Computer Methods and Programs in Biomedicine 181 (2019) 104806

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

mQC: A post-mapping data exploration tool for ribosome profiling

Steven Verbruggen*, Gerben Menschaert*

BioBix, Lab of Bioinformatics and Computational Genomics, Department of Mathematical Modelling, Statistics and Bioinformatics, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, Ghent 9000, Belgium

ARTICLE INFO

Article history:
Received 30 March 2018
Revised 5 October 2018
Accepted 26 October 2018

Keywords:
Ribosome profiling
NGS
Quality visualization
Triplet periodicity

ABSTRACT

Background and objective: Ribosome profiling is a recent next generation sequencing technique enabling the genome-wide study of gene expression in biomedical research at the translation level. Too often, researchers precipitously start trying to test their hypotheses after alignment of their data, without checking the quality and the general features of their mapped data. Despite the fact that these checks are essential to prevent errors and ensure valid conclusions afterwards, easy-to-use tools for visualizing the quality and overall outlook of mapped ribosome profiling data are lacking.

Methods: We present mQC, a modular tool implemented as a Bioconda package and also available in the Galaxy tool shed. Herewith both bio-informaticians as well as non-experts can easily perform the indispensable visualization of both the quality and the general features of their mapped P-site corrected ribosome profiling reads. The user manual, the raw code and more information can be found on its GitHub repository (https://github.com/Biobix/mQC).

Results: mQC was tested on multiple datasets to assess its general applicability and was compared to other tools that partly perform similar tasks.

Conclusions: Our results demonstrate that mQC can accomplish an unfilled but essential position in the ribosome profiling data analysis procedure by performing a thorough RIBO-Seq-specific exploration of aligned and P-site corrected ribosome profiling data.

1. Introduction

Next generation sequencing (NGS) technologies changed and are still changing the clinical research tremendously. These high-throughput technologies generate large complex datasets with a persistently dropping cost. With this collection of novel techniques, biomedical research can be performed on a wider scale and with much more detail [1]. For instance, RNA-seq gained a lot of popularity in biomedicine for gene expression applications over the last few years. This technology brings the deep sequencing strategy to the level of the transcripts and their isoforms [2].

New types of data are popping out of these new technologies and if biomedical researchers want to analyze their results thoroughly, they mostly face a wall of, for them, inaccessible software. However, with the current emergence of integrative bioinformatics frameworks, reproductive data analysis tools find their way to a broader public [3]. For example, the Galaxy project offers a framework for accessible and transparent computational research in life sciences [4]. Usage statistics of the project [5] clearly outline the need for such a framework as the number of tools and users is in a continuous rise. It is clear that through such projects the accessibility gap between data generation and data analysis is starting to close.

Some of the first and most important steps in the analysis of NGS data comprehend data exploration and quality control (QC). Quality checks are needed in large experiments and in experiments over different sites and platforms, but it can also be very valuable in the context of smaller scale studies [6,7]. Raw sequencing data needs to be filtered for contaminations, over-represented sequences and bad reads. Furthermore, base adaptors and low-quality sequences need to be trimmed off. These steps are necessary for the validity of the downstream analyses afterwards as well as for improving the reproducibility [8]. At the same time, metagenic plots and inspections help to give the researcher an idea of the general outlook of his data.

* Corresponding authors.
E-mail addresses: Steven.Verbruggen@UGent.be (S. Verbruggen), Gerben.Menschaert@UGent.be (G. Menschaert).

https://doi.org/10.1016/j.cmpb.2018.10.018

Nomenclature

NGS	next generation sequencing
ORF	open reading frame
QC	quality control
RPF	ribosome-protected fragment

For raw data, there are already a lot of quality inspection tools available. Both FastQC [9] and HTQC [10] perform a general visual assessment of read quality, read length composition, GC bias and base composition. A big advantage of FastQC is that it works on both the raw FASTQ files as well as on SAM and BAM alignment files. Therefore, quality assessment can be done before and after mapping to the reference information (genome or transcriptome). Also the HTSeq package has some functionalities to analyze the base qualities and base compositions of both FASTQ and SAM files [11].

Next, there are some QC tools specially devised for mapped RNA-Seq reads. As RNA-Seq specific features (e.g. splice junctions) are only available for investigation after mapping, these tools work on SAM or BAM files. First, RSeQC [12] is a tool to investigate read-clipping, read distribution, coverage uniformity over the gene body, reproducibility and strand specificity. The saturation status, expression and annotation of splice junctions is also checked next to some basic features that were included in earlier mentioned QC tools. Second, RNA-SeQC [13] offers a broad analysis of, among other things, read counts, read count annotation, duplication rates, GC bias, alignment regions, coverage and transcript counts. Thirdly, Qualimap2 [14] is quite similar to RNA-SeQC but provides some additional features like extended multi-sample comparison. Last, EDASeq [15] is an R package for exploratory investigation of reads and read quality but combines it with some RNA-Seq specific data normalization.

In our opinion, it is optimal to do quality control at least twice for RNA-Seq: initially a basic QC on the raw reads and later a more specific QC on the alignment files.

