

# Ultra-Conserved Non-Coding Elements in the human genome. Their possible functions and association with diseases

Fedorova L.<sup>1</sup>, Fedorov A.<sup>1, 2, 3\*</sup>

<sup>1</sup> CRI Genetics LLC, Santa Monica, CA 90404, USA

<sup>2</sup> Program of Bioinformatics and Proteomics/Genomics, University of Toledo, Toledo, OH 43606, USA

<sup>3</sup> Department of Medicine, University of Toledo, Toledo, OH 43606, USA

\* alexei.fedorov@utoledo.edu

**Key words:** nearest neighbor; DNA conformation; meiosis; bioinformatics; ultraconserved

Thousands of prolonged sequences of human ultra-conserved non-coding elements (UCNEs) share only one common feature: peculiarities in the unique composition of their dinucleotides. Here we investigate whether the numerous weak signals emanating from these dinucleotide arrangements can be used for computational identification of UCNEs within the human genome. For this purpose, we analyzed 4272 UCNE sequences, encompassing 1,393,448 nucleotides, alongside equally sized control samples of randomly selected human genomic sequences. Our research identified nine different features of dinucleotide arrangements that enable differentiation of UCNEs from the rest of the genome. We employed these nine features, implementing three Machine Learning techniques -- Support Vector Machine, Random Forest, and Artificial Neural Networks -- to classify UCNEs, achieving an accuracy rate of 82-84%, with specific conditions allowing for over 90% accuracy. Notably, the strongest feature for UCNE identification was the frequency ratio between GpC dinucleotides and the sum of GpG and CpC dinucleotides. Additionally, we investigated the entire pool of 31,046 SNPs located within UCNEs for their representation in the ClinVar database, which catalogs human SNPs with known phenotypic effects. The presence of UCNE-associated SNPs in ClinVar aligns with the expectation of a random distribution, emphasizing the enigmatic nature of UCNE phenotypic manifestation. We propose that the key to the cryptic properties of UCNEs is hidden in their specific DNA conformations. Our hypothesis is that specific non-canonical DNA conformation of UCNEs may be integral to the homologous pairing of double-stranded DNAs during meiosis.

**Funding:** This research received no external funding.

## References

1. Crossley E.R., Fedorova L., Mulyar O.A., Freeman R., Khuder S., Fedorov A. Computational Identification of Ultra-Conserved Elements in the Human Genome: A Hypothesis on Homologous DNA Pairing. *NAR Genom Bioinform.* 2024;6(3):lqae074. doi 10.1093/nargab/lqae074
2. Fedorova L., Crossley E.R., Mulyar O.A., Qiu S., Freeman R., Fedorov A. Profound Non-Randomness in Dinucleotide Arrangements within Ultra-Conserved Non-Coding Elements and the Human Genome. *Biology.* 2023;12:1125. doi 10.3390/biology12081125
3. Fedorova L., Mulyar O.A., Lim J., Fedorov A. Nucleotide Composition of Ultra-Conserved Elements Shows Excess of GpC and Depletion of GG and CC Dinucleotides *Genes (Basel).* 2022;13(11):2053. doi 10.3390/genes13112053