

oRNAmEnt: a database of putative RNA binding protein target sites in the transcriptomes of model species

Louis Philip Benoit Bouvrette^{1,2}, Samantha Bovaird^{1,3}, Mathieu Blanchette^{1,4} and Eric Lécuyer^{1,2,3,*}

¹Institut de Recherches Cliniques de Montréal (IRCM) Montréal, Québec, Canada, ²Département de Biochimie et Médecine Moléculaire, Université de Montréal, Montréal, Québec, Canada, ³Division of Experimental Medicine, McGill University, Montréal, Québec, Canada and ⁴School of Computer Science, McGill University, Montréal, Québec, Canada

Received August 15, 2019; Revised October 11, 2019; Editorial Decision October 14, 2019; Accepted October 28, 2019

ABSTRACT

Protein–RNA interactions are essential for controlling most aspects of RNA metabolism, including synthesis, processing, trafficking, stability and degradation. *In vitro* selection methods, such as RNAcompete and RNA Bind-n-Seq, have defined the consensus target motifs of hundreds of RNA-binding proteins (RBPs). However, readily available information about the distribution features of these motifs across full transcriptomes was hitherto lacking. Here, we introduce oRNAmEnt (o RNA motifs enrichment in transcriptomes), a database that catalogues the putative motif instances of 223 RBPs, encompassing 453 motifs, in a transcriptome-wide fashion. The database covers 525 718 complete coding and non-coding RNA species across the transcriptomes of human and four prominent model organisms: *Caenorhabditis elegans*, *Danio rerio*, *Drosophila melanogaster* and *Mus musculus*. The unique features of oRNAmEnt include: (i) hosting of the most comprehensive mapping of RBP motif instances to date, with 421 133 612 putative binding sites described across five species; (ii) options for the user to filter the data according to a specific threshold; (iii) a user-friendly interface and efficient back-end allowing the rapid querying of the data through multiple angles (i.e. transcript, RBP, or sequence attributes) and (iv) generation of several interactive data visualization charts describing the results of user queries. oRNAmEnt is freely available at <http://rnabiology.ircm.qc.ca/oRNAmEnt/>.

INTRODUCTION

Throughout their life-cycle, RNA molecules undergo a variety of co- and post-transcriptional regulatory events that control their maturation, function and fate (1–3). By modulating the assembly and function of ribonucleoprotein machineries, protein–RNA interactions play critical roles in virtually all facets of RNA metabolism. Indeed, RNA-binding proteins (RBPs) form an essential class of regulatory factors, which encompass among the most deeply evolutionarily conserved protein families (1,4). These proteins are primarily classified by the type of RNA-binding domain (RBD) they contain, which confers to them the capacity to interact with RNA molecules through binding sites defined by their sequence and/or structural properties (1,4). Recent studies, combining RNA-capture and mass spectrometry profiling, have characterized ~1500 RBPs in human cells, hinting at the staggering complexity of post-transcriptional regulation (5–8).

To characterize the binding specificities of candidate RBPs, binding site selection approaches, in particular RNAcompete and RNA Bind-n-Seq (RBNS) methodologies, have been systematically applied to a growing proportion of eukaryotic RBPs (9–12). Both of these methods involve *in vitro* binding assays combining a recombinantly purified RBP (or its RBD) and a randomized pool of RNA, followed by the biochemical purification of bound RNA molecules and their identification via microarray or RNA-sequencing (9–12). These approaches have enabled the identification of primary sequence consensus binding site motifs for a few hundred RBPs.

Several tools exist to scan user-provided RNA sequences for matches to these *in vitro* motifs, including servers such as CISBP-RNA, RBPmap, ATTRACT and MotifMap-RNA (12–15). However, to date, no resources have been developed for identifying and cataloguing putative RBP motif instances across full transcriptomes. Herein, we describe the

*To whom correspondence should be addressed. Tel: +1 514 987 5646; Fax: +1 514 987 5752; Email: Eric.Lecuyer@irem.qc.ca

oRNAment (**o** RNA motifs enrichment in transcriptomes) database, which catalogues the motif instances of 223 RBPs previously defined via the RNAcompete and RBNS platforms, across the coding and non-coding transcriptomes (excluding introns) of humans and four major model organisms. oRNAment is accessible at <http://rnabiology.ircm.qc.ca/oRNAment/>.

oRNAment ANALYSIS PIPELINE

Pre-processing of the oRNAment input data

oRNAment was created to characterize the distribution properties of potential RBP target sites across model organism transcriptomes from the most up-to-date RBP motif data available (Figure 1).

