사람 유전자 database에 기록된 타액선 유전자 검색과 타액선 subtracted cDNA library에서 얻은 새로운 유전자의 비교 확인

강릉대학교 치과대학 구강병리학 교실[1], 구강미생물 및 면역학 교실[2]
김연숙[1], 최영님[2], 이석근[1]

ABSTRACT

Genes of salivary gland in human gene database and cloning and identification of novel genes from subtracted cDNA library of human salivary gland

[1]Department of Oral Pathology, [2]Department of Oral Microbiology and Immunology, College of Dentistry, Kangnung National University
Yeon Sook Kim[1], Young Nim Choi[2], Suk Keun Lee[1]

This study is aimed to discover new genes from subtracted human salivary gland cDNA library and to evaluate the new genes by the database search using various website tools. We have performed molecular cloning to get new cDNA clones and followed by Genbank search and *in situ* hybridization using human salivary gland tissue sections. Up to date, world wide Genbank database showed 25 unique unknown genes and 366 non-unique unknown genes of human salivary gland, while we found 26 novel genes which were non-redundant in the Genbank search. Among the 26 novel genes 8 clones were intensely expressed in the acinar cells of human salivary gland, 7 clones were intensely expressed in the ductal cells, 6 clones were diffusely expressed both in the acinar cells and ductal cells, and 2 clones were intensely expressed in the salivary mesenchymal cells. Meanwhile, 3 clones were identified their loci in the human genomic DNA through the Genbank search. These data suggest that the 26 clones obtained in this study are novel ones which are not registered at all in the world wide Genbank, and we thought, so far, there may exist many novel genes of human salivary gland remained to be elucidated for their sequences and functions.

Key words: Molecular Cloning, Human, Salivary Gland, Novel Genes, Genbank Search

※ 본 연구는 한국과학재단 목적기초연구지원사업 핵심전문연구(과제번호 981-0709-074-2)의 지원으로 이루어졌음.

서 론

1970년대 후반부터 비약적으로 발전한 염기서열 분석 및 결정방법에 의해 인간을 비롯하여 많은 종의 DNA database가 확립되었으며, 1982년 유럽에서는 EC의 지원으로 독일의 하이델베르그에 있는 European Molecular Biology Laboratory (EMBL)에 세계 최초의 nucleotide sequence database인 EMBL Nucleotide sequence Data Library를 설립하면서 data library 서비스를 시작했고, 1992년 국제사회의 변화에 따라 European Bioinformatics Institute (EBI : http://www.ebi.ac.uk/)를 영국에 설립하면서 풍부한 database를 구축하고 연구자들에게 정보를 개방하고 있다.

미국에서는 1982년 NIH의 지원으로 Los Alamos 국립연구소에 GenBank를 설치했으며, 1988년 NIH에 있는 National Library of Medicine (NLM)의 division으로 National Center for Biotechnology Information (NCBI : http://www.bcbi.nlm.nih.gov/)를 설립하여 다양한 정보를 제공할 수 있게 되었다. NIH genetic sequence database는 2001년 4월 현재, 11,546,000 종류의 유전자에서 약 12,419,000,000 개의 염기서열을 보유하고 있으며, 2개월마다 자료를 갱신하고 있다. 일본에서도 1986년 문부성의 지원으로 National Institute of Genetics (NIG)에 DNA databank(DDBJ : http://www.ddbj.nig.ac.jp/)가 설립되었다.

