Isochore chromosome maps of the human genome

José L Oliver[a], , Pedro Carpena[b], Ramón Román-Roldán[c], Trinidad Mata-Balaguer[a], Andrés Mejías-Romero[a], Michael Hackenberg[a], Pedro Bernaola-Galván[b]

[a] Departamento de Genética, Instituto de Biotecnología, Universidad de Granada, Granada, Spain

[b] Departamento de Física Aplicada II, Universidad de Málaga, Málaga, Spain

[c] Departamento de Física Aplicada, Universidad de Granada, Málaga, Spain

Abstract

The human genome is a mosaic of isochores, which are long DNA segments ($\square$300 kbp) relatively homogeneous in G+C. Human isochores were first identified by density-gradient ultracentrifugation of bulk DNA, and differ in important features, e.g. genes are found predominantly in the GC-richest isochores. Here, we use a reliable segmentation method to partition the longest contigs in the human genome draft sequence into long homogeneous genome regions (LHGRs), thereby revealing the isochore structure of the human genome. The advantages of the isochore maps presented here are: (1) sequence heterogeneities at different scales are shown in the same plot; (2) pair-wise compositional differences between adjacent regions are all statistically significant; (3) isochore boundaries are accurately defined to single base pair resolution; and (4) both gradual and abrupt isochore boundaries are simultaneously revealed. Taking advantage of the wide sample of genome sequence analyzed, we investigate the correspondence between LHGRs and true human isochores revealed through DNA centrifugation. LHGRs show many of the typical isochore features, mainly size distribution, G+C range, and proportions of the isochore classes. The relative density of genes, Alu and long interspersed nuclear element repeats and the different types of single nucleotide polymorphisms on LHGRs also coincide with expectations in true isochores. Potential applications of isochore maps range from the improvement of gene-finding algorithms to the prediction of linkage disequilibrium levels in association studies between marker genes and complex traits. The coordinates for the LHGRs identified in all the contigs

longer than 2 Mb in the human genome sequence are available at the online resource on isochore mapping: http://bioinfo2.ugr.es/isochores.