Establishment of Transcription Factor Database and Human Mutation Database

H. Mizushima [1] hmizushi@gan.ncc.go.jp

K. Hayashi [2] khayashi@gen.kyushu-u.ac.jp

[1] Cancer Information Division, National Cancer Center Research Institute
5-1-1 Tsukiji Chuo-ku, Tokyo 104, Japan

[2] Division of Genome Analysis, Institute of Genetic Information
Kyushu University
3-1-1 Maidasi, Higashi-ku, Fukuoka-shi 812 Japan

Abstract

We started two databases 'TFDB' and 'HMDB'. Transcription Factor DataBase (TFD) was originally maintained by D.Gohsh at National Center for Biotechnology Informaion(NCBI), National Library of Medicine, National Institutes of Health. As NCBI stopped its maintenance since last year, we started a new database, TFDB, to maintain some parts of the database mainly focusing to the DNA binding sequence data. HMDB(Human Mutation DataBase) is a new database collecting Information about mutation in the human genome. As both databases are started very recently, they are still at preliminary stages. We will continue to put more informaions in the future.

1 Introduction

TFD [1, 2, 3, 4] maintained by D.Gohsh at NCBI was a database with some subparts. We made a program using 'sites' table to search potential DNA binding sites in a promoter region of a DNA sequence [5]. Some other researchers and companies also made similar programs which for easier TFD search. For molecular biologists working at the bench, especially those who are analyzing transcription regulation mechanizms, these systems are important and useful. However, NCBI stopped its maintenance last year, and updating has not been done. So, we started to maintain a new database TFDB using TFD(sites) data as a starting point. HMDB is a database collecting Information about mutations in human genome. Although GenBank and GDB contains some of the mutation data, searching them is not always easy for researchers working at the bench. As a starting subset of the HMDB, we will focus on mutations detected by Single Strand Conformation Polymorphism (SSCP) analysis, which is currently the most widely used technique for mutation detection in the clinical field [6, 7].

[1]水島　洋：国立がんセンター研究所がん情報研究部，〒104 中央区築地 5-1-1
[2]林　健志：九州大学遺伝情報実験施設ゲノム解析分野，〒812 福岡市東区馬出 3-1-1

2 Previous Work

TFD(Rel.7.2 at Aug.93) includes Tables with following numbers of records below.

Table	Records
clones	2106
domains	1016
factors	523
polypeptides	1626
sites	2155
methods	38
n_pointers	2876
references	7391
x_pointers	5757

3 Data

All data in TFD sites table with consensus sequence informaion was converted into TFDB. Some new records are added to them. This time, mutaions detected by PCR-SSCP method [6, 7] alone will be included in HMDB, with information on primer sequences for PCR, conditions of electrophoresis etc. The database structre and the record numbers for both databases will be announced at the meeting. The data will be opened for use near future with www/gopher systems at the National Cancer Center Server.

4 Summary

We started to maintain two databases 'TFDB (Transcription Factor DataBase)' and 'HMDB (Human Mutation DataBase)'. As each of the database is started very recently, it is still a preliminary database. We will continue to put more informaions in future.

References

[1] Ghosh, D. (1990) *Nucleic Acids Research* 18, 1749-1756.

[2] Ghosh, D. (1991) *Trends in Biochemical Sciences* 16, 455-457.

[3] Ghosh, D. (1992) *Nucleic Acids Research* 20S, 2091-2093.

[4] Ghosh, D. (1993) *Nucleic Acids Research* 21S, 3117-3118.

[5] Mizushima, H. (1992) *Proceedings of 15th Japanese Molecular Biology Meeting.*

[6] Hayashi, K. (1992) *Genet.Anal.Tech. Appl.* 9, 73-79.

[7] Hayashi, K. and Yandell, D.W. (1993) *Human Mutation* 2, 338-346.