

Conservation of the Regulatory Region and the Genes they Regulate within a Network



Chinmay P. Rele & Laura K. Reed

Department of Biological Sciences, The University of Alabama, Tuscaloosa AL 35404

Introduction

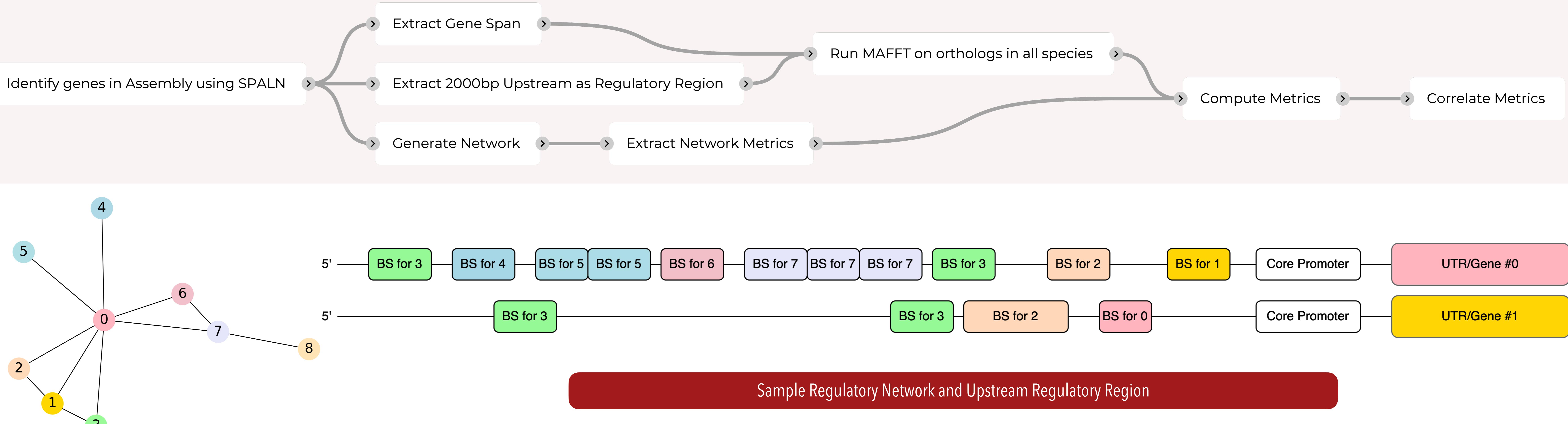
Genes interact with each other in large networks, not simply in linear pathways. Understanding network structure in molecular biology is crucial for unraveling complex interactions between genes, proteins, and other biomolecules, shedding light on regulatory mechanisms, and emergent properties essential for comprehending biological processes at a systems level. **Biological regulatory networks organize molecular interactions governing gene expression and cellular processes.** Biological regulatory networks are large, multifaceted, and have a broad range in regulatory effects.

Understanding the association between the evolution of the coding sequences of genes and their regulatory framework within the confines of a network can inform us of **large-scale evolutionary constraint on these nodes**. Nodes are genes that have a protein product, or those that regulate other genes. They regulate other genes by associating directly or indirectly with their regulatory frameworks.

Genes are **modularly regulated** – they can be regulated by a combination of multiple factors. Appropriate gene regulation is achieved by the *combinatorial outcome of "unit operations"*. There are a modest number of these individual regulatory factors/operations, but a large number of combinations of factors that execute these operations to fine-tune regulatory frameworks.

Nodes with more diverse function/high pleiotropy ought to also have a more diverse regulatory framework.

Method



Conclusion/Takeaways

The more a node is connected in the network, the more the gene span is conserved, and the more fragmented its corresponding regulatory region is.