We acquired the data for 223 unique RBPs, totalling 453 consensus motifs in the form of position weight matrices (PWMs) obtained by either RNAcompete or RBNS (9–12,16,17) (Figure 1i). More precisely, we obtained 218 RNAcompete PWMs (172 RBPs) from the CISBP-RNA resource (12). By design, most motifs determined by RNAcompete were of length 7 nucleotides. In parallel, we derived an additional 235 PWMs (78 RBPs) by executing the RBNS computational analysis pipeline for 7-mer enrichment on RBNS data available from the ENCODE resource (9,10,16,17). Therefore, all motifs in the database are 7 nucleotides in length and are, as such, comparable. Overall, only 27 RBPs were profiled by both methods (Figure 1i, light grey lines). RBPs and their motifs were flagged for their species-specificities as defined by Ray and al. (12) (Figure 1ii). Scans were performed for each PWM individually, regardless of similarities or discrepancies between RNAcompete and RBNS PWMs of the same RBP. Furthermore, motif scans were executed for motifs assigned to each RBP across all five species, regardless of the species representation of a given RBP. However, since RBP orthologs are expected to exhibit similar binding motif specificities if their RBD show >70% identity in amino acid sequence (12), we have flagged the species specificity of each factor so the user can take this information into consideration.

oRNAment is based upon a custom pipeline to perform efficient transcriptome-wide scans for instances of all 453 RBP motifs collected above (Figure 2). We based our pipeline on the widely used MATCH algorithm developed to scan for putative transcription factor binding sites across DNA sequences (18,19). This tool takes as input a motif, in the form of a PWM, and returns the position of the subsequences above a given score (18,19) (Figure 2i-iii). This is conceptually similar to scanning for RBP target motif instances, also taking a PWM as input, across RNA transcripts. Through the use of high-performance python 3.7 libraries (i.e. NumPy, Pandas) and data structures (pre-constructed hash tables of all heptanucleotides and score pairs), we developed a scanning algorithm that allows great efficiency, in terms of memory and speed, permitting timely execution across full transcriptomes. Nevertheless, due to the computational limitations imposed by the large intronic sequence space, the analyses herein only consider exonic regions of target transcriptomes.

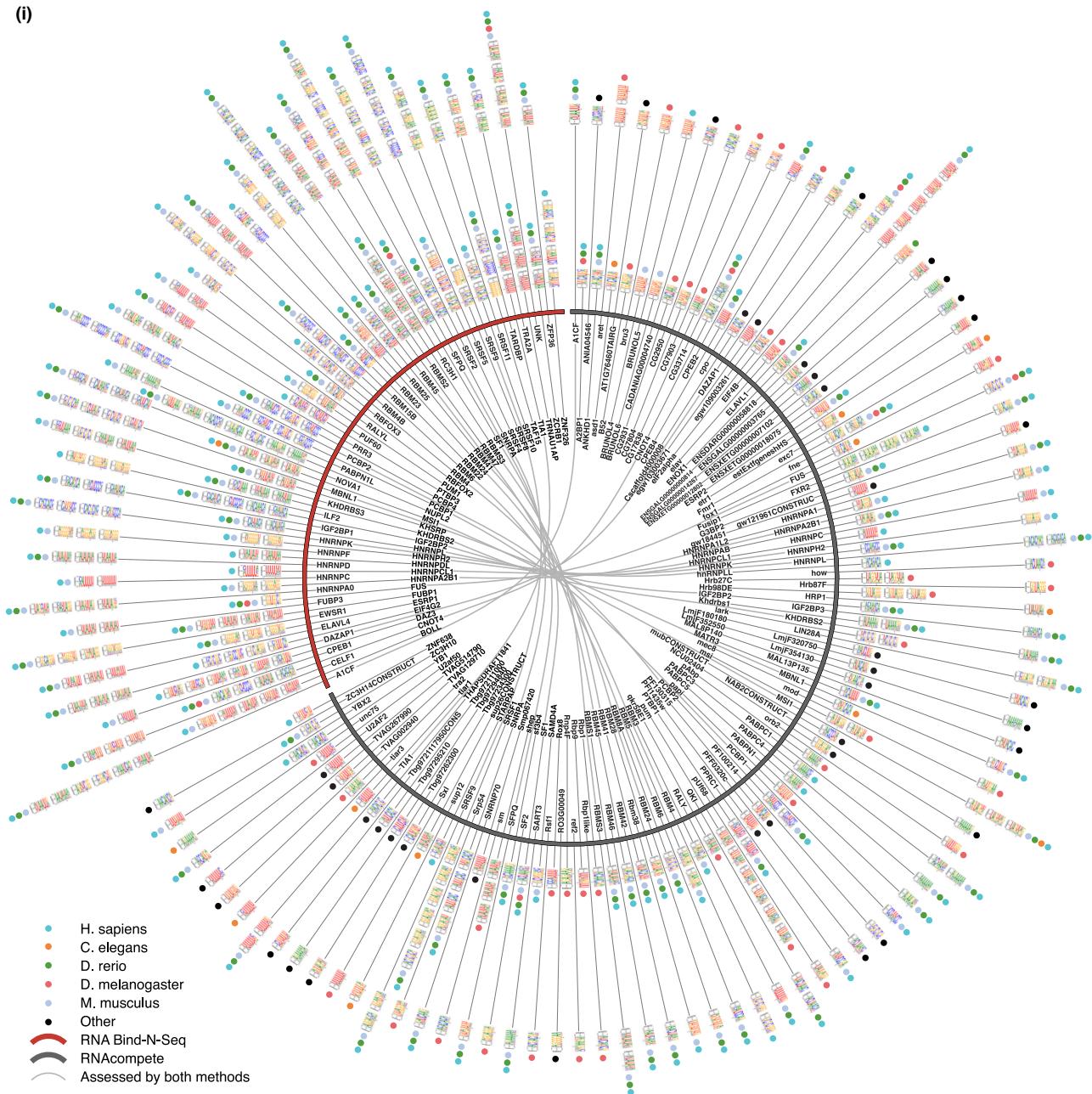
The search algorithm is based on the matrix similarity score (MSS), which measures the correspondence of