현재 European Molecular Biology Laboratory (EMBL), GenBank at NCBI, DNA DataBank of Japan (DDBJ)는 International Nucleotide Sequence Database Collaboration을 구성하여 정보를 교환하고 있으며, Taxonomy Project, Feature Table, db_xref Qualifier, Country Qualifier와 같은 project를 협력하여 운영하고 있다. 최근에는 많은 기업에서도 DNA database를 구축하여 유료 website를 운영하고 있다. 또한 컴퓨터 시스템에 의한 다양한 기능을 통하여, 축적된 정보를 다각적으로 활용할 수 있게 되었다. 이 보고서에서는 미국 NCBI에서 제공하는 Cancer Genome Anatomy Project (CGAP) database와 tool을 주로 사용하여 타액선의 유전자를 검색하였으며, 본 실험실에서 제작한 사람 타액선의 subtracted cDNA library에서 발견한 새로운 유전자와 human gene database를 통한 검색결과를 비교 분석하였다. (2001.9.1. 현재 기준 data)

실제로 생체 조직에서 얻은 cDNA library에서 새로운 clone을 얻는 방식의 cloning이 잘 알려져 있는데 최근에는 사람의 genomic DNA가 대부분 밝혀지고 있으므로 앞으로는 Genbank database의 활용이 우선적으로 요구되고 있다. 따라서 본 연구에서는 사람의 타액선 조직의 molecular cloning과 더불어 Genbank database 검색의 방법을 조사하여 앞으로 신속하고 효과적인 cDNA cloning의 방법을 개발하고 이 방법의 유의성을 평가하고자 한다.

실험재료 및 방법

1. CGAP Data 검색 도구

1) 유전자를 찾기 위한 도구

Gene Finder (http://cgap.nci.nih.gov/Genes/GeneFinder) : NCBI와 NCI databases를 통하여 유전자를 검색한다.

GO Browser (http://cgap.nci.nih.gov/Genes/GOBroswer) : 사람과 생쥐의 유전자를 단백질 분자기능, 생체 내 대사 과정 및 세포 성분 분석을 통하여 검색한다.

Nucleotide BLAST (http://cgap.nci.nih.gov/Genes/SeqFinder) : Nucleotide sequence의 유사성을 검색한다.

2) cDNA Library를 찾기 위한 도구 :

Library Finder (http://cgap.nci.nih.gov/Tissues/LibraryFinder) : EST와 SAGE libraries를 통하여 검색한다.

3) 유전자의 발현을 관찰하기 위한 도구

Gene Library Sorter (GLS) (http://cgap.nci.

nih.gov/Tissues/LibrarySummarizer) : Single cDNA library나 library group에서 unique 또는 non-unique gene을 검색한다.

cDNA XProfiler (http://cgap.nci.nih.gov/Tissues/xProfiler) : 두 개의 cDNA libraries 사이에 유전자 발현을 비교한다.

Digital Gene Expression Displayer(DGED) (http://cgap.nci.nih.gov/Tissues/GXS) : 두 개의 libraries 사이에 유전자 발현 양상의 유의성을 판별한다.

SAGEmap XProfiler (http://www.ncbi.nlm.nih.gov/SAGE/sagexpsetup.cgi) : SAGE (Serial Analysis of Gene Expression) libraries 사이에 유전자 발현을 비교한다.

SAGEmap vNorthern (http://www.ncbi.nlm.nih.gov/SAGE/sagevn.cgi) : SAGE tags를 확인하고 전산화된 유전자 발현을 찾는다.

4) 염색체 검사를 위한 도구

Mitelman Databaseer (http://cgap.nci.nih.gov/Chromosomes/Mitelman) : CGAP가 개발한 사람 암세포의 chromosomal breakpoints 유전자 지도이다.

Recurrent Aberrations (http://cgap.nci.nih.gov/Chromosomes/RecurrentAberrations) : CGAP가 개발한 암세포의 Recurrent Aberrations 검색 도구이다.

FISH-mapped BACs (http://cgap.nci.nih.gov/Chromosomes/CCAP_BAC_Clones) : 사람 유전자의 BAC clones에 대한 염색체 지도를 제공한다.

Expression-Based SNP Imagemaps (http://lpg.nci.nih.gov/html-snp/ts.html) : 암과 염색체 종류에 따라 SNPs를 검색한다.

Genetic and Physical SNP Maps (http://lpg.nci.nih.gov/html-snp/imagemaps.html) : SNPs를 확인, 평가 및 예상하기 위한 유전자 정보를 제공한다.

