Bioinformatics for Oncogenomic Target Identification

Masaru Katoh
mkatoh@ncc.go.jp
Genetics and Cell Biology Section, National Cancer Center Research Institute, Tokyo 104-0045, Japan

Keywords: human genome, amplicon, array CGH, LOH, reverse genetics, LCM microarray

1 Introduction

Activation of proto-oncogenes and inactivation of tumor suppressor genes (TSGs) occur during multistage carcinogenesis. Proto-oncogenes are activated by gene amplification, point mutation and chromosomal translocation, while TSGs are inactivated by promoter CpG hypermethylation, point mutation, chromosomal translocation, and deletion. Array CGH combined with mRNA microarray analysis is applied for genome-wide screening of proto-oncogenes as well as TSGs [2]. Microarray analysis for lase-captured microdissection (LCM) sample is applied for screening of novel tumor markers specifically expressed in tumor cells [1]. Because huge amounts of data concerning genome sequences, expression profile, copy-number changes of human chromosomal regions in tumors are available in the post-genomic era, we applied bioinformatics for oncogenomic target identification.

2 Methods

Human genome sequences, uncharacterized cDNAs, and expressed sequence tags (ESTs) homologous to gene fragments or microsatellite markers were searched for with the Blastn program. Domain structures of novel gene products were determined by using RPS-BLAST, Blastp, and Genetyx programs.

3 Results

3.1 Structural Analyses of Amplified Regions (amplicons), and Identification of Novel Amplified Genes

We determined the structure of *FGFR2* amplicon. *FGFR2* and *WDR11* genes were located at human chromosome 10q26 in the tail-to-tail manner with an interval about 560 kb. *WDR11* was excluded from *FGFR2* amplicon in KATO-III, OCUM-2M, and HSC39 cells due to breakage-fusion-bridge (BFB) process during gene amplification [5].

We next determined the structures of amplicons at human chromosome 11q13.3, and 17q12 , consisting of multiple amplified genes. The 17q12 amplicon in human gastric cancer was found to consist of *PPP1R1B, STARD3, TCAP, PNMT, MGC9753, ERBB2, MGC14832*, and *GRB7* genes [6]. *MGC9753* gene encoded evolutionary conserved seven-transmembrane protein with extracellular cysteine-rich domain. MGC9753 protein showed 90.6% total-amino-acid identity with CAB2 aberrant protein, which lack the third transmembrane-domain of MGC9753 due to cloning artifact or sequencing error. At least *PPP1R1B, STARD3, MGC9753, ERBB2,* and *GRB7* genes were overexpressed due to gene amplification.

We also identified novel amplified genes, including *FNBP2*, *TIPARP,* and *DDX40*. *FNBP2* gene within 1q32 amplicon encoded Formin-binding protein with RhoGAP domain. *TIPARP* gene within 3q25 amplicon encoded TCDD-induced PARP with TPH and WW domains [8]. *DDX40* gene within 17q23 amplicon encoded RNA helicase with DEAD-box domain.

3.2 Structural Analyses of Deleted Regions, and Identification of Novel Candidate TSGs

Several TSGs are located within human chromosome 11q23, because several regions within human chromosome 11q23 are frequently deleted in a variety of human tumors. We have identified and characterized *KIAA1391* and *MTHDIX* genes within 11q23.1 commonly deleted region (CDR) of breast cancer, *TPARM* genes within 11q23.1-q23.2 CDR of malignant melanoma, *LL5A, BCL9L, RNF26* and *MFRP* genes within 11q23.3 CDR of neuroblastoma ([7], and references therein).

3.3 Gene Hunting based on LCM Microarray Data

IMSUT method using large amounts of LCM samples for microarray analyses gives rise to more reliable data than TALPAT method using small amounts of LCM samples amplified with adopter-ligated PCR. By using ESTs reported in the LCM microarray paper from IMSUT [1], we have identified and characterized the following novel genes, specifically regulated in gastric cancer: *IGSF11, CLDN23*, and *MGC29506* [4].

4 Discussion

Cancer research is the process to extract facts on cancer out of private data or public data to establish knowledge for prevention and treatment of cancer. During 1998∼2002, we engaged in the experimental science for molecular cloning, expression analyses and functional analyses of 35 novel human genes implicated in the WNT signaling pathway ([3], and references therein). In 2003, we engaged in the information science, and identified and characterized more than 40 novel human genes as the oncogenomic targets ([4, 5, 6, 8, 7], and references therein). More than 75 novel human genes identified in my laboratory as the outcome of "Medical Genome Project (MGP)" should be applied for early diagnosis, prognostics, prevention, and treatment of human cancer.

References

[1] Hasegawa, S., *et al.*, Genome-wide analysis of gene expression in intestinal-type gastric cancers using a complementary DNA microarray representing 23,040 genes, *Cancer Res.*, 62:7012–7017, 2002.

[2] Hyman, E., *et al.*, Impact of DNA amplification on gene expression patterns in breast cancer, *Cancer Res.*, 62:6240–6245, 2002.

[3] Katoh, M., Paradigm shift in gene-finding method: from bench-top approach to desk-top approach (Review), *Int. J. Mol. Med.*, 10:677–682, 2002.

[4] Katoh, M. and Katoh, M., *CLDN23* gene, frequently down-regulated in intestinal-type gastric cancer, is a novel member of *CLAUDIN* gene family, *Int. J. Mol. Med.*, 11:683–689, 2003.

[5] Katoh, M. and Katoh, M., *FGFR2* and *WDR11* are neighboring oncogene and tumor suppressor gene located on human chromosome 10q26, *Int. J. Oncol.*, 22:1155–1159, 2003.

[6] Katoh, M. and Katoh, M., *MGC9753* gene, located within *PPP1R1B - STARD3 - ERBB2 - GRB7* amplicon on human chromosome 17q12, encodes seven-transmembrane receptor with extracellular six-cystein domain, *Int. J. Oncol.*, 22:1369–1374, 2003.

[7] Katoh, M. and Katoh, M., Identification and characterization of human *LL5A* gene and mouse *Ll5a* gene *in silico*, *Int. J. Oncol.*, 23:1477–1483, 2003.

[8] Katoh, M. and Katoh, M., Identification and characterization of human *TIPARP* gene within the *CCNL* amplicon at human chromosome 3q25.31, *Int. J. Oncol.*, 23:541–547, 2003.