a transcript region to a given RBP motif of the same length. This is defined as $MSS = (current_score - minimum_score) / (maximum_score - minimum_score)$, where *current.score* is the product of each nucleotides probability at its respective position in the PWM, and the *maximum.score* and *minimum.score* are the product of each maximum or minimum probability value, respectively, in the PWM at each position. This provides a value between 0 and 1, where 1 is a perfect match to the top canonical binding motif of a given RBP (Figure 2).

In order to identify putative RBP motif instances, it is necessary to select an appropriate threshold for the MSS, which can vary depending on the user's objectives. This threshold allows the user to include motifs with varying degrees of similarity to the most probable in vitro defined consensus motif. For a given percentile P (e.g. P = 50%) and a given PWM, the threshold TP is chosen so that the probability that a 7-mer randomly generated based on the probabilities specified by the PWM obtains an MSS greater or equal to TP is P. In other words, the fraction of sensitivity (or recall) of the search is P. In practice, TP is obtained by calculating the MSS score of each of the 16,384 possible heptanucleotides, sorting them in decreasing order of MSS, and going down the sorted list until the sum of MSS of the selected heptanucleotides reaches P% of the total (Figure 2ii, dash line). oRNAment contains motif matches for 10 different thresholds, for P ranging from 50% to 95% in increments of 5%, as well as for the special threshold MSS = 1 (canonical motif) (Figure 2).

We observed that the analysis pipeline reasonably predicts RBP binding sites observed by eCLIP in human cells (16,20). We used as a validation set the group of 24 RBPs where eCLIP data was also available and compared the genomic coordinates of oRNAment motif instances, at a 50% threshold, to eCLIP peaks, at a ≥ 3 -fold change and a P -value ≤ 0.001 . For this, we first downloaded the bed narrowPeak files from ENCODE for both HepG2 and K562 cell lines and filtered them in order to only keep peaks in an annotated exon. This allowed a one-to-one comparison with the dataset scanned by oRNAment. We then collapsed peak regions from replicates when they showed any overlap. As the peak region rarely had the exact same coordinates, we kept as one region the coordinates englobing the shortest region between the two replicates (i.e. if replicate 1 had a peak between nucleotides 100–109 and replicate 2 a peak between 102–110, we kept as a peak a region between 102–109). We only kept peak regions of at least seven nucleotides. As eCLIP results tend to be cell dependent and we aimed to have a global dataset, we, on one hand, pooled all the data, replicated or not, from both cell lines and, on another hand, pooled only the data that was replicated within a cell line. We considered an oRNAment motif instance as matching an eCLIP peak when there was any type of overlap between the two coordinates. This revealed a good correspondence, as defined by the ratio of motif instances identified by oRNAment that are in an eCLIP peak. Furthermore, motif instances defined by oRNAment are generally better enriched in eCLIP peaks compared to an equal number of random coordinates taken from the same transcriptomic space that was scanned by oRNAment. As an example, the five motifs recognized by HNRNPK are more

(i)



(ii)

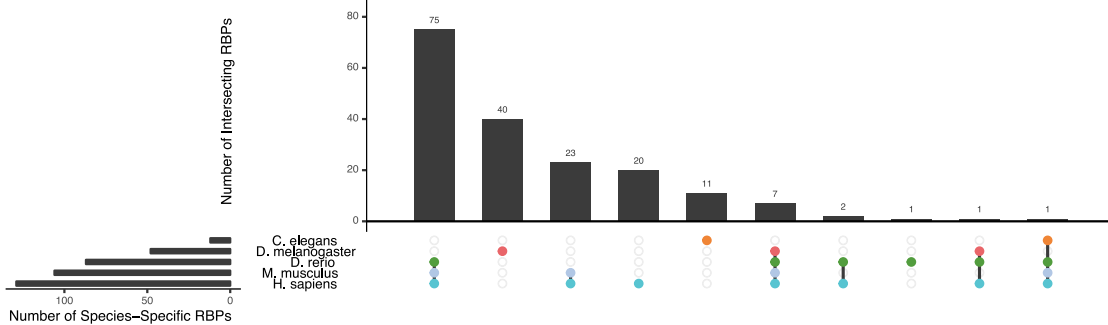
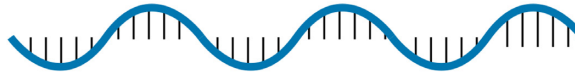
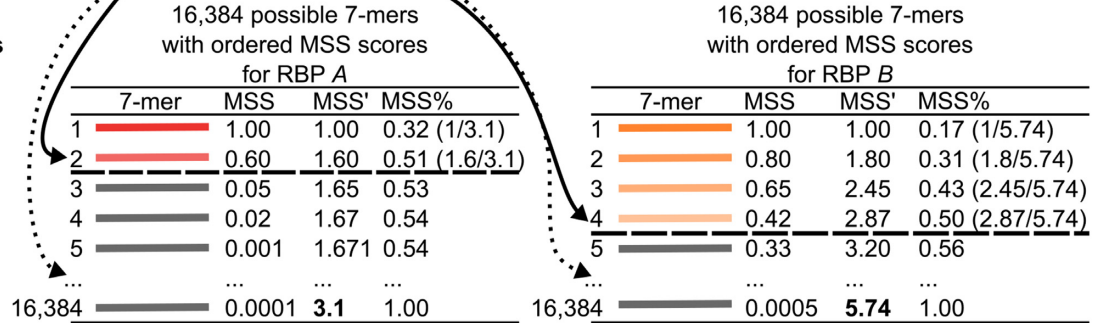


