Systematic Analyses on Mouse Genome Encyclopedia

Toshinori Endo [1]
tendo@rtc.riken.go.jp
Itaru Yamanaka [1]
yitaru@rtc.riken.go.jp
Hideaki Konno [12]
hkonno@rtc.riken.go.jp
Yoshifumi Fukunishi [12]
fukunisi@rtc.riken.go.jp
Jun Kawai [1]
kawai@rtc.riken.go.jp
Harukazu Suzuki [1]
harukazu@rtc.riken.go.jp
Yasuhiro Ozawa [1]
yozawa@rtc.riken.go.jp
Kazuhiro Shibata [1]
shibata@rtc.riken.go.jp
Masayasu Yoshino [1]
yoshino@rtc.riken.go.jp
Masayoshi Ito [1]
maitoh@rtc.riken.go.jp
Piero Carninci [1]
carninci@rtc.riken.go.jp
Yasushi Okazaki [1]
fukunisi@rtc.riken.go.jp
Yoshihide Hayashizaki [1]
yosihide@rtc.riken.go.jp

[1] Laboratory for Exploration Research, Genomic Sciences Center, The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0053, Japan
[2] CREST, Japan Science and Technology Corporation

1 Introduction

Our Mouse Encyclopedia Project, which performs exhaustive collection and sequencing of all mRNA species expressed in mouse, successfully collected more than 70,000 different kind of cDNA clones classifiable by 3'-end sequence tags. By Using those clones, we are completing sequence of the isolated cDNA clones in concurrent manner, which piled up to over 10,000 sequences. This number tells that our project reached to the new realm where other genome project could not have reached in terms of the total number of genes sequenced. In contract to the genomic sequencing, our strategy choosing cDNA as the sequencing material brought us big advantages that the amount of sequencing required to obtain all gene information requires much less effort in terms of the number sequence runs, and each gene information are ready for application, upon completion of individual sequence with out further artificial processing of data, i.e. prediction of gene region and exon structure. By taking this advantage, we have started several projects such as analyses of expression profiles and protein-protein interaction. In the current study, we report results of one of such application, produced by bioinformatic approaches.

2 Sequence data

The sample sequence data used in this study were cloned and sequence in full by our Mouse Encyclopedia Project. The sequence information will be published elsewhere.

3 Nucleotide and amino acid sequence database

For the aim of function assignment and prediction, the following databases are used as the database for homology search: GenBank non-redundant nucleotide and amino acid sequence databases, TIGR function-classified human nucleotide sequence database, and Genome sequences of *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Rattus norvegicus*. Homology search was performed by *blast 2.0.10* [1] on AlphaServer ES40 running Tru64 Unix operating system. The criteria to search for identical/counter-part genes and family gene were set to the E-values as 1e-50 and 1e-5, respectively.

4 Chromosomal mapping of clones *in silico*

Information for chromosomal location of genes and sequence fragments was obtained from DDBJ/ EMBL/ GenBank international database for nucleotide sequences, radiation hybrid panel databases and genetic map databases. Mapping was performed by blast homology search at the criterion of E-value being 1e-50. In addition, by linking the mapping information with Online Mendelian Inheritance in Man (OMIM), the candidate clones which many be related to human diseases were shown.

5 Expression profiles

Some set of library cDNA libraries are created without subtraction process which removes precloned sequences. The frequency pattern of clones observed in the non-subtracted library, can approximate the expression profile in each library. In addition, by comparing the profile of subtracted library with that from non-subtracted one, the efficiency of subtraction was evaluated.

6 Evolutionary studies

The sequences classified by homology were further utilized for evolutionary studies, such as phylogenetic analyses, and search for the intra-gene region where positive natural selection may be operating [2]. In addition, the clones for duplicated genes which scattered on chromosomes were identified for the study of chromosome evolution [3].

Acknowledgments

This study has been supported by Special Coordination Funds and a Research Grant for the Genome Exploration Research Project from the Science and Technology Agency of the Japanese Government, CREST (Core Research for Evolutionary Science and Technology) of Japan Science and Technology Corporation (JST), a Grant-in-Aid for Scientific Research on Priority Areas and Human Genome Program from the Japan Ministry of Education and Culture to Y.H., and Grant-in-Aid for Scientific Research by Japan Ministry of Education and Culture to T.E.

References

[1] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 25:3389–3402, 1997.

[2] Endo, T., Ikeo, K., and Gojobori, T., Large-scale search for genes on which positive selection may operate, *Mol. Biol. Evol.*, 13:685–690, 1996.

[3] Endo, T., Imanishi, T., and Gojobori, T., and Inoko, H. Evolutionary significance of intra-genome duplications on human chromosomes, *Gene*, 205:19–27, 1997.