Future Directions

Identify motifs in upstream regulatory region using MEME/XTREME/GLAM2

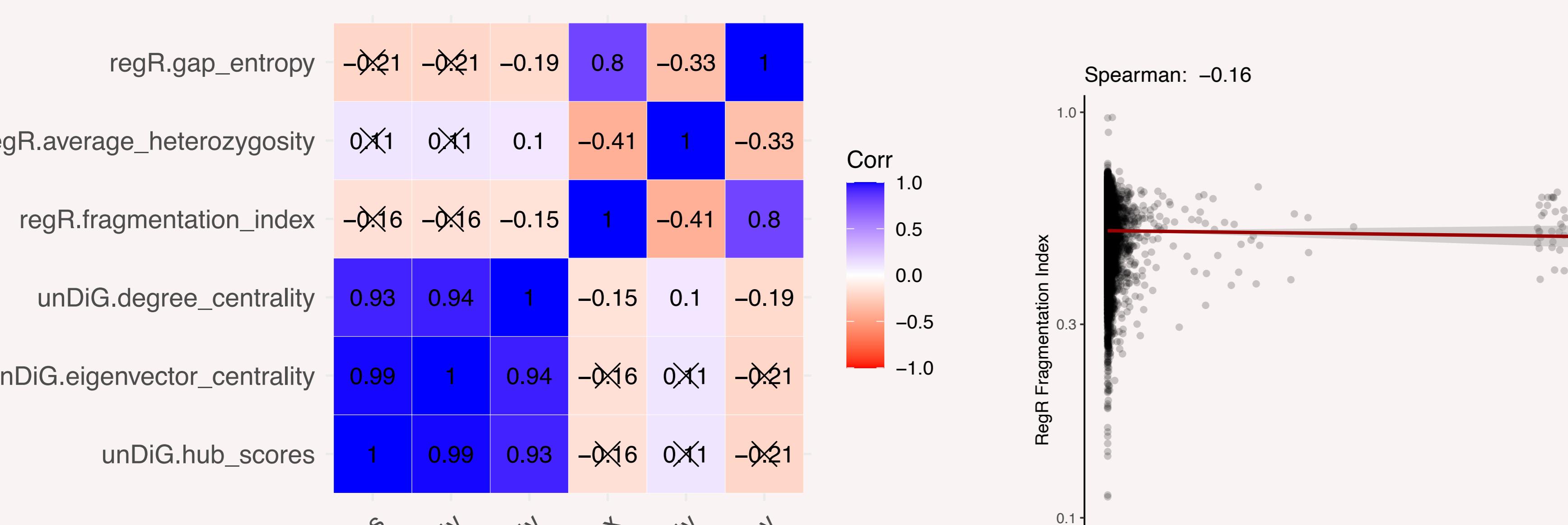
References/Acknowledgements

- I would like to thank all members of the Rele Lab for their input in crucial stages of development. I would also like to thank Daryl Lam, Michael McKinlay and Kevin M. Kocot to help run the analysis on UAHPC.
- Baheri, S. & Vingron, M. Studying the Evolution of Promoter Sequences: A Waiting Time Problem. *J. Comput. Biol.* 17, 1591-1604 (2010).
- Yona, A. H., Alm, E. J. & Goren, J. Random sequences rapidly evolve into de novo promoters. *Nat. Commun.* 9, 1503 (2018).
- Witkopp, P. J. Evolution of cis-regulatory sequence and function in Drosophila. *Heredity* 97, 139-147 (2006).
- Annone, M. I. & Davidson, E. H. The hardening of development: organization and function of genomic regulatory systems. *Development* 124, 1851-1864 (1997).
- Witkopp, P. J. *Evolution across 20 Mammalian Species*. *Cell* 160, 554-566 (2015).
- Yu, H. & Gerstein, M. Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl. Acad. Sci.* 103, 14731-14737 (2006).
- Becker, M. et al. Large-scale correlation network construction for unraveling the coordination of complex biological systems. *Nat. Comput. Sci.* 3, 346-359 (2023).
- Davidson, E. H. *cis-Regulatory Modules, and the Structure/Function Basis of Regulatory Logic*. In *The Regulatory Genome* (ed. Davidson, E. H.) 31-60 (Academic Press, Burlington, 2006). doi:10.1016/B978-012088563-3.50020-1.