GenMap99 (http://www.ncbi.nlm.nih.gov/genemap99/) : International Radiation Hybrid Mapping Consortiu을 통하여 35,000bp 보다 큰 사람 유전자 표식자를 검색한다.

5) Single nucleotide polymorphism (SNP)을 찾기 위한 도구

Expression-Based SNP Imagemaps (http://lpg.nci.nih.gov/html-snp/ts.html)

Genetic and Physical SNP Maps (http://lpg.nci.nih.gov/html-snp/imagemaps.html) : SNPs를 확인, 평가 및 예상하기 위한 유전자 정보를 제공한다.

2. Subtracted cDNA library 제작

성인의 타액선에서 얻은 약 1g의 조직을 통법에 따라서 guanidine thiocyanate 용액 (4.0 M GTC , 0.1 M Tris-Cl (pH 7.5), 1% β-mercaptoethanol)에 넣고 homogenize시킨 후 phenol-chloroform-isoamyl alcohol (25:24:1)을 넣어 섞은 후 15,000rpm에서 20분간 centrifuge하여 RNA를 일차적으로 분리하였다.

그리고 cesium chloride-EDTA 용액 (4.78 M CsCl2 in 100㎖ DEPC water, 0.01 M EDTA (pH 7.5))을 이용한 ultra centrifuge (32,000rpm, 24 hours) 방법으로 이차적으로 순수한 RNA를 얻었다. 이렇게 얻은 total RNA를 Oligotex (Qiagen, CA, USA)를 이용해서 mRNA를 검출하였다. 순수 분리된 mRNA 5㎍을 이용하였는데, MMLV reverse transcriptase (Stratagene, CA, USA)로 first strand cDNA를 만들었으며 RNase H(1.5 U/㎕)와 DNA polymerase I (9.0 U/㎕)을 이용해서 second strand cDNA를 제작하였다. 한편 사람의 타액선에서 얻은 cDNA를 subtraction하기 위해 RHEK (Rhesus Human Epithelial Keratocyte, NIH, USA) 세포주를 충분히 배양해서 같은 방법으로 mRNA를 추출하고 double strand cDNA를 제작하였다. RHEK 세포주의 cDNA에 Klenow (NewEngland Biolabs, Hertfor-dshire, UK) 효소를 이용해서 dUTP-digoxigenin을 labeling한 DNA probe를 제작한 후 사람의 타액선 조직에서 얻은 cDNA와 반응시켜서 hybridization 시켰다. 그리고 biotin이 붙어있는 anti-digoxigenin 항체 (Boehringer Mannheim, GmbH,

Germany)로 반응시킨 후에 magnetic particle이 붙어있는 streptoavidin (Boehringer Mannheim, GmbH, Germany)으로 복합체를 형성시킨 후에 magnetic chamber 내에서 RHEK 세포주의 DNA probe에 hybridization된 cDNA를 제거하였다. 이렇게 얻은 subtracted double strand cDNA의 양끝을 pfu DNA polymerase (2.5 U/㎕)를 이용해 blunting 시켰으며 EcoRI adaptor를 붙인 후 EcoRI adaptor 부위는 T4 polynucleotide kinase로 phosphorylation 시키고, 다시 cDNA를 XhoI으로 digest시켰다 (Stratagene, CA, USA). 이 cDNA는 T4 DNA ligase(4 U/㎕)를 이용해서 Uni-ZAP XR vector에 ligation시켰으며 Vector는 packaging extract (Stratagene, CA, USA)에 packaging시켜서 Lambda phage의 증식을 XL1-Blue MRF' 세포에 감염시켜서 agar plate에서 plaque의 발현을 확인하였다.[1]

