

ExtRamp Online: Enabling ramp sequence calculations via an intuitive web interface.

Justin B. Miller

University of Kentucky

Matthew W. Hodgman

University of Kentucky

Kyle N. Miller

Utah Valley University

Taylor E. Meurs

Brigham Young University

Mark T. W. Ebbert

University of Kentucky

John S. K. Kauwe

Brigham Young University

Perry G. Ridge (✉ perry_ridge@byu.edu)

Brigham Young University

Data Note

Keywords: ramp sequence, web interface, translation, codon usage bias, codon

Posted Date: September 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-908954/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Motivation: Ramp sequences are an understudied evolutionarily-conserved mechanism for regulating protein translational efficiency. Slowly-translated codons concentrated at the 5' end of genes form ramp sequences that counterintuitively increase overall translational efficiency by evenly spacing ribosomes at initiation, which limits downstream ribosomal collisions. We previously developed ExtRamp, which is the only algorithm to identify translational ramp sequences in single genes. ExtRamp currently lacks a web interface to facilitate wider adoption and application for non-programmers. Additionally, ExtRamp currently identifies ramp sequences using only species-wide codon efficiencies that may lack the specificity of tissue and cell type-specific codon usage biases.

Results: We present an online interface for ExtRamp to facilitate wider adoption and application for non-programmers, along with a significant improvement to the underlying algorithm to calculate tissue and cell type-specific ramp sequences (<https://ramps.byu.edu/ExtRampOnline>). ExtRamp Online contains all options available in the original ExtRamp algorithm with additional pre-set default values to enable researchers to calculate human tissue-specific or genome-wide ramp sequences on any web browser. Human tissue and cell type-specific codon usage biases have been precomputed and can be applied with a simple drop-down menu. Hover-over hints provide users with detailed information on all available options, which will help facilitate future creative analyses using ramp sequences.

Availability: ExtRamp Online is publicly available at <https://ramps.byu.edu/ExtRampOnline>. All associated scripts are publicly available at <https://github.com/ridgelab/ExtRampOnline>.

1 Motivation

Ramp sequences are an important evolutionary-conserved mechanism that plays an essential role in regulating translational efficiency. Synonymous codon usage biases directly affect protein translation efficiency (i.e., expression) because codons with more common cognate anticodons are generally translated faster than suboptimal codons with rare cognate anticodons (Dana and Tuller, 2014). Surprisingly, suboptimal codons are over-represented at the 5' end of highly expressed genes, forming a ramp sequence that limits downstream ribosomal collisions by evenly spacing the ribosomes at translation initiation (Tuller, et al., 2010). We previously developed ExtRamp (Miller, et al., 2019), which is the only algorithm to identify gene-specific translational ramp sequences *in silico*. Notably, ExtRamp has been used to identify ramp sequence conservation across all domains of life (McKinnon, et al., 2021; Miller, et al., 2019) and population-specific ramp sequences within different human populations (Hodgman, et al., 2020). However, ExtRamp currently is limited to command-line applications, which require users to navigate various options that are tailored toward genome-wide analyses. Here, we present ExtRamp Online, which is an intuitive web interface that allows researchers to calculate translational ramp sequences on any web browser. We fully integrate tissue and cell type-specific codon optimality into ExtRamp Online to facilitate creative ramp sequence analyses that capitalize on underlying differences in tissue-specific codon usage biases (Dittmar, et al., 2006; Kames, et al., 2020) caused by local tRNA pools

and tissue-specific RNA binding proteins (Dittmar, et al., 2006; Payne and Alvarez-Ponce, 2019). Therefore, we anticipate that ExtRamp Online will allow researchers to easily calculate both genome-wide and tissue-specific ramp sequences, compare the effects of tissue-specific translational ramp sequences on gene expression using their own datasets, and utilize ExtRamp more fully without requiring any bioinformatics expertise.

2 Implementation

We built ExtRamp Online using an ASP.NET Core 3.1 (<https://dotnet.microsoft.com/apps/aspnet>) framework on a Microsoft Windows server hosted at Brigham Young University. ExtRamp Online runs a modified version of ExtRamp within the web browser using Pyodide (<https://pyodide.org/en/stable/>). All parameters are customizable and preset to a default test case that allows users to see the expected output. Additionally, scroll-over hints teach users how to change parameters and effectively use those parameters in ramp sequence calculations. All associated scripts and information about the website are available at <https://github.com/ridgelab/ExtRampOnline>.

2.1 Tissue and cell type-specific codon efficiencies

The ExtRamp algorithm requires the relative adaptiveness of each codon, which is generally calculated from a set of highly-expressed genes (Miller, et al., 2019). ExtRamp determines the relative synonymous codon usage (i.e., codon adaptiveness) by counting the occurrences of each codon that encodes for the same amino acid. The relative synonymous codon usage is a metric from 0-1 that divides the number of codon occurrences for a specific codon by the number of codon occurrences for the most common codon that encodes the same amino acid and is usually calculated from a FASTA file of highly expressed genes because codon usage biases from highly expressed genes correspond with tRNA levels (Post, et al., 1979).