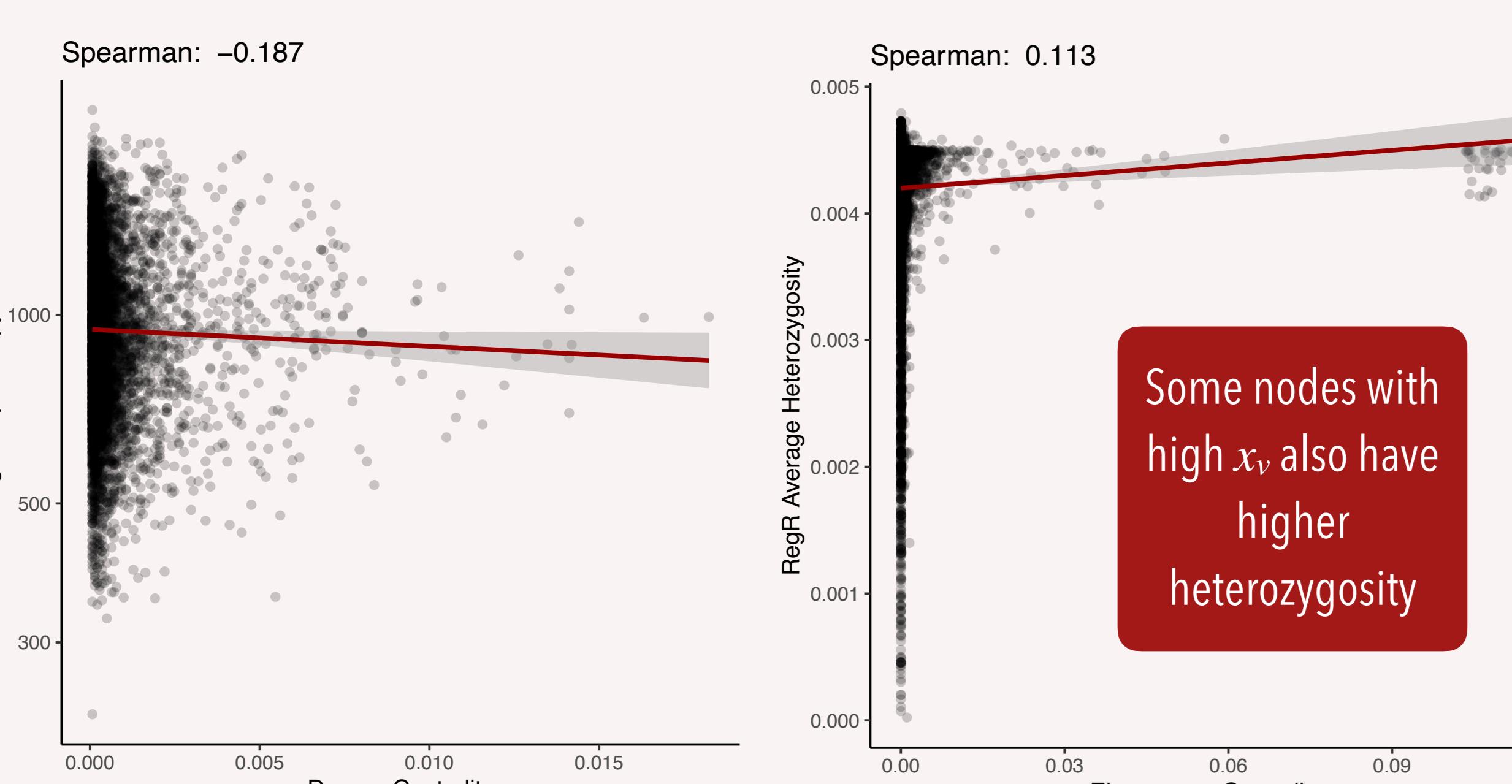
Network vs. Regulatory Region (RegR)

Hypothesis

Highly connected nodes need more TFBs in their upstream regulatory regions, so the upstream regulatory region is going to have more motifs/the MSA will be more conserved.



Nodes with higher **hub scores** have more connections to other nodes in the network, indicating their importance as central hubs or key connectors within the network structure.



Degree centrality refers to the measure of how connected a node is to other nodes in a network. A higher degree centrality for a node indicates that it is more connected to other nodes in the network.