Cloning 및 Sequencing

위의 방법으로 얻어진 cDNA library에서 클론을 얻기 위하여, SOLR strain (Stratagene, CA, USA)의 세포와 Exassist helper phage (Stratagene, CA, USA)의 도움으로 *in vivo* mass excision을 수행하였다. 이때 X-Gal (Sigma, SL, USA) 반응에 의해서 파란색 콜로니는 버리고 백색 콜로니를 취해서 ampicillin-LB broth에 배양하였다. 모두 200례의 cDNA 클론을 얻었으며 충분히 배양된 세균을 resin bead (Promega, WI, USA)를 이용해서 cDNA를 포함하는 플라스미드로 분리하였고, 분리된 cDNA를 polymerase chain reaction(PCR)으로 삽입된 cDNA 크기를 확인한 후 염기서열을 분석하였다. 염기서열은 ALF-express (Phamacia Biotech, CA, USA)의 auto sequencer를 사용해서 부분적 염기서열 배열을 시행하였는데 여기서 얻은 sequencing 결과는 NCBI(National Center for Biotechnology Information)의 blast search program을 통하여 유사성 검색을 하였으며, 유사성 검색에서 비반복성 유전인자로 확인된 clone들은 RNA *in situ* hybridization을 시행하기 위하여 보관하였다.

RNA *in situ* hybridization

성인 타액선 조직에 대한 조직화학적 검색은 신선한 조직을 4% 중성 paraformaldehyde에 고정시킨 후 RNase protection된 방법으로 paraffin block을 제작하였으며, 4㎛ 두께의 현미경 절편을 silanized 유리 슬라이드에 부착하여 보관하였다. RNA in situ hybridization에 사용되는 RNA probe를 제작하기 위하여, RNA *in vitro* transcription 반응을 시켰는데, DNA template 6㎕, 10×buffer (400mM Tris-HCl, pH 7.9 at 25℃, 60 mM MgCl2, 100 mM DTT, 10 mM NaCl and 20 mM spermidine) 2㎕, RNase inhibitor 1㎕, NTP (DIG-UTP) mixture (10mM ATP, 10 mM CTP, 10 mM GTP, 6.5 mM UTP, 3.5 mM DIG-11-UTP, pH 7.5 at 20℃) 1㎕, T3 또는 T7 RNA polymerase (20u/㎕) 2㎕와 DEPC H_2O로 20㎕의 용량을 만들어 혼합하여 37℃에서 2시간 동안 반응시킨 후 formamide 용액 20㎕를 첨가하여 -20℃에 보관하였다. RNA probe의 생산을 확인하기 위하여 보관된 반응물 중에서 약 2㎕를 1% agarose gel에 전기 영동하여 band를 조사하였다. 파라핀 절편 슬라이드를 xylene에 20분간 넣어 파라핀을 제거하였고 100% ethanol에서부터 90, 80, 70% ethanol까지 2분씩 수화시킨 후 DEPC H_2O로 혼합한 PBS용액으로 5분간 세척하였다. 단백질 완충액 (10 mM Tris-HCl, 5 mM EDTA)에 proteinase K 용액 (0.01㎍/㎕)을 넣어 실온에서 15분간 반응시킨 후, 4% paraformaldehyde 용액으로 실온에서 10분간 재고정 시킨 후 0.2N HCl용액으로 실온에서 10분 동안 반응시켜 내재성 alkaline phosphatase를 불활성화 시켰다.

DEPC H2O로 처리한 PBS용액으로 5분씩 2회 수세한 후 0.1M triethanolamine (TEA) 용액으로 10분간 반응시키고, 0.25% acetic acid가 포함된 0.1M TEA 용액으로 실온에서 10분간 반응시켰으며, DEPC H_2O로 혼합한 PBS로 5분씩 2회 수세했다. 50℃로 예열된 100㎕의 hybridization용액 (50%

