iNucs

Citation for published version:

Digital Object Identifier (DOI):
10.1093/bioinformatics/btab698

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher’s PDF, also known as Version of record

Published In:
Bioinformatics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
iNucs: inter-nucleosome interactions

Mehrdad Oveisì,1†, Manu Shukla2†, Nogayhan Seymen1, Masae Ohno3,4, Yuichi Taniguchi4,4, Sunil Nahata5, Remco Loos6, Ghulam J. Mufti1, Robin C. Allshire2, Stefan Dimitrov5,7 and Mohammad M. Karimi1,1.*

Abstract

Motivation: Deciphering nucleosome–nucleosome interactions is an important step toward mesoscale description of chromatin organization but computational tools to perform such analyses are not publicly available.

Results: We developed iNucs, a user-friendly and efficient Python-based bioinformatics tool to compute and visualize nucleosome-resolved interactions using standard pairs format input generated from pairtools.

Availability and implementation: https://github.com/Karimi-Lab/inucs/.

Contact: mohammad.karimi@kcl.ac.uk

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

In eukaryotes, DNA is packaged in the nucleus as a nucleoprotein complex called chromatin. The nucleosome is the fundamental unit of chromatin, around which 147 bp of DNA is wound. The packaging of chromatin in the nucleus is not linear. This results in different levels of 3-D genome folding and interaction. Deciphering genome organization within the nucleus is a multi-scale problem and while higher-order tertiary genome structures have been extensively studied using chromatin conformation capture technologies (e.g. 3C, Hi-C etc.), the secondary structure of chromatin, i.e. nucleosomal interactions, remains elusive. Recently, new genome-wide chromosome conformation approaches have been introduced, allowing interactions at nucleosome level. However, the commonly utilized visualization tools such as HiGlass and Juicebox (Kerpedjiev et al., 2018; Robinson et al., 2018) are limited in their ability to provide nucleosome-resolved interactions due to their underlying bin-based methods that do not take into account the nucleosome positions (Supplementary Fig. S1). Therefore, a dedicated bioinformatics tool which allows analysis and visualization of nucleosome interaction networks at individual loci is needed. In this study, we describe iNucs, a user-friendly and efficient Python-based tool for computing and visualizing nucleosome-level interaction networks. iNucs is designed to receive ligation junctions in pairs format generated by pairtools (https://pairtools.readthedocs.io/en/latest/) and provide nucleosome resolution view consistent with pairsQC (https://github.com/4dn-dcic/pairsqc) output and 4D nucleosome (4DN)-processed files. This allows iNucs to be integrated as a plugin to 4DN analysis pipeline in the future. In addition, the tool provides division and visualization of interaction pairs based on their orientation providing further spatial information about the conformation of chromatin fiber at genomic loci.

2 Materials and methods

iNucs consists of prepare and plot modules for which the processes, input and output files and formats are shown in Figure 1A. For a detailed, step-by-step examples of this tool, please see iNucs GitHub page.

Prepare module: This module takes in nucleosome genomic coordinates, ligation junctions (in pairs format) and chromosome names as input, and it produces a nucleosome interaction matrix (NIM) as output. The pairs’ input data are typically very large, particularly for human samples (Fig. 1B). Thus, the running time of the underlying
algorithms to produce NIM can be potentially very high. To overcome this problem, we reduced the problem of matching DNA interactions with nucleosome coordinates into a sorting algorithm. This in turns allows us to leverage the power of existing sorting algorithms in Python to find an optimized solution for generating NIM. The underlying algorithm in iNucs has a time complexity (n log n), where n is the larger of nucleosome counts and ligation junction counts reported in pairs file (Fig. 1B). One substantial benefit of reducing our problem into sorting is that iNucs can utilize the hardware-supported vectorization capabilities provided by Python libraries NumPy/Pandas that significantly speed up our program. Moreover, iNucs uses advanced parallelization techniques which exploits the power of CPU cores in computational resource and respective times required to perform the analysis.

**3 Evaluation and conclusion**

We used recently published *Saccharomyces cerevisiae* Hi-CO (Ohno et al., 2019) and H1 human embryonic stem cell (hESc) line Micro-C (Krietenstein et al., 2020) data to validate iNucs results. The genomic coordinates of nucleosomes for *S. cerevisiae* and ligation junctions (pairs file) were provided by the authors of Hi-CO data. In the case of human data, the pairs file was publicly available in 4DN portal, but there were no matched nucleosome coordinates. Therefore, we downloaded nucleosome coordinates from NucMap repository (https://nsgd.cncb.cn/nucmap/NucMap_FTP_Directory/Homo_sapiens/byDataType/Nucleosome_peaks_DANPOS/) for H1-hESC (Sample ID hsNuc0070101) (Yazdi et al., 2015; Zhao et al., 2019). NucMap (Chen et al., 2013) was the main tool utilized by NucMap to call nucleosome peaks and coordinates from H1-hESC MNase-seq dataset. We plotted nucleosome-resolved interaction heatmap matrices for specific regions in the respective genomes and compared them with the related heatmaps published in original studies for those regions (Supplementary Fig. S2A and B). As nucleosome resolved heatmap was not available for H1-hESC, we focused on general chromatin organizational features. As expected, local chromatin features such as self-associated domains and boundaries were accurately reproduced by iNucs in the case of H1-hESC (Supplementary Fig. S2A). Also, iNucs generated nucleosome-resolved interaction matrices, which closely resembled nucleosome interaction profiles reported previously for yeast (Ohno et al., 2019) (Supplementary Fig. S2B).

**Acknowledgement**

The authors thank Oliver Rando for helpful suggestions during the course of this study.

**Funding**

R.C.A. is a Wellcome Principal Research Fellow [095021, 200885]; the Wellcome Centre for Cell Biology is supported by core funding from Wellcome [203149]. This work was supported by grants-in-aid for Scientific Research (A) [20H00460], Japan Society for the Promotion of Science. Astellas Foundation for Research on Metabolic Disorders and Suntory Rising Stars Encouragement Program in Life Sciences (SunRiSE). M.O., N.S. and R.C.A. are supported by the Japan Society for the Promotion of Science Postdoctoral Fellowship Program. M.O. and N.S. are supported by the Japan Society for the Promotion of Science Postdoctoral Fellowship Program.

**Conflict of Interest:** none declared.

**References**