Of all RegR/Network correlations, **average gaps in the RegR alignment were present in 66% of all positive correlations, and gap metrics of the MSA were present in 67% of all correlations.**

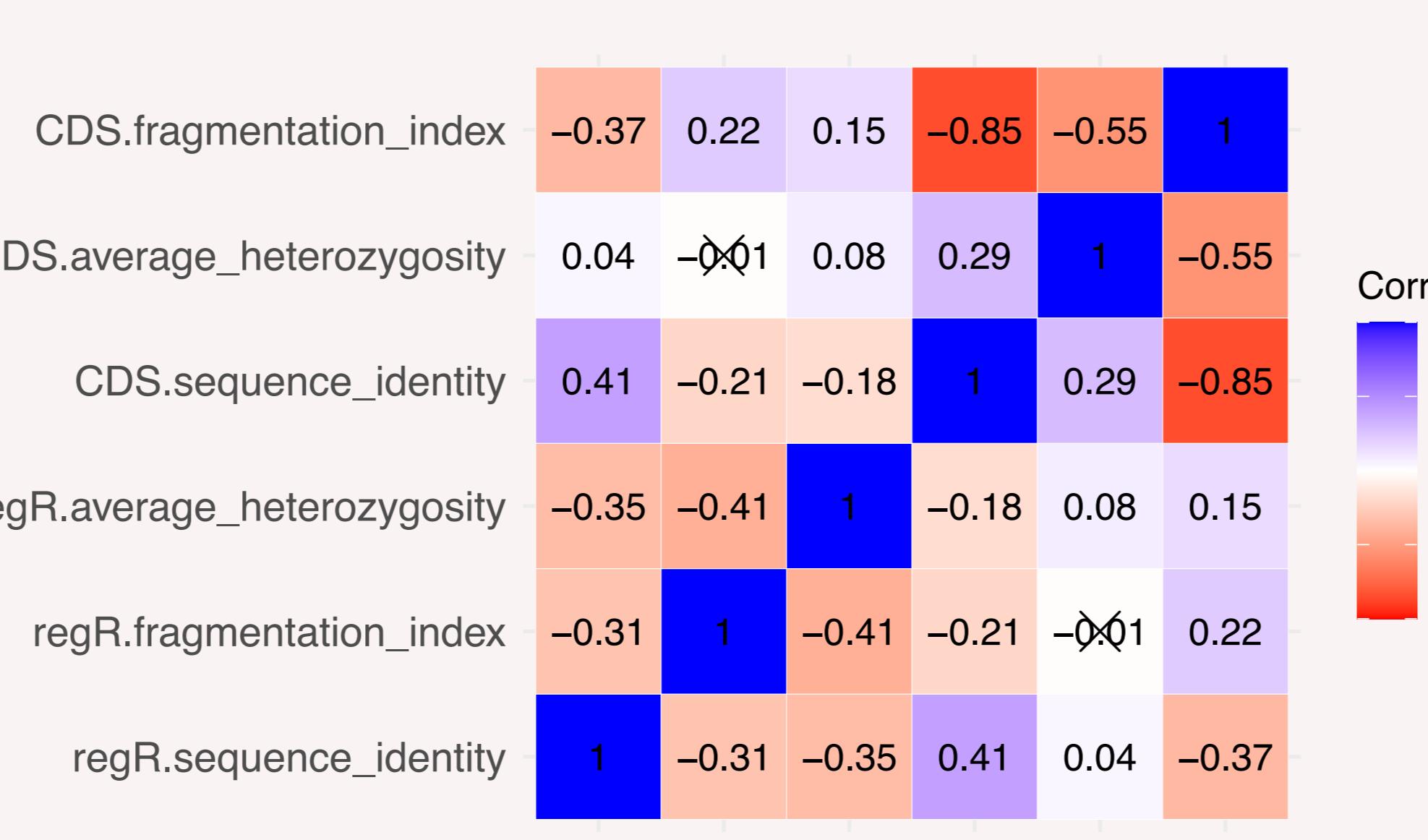


Results

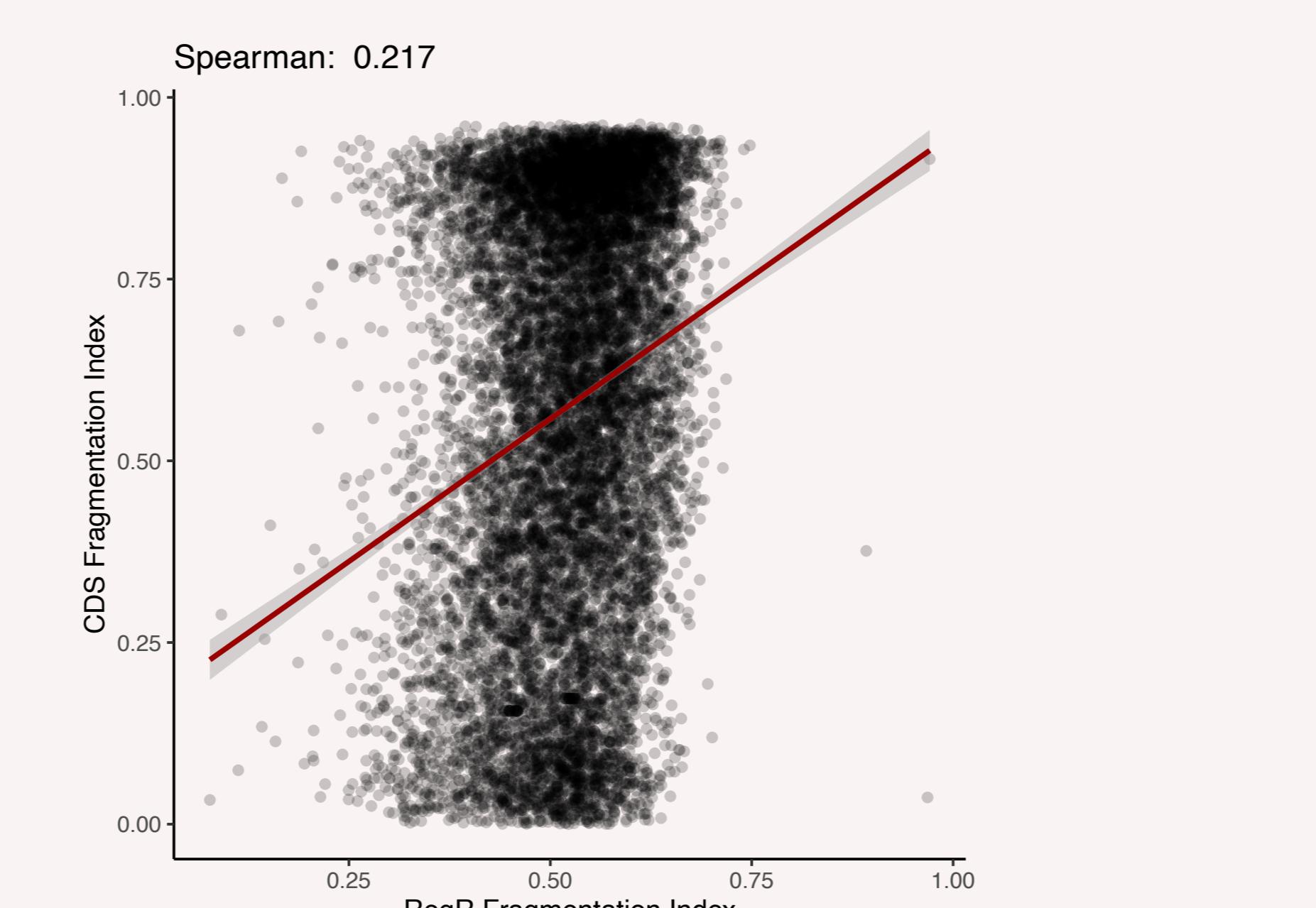
Regulatory Region (RegR) vs. CoCoding Sequence (CDS)

Hypothesis

More conserved genes are going to have equally conserved regulatory regions because their regulatory framework also needs to be constrained to maintain similar function.



Sequence Identity is calculated by looking at each position in the MSA. If the same residue is in the position in all sequences, the identity is 1. This is then averaged over the length of the entire alignment. Larger the sequence identity of an alignment, more conserved the alignment was at specific base positions.