Figure 1. The oRNAment database contains 453 motifs attributable to 223 RBPs in 5 species. (i) Motifs obtained for each RBP come from RNAcompete (red segment) and RBNS (dark grey segment) experiments. Links shown between RBPs (light grey lines) denote those that were assessed by both methods. Coloured dots show the species-specificity of each motif according to Ray *et al.* (12). There are 181 RBPs with binding specificities in the species included in the database and 42 from external species. (ii) Upset plot showing the distribution of interrogated species-specific RBPs across all five species.

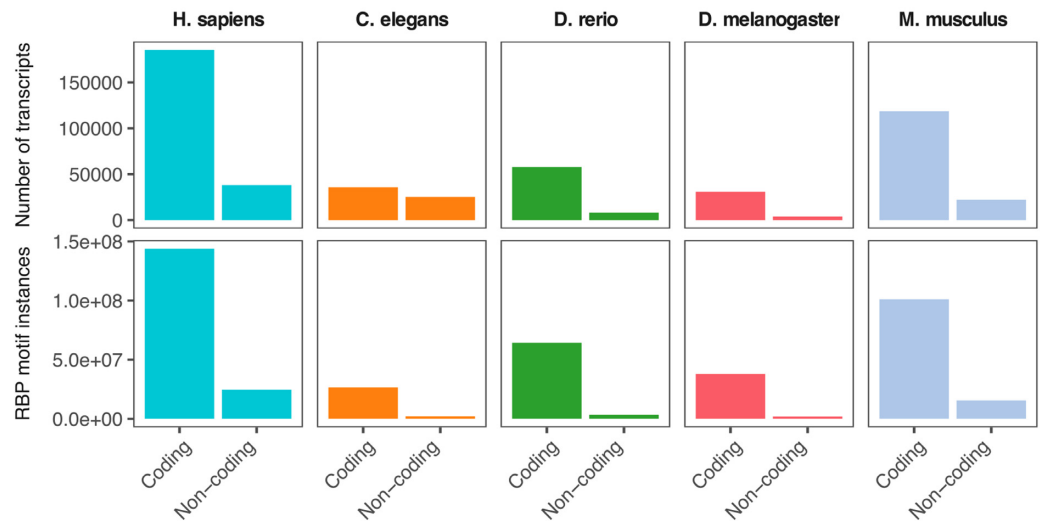
(i) Scanned 7-mers in transcript



(ii) Look up 7-mers in all dictionaries of k-mers



(iii) Keep all positions of 7-mers passing 50% MSS threshold



(iv) Validate predictions with eClip

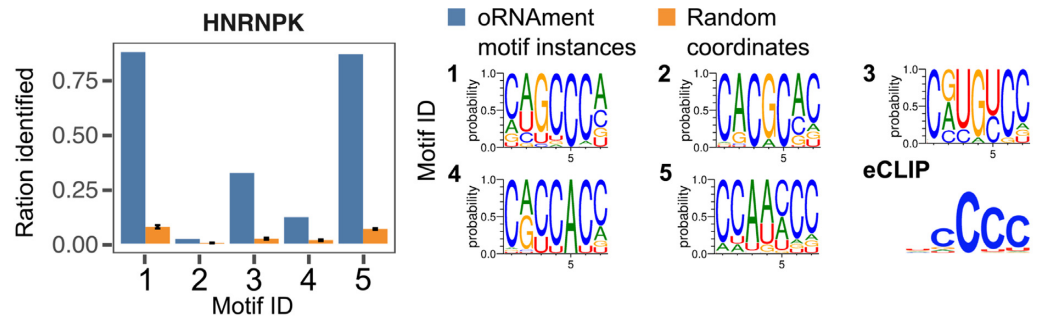


Figure 2. The oRNAment computational pipeline. (i) For a given transcript, the algorithm linearly scans for subsequences of length 7 and (ii) reports only those that have an MSS higher than the threshold, represented by the dashed line (table look-up, exemplified by the arrows, only shown for the second and fourth sequentially scanned 7-mers; sum of MSS' used as denominator for MSS% computation in bold). (iii) oRNAment reports all motif instances in all transcripts across five species. (iv) oRNAment reasonably predicts RBP binding sites observed by eCLIP in human K562 and HepG2 cells (blue bars in histogram), as shown for the five motifs bound by the HNRNPK RBP, in comparison to the same number of random sequences (orange bars).

highly enriched in HNRNPK eCLIP peaks compared to random coordinates (Figure 2iv), while additional examples are shown in Supplemental Figure S1A and B and Supplemental Table S1 (20). Furthermore, oRNAmnt displays reasonable false negative rates and precision (Supplemental Figure S1C, D, E, F and Table S1).

