

Describing the genes social networks relying on chromosome conformation capture data

Ivan Merelli¹✉, Pietro Liò², Luciano Milanese¹

¹Institute for Biomedical Technologies - National Research Council of Italy, Italy

²University of Cambridge, United Kingdom

Received 31 July 2013; Accepted 6 September 2013; Published 14 October 2013

Competing interests: the authors have declared that no competing interests exist.

Abstract

The analysis of the chromosomal conformation in the nucleus is critical for understanding mechanisms that regulate genes colocalisation and coexpression. As for social networks that combine information about positions and activities of users, we show how it is possible, by using an R package developed for this purpose, to correlate the three-dimensional organisation of chromosome in the nucleus and gene activity during stem cell differentiation.

Motivation and Objectives

In the social network society it is often difficult to organise information in a systemic view. In all fields, from information technology and sociology to biomedical science, and in particular genetics, the description of the interactions established by vertices in a network is incredibly important (Barabási *et al.*, 1999). Nowadays, through mobile devices localization it is possible to provide suitable information about facilities and opportunities in the neighbourhood of the users. Social networks such as FourSquare make of positioning the core of their business. Information about localizations combined with a constant encouragement of people to describe their activities, allows tracing mass movement and behavior (Noulas *et al.*, 2011). The collection of such information can be extremely useful, for example, in marketing, in order to target advertisements and promotions.

These concepts have been already applied to medicine, which in future will be participatory and personalised (Hood, 2013) by mean of social networks, such as PatientsLikeMe. Moreover, future medicine is expected to be predictive and preventive (Hood, 2013), by fully exploiting the integration with omics science, in particular for understanding how the 3D nuclear maps of genes can be exploited for precisely targeting new drugs. Noteworthy, both social networks and molecular networks could use semantic information, part of which could be shared across these domains. From a human cognitive behavior, it would be ideal to have similar ways for querying different networks and perhaps molecular and cellular information could be even linked to patients' social networks. Can we design 3D nu-

clear information the same way we design social networks?

Despite several hundreds human genomes have been sequenced, we know very little about three-dimensional chromosome conformation beyond the scale of the nucleosome. Considering the number of evidences about colocalisation and coregulation of genes, this is very important in describing the social behavior of genomic actors (Di Stefano *et al.*, 2013). In particular, recent advances in high throughput molecular biology techniques and bioinformatics have provided insights into chromatin interactions on a larger scale (Lieberman-Aiden *et al.*, 2009). A novel sequencing technique called Chromosome Conformation Capture (3C) allows analysing the organisation of chromosomes in a cell's natural state (Duan *et al.*, 2012). While performed genome wide, this technique is usually called Hi-C. Clearly, studying the structural properties and spatial organisation of chromosomes is important for the understanding and evaluation of the regulation of gene expression, DNA replication and repair, and recombination (Lin *et al.*, 2012).

Methods

Inspired by social networks like FourSquare, we developed NuChart (Merelli *et al.*, PLoS One, accepted), an R package that integrates Hi-C information, describing the chromosomal neighborhood, with predicted CTCF binding sites (Botta *et al.*, 2010), isochores (Marculescu, *et al.*, 2006), potential cryptic RSSs (Varriale and Bernardi, 2010), and other user desired genomics features, such as methylation and chromatin conformation, to infer how the three-dimensional organisa-