A few years ago, a new technique called ribosome profiling (RIBO-Seq) broke new ground in gene expression research [16,17]. In this technique, ribosomal arrest is achieved by applying antibiotics. Afterwards, all ribosome-protected RNA fragments that are not digested by a nuclease treatment, are analyzed with deep sequencing. In that way, only the small portion of mRNA that is in the mRNA channel of the ribosome during translation, is sequenced. Based on the signal around canonical initiation events, we can estimate the position of the ribosomal P-site within the read alignments as a function of the ribosome-protected fragment (RPF) length [18]. In that way, it is possible to attain the exact positions of the translating ribosomes at base resolution. As such, it opens the way for a genome-wide analysis of translation on subcodon resolution. It allows to investigate the gene expression in a whole new fashion, looking at the stage of the translation process but still with the advantages of NGS [16,17].

Post-transcriptional processes explain as much variation in the final protein abundance as the mRNA levels [19,20], so for gene expression studies in biomedical contexts it is important to include this post-transcriptional information by introducing ribosome profiling in this research field. The first steps to achieve this introduction are taken as ribosome profiling becomes ever easier and cheaper to perform [21]. Located within the fields of virology [22,23], neurology [24] and cancer research [25–29], a first series of recent biomedical studies benefited from ribosome profiling to obtain an improved mapping of gene expression. Quality inspection of ribosome profiling data was however sparsely mentioned in these papers, but it is clear that with a growing interest in RIBO-Seq analysis from the biomedical field, the availability of specific QC tools is more important and more urgent than ever.

Generating data for biomedical research is one part, but the analysis is of course as important. In recent years, a wide range of computational tools was developed for RIBO-Seq, but in our opinion, none of these public tools allows a focused and complete quality check. Nevertheless, drawing conclusions and making discoveries with downstream tools is really dependent on the premise that we can assure the quality of the data [30]. And as the amount of generated RIBO-Seq data grows continuously [31], quality control tools become ever more essential to complete and justify the general analysis strategy [32,33].

The community is still heavily investigating all protocol factors (e.g. buffer conditions, cell lysis methods, nuclease treatments and antibiotics) which could have a possible influence on the obtained data [34]. Up until more is known about these possible confounders, standardized quality control tests are difficult to realize [35]. Nonetheless, for the downstream analysis pipeline, it is indispensable to have a clear overview of the general quality and features of the data. In our opinion, researchers pass too quickly from mapping onto hypothesis testing without having a general inspection of their aligned data. Therefore, tools to visualize the general features of the obtained data are needed in order to prevent errors and ensure downstream validity.

Standard RNA-Seq QC tools can be applied to analyze the general outlook of the RIBO-Seq data. But on top of that, additional quality checks and feature visualizations are necessary because of ribosome profiling specific characteristics as triplet periodicity (phasing) [16], the usage of translation inhibitors [36] and the

Table 1
Comparison of mQC with existing tools. (Abbreviations: ORF = open reading frame, RPF = ribosome protected fragment, QC = quality control).

	FastQC	mQC	RUST	RiboSeqR	Riboviz	Ribo-TISH
Technical features						
Ribosome profiling specific	No	Yes	Yes	Yes	Yes	Yes
For genome/transcriptome/ORF based alignment files	Raw/aligned	Genome	Transcriptome	Transcriptome	ORF	Genome
Metagenic annotation plots	No	Yes	No	No	No	No
P-site offsets	No	Calculated before plotting	No	No	No	Calculated after plotting
Frame analysis over RPF lengths	No	Yes	Yes ⋆	Yes ⋆	No	Yes ⋆
Analysis of framing relative to ORF location	No	Yes	No	Yes ⋆°	Yes ⋆°	Yes ⋆°
Analysis of framing relative to codon identity	No	Yes	No	No	No	No
Implementation and availability						
Availability of scripts	Own download page	GitHub	Own download page	Only with package manager	GitHub	GitHub
Package manager	Conda	Conda	No	Bioconductor	No	Pypi
Availability to non-experts	Galaxy tool shed	Galaxy tool shed	RiboGalaxy	RiboGalaxy	No	No
References			[31,34]	[38]	[39]	[40]
Notes	Available on https://www.bioinformatics.babraham.ac.uk/projects/fastqc/		RUST works only at the codon level whereas other tools are at the subcodon level		No separate QC possible, pipeline only executable as one total process	

⋆: Not P-site corrected
°: Only at the initiation and termination region

ribosome-protected fragment length distribution [37]. Therefore, we developed 'mQC', a feature and quality visualization tool that allows a thorough inspection of the mapped ribosome profiling data.

2. Background

An overview of existing tools for the visualization of ribosome profiling data is given in Table 1. Ribo-Seq unit step transformation (RUST) [34] started as a simple normalization method for RIBO-Seq but with its inclusion in GWIPS-viz [31], it was used for presenting triplet periodicity of the datasets included in the genome browser. Another package, called RiboSeqR [38], parses and aligns ribosomal data to coding sequences in order to obtain non P-site corrected frame plots. RiboSeqR and RUST are both included in RiboGalaxy (https://ribogalaxy.ucc.ie), a Galaxy-based ribosome profiling platform. Furthermore, Riboviz [39] provides tools for some RIBO-Seq data exploration and visualization, although these tools cannot be executed separately from the rest of the pipeline and the user-friendly web application only allows viewing pre-executed analyses of certain datasets. Finally, Ribo-TISH [40] is a ribosome profiling toolkit that also contains some quality visualization capabilities.

