

## Abstract

# DeepG4 : A deep learning approach to predict active G-quadruplexes

ROCHER V.<sup>1</sup>, GENAIS M.<sup>1</sup>, NASSEREDDINE E.<sup>1</sup> and MOURAD R.<sup>1\*</sup><sup>1</sup>LBCMCP, Centre de Biologie Intégrative (CBI), CNRS, Université de Toulouse, UT3

\*Corresponding author: raphael.mourad@ibcg.biotoul.fr

## Abstract

G-Quadruplex (G4) are alternative DNA secondary structures composed of Guanine-rich DNA sequences which can form a four-stranded structure based on a simple strand, and let the second one free [1]. These structures have been found initially on telomeres, but more recent studies found an enrichment of these structures on promoters of active genes, and suggest an active role in transcription of these genes [2]. Former in-silico methods to detect and study G4 remained mostly on the detection of a specific motif chain [3], but recent methods have been developed to identify G4 at genome-wide scale using Next Generation sequencing approach, like G4-seq [4](in-vitro G4) and BG4-seq [5](in-vivo). Here, we propose a sequence-based computational Deep learning model to predict in-vivo DNA G4 using the DNA sequences of BG4-seq peaks, in order to detect new motifs involved in the G4 prediction. Deep learning is a recent and popular Machine learning set of approaches where model learn features directly from the data, meaning that we could identify de-novo motifs that are related to G4 prediction. This model can be applied to any DNA sequence to predict the G4 formation, and be used in genetics to study the impact of SNP's on the DNA G4 formation propensities.

## References

- [1] D. Sen and W. Gilbert. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, 334(6180):364–366, Jul 1988.
- [2] J. Spiegel, S. Adhikari, and S. Balasubramanian. The Structure and Function of DNA G-Quadruplexes. *Trends Chem*, 2(2):123–136, Feb 2020.
- [3] J. L. Huppert and S. Balasubramanian. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, 33(9):2908–2916, 2005.
- [4] G. Marsico, V. S. Chambers, A. B. Sahakyan, P. McCauley, J. M. Boutell, M. D. Antonio, and S. Balasubramanian. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.*, 47(8):3862–3874, 05 2019.
- [5] R. Hänsel-Hertsch, D. Beraldi, S. V. Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, D. Tannahill, and S. Balasubramanian. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, 48(10):1267–1272, 10 2016.