ExtRamp Online provides options to integrate pre-computed genome-wide relative synonymous codon usages, as well as tissue and cell type-specific relative synonymous codon usages. We used the Human Protein Atlas (Ponten, et al., 2011; Thul and Lindskog, 2018; Uhlén, et al., 2015), GTEx Project (Consortium, 2015), and FANTOM5 datasets (Lizio, et al., 2019; Lizio, et al., 2015), as well as a consensus file combining information from each database to calculate the relative synonymous codon usage within each tissue. All isoforms for genes with multiple annotated isoforms were required to have expression values within the same quartile for a given tissue. In total, we included 18,388 genes that spanned 45 tissues from the FANTOM5 dataset, 34 tissues from GTEx, 43 tissues from the Human Protein Atlas, and 62 tissues from the consensus dataset. We also downloaded 66 tissue-stratified cell type-specific expression datasets with annotated expression levels for each gene from the Human Protein Atlas database in April 2021.

For each tissue in each dataset, the relative synonymous codon usage was calculated from a FASTA file containing the longest GRCh38 isoform of highly expressed genes (i.e., genes with expression in the fourth quartile). Each tissue file contained at least 4,000 highly expressed genes with an average of 9,307 genes per file. Highly expressed genes were previously labeled by the Human Protein Atlas for cell type-specific data, and those labels were used here. We removed genes with uncertain reliability scores, and we removed cell types containing fewer than 900 highly expressed genes. All relative synonymous codon usages for both tissue-specific and cell type-specific analyses are fully integrated in ExtRamp Online with an interactive drop-down menu.

2.2 ExtRamp Online Usage

Users can access all features and options from the original ExtRamp software using ExtRamp Online through an intuitive interface that does not require programming experience. Users can either paste a FASTA formatted gene header and sequence into a text box or upload a small FASTA file of coding sequences. Because ramp sequences are computed using tRNA abundances and codon usage biases, users can upload species tRNA adaptation index values or use provided tissue or cell type-specific relative synonymous codon usage values. By default, relative synonymous codon usage values are calculated from the human reference genome GRCh38, but users can upload any reference genome. All parameters are set to the default values recommended by ExtRamp (see <https://github.com/ridgelab/ExtRamp>). ExtRamp Online generates a variety of output files, including a FASTA file containing the identified ramp sequences, a list or CSV of codon efficiency values, the headers of genes that did not contain a ramp sequence, the headers of genes that could not be processed by ExtRamp, and the segments of genes after identified ramp sequences. Examples of these files are available at ExtRamp Online, with checkboxes allowing users to select which files are downloaded. Because ExtRamp Online runs within the browser, the computer specifications of the user, Internet connection, and web browser may affect performance. For more intensive calculations (e.g., parallel comparisons or genome-wide analyses), we provide an automatic command generator that provides users with the command to run ExtRamp from the command line on their own machines in either a Linux, Mac, or Windows environment.

3 Discussion

We anticipate that ExtRamp Online will help researchers understand how altered tRNA levels affect ramp sequences and contribute to disease development. Many diseases are associated with altered tRNA levels (Gingold, et al., 2014; Goodarzi, et al., 2016; Kirchner, et al., 2017; Kirchner and Ignatova, 2015; Mahlab, et al., 2012; Pavon-Eternod, et al., 2009; Pavon-Eternod, et al., 2013; Polte, et al., 2019; Rudolf, 2019; Torres, et al., 2014). For example, cancer cells increase expression of genes that promote cell proliferation and cancer progression by altering tRNA levels to match genic codon usage biases (Gingold, et al., 2014; Goodarzi, et al., 2016; Pavon-Eternod, et al., 2009; Pavon-Eternod, et al., 2013). These adjusted tRNA levels may increase gene expression by creating ramp sequences. Using tRNA usage values calculated from RNA-seq or single cell RNA-seq data of infected and healthy individuals, ExtRamp Online can identify the extent to which altered tRNA levels affect ramp sequences.

Ramp sequences created or destroyed by altered tRNA levels could significantly change protein expression and become pathogenic. These types of tissue-specific analyses are now possible to anyone, regardless of programming experience, with ExtRamp Online.

We expect that researchers will use ExtRamp Online to investigate the role ramp sequences play in variant pathogenicity and subsequent disease development in specific tissues or cell types. Pathogenic variants identified in genome-wide association studies can easily be run through ExtRamp Online to identify whether they create, destroy, or do not affect ramp sequences. If a genetic variant alters a ramp sequence, it could significantly alter protein expression levels of that gene, which may indicate a functional mechanism for its disease association. Users can also analyze how pathogenic variants affect ramp sequences in different tissues or cell types, which may indicate cell type-specific pathogenicity that can be targeted through personalized therapeutics. Additionally, analyzing the effects of variants of unknown clinical significance on ramp sequences may provide additional insights into their potential pathogenicity by identifying a mechanism that may alter gene expression. Therefore, ExtRamp Online not only enables researchers to understand why some variants are pathogenic, but also to prioritize variants of unknown clinical relevance for further investigation.

