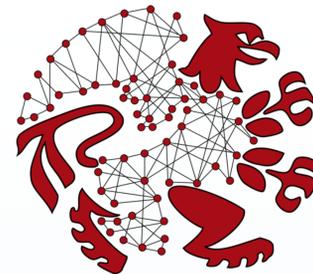


# Prefix/Suffix Variation in Retinoic Acid Response Elements

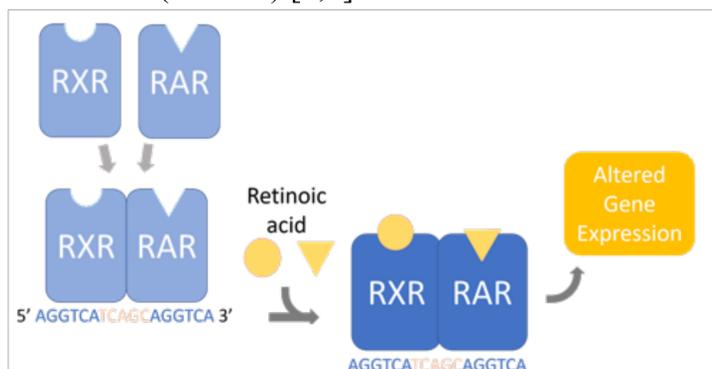
Yuan Zhuang, Kara L. Cervený, and Anna Ritz

Reed College Biology Department, Portland, OR, USA

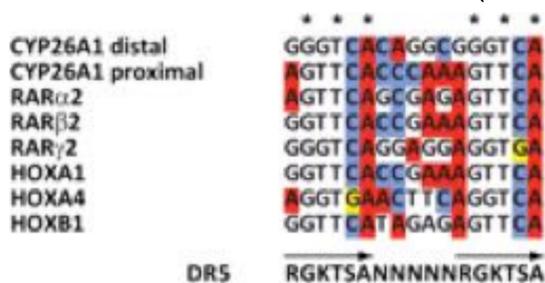


## I. Motivation

Retinoic acid (RA) is a derivative of Vitamin A that is essential for vertebrate growth and development. RA regulates the expression of target genes by converting a heterodimer comprised of a general retinoid receptor (RXR) and an RA-specific receptor (RAR), into an active transcription factor that binds motifs called Retinoic acid response elements (RAREs) [2,3].



The RARE motif, originally thought to be a direct repeat (DR) of AGGTCA, was later found to have variation in the sequence [2]. This motif can be characterized as (A/G)G(G/T)T(C/G)A [2].



We investigate the distribution of the exact repeats that comprise the RARE motif in zebrafish gene promoter regions.

## II. Methodology

We examined the promoter sequences of the zebrafish transcriptome 1Kb and 10Kb 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], totaling 64 different prefix/suffix combinations in the 17-mer.

## References

- [1] Ronald M Evans and David J Mangelsdorf. 2014. Nuclear receptors, RXR, and the big bang. *Cell*.
- [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*.
- [3] David J Mangelsdorf and Ronald M Evans. 1995. The RXR heterodimers and orphan receptors. *Cell*.

## III. Results

We performed the same analysis on the 1Kb and 10Kb upstream regions of each gene in the zebrafish reference genome.

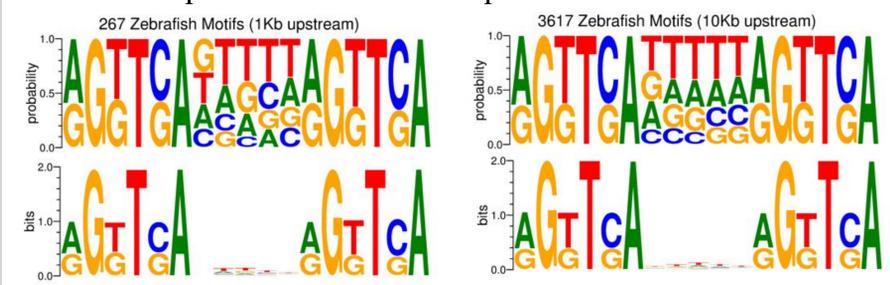
### 1Kb Upstream

- 249 genes had at least one RARE motif instance
- 6 (2.5%) of genes contained more than one motif

### 10Kb Upstream

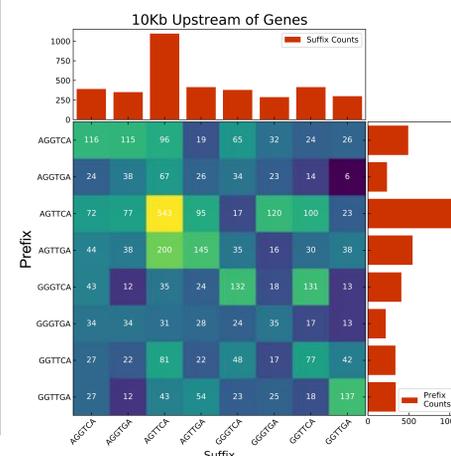
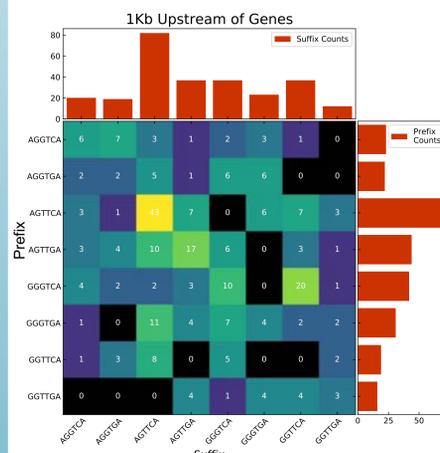
- 3,281 genes had at least one RARE motif instance
- 210 (6.4%) of genes contained more than one motif

The distribution of nucleotides in the 17-mer motifs were similar for the 1Kb and 10Kb datasets, despite the 13-fold increase in genes with the motif. The information content of the 5-mer span between the prefix and suffix was quite low.



- 31% of motifs are DRs (diagonal)
- 50% of DRs are AGGTCA...AGGTCA

- 34% of motifs are DRs (diagonal)
- 44% of DRs are AGGTCA...AGGTCA



## IV. Conclusions & Future Work

- While the AGGTCA...AGGTCA is a prevalent RARE motif, it only comprises 16% and 15% of motifs 1Kb and 10Kb upstream, respectively.
- Our next goal is to cross-reference genes identified in this unbiased search with genes recently identified as RA-responsive by an RNA-Seq experiment.

## Acknowledgements

We thank Tayla Isensee for the pathway diagram of RA signaling. This work was supported by NSF award DBI-1750981 (to AR) and the Reed College Hellen Stafford Summer Research Fellowship Fund (to YZ).



REED COLLEGE