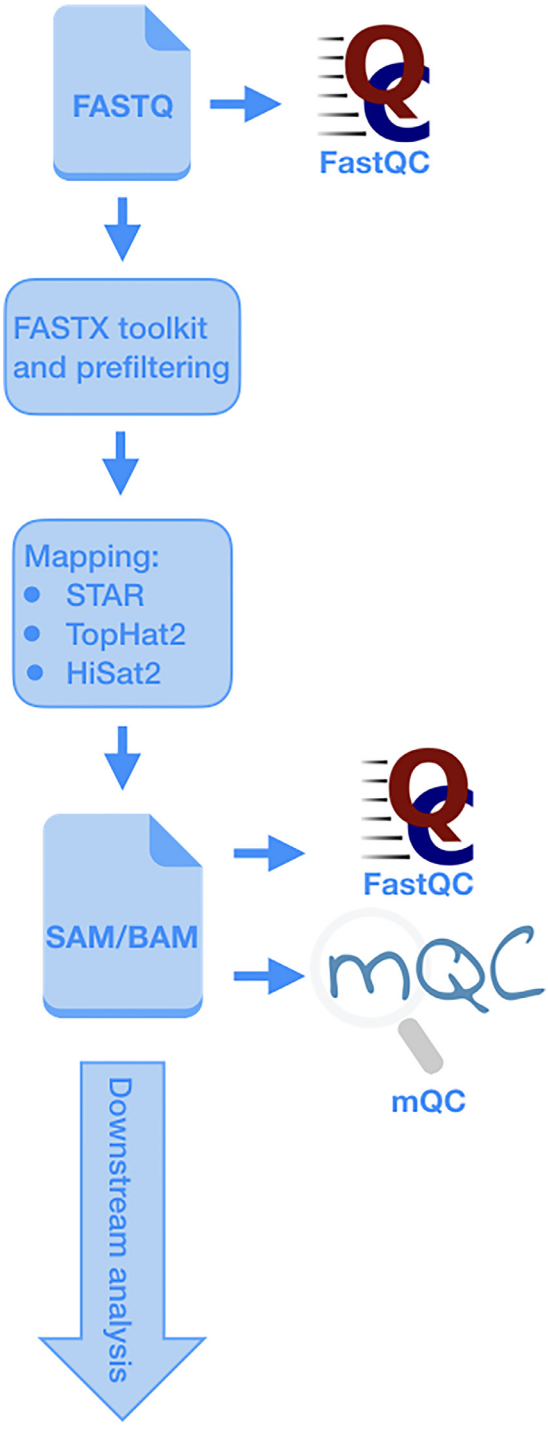
RPFs span around 26 to 32 nucleotides on the translated transcript. Somewhere internal in this RPF, there is the P-site, the site where the growing polypeptide chain is hold during translation. Furthermore, the start codon is also recognized in this site by the initiator tRNA. There is an offset in between the start (the 5' end) of the RPF and the P-site, called the P-site offset, and this offset is related to the length of the RPF [16]. It is important to know this offset for each RPF length as this will allow to pinpoint each read to the exact base position on the transcript where translation was taking place. Offset correction thus reveals the true translation profile at subcodon resolution and exposes the possible triplet periodicity. P-site offsets have to be calculated for each sample separately as they depend on wet-lab protocol conditions. Therefore, the distances between the observed start codon peaks and the canonical start sites need to be verified at the metagenic level, leading to an offset per RPF length. A lot of offset calculation tools are embedded in a bigger pipeline and this mostly obstructs their independent applicability. Plastid [18], on the other hand, can function outside its pipeline and outputs both the results and a metagenic overview plot of its start peak distance calculations per RPF length. Plots generated with all the aforementioned QC tools are not P-site corrected though, but due to their relevance for obtaining a correct translation profile, it is indispensable to apply the offsets before visualizing the RIBO-Seq specific features.

At the same time, the existing tools sometimes lack an intuitive interface for non-experts. We want our tool to be available to a broad public and therefore we included mQC both as a package in Bioconda and as a tool in the Galaxy framework.

3. Description of the method

Our tool, mQC, can be easily integrated in a general ribosome profiling workflow (Fig. 1). After basic quality control, adaptor trimming, base filtering and RNA filtering, the reads can be aligned to the reference genome with TopHat2 [41], HiSat2 [42] or STAR [43] mapping algorithms. Afterwards, the resulting BAM or SAM file can be entered into mQC.

The tool is written in Perl, complemented with Python and R for their plotting abilities. P-site offsets (i) can be used from a set of standard offsets [16], (ii) can be calculated from the inputted data using Plastid automatically during mQC runtime [18], (iii) can be set to constant 3' offsets [44] or (iv) can be suggested by the



Fi. 1. How mQC can be built into a general RIBO-Seq workflow. Raw ribosome profiling data in a FASTQ file can be checked for general quality with FastQC. Afterwards, adaptors can be clipped and reads can be filtered based on base quality with the FASTX toolset (http://hannonlab.cshl.edu/fastx_toolkit/). Several classes of RNA (e.g. rRNA) can be filtered out with any of the further mentioned mapping algorithms. Afterwards, mapping against the reference genome can be performed resulting in alignment files in SAM or BAM format. The general quality, again checked with FastQC, of the trimmed, filtered and aligned reads should have been improved compared to the raw reads. Ribosome profiling specific quality features can then be checked with mQC. If the quality is satisfactory, then one can continue with downstream analysis tools.