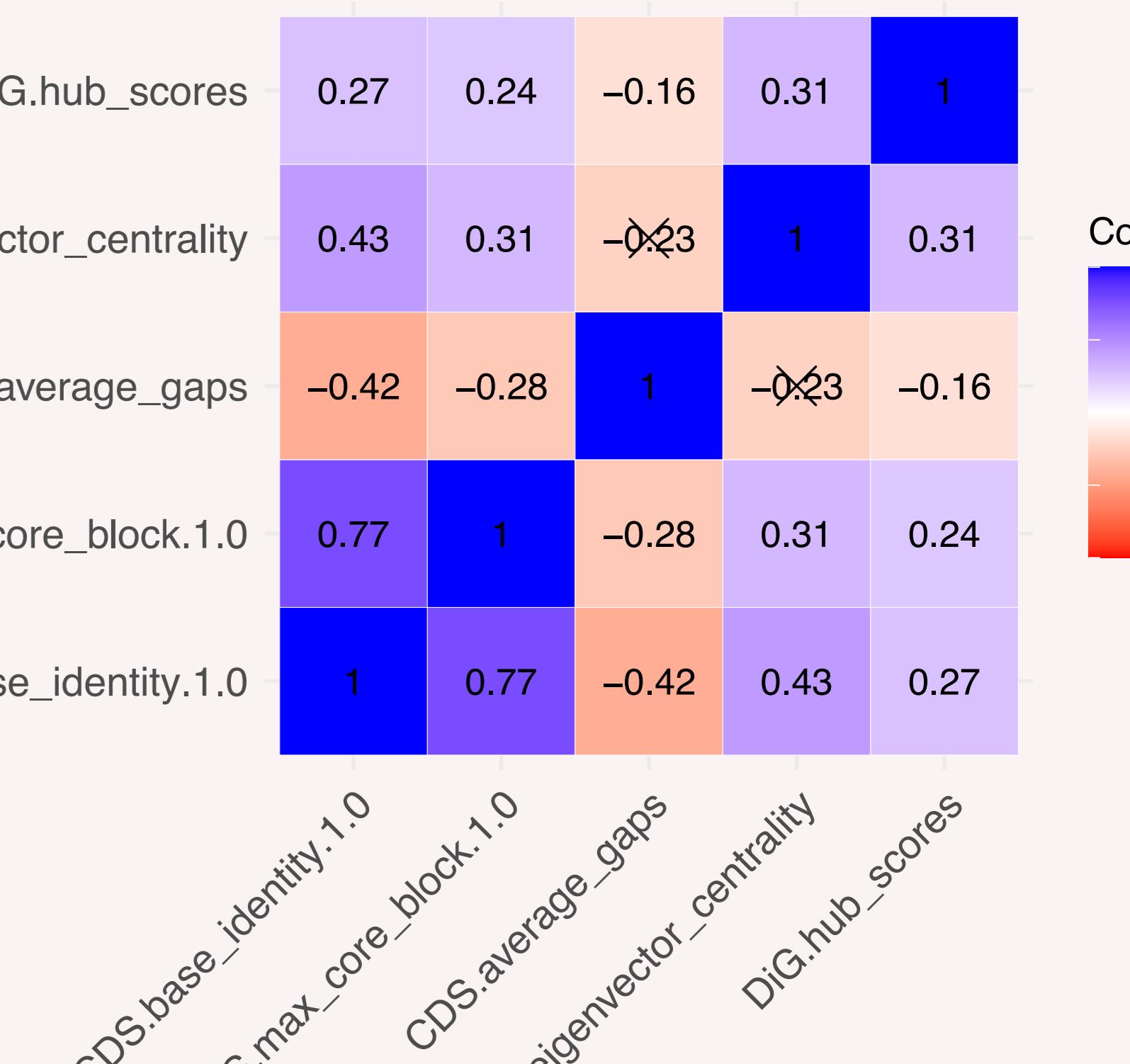
Average Heterozygosity (genetic variation within an individual's DNA at specific loci) is calculated by identifying the distribution of all residues in a single position in the alignment. Higher the average heterozygosity, higher the variation in the component sequences used to generate the alignment.