formamide, 10mM tris-HCl pH 7.6, 200㎍/㎖ tRNA, 1× Denhardt's solution, 10% Dextran sulfate, 600mM NaCl, 0.25% SDS, 1mM EDTA, DEPC H_2O)에 5㎕의 RNA probe를 첨가한 후 조직 위에 기포가 생기지 않도록 hybrislip으로 봉입시켜 55℃ chamber에서 12시간 동안 hybridization 시켰다. 50℃로 예열된 2×SSC(sodium chloride sodium citrate)-50% formamide용액에서 30분간 반응시킨 후 hybrislip을 제거하고 다시 50℃로 예열된 2×SSC-50% formamide 용액에서 30분간 반응시켰다. DIG 1 (0.1M Tris, 0.15M NaCl, pH 7.5)용액으로 37℃에서 30분간 반응시킨 후 blocking reagent가 들어 있는 DIG 2 (0.1M Tris, 0.15M NaCl, 2% blocking, pH 7.5)용액으로 실온에서 2시간 동안 반응시켰다. 여분의 용액을 제거하고 anti-DIG antibody(Boehringer Mannheim, GmbH, Germany)를 DIG 1 용액에 1:1,000으로 희석하여 실온에서 2시간 반응시켰다. 그리고 DIG 1용액으로 30분씩 4회 수세하고 BCIP-NBT(5-bromo-4-chloro-3-indolyl phosphate - nitroblue tetrazolium salt ; Kirkegaard & Perry Lab, MA, USA)로 실온에서 20분간 발색시켰다. 발색이 끝난 후 DIG 4 (10mM Tris, 1mM EDTA, pH 7.5)용액으로 10분간 수세하여 발색을 중지시킨 후, glycerol gel (DAKO, CA, USA)로 봉입하였다. RNA *in situ* hybridization에서 발현된 양성반응의 결과는 Image Pro-4.0 (Media Cybernetics, MA, USA)의 영상 분석 처리를 통하여 양성반응의 정도를 분석하였다.[2]

연구 결과

Human gene database 검색

1. Library finder

사람의 타액선 cDNA library를 확인하기 위해, Library Finder를 이용하여 모든 EST libraries를 검색한 결과 normal human salivary gland library 3 groups (UniGene Library ID : 141, 488, 753), carcinorma, human salivary gland library 2 groups (UniGene Library ID : 4,875, 6,362), uncharacterized human salivary gland library 1 group (UniGene Library ID : 5,022)이 있었다.

2. cDNA XProfiler

정상인 사람 타액선 (normal human salivary gland) cDNA library들과 타액선 암종의 (carcinorma, salivary gland) cDNA library들의 유전자 발현 비교

ID:141 (14 sequences), ID:488 (102 sequences), ID:753 (11 sequences) 등의 정상인 사람 타액선 cDNA library와 ID:4875 (NIH MGC_59, 9,855 sequences)와 ID:6362 (NIH MGC_102, 1,277 sequences) 등의 타액선 암종의 cDNA library의 유전자 발현을 비교하기 위해 cDNA XProfiler를 이용하여 검색한 결과는 Table 1과 같았다. 즉, 정상인 사람 타액선 cDNA libraries나 타액선 암종의 cDNA library에만 존재하는 유전자는 26종 (known 1, unknown 25)이 있었다 (A or B, unique). 그 외에 정상인 사람 타액선 cDNA library나 타액선 암종 cDNA library에 존재하는 유전자들은 모두 2,539종 (known 2,173, unknown 366)이 있었다 (A or B, non-unique). 정상인 사람 타액선 cDNA library와 타액선 암종 cDNA library에 공통으로 존재하는 유전자는 3종의 non-unique known gene들이 있었다 (A and B, non-unique). 정상인 사람 타액선 cDNA library에서만 발견되는 유전자는 1종 (known ; PRB2(S80905))종이 있었다 (A minus B, unique). 정상인 사람 타액선 cDNA library와 다른 library에서는 발견되지만, 타액선 암종 cDNA library에서는 발견되지 않는 유전자는 17종 (known 16, unknown 1)이 있었다 (A minus B, non-unique). 타액선 암종 cDNA library에서만 발견되는 유전자는 25종 (unknown)이 있었다 (B minus A, unique). 그리고 타액선 암종 cDNA library와 다른 library에서는 발견되지만, 정상인 사람 타액선 cDNA library에서는 발견되지 않는 유전자는 모두 2,522종 (known ; 2,154, unknown ; 365)이 있었다 (B minus A, non-unique).