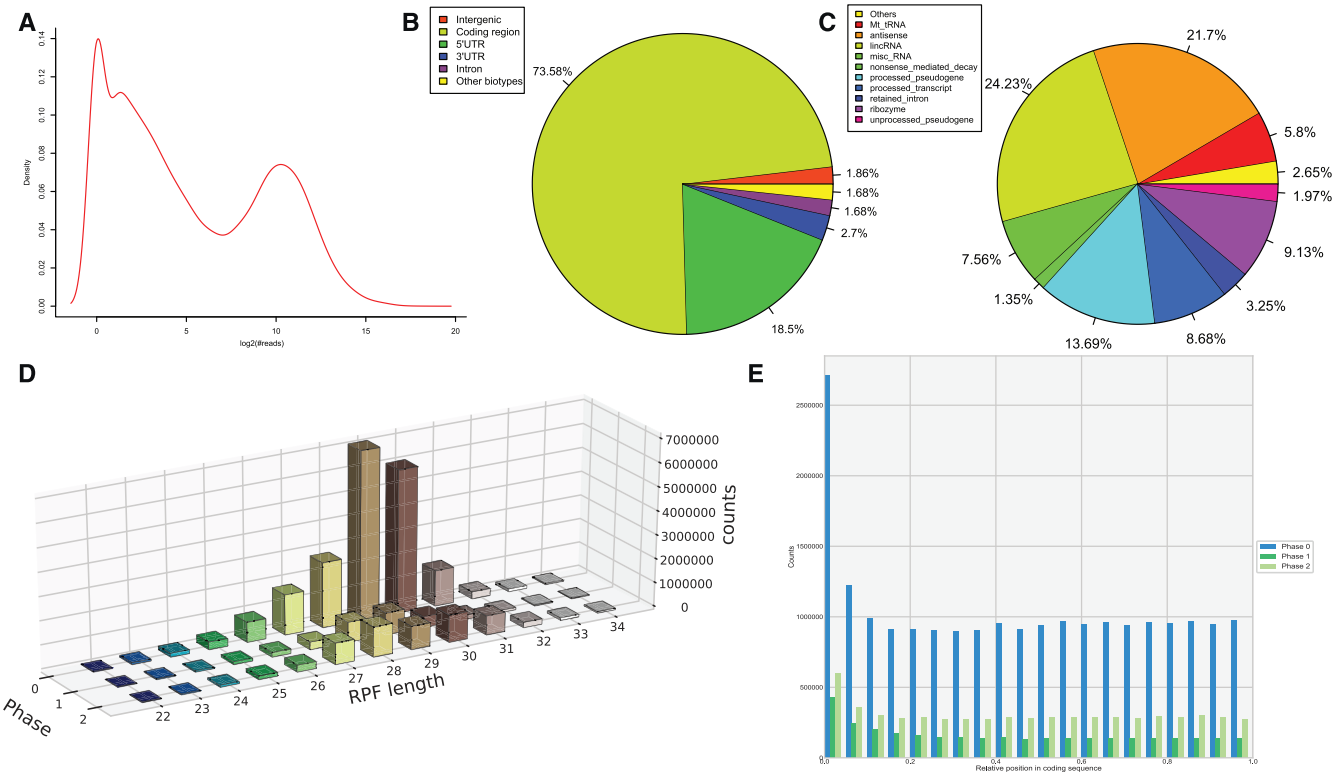


Fig. 2. Overview of the most important metagenic figures generated with mQC for cycloheximide-treated RIBO-Seq data in human colon cancer cell line HCT116. (A) Gene density plot. (B) Annotation pie chart. (C) Detail pie chart of the 'other biotypes' category from subfigure B. (D) 3D plot of the ribosomal phasing against the ribosome protected fragment (RPF) length. (E) Plot of the ribosomal phase against the relative position in the open reading frame.

user in the form of a tab separated list. Based on the annotation information of all protein-coding canonical transcripts from Ensembl [45], a library of canonical open reading frame (ORF) structures, including their ribosomal phase and codon information, is constructed. Alignments will be positioned onto this library, taking the P-site offsets into account, resulting in a detailed count distribution. These count distributions will then be used to plot different features of the aligned data. In the end, all figures are merged into a well-arranged HTML file. A ZIP file allows exporting all results.

Users can choose different input parameters of the mQC program. Besides the input SAM/BAM file and the species, the most important parameters are the Ensembl annotation version and the option to only use unique alignments. An overview of all parameters can be found in the user manual on the mQC GitHub repository (https://github.com/Biobix/mQC).

4. Results

mQC was tested on different datasets to show that the tool is generally applicable (supplemental files 1–15). These files give users a good overview of what can be expected from the different plotting tools. Supplemental file 1 describes in detail what can be learnt from these plots and what can be expected from high-quality data. Next to quality checks, these test runs resulted also in some interesting findings concerning the data itself.

In Fig. 2, results of the mQC analysis of our in-house human colon cancer cell line HCT116 [46], treated with the translation inhibitor cycloheximide, are given as an example set. Fig. 2A shows a gene density plot with two clearly separated peaks. A general difference in functionality between the genes making up these two peaks is observable, as clarified in supplemental figure 1. The biggest part of the reads resides mostly in coding regions and to a lesser extent in upstream untranslated regions (Fig. 2B). A more detailed distribution of the 'other biotypes' category can be useful to check previous read filtering steps for remaining unwanted alignments like mitochondrial tRNAs (Fig. 2C) or for coding evidence in lncRNAs [37]. Fig. 2D allows selecting the RPF lengths with the best phase distribution. In high-quality ribosome profiling datasets, a phase 0 enrichment is often observed (although data without this enrichment could also be of use downstream in the analysis). In Fig. 2E, the phase distribution in relationship to the overall location in the ORF is given and the influence of translation inhibitors like cycloheximide and harringtonine is clearly emerging at the initiation sites. Phase distribution can also give differing patterns for different codons and amino acids as can be seen in Fig. 3. In addition, mQC shows distributions of the total ribosome counts per codon identity in comparison to an averaged reference line constructed over different samples of the given species (Fig. 4). Biases in the protocol or differences in decoding rates can lead to major deviations from this reference line.

