Comparative analysis of mouse and human genomic sequence: implications for software development.

Mallon A-M[1], Denny, P[1], Strivens, MA[1], Bate, R[1], Miller, W[2], Botcherby, M[3], Rhodes, M[3], Brass, A[4], Dangel A[5], Cunningham D[5], Herman GE[5], Brown, S.D.M[1],

1: MRC UK Mouse Genome Centre and Mammalian Genetics Unit, Harwell, Oxon., UK.

2: Department of Computer Science, The Pennsylvania State University, University Park PA 16802, USA.

3: MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, UK

4: School of Biological Sciences, University of Manchester, Manchester, UK.

5: Department of Pediatrics and Children's Hospital Research Foundation, Ohio State University, Columbus, Ohio 43205, USA.

E-mail address: A.Mallon@har.mrc.ac.uk

An increase in genomic sequence from syntenic regions of mouse and human genomes in recent years has enabled us to study the benefits of comparative analysis in identifying putative exons, genes and regulatory regions. Pairwise gapped alignments of homologous genomic sequence from these two species yields a complicated pattern of conservation. A comprehensive study of this pattern in seven previously annotated genomic regions [(1-7)] from mouse and man enabled us to define a simple filtering criteria for the identification of coding exons based on the length and percentage identity of each conserved element. The incorporation of this prediction system with existing gene prediction techniques has enabled us to show an increase in 'predictive power'. This filtering criteria has been utilised within the *Bpa/Str* region[7] of mouse and human X chromosome to identify novel coding exons that have been subsequently tested for expression by RT-PCR.

At present it is unclear how data extracted from the comparative analysis of two species can be stored and used beneficially. A large number of the conserved elements identified above cannot at present be assigned a function, it is therefore proposed to store them in a public database that can be searched against. Currently we are developing software to use this filtering technique and combine other tests such as the identification of ORF for use in identifying putative functional elements.

1: *mnd2* region of chromosome 2p13 (Jang, W. *et.al.* (1999) *Genome Res.* 9, 53-61)

2: Gene rich cluster on mouse chromosome 6 (Ansari-Lari, M.A. *et.al.* (1998) *Genome Res.* 8, 29-40)

3: BTK locus (Oeltjen, J.C., *et.al.* (1997) *Genome Res.* 7, 315-29)

4: ADA gene region (Brickner, A.G., *et.al.* (1999) *Mamm Genome* 10, 95-101)

5: T-cell receptor locus (Koop,B.F. & Hood L. (1994) *Nature Genetics* 7, 48-53)

6: ERCC2 gene regions (Lamerdin, J.E., *et.al.* (1996) *Genomics* 34, 399-409)

7: *Bpa/Str* region of mouse X chromosome