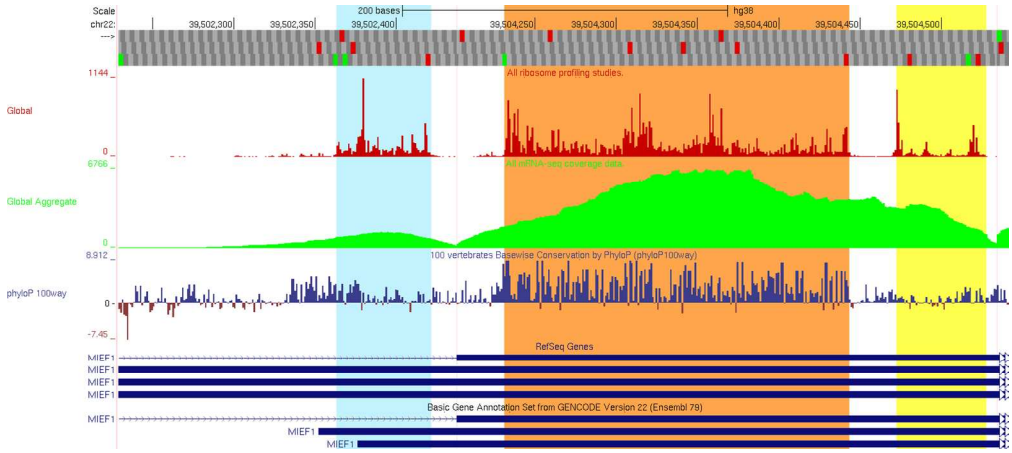


Figure 10

427x189mm (96 x 96 DPI)

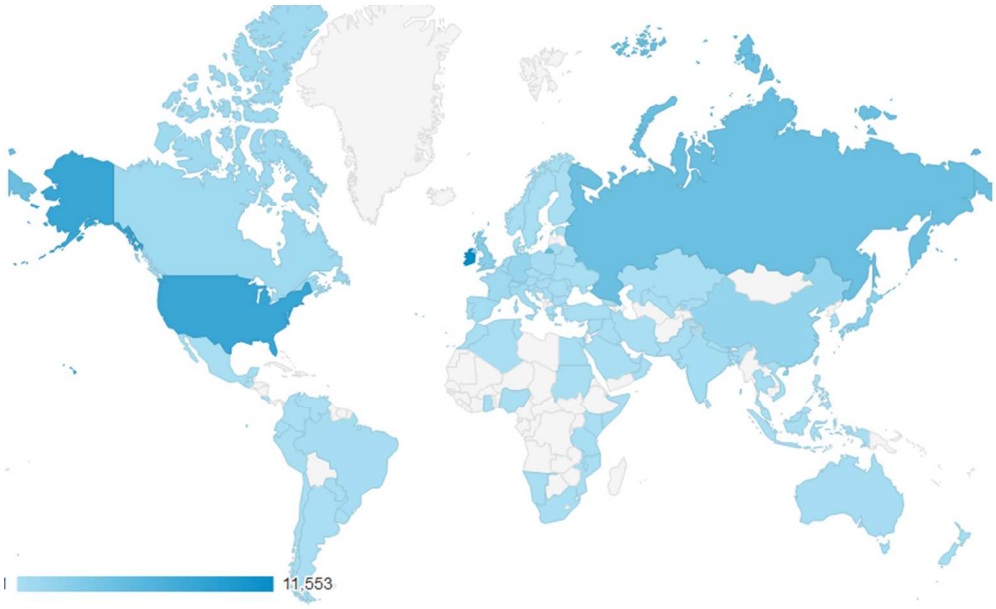


Figure 11

300x181mm (72 x 72 DPI)



Figure 12

221x140mm (96 x 96 DPI)