5. Discussion

As shown in Figs. 2–4, the clear report, outputted by mQC, offers users an instant overview of their mapped ribosome profiling data from different angles. This quality check is indispensable before proceeding with downstream data investigation. Moreover, our findings during test runs on different datasets from different

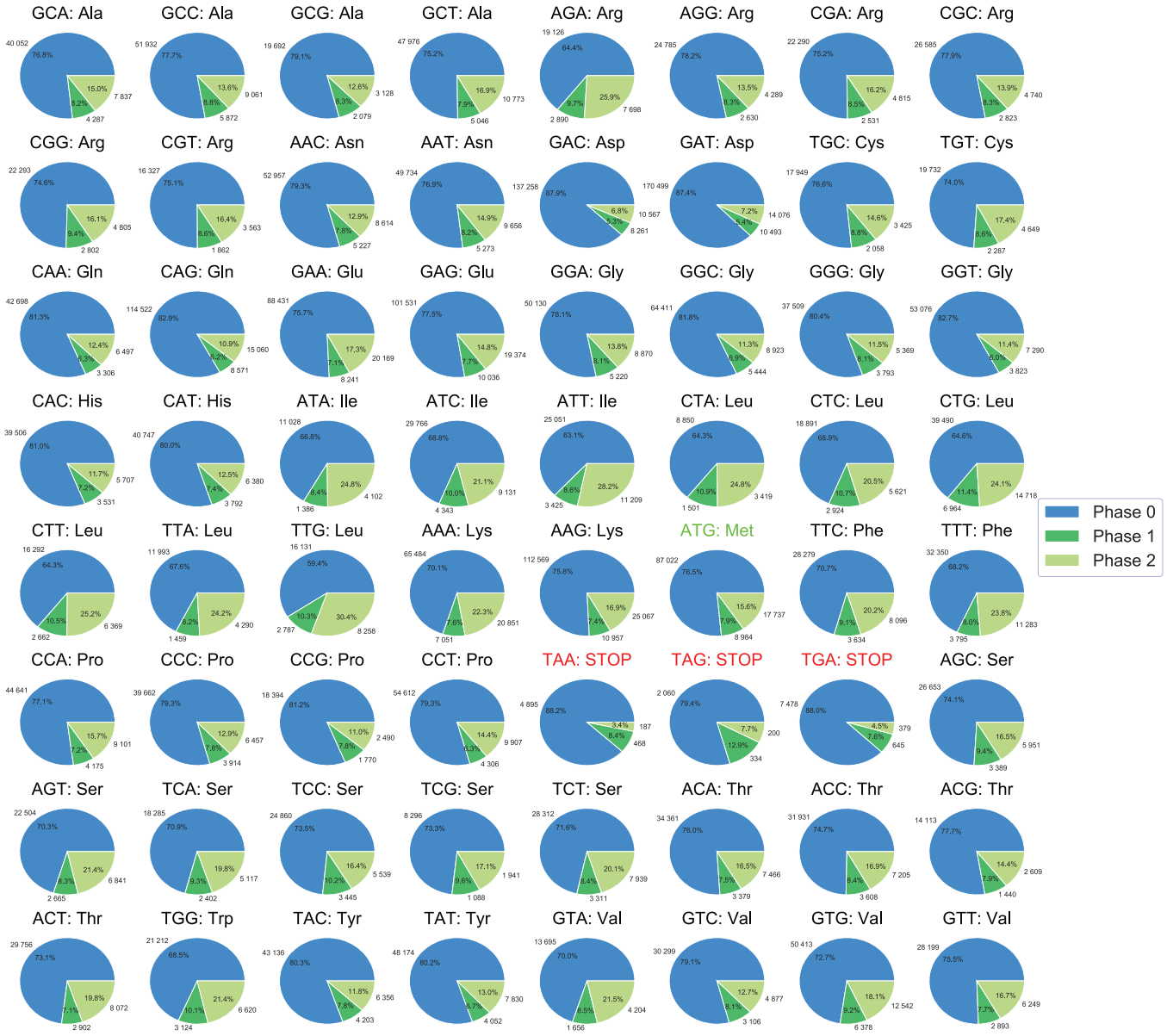


Fig. 3. Plot illustrating the relation between codon identity and the ribosomal phase distribution for cycloheximide-treated RIBO-Seq data in human embryonic kidney cell line HEK293 [49]. Synonymous codons for aspartate, glutamine, histidine and tyrosine show remarkable analogous phase distributions.

species (supplemental files 1–15) show the value and the broad applicability of mQC in ribosome profiling analysis.

Table 1 gives a thorough comparison with other tools available in the field that partly allow to perform similar visualization (sub)tasks on ribosome profiling data. Compared to most other ribosome profiling QC tools, mQC processes genomic alignment files. This is important in the light of (re-)annotation of translation, which is a major application of ribosome profiling. Furthermore, as described earlier, P-site offsets are an essential feature to pinpoint RPFs onto the right position and frame. By default, mQC calculates P-site corrections with Plastid [18], but users can easily change these corrections by inputting a tab-separated list. Unlike other tools, mQC applies the P-site offsets before plotting in order to investigate the principal ribosome profiling hallmarks properly. P-site corrected reads are plotted in function of RPF lengths, relative ORF locations and codon identities. In that way, mQC inspects ribosomal framing and triplet periodicity more elaborately than other existing tools.

