Press Release

EMBL-EBI

Evolution by 'copy-paste'

Study provides new insights into an ancient mechanism of mammalian evolution

Hinxton, 12 January 2012 - A team of geneticists and computational biologists in the UK today reveal how an ancient mechanism is involved in gene control and continues to drive genome evolution. The new study is published in the journal *Cell*.

To function properly, mammalian tissues require the protein CTCF, which has several key activities including the regulation of genes and interaction with proteins in the cell's nucleus to alter gene activity. CTCF acts by binding to DNA and plays a role in diseases such as HIV infection and cancer. However, very little is known about the origin of the DNA sequences that are bound by CTCF.

In this study, the researchers used samples from six mammals (human, macaque, mouse, rat, dog, and short-tailed opossum) to pinpoint where CTCF binds to each genome. They discovered around 5000 sites that are present in most cell types and tissues, and that have not changed over hundreds of millions of years of mammalian evolution. Because these CTCF binding sites are conserved throughout evolution, the researchers believe that many might play an important role in gene regulation.

The team found an even larger number of locations where CTCF binds DNA in only one lineage or a single species. These additional sites represent a signature of important evolutionary changes since our last common ancestor – legacies, in some cases, of the evolutionary path to humans. These newer CTCF sites are embedded inside virus-like stretches of DNA called 'retro-transposons'. Retro-transposons use a copy-paste mechanism to spread copies of themselves throughout the genome.

"We developed a new, integrated model of CTCF evolution, which explains the origin of these 5000 highly conserved CTCF binding events in mammals," said Paul Flicek of the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) and the Wellcome Trust Sanger Institute.

"Taken together, our findings provide fascinating insight into an ancient mechanism of evolution that is still actively changing our genome."

"CTCF is a key regulator involved in chromatin and gene expression remodelling, both of which are perturbed in the development of cancer. The gene expression and chromatin changes in cancer have also recently been relied on to predict the outcome of specific cancer treatments, which is why it is so important to have a detailed understanding of how particular parts of the genome are resistant or plastic to changes," said Duncan Odom of Cancer Research UK and the Wellcome Trust Sanger Institute.

The retro-transposon's copy-and-paste behaviour has long been considered totally self-serving. However, the study showed that when a retro-transposon containing a CTCF-binding sequence spreads around a mammal's genome, it can deposit functional CTCF binding sites in novel locations, altering the activity of distant genes.

"We looked at six mammalian species representing primates, marsupials, rodents and carnivores, and discovered a simple mechanism that they all use to remodel their DNA," explained Petra Schwalie of EMBL-EBI. "We also found that our distant ancestors also experienced the same complicated relationship between CTCF and retro-transposons."

Using molecular palaeontology techniques, the researchers were able to identify fossil traces of older retro-transposon expansions in the DNA around the shared CTCF binding locations, and showed that this process has been active for hundreds of millions of years.

The study combined the efforts of researchers at EMBL-EBI, the Wellcome Trust Sanger Institute, Cancer Research UK, and the Cambridge Hepatobiliary Service at Addenbrooke's Hospital in Cambridge, UK. ●

Source Article

Schmidt D., Schwalie P., Wilson M.D., et al. (2012) Waves of repeat-driven CTCF binding expansions have shaped mammalian genomes. Cell (in press). Published 12 January; DOI: 10.1016/j.cell.2011.11.058

Contact the EMBL Press Office: contactpress@ebi.ac.uk

Mary Todd-Bergman, Outreach Programme Project Leader, Hinxton, UK, Tel: +44 1223 494 665, www.ebi.ac.uk/OTT

Katrina Pavelin, EMBL-EBI Scientific Outreach Officer - Hinxton, UK, Tel: +44 1223 494 452, www.ebi.ac.uk/OTT

Sonia Furtado, EMBL Press Officer, Heidelberg, Germany, Tel: +49 6221 387 8263, www.embl.org

NOTES TO EDITORS

About EMBL-EBI

The European Bioinformatics Institute (EBI) is part of the European Molecular Biology Laboratory (EMBL) and is located on the Wellcome Trust Genome Campus in Hinxton near Cambridge, UK. The EBI grew out of EMBL's pioneering work in providing public biological databases to the research community. It hosts some of the world's most important collections of biological data, including DNA sequences (ENA), protein sequences (UniProt), animal genomes (Ensembl), 3D structures (the Protein Databank in Europe), data from gene expression experiments (ArrayExpress), protein-protein interactions (IntAct) and pathway information (Reactome). EMBL-EBI hosts several research groups and its scientists continually develop new tools for the biocomputing community. www.ebi.ac.uk

About EMBL

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 20 member states (Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom) and associate member state Australia. Research at EMBL is conducted by approximately 85 independent groups covering the spectrum of molecular biology. The Laboratory has five units: the main Laboratory in Heidelberg, and outstations in Grenoble, Hamburg, Hinxton and Monterotondo (near Rome). The cornerstones of EMBL's mission are: to perform basic research in molecular biology; to train scientists, students and visitors at all levels; to offer vital services to scientists in the member states; to develop new instruments and methods in the life sciences and to actively engage in technology transfer activities. Around 190 students are enrolled in EMBL's International PhD programme. Additionally, the Laboratory offers a platform for dialogue with the general public through science communication activities such as lecture series, visitor programmes and the dissemination of scientific achievements. www.embl.org

About the Wellcome Trust Sanger Institute

The Wellcome Trust Sanger Institute is one of the world's leading genome centres. Through its ability to conduct research at scale, it is able to engage in bold and long-term exploratory projects that are designed to influence and empower medical science globally. Institute research findings, generated through its own research programmes and through its leading role in international consortia, are being used to develop new diagnostics and treatments for human disease. www.sanger.ac.uk

About Cancer Research UK

Cancer Research UK is the world's leading cancer charity, dedicated to saving lives through research. Our ground-breaking work into the prevention, diagnosis and treatment of cancer has helped save millions of lives. Our work is funded entirely by the public. We have been at the heart of the progress that has already seen survival rates double in the past 40 years. We support research into all aspects of cancer through the work of over 4000 scientists, doctors and nurses. Together with our partners and supporters, our vision is to beat cancer. www.cancerresearchuk.org