The pipeline was executed on all coding (cDNA) and non-coding (ncRNA) transcripts obtained from the FASTA sequences of Ensembl gene release 97, for *Homo sapiens* (GRCh38), *Caenorhabditis elegans* (WBcel235), *Danio rerio* (GRCz11), *Drosophila melanogaster* (BDGP6) and *Mus musculus* (GRCm38) (21).

Database implementation

oRNAmnt is built upon the column-oriented DBMS *yan-dex ClickHouse* version 19.5.3.1. The server-side back end of the web application makes use of *Django* version 2.1.9 and is written in *Python* 3.7.0. The client interface is implemented in *Django's* HTML template language with the inclusion, for a greater interactive experience, of several *JavaScript* libraries, including *jquery* version 3.3.1, *datatables* version 1.10.19, *charts.js* version 2.0, *ViennaRNA/fornac.js* version 1.1.8, and *IGV.js* version 2.2.13. The layout styling was created with *Bootstrap* 4 and *Bootstrap-material-design* version 4.1.1.

PRIMARY FEATURES OF oRNAmnt

Overall functionality

oRNAmnt contains the position of all motif matches for all PWMs defined by *RNAcompete* or *RBNS*, across the transcriptomes (excluding introns) of all five interrogated species. The user can narrow their search to only motifs for which RBPs are represented in a specific species or group of species.

For each type of search, the database outputs distinct figures summarizing the abundance and distribution of motifs across queried transcripts, subregion types (e.g. coding sequence, UTRs), or RNA biotypes (Figure 3i–ix). It also outputs individual graphs showing the position of all motif instances and their MSS within each transcript for the selected species (Figure 3x). Moreover, a detailed table lists all motif instances along with their associated gene name, transcript ID, biotype, position along the transcript, MSS, genomic coordinates, and probability for the 7-mer region to be structurally unpaired, as assessed by *RNAplfold* predictions (22). Further detailed information, including the predicted RNA secondary structure (Figure 3xi), as assessed by *RNAfold*, can be accessed for a specific transcript from the table (22). For a multifaceted overview of multiple motifs, oRNAmnt also features an embedded *Integrated Genome Browser (IGV)* (Figure 4). All the above information can readily be downloaded as an Excel, CSV, or bed file. This can be achieved either by downloading a subset of the database from the detailed table stemming from a query or by downloading the entire database content.

The database contains a detailed tutorial page to help the user navigate the resource. This section documents the algorithm implemented and the RBP motif data used in oRNAmnt. Furthermore, it provides comprehensive instruc-

tions on how to use each functionality through step-by-step demonstrations using real examples.

Search by transcripts

This functionality allows the user to query the database for a specific gene, transcript, or group of genes or transcripts, in a specified species, and returns all their putative RBP binding sites. The results are visualized with interactive summarizing charts/histograms (Figure 3i–v) and a detailed table. First, a treemap, or histogram, shows the total number of putative instances associated with each RBP. Second, a polar plot, or histogram, illustrates the subregions where these motif instances are observed. Third, a box plot describes the distribution of motif instances within all transcripts searched within oRNAmnt. This is especially useful when searching for multiple transcripts to determine if they have a common RBP binding site.

Search by RBP

This functionality allows a user to query the database for a specific RBP, in a specified species, and returns all its putative binding sites in all coding and non-coding transcripts. The user can restrict or expand their query results by specifying the PWM's sensitivity threshold. The results of this query are visualized with interactive summarizing charts and a detailed table (Figure 3vi–ix). A doughnut plot, or histogram, shows the total number of putative motif instances identified for the queried RBP grouped by gene biotype allowing a user to, for example, predict protein-non-coding RNA interactions. Finally, a radar plot, or histogram, shows the subregions where these putative motif instances are observed.

Search by attributes

This functionality allows the user to query the database for a specific combination of transcript attributes [e.g. 3'untranslated region (UTRs) of mRNA, rRNA] in a given species and returns all associated putative RBP binding sites. When an attribute is incompatible with other selections, it is shown as a blocked option (unclickable and greyed out text displaying "NA"). Contrastingly, when selecting the protein coding biotype, the region NA corresponds to the Ensembl annotation for unavailable information and it is selectable. The results are visualized with a treemap, or histogram, showing the total number of putative instances identified for each RBP and a detailed table.

Interactively visualize motif instances

oRNAmnt offers the possibility for a user to browse the genome of a given species and interactively visualize putative motif instances of up to three RBPs in an embedded *Integrated Genomic Viewer (IGV)* browser (Figure 4) (23,24). Unlike a detailed transcript query, which is designed to describe binding sites for specific and individual RBPs, this functionality allows the users to mine the data in a broader exploratory manner. The user can search for one or multiple loci, querying by genomic positions, and visualizing the RBP binding sites along each annotated exon.