6. Mode of availability of software

Importantly, mQC offers easy and broad availability, tailor-made for different target audiences. The raw code and manual is available on GitHub (https://github.com/Biobix/mQC), but to circumvent installation issues, mQC is also implemented as a Conda package so that all dependencies will be taken care of automatically. The tool is available under GNU GPL license. For non-bioinformaticians, an easy-to-use graphical version is available in the Galaxy tool shed. In that way, non-experts can execute mQC on their data of choose. A link to a pre-executed example of the Galaxy version of mQC can be found on our GitHub page as well.

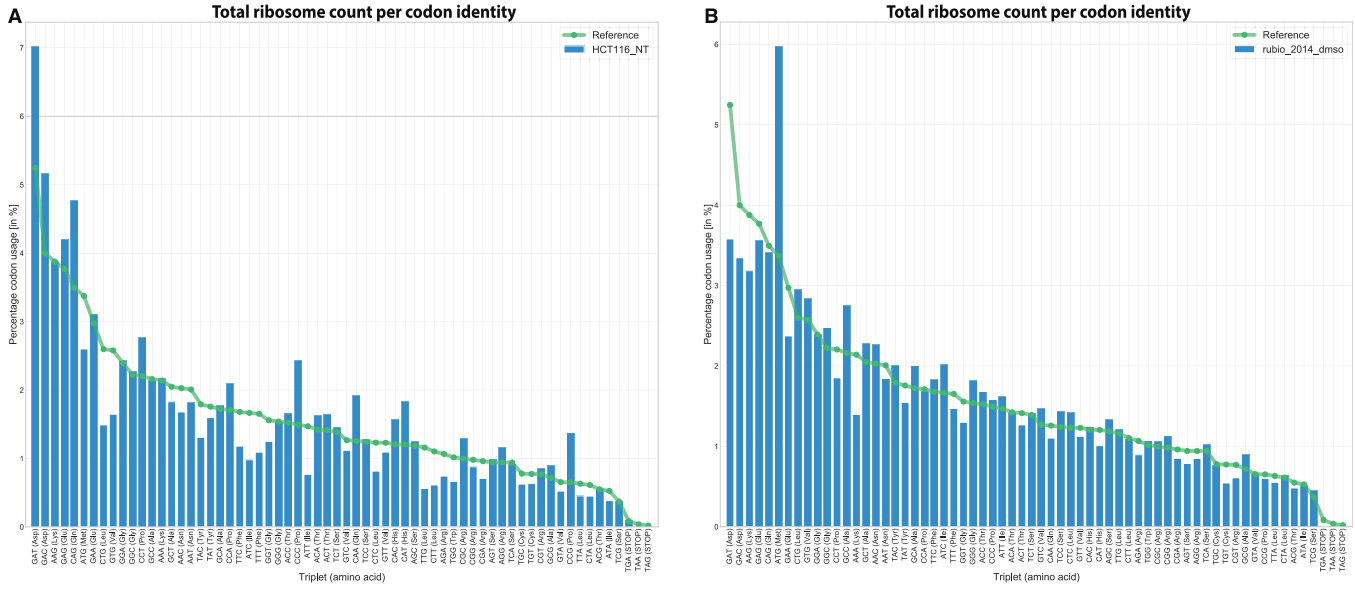


Fig. 4. Codon distribution plots. These plots describe the total ribosome counts over the different codons/amino acids compared to a reference line averaged over different samples of the same species. (A) Plot of human HCT116 data [46]. In this plot, an enrichment of all four codons of proline ('CCC', 'CCG', 'CCA' and 'CCT') can be observed compared to the reference line. (B) Plot of human MDA-MB231 DMSO control data [50]. An 'ATG' peak can be observed. As this peak was also observed in the silvestrol stress data of that same study (supplemental file 7), this points to a possible confounder in the common experimental protocol of both datasets. Analogous results were obtained when the codon counts were first normalized by the total count in the open reading frame (i.e. normalized for expression), as shown in supplemental figure 2.

7. Future plan

We expect to expand mQC further based on the feedback we get from users, although we have already some improvements in mind ourselves. First of all, we want to introduce other P-site offset callers into mQC. Now, users can apply offsets from Plastid [18], from a list of standard offsets [16] or from a user-defined tab-separated list. P-site offset results from other tools can already be introduced in mQC by inputting the results as a tab-separated file but we want to include additional callers [47] more directly. Second, the reference line of the codon usage plot leaves space for optimization. As the influence of more and more experimental factors on the data output will be better understood, these can be integrated in the construction of the reference line. We also plan to keep adding samples to the averaged reference with its further improvement in mind. Third, other annotation bundles than Ensembl (e.g. UCSC, RefSeq) can be included in mQC. Fourth, plots with additional metrics (such as ribosome residence time [48] or other measures of decoding rate) can surely complement the mQC output. Fifth, the relative position plots can be expanded to the immediate 5' and 3' regions just outside the coding sequence to study the transition from uncoding to coding parts and back. And last, we plan to develop an independent graphical user interface for users who prefer a graphical interface outside the Galaxy framework.