CoCoding Sequence (CDS) vs. Network

All correlations reported are significant with $p \leq 1e-2$

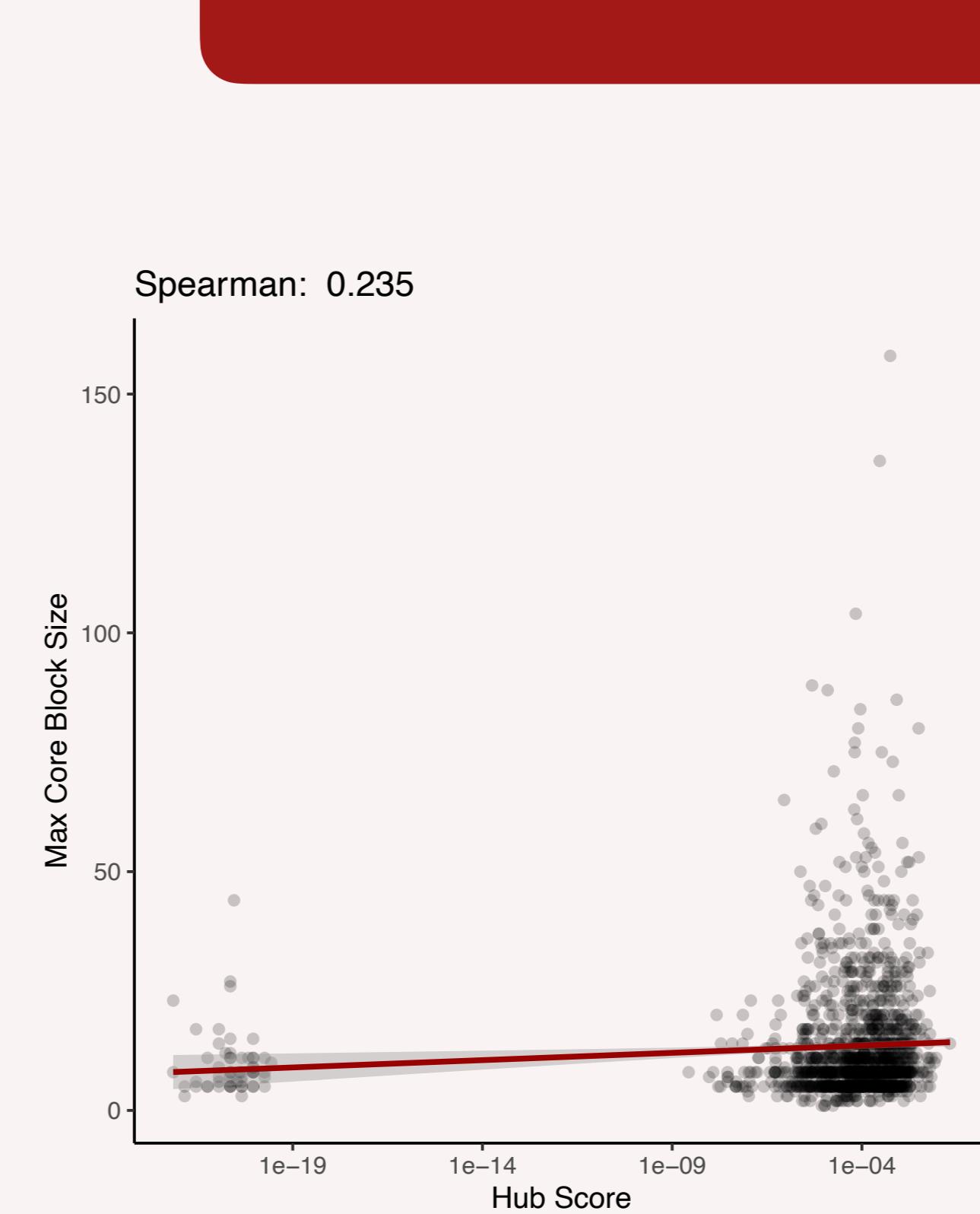
Hypothesis

Highly connected nodes are going to be more conserved in their gene span due to putative pleiotropy.