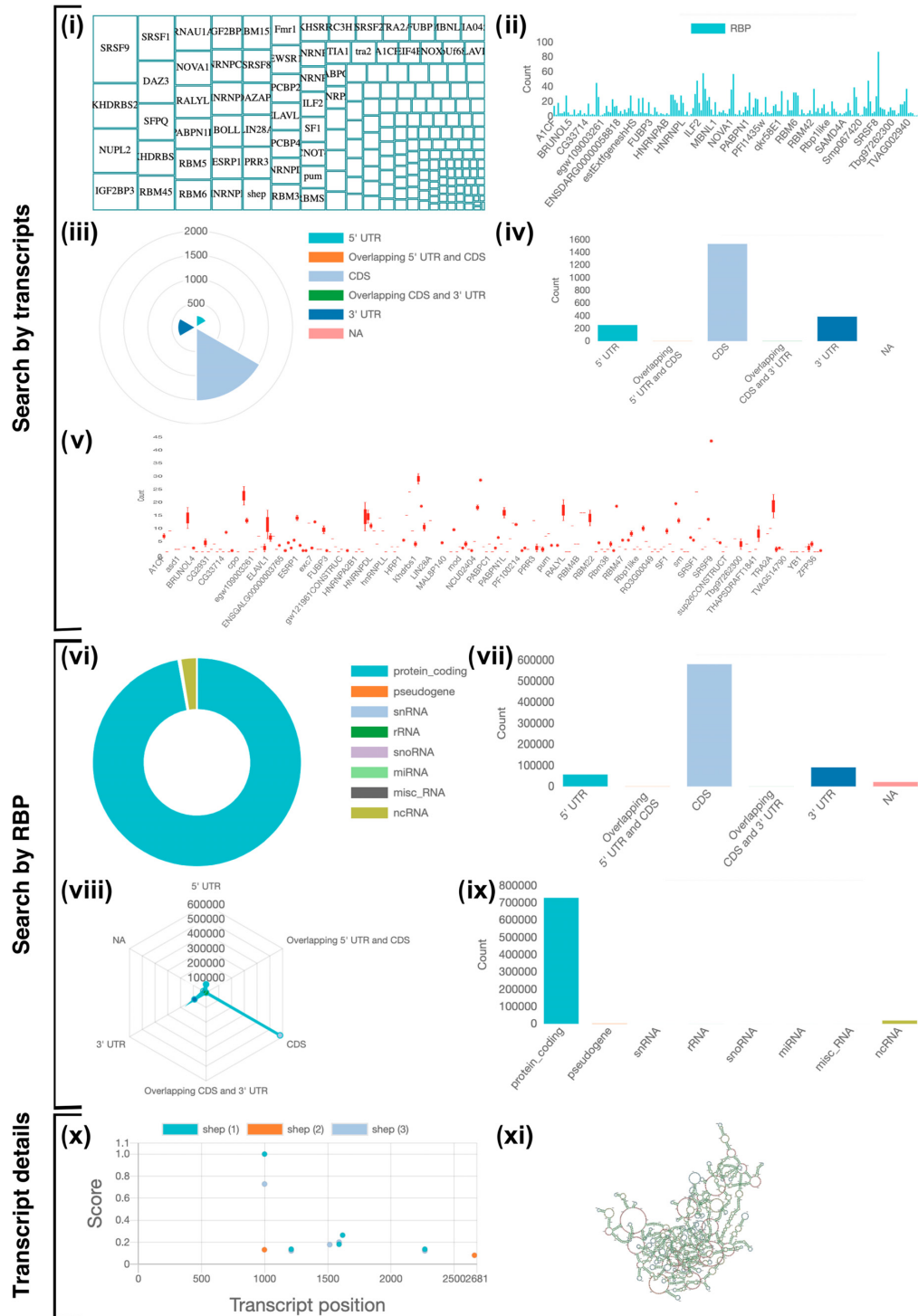
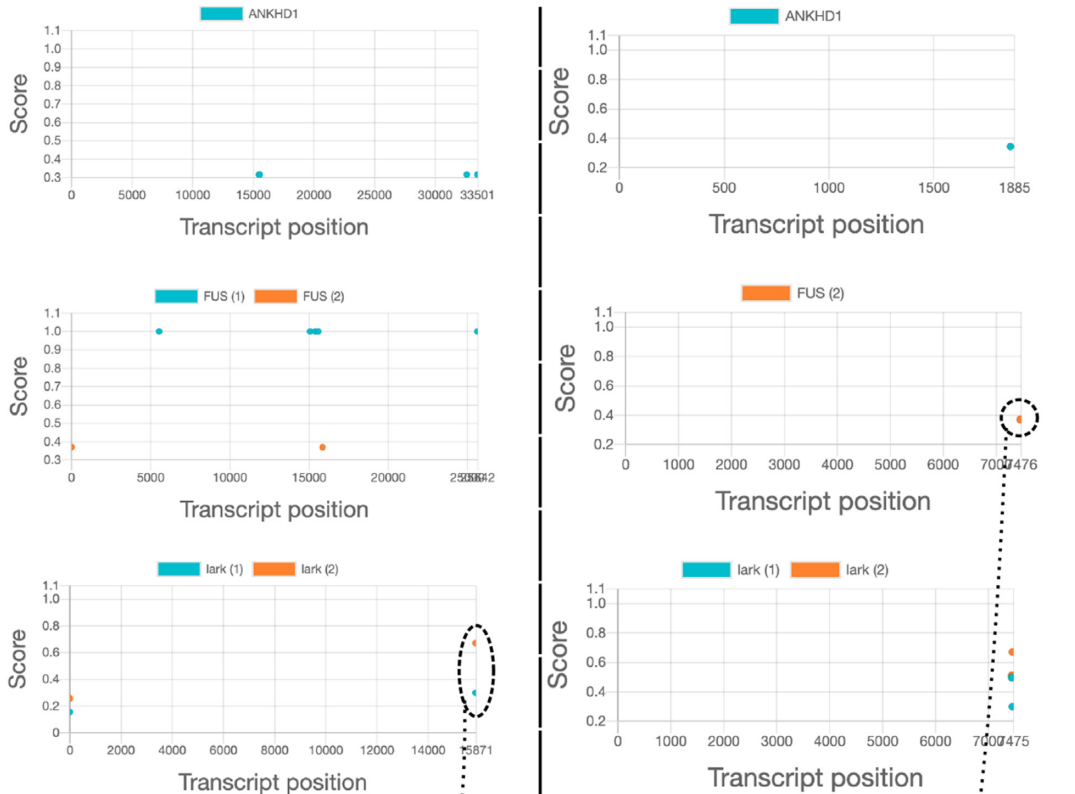


