Robust interaction detection in Capture Hi-C connects enhancers and disease SNPs to target genes

Jonathan Cairns[1], Paula Freire-Pritchett[1], Biola M Javierre[1], Oliver Burren[2], Sven Sewitz[1], Stefan Schoenfelder[1], Csilla Várnai[1], Andrew Dimond[1], Steven W. Wingett[1], Vincent Plagnol[3], Daniel Zerbino[4], Cameron Osborne[5], Willem Ouwehand[6], Mattia Frontini[6], Chris Wallace[7], Peter Fraser*[1], Mikhail Spivakov*[1]

(* - joint senior authors).

[1]*Babraham Institute, Cambridge.* [2]*Diabetes and Inflammation Laboratory, University of Cambridge.* [3]*Institute of Genetics, University College London.* [4]*EMBL-EBI, European Molecular Biology Laboratory.* [5]*King's College London.* [6]*NHS Blood and Transplant, Department of Haematology, University of Cambridge.* [7]*Department of Medicine, University of Cambridge.*

ABSTRACT

Nuclear organisation and, in particular, promoter-enhancer looping interactions have important implications for gene regulation. Capture Hi-C (CHi-C) is a state-of-the-art method for profiling nuclear interactions, involving (at least on one end) a large number of loci of interest, globally and at high resolution [1-3]. However, signal detection in CHi-C data is challenging and cannot be done with established Hi-C analysis methods, primarily due to asymmetry of the CHi-C interaction matrix.

We have developed CHiCAGO (http://regulatorygenomicsgroup.org/chicago), an open-source package for robust interaction detection in CHi-C. CHiCAGO features a noise model and algorithms for background correction that are specifically adapted to CHi-C data. CHiCAGO further implements a weighted false discovery control procedure that enables genome-wide, fragment-level signal detection, despite highly variable signal-to-noise ratios and the enormous number of tested hypotheses.

CHiCAGO has been used in a number of experimental contexts. We focus on one application, involving profiling promoter interactions in primary human blood cells. This analysis has led to insights into the basic principles of gene regulation, as well as enabled connecting disease-associated SNPs with their putative target genes.

[1] Schoenfelder, Stefan, et al. "The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements." *Genome research* 25.4 (2015): 582-597.
[2] Mifsud, Borbala, et al. "Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C." *Nature genetics* 47.6 (2015): 598-606.
[3] Sahlén, Pelin, et al. "Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution." *Genome biology* 16.1 (2015): 1-13.