8. Conclusion

Our results demonstrate that mQC can accomplish an unfilled but essential position in the ribosome profiling data analysis procedure by performing a thorough RIBO-Seq-specific exploration of the mapping results in different perspectives. We ensured that our tool is available for a broad audience, including both bio-informaticians as well as wet lab researchers and that an easy-to-interpret report is outputted. This should allow to include the indispensable quality check in the ribosome profiling analysis pipeline, improving the usability of ribosome profiling for investigating gene expression in different research applications like biomedicine.

Competing interests

None declared.

Acknowledgements

We would like to thank Björn Grüning for helping us setting up the Bioconda package and the Galaxy version of mQC.

Funding

This work was supported by the Special Research Fund (BOF) of Ghent University [01D20615 to S.V.] and by a Postdoctoral Fellowship of the Research Foundation - Flanders (FWO-Vlaanderen) [12A7813N to G.M.].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2018.10.018.

References

[1] Y. Han, S. Gao, K. Muegge, W. Zhang, B. Zhou, Advanced applications of RNA sequencing and challenges, Bioinf. Biol. Insights 9 (2015) 29–46., doi:10.4137/BBI.S28991.
[2] E.T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, Alternative isoform regulation in human tissue transcriptomes, Nature 456 (2008) 470–476., doi:10.1038/nature07509.
[3] A. Nekrutenko, J. Taylor, Next-generation sequencing data interpretation: enhancing reproducibility and accessibility, Nat. Rev. Genet. 13 (2012) 667–672., doi:10.1038/nrg3305.
[4] J. Goecks, A. Nekrutenko, J. Taylor, Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences, Genome Biol. 11 (2010) R86., doi:10.1186/gb-2010-11-8-r86.
[5] Galaxy Project Stats, 2018. https://galaxyproject.org/galaxy-project/statistics/.
[6] S. Li, P.P. Łabaj, P. Zumbo, P. Sykacek, W. Shi, L. Shi, et al., Detecting and correcting systematic variation in large-scale RNA sequencing data, Nat. Biotechnol. 32 (2014) 888–895., doi:10.1038/nbt.3000.

[7] T. Lappalainen, M. Sammeth, M.R. Friedländer, P.A.C. 'T Hoen, J. Monlong, M.A. Rivas, et al., Transcriptome and genome sequencing uncovers functional variation in humans, Nature 501 (2013) 506–511., doi:10.1038/nature12531.

[8] Z. Su, P.P. Łabaj, S. Li, J. Thierry-Mieg, D. Thierry-Mieg, W. Shi, et al., A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium, Nat. Biotechnol. 32 (2014) 903–914, doi:10.1038/nbt.2957.

[9] S. Andrews, FastQC: a quality control tool for high throughput sequence data, 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[10] X. Yang, D. Liu, F. Liu, J. Wu, J. Zou, X. Xiao, et al., HTQC: A fast quality control toolkit for Illumina sequencing data, BMC Bioinf. 14 (33) (2013), doi:10.1186/1471-2105-14-33.

[11] S. Anders, P.T. Pyl, W. Huber, HTSeq-A Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2015) 166–169, doi:10.1093/bioinformatics/btu638.

[12] L. Wang, S. Wang, W. Li, RSeQC: quality control of RNA-seq experiments, Bioinformatics 28 (2012) 2184–2185, doi:10.1093/bioinformatics/bts356.

[13] D.S. DeLuca, J.Z. Levin, A. Sivachenko, T. Fennell, M.D. Nazaire, C. Williams, et al., RNA-SeQC: RNA-seq metrics for quality control and process optimization, Bioinformatics 28 (2012) 1530–1532, doi:10.1093/bioinformatics/bts196.

[14] K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data, Bioinformatics 32 (2015) 292–294, doi:10.1093/bioinformatics/btv566.

[15] D. Risso, EDASeq: exploratory data analysis and normalization for RNA-Seq, 2015. http://bioconductor.org/packages/devel/bioc/vignettes/EDASeq/inst/doc/EDASeq.pdf.

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

[17] N.T. Ingolia, G.a Brar, S. Rouskin, A.M. Mcgeachy, J.S. Weissman, The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments, Nat. Protoc. 7 (2012) 1534–1550, doi:10.1038/nprot.2012.086.

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

[19] J.B. Plotkin, Transcriptional regulation is only half the story, Mol. Syst. Biol. 6 (2010) 1–2., doi:10.1038/msb.2010.63.

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

[21] D.W. Reid, S. Shenolikar, C.V. Nicchitta, Simple and inexpensive ribosome profiling analysis of mRNA translation, Methods 91 (2015) 69–74., doi:10.1016/j.ymeth.2015.07.003.

[22] N. Irigoyen, A.M. Dinan, I. Brierley, A.E. Firth, Ribosome profiling of the retrovirus murine leukemia virus, Retrovirology 15 (2018) 1–8., doi:10.1186/s12977-018-0394-5.

[23] D.W. Reid, R.K. Campos, J.R. Child, T. Zheng, K.W.K. Chan, S.S. Bradrick, et al., Dengue virus selectively annexes endoplasmic reticulum-associated translation machinery as a strategy for co-opting host cell protein synthesis, J. Virol. 92 (2018) e01766-17, doi:10.1128/JVI.01766-17.

[24] M. Kapur, S.L. Ackerman, MRNA translation gone awry: translation fidelity and neurological disease, Trends Genet. 34 (2018) 218–231, doi:10.1016/j.tig.2017.12.007.