Figure 3. Examples of the figures generated by oRNAment when searching for motifs in specific RNAs or RBPs. Upon a user’s query, either by transcript (i–v) or by RBP (vi–ix), multiple figures summarizing the results are provided. (i–v) When searching by transcripts (here the *cen* mRNA in *Drosophila*), oRNAment provides: (i) a treemap of the most abundant RBP motif instances (likewise shown when searching by attributes); or (ii) a histogram of the same results; (iii) a polar plot showing in which subregion of the transcript RBP motif instances are observed (here in *cen*); or (iv) a histogram of the same results; (v) a boxplot of the distribution of RBP motif instances in all transcripts queried (here, the boxplot shows the distribution of the number of motif instances among the two isoforms of *cen*). (vi–ix) When searching by RBP, oRNAment provides: (vi) a doughnut plot showing in which gene biotypes putative binding sites for the queried RBP are observed (here for SRSF9); or (vii) a histogram of the same results; and (viii) a radar plot showing in what transcript subregion putative binding sites for the queried RBP are observed; or (ix) a histogram of the same results. All search functionalities provide a table from which the user can access gene-level or transcript-level details. (x–xi) By selecting a gene/transcript and RBP pair, oRNAment will provide: (x) a scatter plot showing the position of each putative RBP binding site and corresponding MSS scores, here above the 50% MSS’ threshold respectively for each motif of the *shep* RBP on the *cen* mRNA. The transcript positions, on the x-axis, end at the last motif instance + 10 nucleotides; and (xi) a predicted 2D structure of the *cen* transcript as established by RNAfold with default parameters.

(i) ENST00000262160 (protein_coding) ENST00000398417 (protein_coding)

SMAD2
chr18:47783250-47964001

SMAD4
chr18:51000898-51113761



(ii)

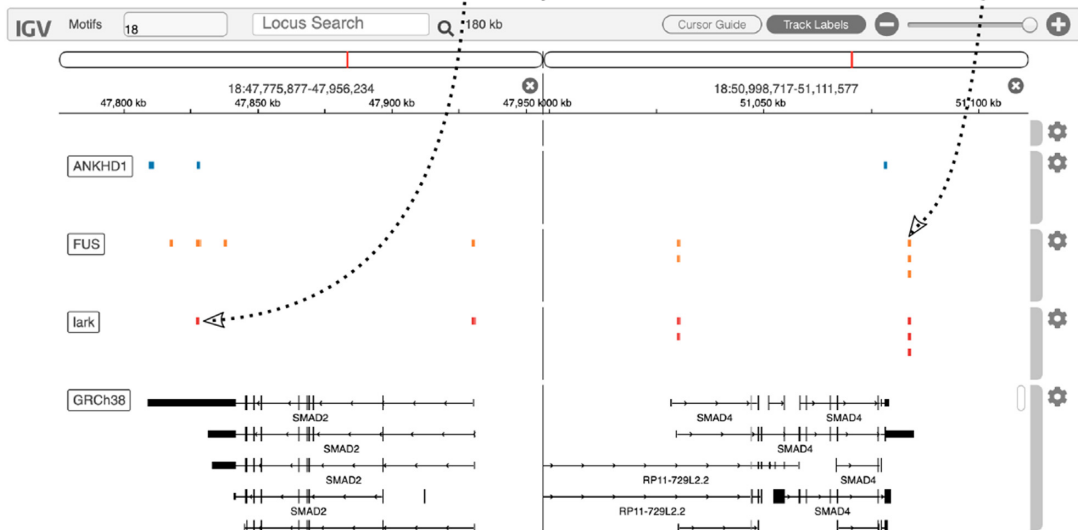


Figure 4. Combined visualization of putative binding sites for three RBPs in two genes through a standard scatter plot and an embedded Integrative Genome Browser. (i) Example of oRNAment transcript-level view scatter plot of three RBPs (ANKHD1, FUS and lark) for two mRNAs (*SMAD2* and *SMAD4*) and (ii) Integrative Genome Browser view incorporating the same results when searching for their loci [IGV *Locus search* input in the form: 18:47783250-47964001 18:51000898-51113761 (i.e. with a space separating the coordinates)]. Two examples of corresponding motif instances (lark in *SMAD2* and FUS in *SMAD4*) between the two types of analysis are shown.