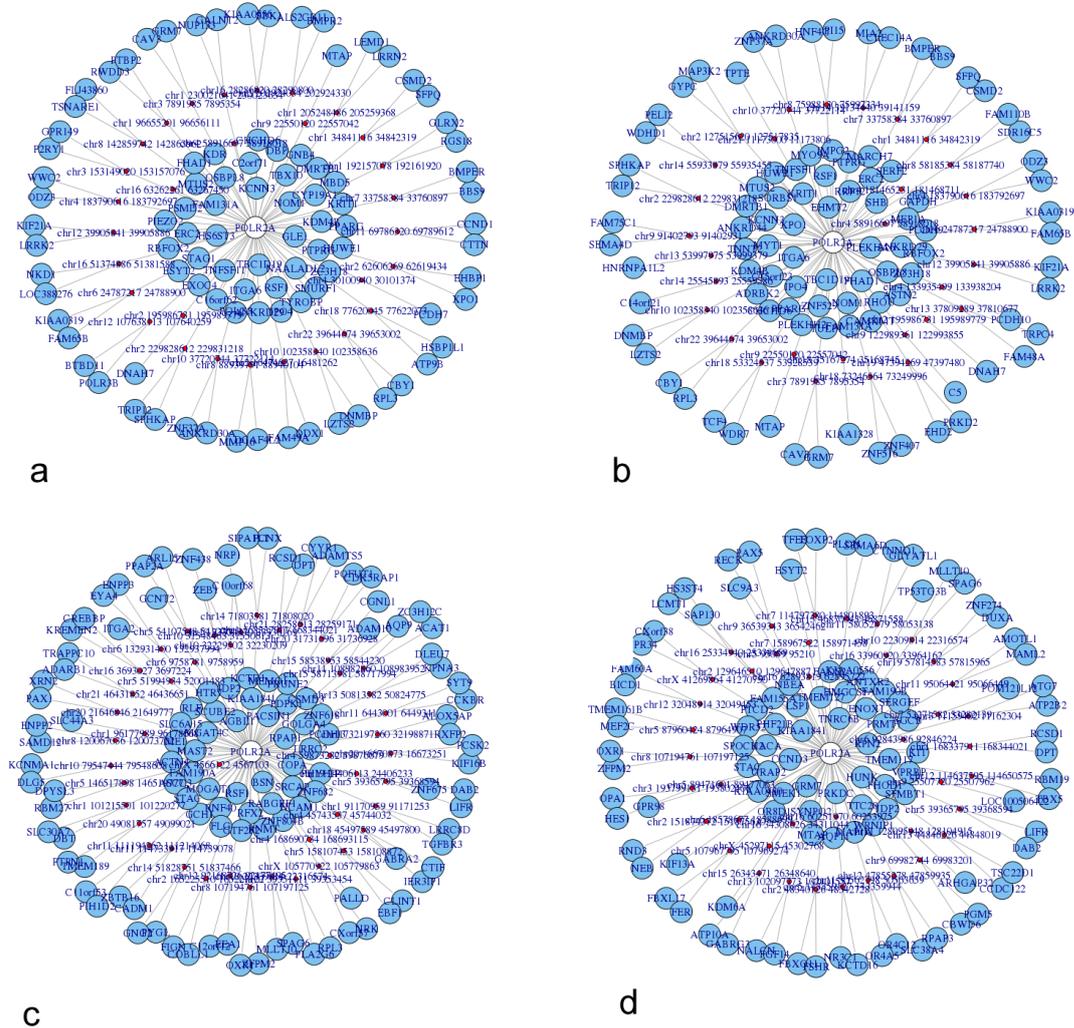


Figure 1. Neighborhood graph of the gene POLR2A in four different runs from the Hi-C experiments of Dixon to show inter and intra run modifications. In the panel a) and b) on the top part of the figure, the sequencing runs are from the same cell line hESC, while panel c) and d) are from IMR90.

tion of DNA works in controlling gene expression. This can be very useful while studying the differentiation of stem cells or to identify chromosomal reorganisations in cancer cells. For example, Philadelphia translocation is a specific chromosomal abnormality that is associated with chronic myelogenous leukemia (CML). It is the result of a reciprocal translocation between chromosome 9 and 22. The software has been designed to answer question such as: are chromosomal translocations occurring between nearby chromosomes?

Beside functions for loading and normalizing data, the core of NuChart is the core of the

neighborhood graph of the user provided list of genes or pathway. This package provides the possibility of analysing Hi-C data in a multi-omics context, by enabling the capability of mapping on the graph vertices expression data, according to a particular transcriptomics experiment, and on the edges genomic features that are known to be involved in chromosomal recombination, looping and stability. At the same way of FourSquare, the relative positions of the network actors and their functional activities are the core information for describing the neighborhood global behavior.

NuChart also provides functions to describe, compare, and analyse statistically the neighborhood graphs after their creation, which can be very important to highlight local and global characteristics of the Hi-C fragment distributions in the nucleus and of the multi-omics features in the context of the DNA three-dimensional topology. The possibility of analysing data to infer structural-activity relationships in a social network is of critical importance (Reagans and McEvily, 2003).