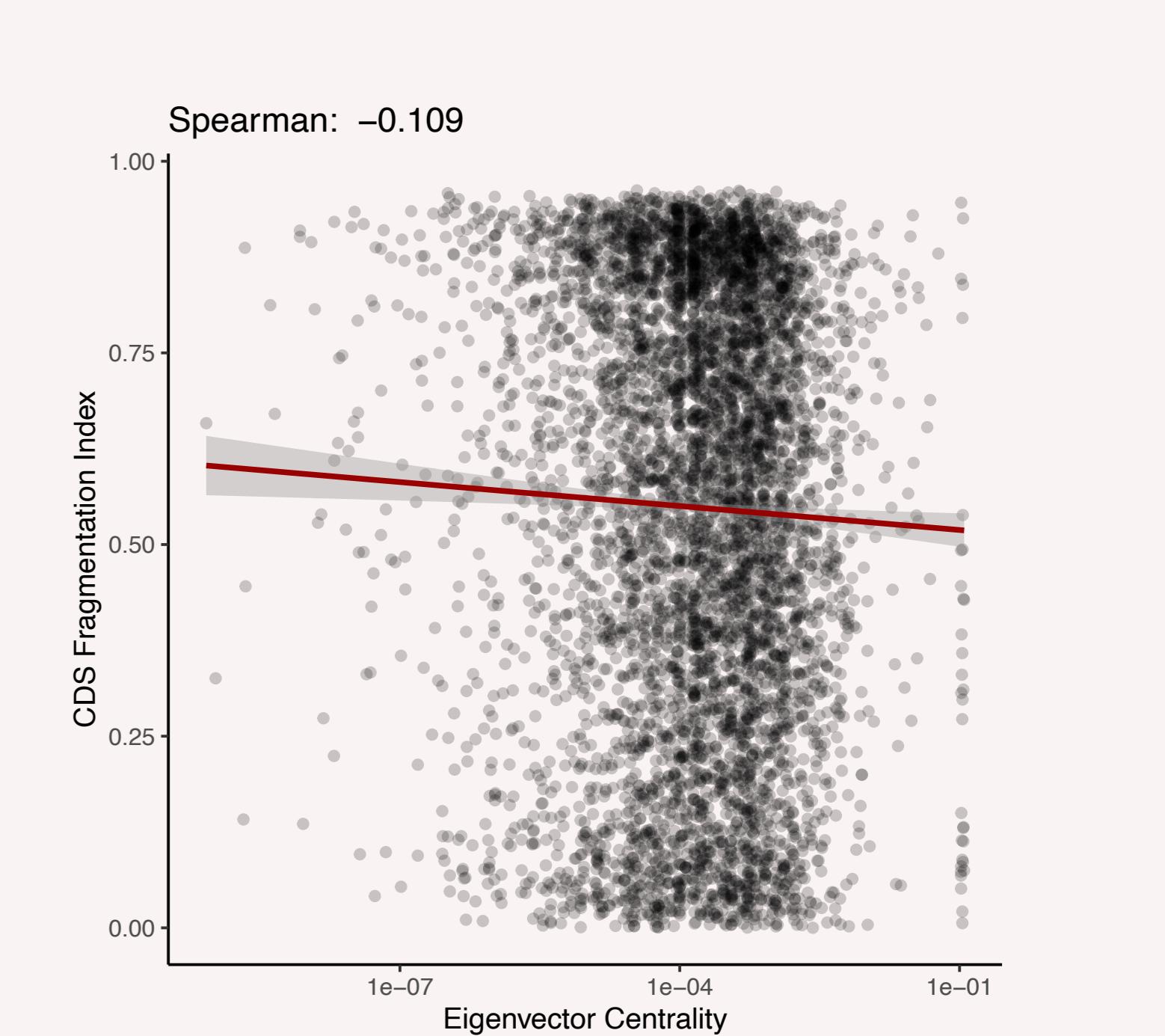
Eigenvector centrality (x_v) is a measure of the importance of a node in a network based on its direct connections but also on the centrality of its neighbors. A higher eigenvector centrality for a node indicates that it is not only well-connected but also connected to other nodes that are themselves well-connected.

Analysis for the CDS sequence was run on the entire gene span, and not only the protein-coding region



Base Identity

Base Identity is the number of residues in that are perfectly conserved at that position in all sequences in the alignment. Higher base identity means higher levels of conservation.



Max Core Block Size is the number of contiguous residues that have perfect identity at the particular position in the alignment. Higher max core block size, higher the contiguous residues that are perfectly conserved in the alignment.