CONCLUSION

oRNAment is a modern platform that offers access to a nucleotide-resolution mapping of putative RBP binding sites across the transcriptomes of human and four important model organisms, namely *C. elegans*, *D. rerio*, *D. melanogaster* and *M. musculus*. The methodology and thresholds employed results in a computationally expensive analysis that produces a large quantity of data. oRNAment palliates this issue by having pre-computed all possible instances through high performance computing resources and by storing the data in a state-of-the-art column-oriented DBMS, which enables efficient retrieval and processing of large quantities of data up to 1000 times faster than traditional data management methods. Altogether, we propose a tool from which the searches and resulting figures are fully interactive and responsive on both desktops and tablets. oRNAment is the first database detailing the transcriptome-wide distribution features of putative RBP target motifs across multiple species. As such, it should prove very useful for users aiming to address hypotheses and to design experiments to study post-transcriptional gene regulation. Future versions will include the complete transcriptome of more species and the addition of other RBPs as their motifs are experimentally defined.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This research was enabled in part by support provided by Calcul Québec (www.calculquebec.ca), Compute Ontario (www.computeontario.ca) and Compute Canada (www.computeCanada.ca).

FUNDING

This work was supported by grants from the Canadian Institutes of Health Research (CIHR) to E.L. and from the Fonds de Recherche Québec – Nature et Technologies (to M.B. and E.L.); as well as scholarships from the Fonds de Recherche Québec – Santé (to E.L.) and from CIHR (to S.B.). Funding for open access charge: Canadian Institutes of Health Research (CIHR).

Conflict of interest statement. None declared.

REFERENCES

- Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A census of human RNA-binding proteins. *Nat. Rev. Genet.*, **15**, 829–845.
- Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.*, **582**, 1977–1986.
- Li, X., Kazan, H., Lipshitz, H.D. and Morris, Q.D. (2014) Finding the target sites of RNA-binding proteins. *Wiley Interdiscip. Rev. RNA*, **5**, 111–130.
- Lunde, B.M., Moore, C. and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.*, **8**, 479–490.
- Baltz, A.G., Munschauer, M., Schwanhauser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M. *et al.* (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell*, **46**, 674–690.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.
- Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.*, **19**, 327–341.
- Benoit Bouvrette, L.P., Blanchette, M. and Lecuyer, E. (2019) Bioinformatics Approaches to Gain Insights into cis-Regulatory Motifs Involved in mRNA Localization. The Biology of mRNA: Structure and Function. *Adv. Exp. Med. Biol.*, **1203**, In press.
- Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L., Pratt, G.A. *et al.* (2018) Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell*, **70**, 854–867.
- Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.A. and Burge, C.B. (2014) RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell*, **54**, 887–900.
- Ray, D., Kazan, H., Chan, E.T., Pena Castillo, L., Chaudhry, S., Talukder, S., Blencowe, B.J., Morris, Q. and Hughes, T.R. (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat. Biotechnol.*, **27**, 667–670.
- Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.
- Giudice, G., Sanchez-Cabo, F., Torroja, C. and Lara-Pezzi, E. (2016) ATTRACT—a database of RNA-binding proteins and associated motifs. *Database*, **2016**, baw035.
- Paz, I., Kosti, I., Ares, M. Jr, Cline, M. and Mandel-Gutfreund, Y. (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.*, **42**, W361–W367.
- Liu, Y., Sun, S., Bredy, T., Wood, M., Spitale, R.C. and Baldi, P. (2017) MotifMap-RNA: a genome-wide map of RBP binding sites. *Bioinformatics*, **33**, 2029–2031.
- VanNostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Blue, S.M., Dominguez, D., Cody, N.A.L., Olson, S., Sundararaman, B. *et al.* (2018) A large-scale binding and functional map of human RNA binding proteins. bioRxiv doi: <http://doi.org/10.1101/179648>, 05 October 2018, preprint: not peer reviewed.
- Lambert, N.J., Robertson, A.D. and Burge, C.B. (2015) RNA Bind-n-Seq: measuring the binding affinity landscape of RNA-binding proteins. *Methods Enzymol.*, **558**, 465–493.
- Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V. and Wingender, E. (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.*, **31**, 3576–3579.
- Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.*, **23**, 4878–4884.
- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K. *et al.* (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*, **13**, 508–514.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C.G. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.
- Lorenz, R., Bernhart, S.H., Honer, Z., Siederdisen, C., Tafer, H., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
- Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.
- Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.