

Prefix/Suffix Variation in Retinoic Acid Response Elements

Yuan Zhuang
Biology Department
Reed College
frzhuang@reed.edu

Kara L. Cervený
Biology Department
Reed College
cervenky@reed.edu

Anna Ritz
Biology Department
Reed College
aritz@reed.edu

CCS CONCEPTS

• Applied computing → Recognition of genes and regulatory elements.

KEYWORDS

Retinoic acid receptor, Retinoic acid response element, sequence motifs

ACM Reference Format:

Yuan Zhuang, Kara L. Cervený, and Anna Ritz. 2020. Prefix/Suffix Variation in Retinoic Acid Response Elements. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '20)*, September 21–24, 2020, Virtual Event, USA. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3388440.3414914>

1 INTRODUCTION

Retinoic acid (RA) is a derivative of Vitamin A that is essential for vertebrate growth and development. RA regulates the expression of target genes when it binds to a heterodimer of a general receptor (RXR) and an RA-specific receptor (RAR), binding to motifs called Retinoic acid response elements (RAREs) [2, 3]. Early observations of this motif identified a perfectly conserved 6-mer direct repeat (DR) of AGGTCA separated by a small number of nucleotides [3]. Later analysis revealed additional variable points, resulting in a RARE motif represented as a repeat of (A/G)G(G/T)T(C/G)A with a spacer of 1, 5, or 8 nucleotides [2]. Further, RXR is known to interact with other binding partners such as the Vitamin D Receptor and the Thyroid Receptor via this motif [1]. In this work, we investigate the distribution of the exact repeats that comprise the RARE motif in sequences immediately upstream of zebrafish genes.

2 METHODS

We examined the promoter sequences of the zebrafish transcriptome 1Kb upstream of each of the 65,171 genes in the danRer11 (GRCz11) reference genome. We searched for 17-mers that contained a 6-mer prefix and a 6-mer suffix with five wildcard nucleotides in between. For each prefix and suffix, we looked for all possible combinations the 6-mer RARE motif reported in previous work [2], totalling 64 different prefix/suffix combinations in the 17-mer.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
BCB '20, September 21–24, 2020, Virtual Event, USA
© 2020 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-7964-9/20/09.
<https://doi.org/10.1145/3388440.3414914>

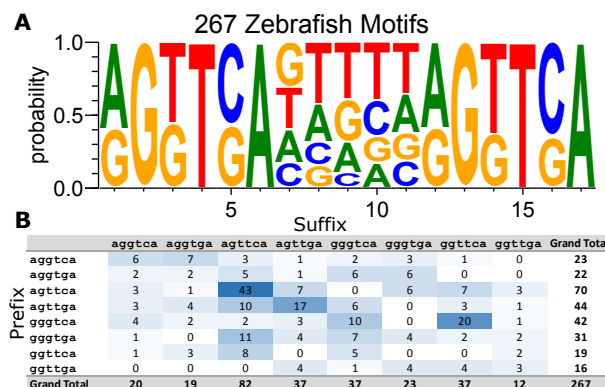


Figure 1: The variable RARE motif. A: Weblogo of RARE motif instances found 1Kb upstream of zebrafish transcripts. B: Table of the number of (prefix,suffix) pairs for each of the 64 motif combinations.

3 RESULTS AND DISCUSSION

We identified a total of 249 genes that contained at least one instance of a RARE motif in the promoter sequence. Of these, 243 (98%) genes contained a single motif, and the remaining genes contained multiple 17-mers, totalling 267 instances of the motif (Figure 1A). The information content of the 5-mer span between the prefix and suffix is quite low despite the prevalence of T's. Strikingly, one gene (XR_661516.3) contained ten exact 17-mer repeats of GGGTCA GTGTGGGTCA, which includes a conserved middle 5-mer. XR_661516 is an uncharacterized RNA transcript (LOC103911034), and its expression in the presence of RA has yet to be explored.

We then examined the distribution of (prefix,suffix) pairs in the 17-mer instances (Figure 1B). Thirty-one percent of the instances are DRs (diagonal), but this varies by the specific 6-mer. Two of the eight 6-mers are most commonly found as DRs, whereas the other 6-mers have variable (but non-random) patterns. We plan to interrogate existing RNA-seq data to investigate whether this (prefix,suffix) variation is biologically relevant to RA signaling.

ACKNOWLEDGMENTS

This work was supported by NSF award DBI-1750981 (to AR) and the Reed College Helen Stafford Summer Research Fellowship Fund.

REFERENCES

- [1] Ronald M Evans and David J Mangelsdorf. 2014. Nuclear receptors, RXR, and the big bang. *Cell* 157, 1 (2014), 255–266.
- [2] Sébastien Lalevée et al. 2011. Genome-wide in silico identification of new conserved and functional retinoic acid receptor response elements (direct repeats separated by 5 bp). *Journal of Biological Chemistry* 286, 38 (2011), 33322–33334.
- [3] David J Mangelsdorf and Ronald M Evans. 1995. The RXR heterodimers and orphan receptors. *Cell* 83, 6 (1995), 841–850.