ExtRamp Online significantly improves the accessibility of ramp sequence calculations, which will enable future innovative adaptations of ramp sequences in diverse analyses. By employing pre-computed tissue and cell type-specific relative synonymous codon usage values, ExtRamp Online can identify ramp sequences more precisely than using whole genome relative synonymous codon usage values. Additionally, ExtRamp Online computes the codon landscape by calculating translational efficiencies along the gene, which may support more complete protein folding predictions, since varied translational rates assist in protein folding (Zhang, et al., 2009) and mutations that alter the codon landscape can significantly alter protein functionality (Kimchi-Sarfaty, et al., 2007). We developed ExtRamp Online as a hypothesis-generating tool that can be used for a variety of studies ranging from species-wide ramp sequence calculations to single-cell disease pathogenicity analyses.

Declarations

Acknowledgements

We appreciate the Office of Research Computing and Life Sciences IT at Brigham Young University for providing the resources to support the online interface and dedicated web server to run ExtRamp Online.

Funding

This work was supported by Brigham Young University, the University of Kentucky Sanders-Brown Center on Aging, and BrightFocus grant #A2020118F (PI: Miller).

Conflict of Interest: none declared.

References

- Consortium, G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348(6235):648-660.
- Dana, A. and Tuller, T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Research* 2014;42(14):9171-9181.
- Dittmar, K.A., Goodenbour, J.M. and Pan, T. Tissue-specific differences in human transfer RNA expression. *PLoS Genet* 2006;2(12):e221.
- Gingold, H., *et al.* A dual program for translation regulation in cellular proliferation and differentiation. *Cell* 2014;158(6):1281-1292.
- Goodarzi, H., *et al.* Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell* 2016;165(6):1416-1427.
- Hodgman, M.W., *et al.* CUBAP: an interactive web portal for analyzing codon usage biases across populations. *Nucleic Acids Research* 2020;48(19):11030-11039.
- Kames, J., *et al.* TissueCoCoPUTs: Novel Human Tissue-Specific Codon and Codon-Pair Usage Tables Based on Differential Tissue Gene Expression. *J Mol Biol* 2020;432(11):3369-3378.
- Kimchi-Sarfaty, C., *et al.* A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 2007;315(5811):525-528.
- Kirchner, S., *et al.* Alteration of protein function by a silent polymorphism linked to tRNA abundance. *PLoS Biol* 2017;15(5):e2000779.
- Kirchner, S. and Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat Rev Genet* 2015;16(2):98-112.
- Lizio, M., *et al.* Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res* 2019;47(D1):D752-D758.
- Lizio, M., *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* 2015;16:22.
- Mahlab, S., Tuller, T. and Linial, M. Conservation of the relative tRNA composition in healthy and cancerous tissues. *Rna* 2012;18(4):640-652.
- McKinnon, L.M., *et al.* A comprehensive analysis of the phylogenetic signal in ramp sequences in 211 vertebrates. *Scientific Reports* 2021;11(1):622.

- Miller, J.B., Brase, L.R. and Ridge, P.G. ExtRamp: a novel algorithm for extracting the ramp sequence based on the tRNA adaptation index or relative codon adaptiveness. *Nucleic Acids Research* 2019;47(3):1123-1131.
- Miller, J.B., Brase, L.R. and Ridge, P.G. ExtRamp: a novel algorithm for extracting the ramp sequence based on the tRNA adaptation index or relative codon adaptiveness. *Nucleic Acids Res* 2019.
- Pavon-Eternod, M., *et al.* tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res* 2009;37(21):7268-7280.
- Pavon-Eternod, M., *et al.* Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. *RNA* 2013;19(4):461-466.
- Payne, B.L. and Alvarez-Ponce, D. Codon Usage Differences among Genes Expressed in Different Tissues of *Drosophila melanogaster*. *Genome Biol Evol* 2019;11(4):1054-1065.
- Polte, C., *et al.* Assessing cell-specific effects of genetic variations using tRNA microarrays. *BMC Genomics* 2019;20(Suppl 8):549.
- Ponten, F., *et al.* The Human Protein Atlas as a proteomic resource for biomarker discovery. *J. Intern. Med.* 2011;270(5):428-446.
- Post, L.E., *et al.* Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 1979;76(4):1697-1701.
- Rudorf, S. Efficiency of protein synthesis inhibition depends on tRNA and codon compositions. *PLoS Comput Biol* 2019;15(8):e1006979.
- Thul, P.J. and Lindskog, C. The human protein atlas: A spatial map of the human proteome. *Protein Sci.* 2018;27(1):233-244.
- Torres, A.G., Batlle, E. and Ribas de Pouplana, L. Role of tRNA modifications in human diseases. *Trends Mol Med* 2014;20(6):306-314.
- Tuller, T., *et al.* An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* 2010;141(2):344-354.
- Uhlén, M., *et al.* Tissue-based map of the human proteome. *Science* 2015;347(6220):1260419.
- Zhang, G., Hubalewska, M. and Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* 2009;16(3):274-280.