with experimental laboratories. The results are widely disseminated via the internet, peer reviewed scientific publications and personal interactions. Our most recent research concentrates on mechanisms of retroposon integration in mammals (Jurka, J., PNAS, in press; Jurka, J and Klonowski, P., J. Mol. Evol. 43:685-689).

We continue to develop reference collections of mammalian repeats which became a worldwide resource for annotation and study of newly sequenced DNA. The reference collections are being revised annually as part of a larger database of repetitive DNA, called Repbase. The recent influx of sequence data to public databases created an unprecedented need for automatic annotation of known repetitive elements. We have designed and implemented a program for identification and elimination of repetitive DNA known as CENSOR.

Reference collections of mammalian repeats and the CENSOR program are available electronically (via anonymous ftp to ncbi.nih.gov; directory repository/repbase). CENSOR can also be run via electronic mail (mail "help" message to censor@charon.lpi.org).

DOE Grant No. DE-FG03-95ER62139.

*TRRD, GERD and COMPEL: Databases on Gene-Expression Regulation as a Tool for Analysis of Functional Genomic Sequences

A.E. Kel, O.A. Podkolodnaya, O.V. Kel, A.G. Romaschenko, E. Wingender,[1] G.C. Overton,[2] and **N.A. Kolchanov**
Institute of Cytology and Genetics; Novosibirsk, Russia
Kolchanov: +7-3832/353-335, Fax: -336 or /356-558,
kol@benpc.bionet.nsc.ru
http://transfac.gbf-braunschweig.de
[1]Gesellschaft für Biotechnologische Forschung; Braunschweig, Germany
[2]Department of Genetics; University of Pennsylvania School of Medicine; Philadelphia, PA 19104-6145

The database on transcription regulatory regions in eukaryotic genomes (TRRD) has been developed [1] (http://www.bionet.nsk.su/TRRD.html; ftp://ftp.bionet.nsk.su/pub/trrd/). The main principle of data representation in TRRD is modular structure and hierarchy of transcription regulatory regions. TRRD entry corresponds to a gene as entire unit. Information on gene regulation is provided (cell-cycle and cell type specificity, developmental stage-specificity, influence of various molecular signals on gene expression). TRRD database contains information about structural organization of gene transcription regulatory region. TRRD contains description of known promoters and enhancers in 5', 3' regions and in introns. Description of binding sites for transcription factors includes nucleotide sequence and precise location, name of factors that bind to the site, experimental evidences for the binding site revealing. We provide cross-references to TRANSFAC database [2] for both sites and factors as well as for genes. TRRD 3.3 release includes 340 vertebrate genes.

The Gene Expression Regulation Database (GERD) collects information on features of genes expression as well as information about gene transcription regulation. The current release of GERD contains 75 entries with information on expression regulation of genes expressed in hematopoietic tissues in the course of ontogenesis and blood cells differentiation. COMPEL database contains information about composite elements which are functional units essential for highly specific transcription regulation [3]. Direct interactions between transcription factors binding to their target sites within composite elements result in convergence of different signal transduction pathways. Nucleotide sequences and positions of composite elements, binding factors and types of their DNA binding domains, experimental evidence confirming synergistic or antagonistic action of factors are registered in COMPEL. Cross-references to TRANSFAC factors table are given. TRRD and COMPEL are provided by cross-references to each other. COMPEL 2.1 release includes 140 composite elements.

We have developed a software for analysis of transcription regulatory region structure. The CompSearch program is based on oligonucleotide weight matrix method. To collect sets of binding sites for the matrixes construction we have used TRANSFAC and TRRD databases. The CompSearch program takes into account the fine structure of experimentally confirmed NFATp/AP-1 composite elements collected in COMPEL (distances between binding sites in composite elements, their mutual orientation). By means of the program we have found potential composite elements of NFATp/AP-1 type in the regulatory regions of various cytokine genes. Analysis of composite elements could be the first approach to reveal specific patterns of transcription signals encoding regulatory potential of eukaryotic promoters.

References

1. Kel O.V., Romaschenko A.G., Kel A.E., Naumochkin A.N., Kolchanov N.A. Proceedings of the 28th Annual Hawaii International Conference on System Sciences [HICSS]. (1995), v.5, Biotechnology Computing, IEE Computer Society Press, Los Alamos, California, p. 42-51.
2. Wingender E., Dietze P., Karas H., and Knuppel R. TRANSFAC: a database on transcription factors and their DNA binding sites (1996). Nucl. Acids Res., 1996, v. 24, pp. 238-241.
3. Kel O.V., A.G. Romaschenko, A.E. Kel, E. Wingender, N.A. Kolchanov. A compilation of composite regulatory elements affecting gene transcription in vertebrates (1995). Nucl. Acids Res., v. 23, pp. 4097-4103.

(abstract continued)