Results and Discussion

In the following example, we discuss an interesting analysis related to the data of Dixon *et al.* (Dixon *et al.*, 2012) experiments. The idea is to show the different chromosomal organisations that occur in the nucleolus, while gene expression is heavily characterising the differentiation of stem cells. Respectively, the graphs in Figure 1 in the top part are from two different sequencing runs performed on human embryonic stem cells (SRA:SRR400261 and SRA:SRR400262), while the graphs in the bottom part are from human lung embryonic fibroblast (SRA:SRR400264 and SRA:SRR400265).

In particular, these graphs show the neighborhood of the POLR2A gene, which catalyses the transcription of DNA to synthesize precursors of mRNA and most snRNA and microRNA, in the different cell lines. Noteworthy, the variability in the neighborhood of this gene, computed as correlation between the lists of adjacent genes, in the different cell lines is significant. While the similarity between two different runs of sequencing performed on the same cell type is quite high (respectively 60% and 67%), there are very importance differences between the two different cell lines (correlation below 30%), which witness the importance that chromosomal re-organisations at nucleolus level has for co-expression. This is very important to understand, in a particular moment, what the cell is going to express, by re-organising its internal chromosomal structure in the three dimensional space.

Recalling the parallelism with FourSquare, the power of these tools relies in capturing and describing the colocalisation and co-activation of entities in the social network. Moreover, the interaction of the social actors with the environment is of critical importance for understanding dynamics of the whole system. Future medicine will require the integration of different social and

genetic networks in a multilevel approach: therefore, the possibility of having topological coherent graph descriptions and overlapping semantics for annotations across these two domains will be very important.

Acknowledgements

This work has been supported by the Italian Ministry of Education and Research (MIUR) through the Flagship (PB05) InterOmics, HIRMA (RBAP11YS7K) and the European MIMOMICS projects.

References

- Barabási A-L, Albert R (1999) Emergence of Scaling in Random Networks. *Science* **286**(5439), 509-512. doi:10.1126/science.286.5439.509.
- Botta M, Haider S, *et al.* (2010) Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide. *Mol Syst Biol.* **6**, 426. doi:10.1038/msb.2010.79.
- Di Stefano M, Rosa A, *et al.* (2013) Colocalization of Coregulated Genes: A Steered Molecular Dynamics Study of Human Chromosome 19. *PLoS Comput Biol* **9**(3), e1003019. doi:10.1371/journal.pcbi.1003019.
- Dixon JR, Selvaraj S, *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**(7398), 376-380. doi:10.1038/nature11082.
- Duan Z, Andronescu M, *et al.* (2012) A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. *Methods* **58**(3), 277-288. doi:10.1016/j.ymeth.2012.06.018.
- Hood L (2013) Systems Biology and P4 Medicine: Past, Present, and Future. *Rambam Maimonides Med J.* **4**(2), e0012. doi:10.5041/RMMJ.10112.
- Lieberman-Aiden E, van Berkum NL, *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**(5950), 289-293. doi:10.1126/science.1181369.
- Lin YC, Benner C, *et al.* (2012) Global changes in the nuclear positioning of genes and intra- and inter-domain genomic interactions that orchestrate B cell fate. *Nat Immunol.* **13**(12), 1196-1204. doi:10.1038/ni.2432.
- Marculescu R, Vanura K, *et al.* (2006) Recombinase, chromosomal translocations and lymphoid neoplasia: targeting mistakes and repair failures. *DNA Repair (Amst)* **5**(9-10), 1246-1258.
- Noulas A, Scellato S, *et al.* (2011) An Empirical Study of Geographic User Activity Patterns in Foursquare. *Proceedings of the Fifth International Conference on Weblogs and Social Media ICWSM 2011*, Barcelona, Catalonia, Spain, July 17-21, 2011. The AAAI Press. Palo Alto, CA, USA. pp. 70-73.
- Reagans R, McEvily B (2003) Network structure and knowledge transfer: The effects of cohesion and range. *Administrative science quarterly* **48**(2), 240-267.
- Varriale A, Bernardi G (2010) Distribution of DNA methylation, CpGs, and CpG islands in human isochores. *Genomics* **95**(1), 25-28. doi:10.1016/j.ygeno.2009.09.006.