Table 1. Comparison of gene expression between normal human salivary cDNA library and salivary carcinoma cDNA library by XProfiler

Subset	Unique Genes		Non-Unique Genes	
	Known	Unknown	Known	Unknown
A	1	0	19	1
B	0	25	2,157	365
A or B	1	25	2,173	366
A and B	0	0	3	0
A minus B	1	0	16	1
B minus A	0	25	2,154	365

A : normal human salivary gland cDNA libraries, ID:141 (14 sequences), ID:488 (102 sequences), ID:753 (11 sequences)
B : carcinoma, salivary gland cDNA library, ID:4875 (NIH MGC_59, 9855 sequences), ID:6362 (NIH MGC_102, 1,277 sequences)

사람 타액선 cDNA library들과 사람 전립선 (human prostate) cDNA library들의 유전자 발현 비교

ID:141 (14 sequences), ID:488 (102 sequences), ID:753 (11 sequences) 등 5개의 사람 타액선 cDNA libraries와 ID:138, ID:319 (LNCAP cells), ID:716 (Barstead prostate BPH HPLRB4 1) 등 261개의 사람 전립선 cDNA library의 유전자 발현을 비교하기 위해 cDNA xProfiler를 이용하여 검색한 결과는 Table 2와 같았다. 즉, 사람 타액선 cDNA library나 사람 전립선 cDNA library에만 존재하는 유전자는 1,425종 (known 8, unknown 1,417)이 있었다 (A or B, unique). 그 외에 사람 타액선 cDNA library나 사람 전립선 cDNA library에 존재하는 유전자들은 모두 15,505종 (known 9,688, unknown 5,817)이 있었다 (A or B, non-unique). 사람 타액선 cDNA library와 사람 전립선 cDNA library에 공통으로 존재하는 유전자는 2,255종 (known : 2,011, unknown : 244)이 있었다 (A and B, non-unique). 사람 타액선 cDNA library에서만 발견되는 유전자는 26 종(known : 1, unknown : 25)이 있었다 (A minus B, unique). 사람 타액선 cDNA library와 다른 library에서는 발견되지만, 사람 전립선 cDNA library에서는 발견되지 않는 유전자는 284종 (known 162, unknown 122)이 있었다 (A minus B, non-unique). 사람 전립선 cDNA library에서만 발견되는 유전자는 1,399종 (known : 7, unknown : 1,392)이 있었다 (B minus A, unique). 그리고 사람 전립선 cDNA library와 다른 library에서는 발견되지만, 사람 타액선 cDNA library에서는 발견되지 않는 유전자는 모두 12,966종 (known : 7,515, unknown : 5,451)이 있었다 (B minus A, non-unique).

Table 2. Comparison of gene expression between human salivary cDNA library and human prostate gland cDNA library by XProfiler

Subset	Unique Genes		Non-Unique Genes	
	Known	Unknown	Known	Unknown
A	1	25	2,173	366
B	7	1,392	9,526	5,695
A or B	8	1,417	9,688	5,817
A and B	0	0	2,011	244
A minus B	1	25	162	122
B minus A	7	1,392	7,515	5,451

A : human salivary gland cDNA libraries, ID:141 (14 sequences), ID:488 (102 sequences), ID:753 (11 sequences), and other 2 libraries.
B : human prostate cDNA library, ID:138, ID:319 (LNCAP cells), ID:716 (Barstead prostate BPH HPLRB4 1), and other 259 libraries

사람 타액선 cDNA library들과 사람 유선(human mammary gland) cDNAlibrary들의 gene expression의 비교

