

RECOMB 2004

Proceedings of the
Eighth Annual International Conference on Research
in Computational Molecular Biology

Dan Gusfield, Philip Bourne, Sorin Istrail,
Pavel Pevzner, Michael Waterman
Editors

March 27-31, 2004 • San Diego, California, USA



Sponsored by
SIGACT

with support from

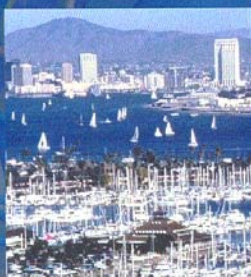
The University of California-San Diego,
The San Diego Supercomputer Center,
and The International Society for Computational Biology
Accelrys, Applied Biosystems, Apple Computer,
Cambridge University Press, Cornell Theory Center,
CRC Press - Taylor and Francis Group, The Department of Energy,
Geneart GmbH, Kluwer Academic Publishers, Microsoft Research,
The MIT Press, The National Science Foundation, Paracel,
Partek Incorporated, Springer-Verlag New York Inc., and Timelogic

RECOMB

2004 — SAN DIEGO

Currents in Computational Molecular Biology 2004

Edited by:
Apostol Gramada
Philip E. Bourne



H10. Alu Clustering in the Human Genome: Origins and Consequences

Michael Hackenberg¹ and Jose L. Oliver¹

Keywords: Alu, Isochores, Recombination, Evolution, Negative Selection, Alu Clustering

1 Introduction

Approximately 10% of the human genome is made up of Alu elements. Although the Alu repeats have been the focus of investigation over the last few years, many aspects of their genomic role still remain unsolved. For example, they appeared in the genome some 50-80 MYA and had their major spread approx. 30 MYA, which is about two orders of magnitude higher than the present amplification rate [1]. Other topics which are still controversial deal with the Alu distribution over the human genome. In general, Alu density is positively correlated with the GC content of the genomic region (isochore) in which they reside [2]. However, recently inserted Alus present a different pattern. As they depend on the LINE-1 transposition machinery, they show an initial density maximum in GC-poor regions. In [3], we analysed the Alu distribution as a function of evolutionary age and isochore membership, as well as the densities of possible recombination outcomes, inferring that recombination is probably the most important mechanism driving the density shift from GC-poor (L) to GC-rich (H) isochores. Here, we present an analysis of the Alus as a function of isochore membership and physical distance to the next Alu repeat (Distance to the Nearest Neighbour - DNN), which reinforces our argument and sheds light on this controversy from a different side.

We analysed the human reference sequence (April 2003 freeze, UCSC version hg15), based on NCBI Build 34 and produced by the International Human Genome Sequencing Consortium (IHGSC), downloaded from (<ftp://genome.ucsc.edu/goldenpath/10april2003/bigZips/chromFa.zip>). The partition of the chromosomes into isochores was performed using the IsoFinder segmentation algorithm (see [4] and references therein). The Alu densities indicated have been calculated as the number of elements per 10 kb.

2 Results and Discussion

As a measure for Alu clustering, we used the distance to the next Alu (DNN). The Alu density ratios (densities in H isochores divided by densities in L isochores) in the different isochores as a function of DNN are shown in Figure 1. For short distances, the densities are up to 8 times higher in H than in L isochores, which corresponds to an extremely higher clustering in the H isochores. For example, in the GC-richest isochore H4, almost 40% of all Alus are closer than 20 bp to each other, whereas in the GC-poorest isochore L1 this fraction declines to 5%. With growing distances to the closest Alu, however, the density maximum shifts gradually towards the L isochores; thus, Alus with distances greater than 2000 bp to the next Alu show a density maximum in L isochores. This means that interactions with other Alus are crucial for the density shift, as single Alus with long distances to the next Alu remain in L isochores. There are two possible explanations: Alu/Alu recombination and preferential insertion in or near preexisting Alus, both leading to pronounced Alu clustering. It is known that the poly-A tail of an Alu may constitute a good insertion target for

¹ Departamento de Genética, Facultad de Ciencias, Universidad de Granada, E-18071 Granada, Spain.
E-mail: genmol@ugr.es, oliver@ugr.es

new insertions [5]. Analysing the densities of Alu/Alu insertions (i.e., insertions in the AT-rich linker, unpublished), we found that the insertion probability is positively correlated with the isochore GC content. We believe, nevertheless, that the preferential targeting on preexisting Alus in H isochores can only partially contribute to form the pronounced overall Alu density maximum in H2/H3 isochores, but that the density shift and the formation of the biased distribution over isochores is driven mainly by recombination [3].

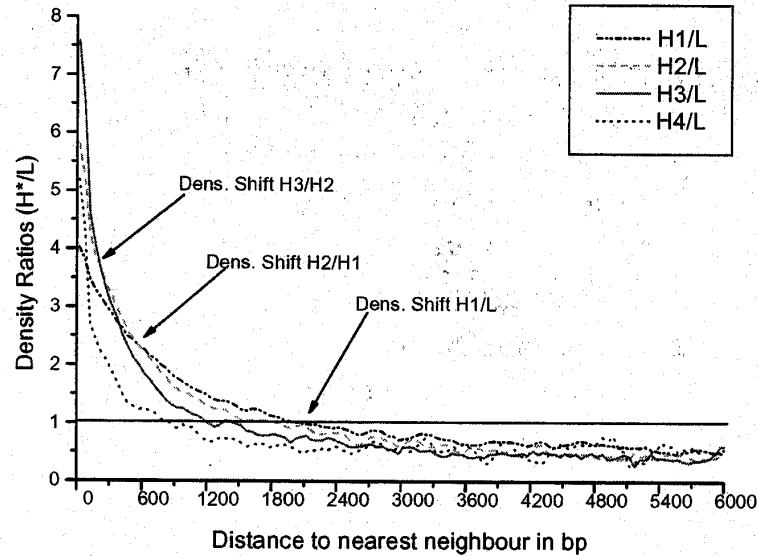


Figure 1: Alu density ratios as a function of DNN (Distance to Nearest Neighbour) and isochore membership are shown. It can be seen that very closely spaced Alus have a pronounced density maximum in the H3 isochore. With growing distance, the density maximum shifts gradually towards L isochores.

3 References and bibliography

- [1] Batzer M.A., Deininger P.L. 2002. Alu repeats and human genomic diversity. *Nature Reviews Genetics* 3:1-10.
- [2] Bernardi G. 2001. Misunderstandings about isochores. Part 1. *Gene* 276:3-13.
- [3] Hackenberg M, and Oliver J.L. 2004. The biased distribution of Alus in the human isochores might be driven by recombination. Submitted.
- [4] Oliver J.L, Carpena P, Hackenberg M, Bernaola-Galván P. IsoFinder: computational prediction of isochores in genome sequences, in preparation
- [5] Stenger, J.E. et al. 2001. Biased Distribution of Inverted and Direct *Alus* in the Human Genome: Implications for Insertion, Exclusion, and Genome Stability. *Genome Research* 11:12-27.