This study is published by the prestigious magazine ***Genome Research***

Comparison of the human genome with genomes of other animals helps to identify new functional repetitive structures in proteins

These structures may be involved in neurodegenerative and developmental illnesses

Barcelona, 2 June 2010. A project executed by the research group on **Evolutionary Genomics of the Research Programme in Biomedical IT (GRIB) by IMIM and UPF** employed comparison of the human genome with the genomes of other species of vertebrates to verify that the repetitive **motifs** found in human proteins are important for the good operation of the body and which could correspond to the fraction of the genome called 'rubbish' (with no function).

Motifs are structures in which a single amino acid repeats several times in a row and are found in some 20% of human proteins. Until recently, it was thought that the majority of them were 'rubbish', but recent studies have proven that some have an important function. When they mutate, they cause developmental or neurodegenerative diseases, such as Huntington's disease. This makes it even more important to identify which motifs are functional, as they could be involved in diseases for which the genetic origin is still unknown.

The principal novelty of this study was to critically compare the level of conservation of tandem repeats that are found in human proteins –those that we don't know whether they are functional or 'rubbish'- with the level of conservation of the motif collection that we know are basically 'rubbish'. The latter were selected based on their localisation in the genome, outside the region encoded for proteins. To measure the evolutionary conservation, genomes from 11 species of vertebrates were used, including species relatively close to humans, such as rats and cows, and more distant species, such as fish. If a repetitive motif in a human protein is found to be much better conserved than typically for the 'rubbish' motifs, we can conclude that natural selection played a role in its preservation. The study calculated that some 90% of repetitive structures in human proteins that are conserved in other mammal species have been maintained by selection. Finding that a motif is well conserved in mammals is therefore enough to suspect that the motif is functional.

Mar Albà, ICREA researcher and coordinator of the research group of Evolutionary Genomics at GRIB (UPF-IMIM) explains that '*We have observed that if the motif is located in a zone that is often encoded for a protein, it is also found in many other species. Conversely, if the motif is found in a region that is not encoded for proteins, it is not functional. This implies that the majority of repetitive motifs in human proteins could have a function, as we observe an important symptom of natural selection*'.

Among the repetitive structures analysed, the study selected a group of 92 that, due to their length, and significant degree of conservation in different vertebrate species, have a very high probability of playing an important role in cells. These groups of motifs include two repetitions of alanine in the HOXD13 and PHOX2B genes. When they mutate and become longer, they cause polydactylism and congenital central hypoventilation syndrome, respectively. Two repetitions of histidine were also included in the FAM76B and DYRK1A proteins that, as seen in a previous project led by Susana de la Luna (ICREA researcher at the Centre of Genomic Regulation) are important for proper localisation of cell protein. Mutations in other motifs from the list of 92 may also lead to important alterations, which should be studied.

The work published in Genome Research was completely performed by members of the Evolutionary Genome Group, as part of the Research Programme in Biomedical IT of IMIM/UPF. Concretely, Loris

Mularoni participated, who is currently at Johns Hopkins University, Alice Ledda and Macarena Toll-Riera, doctoral students in the group, and M. Mar Albà, the group coordinator.

Reference article:

Mularoni, L., Ledda, A., Toll-Riera, M., Albà, MM (2010) Natural selection drives the accumulation of amino acid Tandem Repeats in human proteins. Genome Research, June 2010.

For further information, please contact the IMIM Communications Service: Rosa Manaut, Tel: 618 509 885 or Marta Calsina Tel: 933 160 680 or 638 720 000.