ID:141 (14 sequences), ID:488 (102 sequences), ID:753 (11 sequences) 등의 5개의 사람 타액선 cDNA library와 ID:634 (NCI_CGAP_Br7, 379 sequences), ID:726 (NCI_CGAP_Br14, 1,277 sequences) 등 906개의 사람 유선 cDNA library의 유전자 발현을 비교하기 위해 cDNA xProfiler를 이용하여 검색한 결과는 Table 3과 같았다. 즉, 사람 타액선 cDNA library나 사람 유선 cDNA library에만 존재하는 유전자는 489종 (known 11, unknown 478)이 있었다 (A or B, unique). 그 외에 사람 타액선 cDNA library나 사람 유선 cDNA library에 존재하는 유전자들은 모두 13,937종 (known 9,405, unknown 4,532)이 있었다 (A or B, non-unique). 사람 타액선 cDNA library와 사람 유선 cDNA

library에 공통으로 존재하는 유전자는 2,219종 (known : 1,978, unknown : 229)이 있었다 (A and B, non-unique). 사람 타액선 cDNA library에서만 발견되는 유전자는 26종 (known : 1, unknown : 25)이 있었다(A minus B, unique). 사람 타액선 cDNA library와 다른 library에서는 발견되지만, 사람 유선 cDNA library에서는 발견되지 않는 유전자는 332종 (known 195, unknown 137)이 있었다 (A minus B, non-unique). 사람 유선 cDNA library에서만 발견되는 유전자는 463종 (known : 10, unknown : 453)이 있었다 (B minus A, unique). 그리고 사람 유선 cDNA library와 다른 library에서는 발견되지만, 사람 타액선 cDNA library에서는 발견되지 않는 유전자는 모두 11,416종 (known : 7,232, unknown : 4,166)이 있었다 (B minus A, non-unique).

Table 3. Comparison of gene expression between human salivary cDNA library and human mammary gland cDNA library by XProfiler

Subset	Unique Genes		Non-Unique Genes	
	Known	Unknown	Known	Unknown
A	1	25	2,173	366
B	10	453	9,210	4,395
A or B	11	478	9,405	4,532
A and B	0	0	1,978	229
A minus B	1	25	195	137
B minus A	10	453	7,232	4,166

A : human salivary gland cDNA libraries, ID:141 (14 sequences), ID:488 (102 sequences), ID:753 (11 sequences), and other 2 libraries
B : human mammary gland cDNA library, ID:634 (NCI_CGAP_Br7, 379 sequences), ID:726 (NCI_CGAP_Br14, 1,277 sequences), and other 904 libraries

사람 전립선 cDNA library들과 사람 유선 (human mammary gland) cDNA library들의 유전자 발현 비교

ID:141 (14 sequences), ID:488 (102 sequences), ID:753 (11 sequences) 등의 사람 전립선 cDNA library와 ID:4875 (NIH MGC_59, 9,855 sequences)와 ID:6362 (NIH MGC_102, 1,096 sequences) 등, 906개의 사람 유선 cDNA library의 유전자 발현을 비교하기 위해 cDNA xProfiler를 이용하여 검색한 결과는 Table 4와 같았다. 즉, 사람 전립선 cDNA library나 사람 유선의 cDNA library에만 존재하는 유전자는 1,887종 (known 18, unknown 1,869)이 있었다 (A or B, unique). 그 외에 사람 전립선 cDNA library나 사람 유선 cDNA library에 존재하는 유전자는 모두 19,210종 (known 10,930, unknown 8,280)이 있었다 (A or B, non-unique). 사람 전립선 cDNA library와 사람 유선 cDNA library에만 공통으로 존재하는 유전자는 25종 (known : 1, unknown : 24)의 unique 유전자 있었고 (A and B unique), 사람 전립선 cDNA library와 사람 유선 cDNA library에 공통으로 존재하며 다른 library에서도 발견되는 유전자는 9,566종 (known : 7,804, unknown : 1,762)이 있었다 (A and B, non-unique). 사람 전립선 cDNA library에서만 발견되는 유전자는 1,399종 (known : 7, unknown : 1,392)이 있었다 (A minus B, unique). 사람 전립선 cDNA library와 다른 library에서는 발견되지만, 사람 유선 cDNA library에서는 발견되지 않는 유전자는 5,630종 (known 1,721, unknown 3,909)이 있었다 (A minus B, non-unique). 사람 유선 cDNA library에서만 발견되는 유전자는 463종 (known : 10, unknown : 453)이 있었다 (B minus A, unique). 그