[25] B. Cao, Z. Fang, P. Liao, X. Zhou, J. Xiong, S. Zeng, et al., Cancer-mutated ribosome protein L22 (RPL22/eL22) suppresses cancer cell survival by blocking p53-MDM2 circuit, Oncotarget 8 (2017) 90651–90661, doi:10.18632/oncotarget.21544.

[26] P. Martin-Marcos, F. Zhou, C. Karunasiri, F. Zhang, J. Dong, J. Nanda, et al., eIF1A residues implicated in cancer stabilize translation preinitiation complexes and favor suboptimal initiation sites in yeast, Elife 6 (2017) e31250, doi:10.7554/eLife.31250.

[27] D.L. Kiss, W. Baez, K. Huebner, R. Bundschuh, D.R. Schoenberg, Impact of FHIT loss on the translation of cancer-associated mRNAs, Mol. Cancer 16 (2017) 179, doi:10.1186/s12943-017-0749-x.

[28] K.A. Kruth, M. Fang, D.N. Shelton, O. Abu-Halawa, R. Mahling, H. Yang, et al., Suppression of B-cell development genes is key to glucocorticoid efficacy in treatment of acute lymphoblastic leukemia, Blood 129 (2017) 3000–3008, doi:10.1182/blood-2017-02-766204.

[29] A. Sendoel, J.G. Dunn, E.H. Rodriguez, S. Naik, N.C. Gomez, B. Hurwitz, et al., Translation from unconventional 5′ start sites drives tumour initiation, Nature (2017), doi:10.1038/nature21036.

[30] H. Wang, Y. Wang, Z. Xie, Computational resources for ribosome profiling: from database to Web server and software, Brief. Bioinf. (2017) 1–12, doi:10.1093/bib/bbx093.

[31] A.M. Michel, S.J. Kiniry, P.B.F. O'Connor, J.P. Mullan, P.V. Baranov, GWIPS-viz: 2018 update, Nucleic Acids Res. (2017) 1–8, doi:10.1093/nar/gkx790.

[32] J.T. Leek, R.B. Scharpf, H.C. Bravo, D. Simcha, B. Langmead, W.E. Johnson, et al., Tackling the widespread and critical impact of batch effects in high-throughput data, Nat. Rev. Genet. 11 (2010) 733–739, doi:10.1038/nrg2825.

[33] M.A. Taub, H. Corrada Bravo, R.A. Irizarry, Overcoming bias and systematic errors in next generation sequencing data, Genome Med. 2 (2010) 87, doi:10.1186/gm208.

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

[35] A. Bartholomäus, C. Del Campo, Z. Ignatova, Mapping the non-standardized biases of ribosome profiling, Biol. Chem. 397 (2015) 23–35, doi:10.1515/hsz-2015-0197.

[36] M.V Gerashchenko, V.N. Gladyshev, Translation inhibitors cause abnormalities in ribosome profiling experiments, Nucleic Acids Res. 42 (2014) 1–7, doi:10.1093/nar/gku671.

[37] N.T. Ingolia, G.A. Brar, N. Stern-Ginossar, M.S. Harris, G.J.S. Talhouarne, S.E. Jackson, et al., Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes, Cell Rep. (2014) 1–15, doi:10.1016/j.celrep.2014.07.045.

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

[39] O. Carja, T. Xing, E.W.J. Wallace, J.B. Plotkin, P. Shah, riboviz: analysis and visualization of ribosome profiling datasets, BMC Bioinf. 18 (2017) 461, doi:10.1186/s12859-017-1873-8.

[40] P. Zhang, D. He, Y. Xu, J. Hou, B. Pan, Y. Wang, et al., analysis of translational initiation, Nat. Commun. 8 (2017) 1749, doi:10.1038/s41467-017-01981-8.

[41] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions., Genome Biol. 14 (2013) R36, doi:10.1186/gb-2013-14-4-r36.

[42] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (2015) 357–360, doi:10.1038/nmeth.3317.

[43] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21, doi:10.1093/bioinformatics/bts635.

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

[45] B.L. Aken, P. Achuthan, W. Akanni, M.R. Amode, F. Bernsdorff, J. Bhai, et al., Ensembl, Nucleic Acids Res. 45 (2017) D635–D642 (2017), doi:10.1093/nar/gkw1104.

[46] J. Crappé, E. Ndah, A. Koch, S. Steyaert, D. Gawron, S. De Keulenaer, et al., PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration, Nucleic Acids Res. (2014) 1–10, doi:10.1093/nar/gku1283.

[47] A. Popa, K. Lebrigand, A. Paquet, N. Nottet, K. Robbe-Sermesant, R. Waldmann, et al., RiboProfiling: a Bioconductor package for standard Ribo-seq pipeline processing, F1000Research. 5 (2016) 1309. doi:10.12688/f1000research.8964.1.

[48] B.F.Gardin Justin, Yeasmin Rukhsana, Yurovsky Alisa, CaiYing Steve Skiena, Measurement of average decoding rates of the 61 sense codons in vivo, Elife (2014) 3, doi:10.7554/eLife.03735.001.

[49] X. Gao, J. Wan, B. Liu, M. Ma, B. Shen, S.-B. Qian, Quantitative profiling of initiating ribosomes in vivo, Nat. Methods 12 (2014) 147–153, doi:10.1038/nmeth.3208.

[50] C.A. Rubio, B. Weisburd, M. Holderfield, C. Arias, E. Fang, J.L. Derisi, et al., Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation, Genome Biol. 15 (2014) 476, doi:10.1186/s13059-014-0476-1.