# **Paperfly: ab initio binding site reconstruction**

Faltejsková Kateřina1,2, Vondrášek Jiří1

*1 Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo náměstí*

*542/2, 160 00 Praha 6, Czech Republic*

*2 Charles University, Faculty of Mathematics and Physics, Computer Science Institute, Malostranské náměstí*

*25, 118 00 Praha 1, Czech Republic*

The specific recognition of a DNA locus is a widely studied issue. It is generally agreed that the recognition can be influenced not only by the binding motif, but by the larger context of the binding site. In order to study the binding site including the sequential context of the binding motif, we introduce PAPerFly: the Partial Assembly-based Peak Finder, a new tool capable of reconstructing the binding site from ChIP-seq or similar experimental data.

Using a novel heuristic algorithm that utilizes approaches used in the genome assembly, Paperfly can reconstruct the unique binding sites captured in a sequencing experiment without using the reference genome. Additionally, we show that Paperfly can be combined with the standard data processing pipeline to link the unique sequences of the binding site and their respective abundance over the genome.

The source code of the tool is freely available at https://github.com/Caeph/paperfly or at

https://doi.org/10.5281/zenodo.6379332.