Table 4. Comparison of gene expression between human prostate gland cDNA library and human mammary gland cDNA library by XProfiler

Subset	Unique Genes		Non-Unique Genes	
	Known	Unknown	Known	Unknown
A	7	1,392	9,526	5,695
B	10	453	9,210	4,395
A or B	18	1,869	10,930	8,280
A and B	1	24	7,804	1,762
A minus B	7	1,392	1,721	3,909
B minus A	10	453	1,405	2,609

A : human prostate cDNA library, ID:138, ID:319 (LNCAP cells), ID:716 (Barstead prostate BPH HPLRB4 1), and other 259 libraries
B : human mammary gland cDNA library, ID:634 (NCI_CGAP_Br7, 379 sequences), ID:726 (NCI_CGAP_Br14, 1,277 sequences), and other 904 libraries

리고 사람 유선 cDNA library와 다른 library에서는 발견되지만, 사람 전립선 cDNA library에서는 발견되지 않는 유전자는 모두 4,014종 (known ; 1,405, unknown ; 2,609)이 있었다 (B minus A, non-unique).

4. Digital Gene Expression Displayer (DGED)

위에서 사용한 정상인 사람 타액선 cDNA library와 타액선 암종 cDNA library에서, 각각의 유전자 발현의 차이를 알아보기 위해 Digital Gene Expression Displayer (DGED) 검색을 시행한 결과 (변수 : 0, 0.1) Table 5와 같았다. 즉, 정상인 타액선 cDNA library에서 가장 많이 발견된 유전자는 HTN3 (43개) 이었으며, 2번 이상 발견되는 유전자는 모두 14종류이었고, 이 중에서 특히 heterogeneous nuclear ribonuclearprotein C (HNRPC ; NM_004500, NM_031314)는 정상인 사람 타액선 cDNA libraries에서 보다 타액선 암종 cDNA library에서 빈번하게 발견되었다. 한편, 정상인 사람 타액선 cDNA libraries에서는 발현되지 않는 retinoblastoma binding protein 7 (RBBP7 ; NM_ 002893)과 calmodulin 2 (CALM2 ; NM_001743)의 발현이 타액선 암종 cDNA library에서는 현저하게 증가되었다.

5. 사람 타액선의 subtracted cDNA library에서 새로운 유전자의 검색

사람 타액선의 subtracted cDNA library에서 얻은 200개의 clone을 sequencing 한 후 Genbank 검색한

Table 5. Comparaison of gene expression between normal human salviary gland cDNA library and salivary carcinoma cDNA library by Digital Gene Expression Displayer (DGED) (변수 : 0, 0.1)

Subset	Gene Name	Accession	Libraries		Sequences		Seq Odds	Chi Sqr
			A	B	A	B	A:B	P(
HTN3	histatin 3	NM_000200	2	0	43	0	NaN	0.00e+00
HTN1	histatin 1	NM_002159	2	0	17	0	NaN	2.24e-297
STATH	statherin	NM_003154	2	0	11	0	NaN	4.83e-193
PRB1	proline-rich protein BstNI subfamily 1	NM_005039	2	0	4	0	NaN	2.19e-71
PRH2	proline-rich protein HaeIII subfamily 2	NM_005042	2	0	4	0	NaN	2.19e-71
PRB3	proline-rich protein BstNI subfamily 3	NM_006249	1	0	4	0	NaN	2.19e-71
PRB2	proline-rich protein BstNI subfamily 2	S80905	2	0	3	0	NaN	5.43e-54
AMY1A	amylase, alpha 1A	NM_004038	1	0	3	0	NaN	5.43e-54
FLJ20073	hypothetical protein FLJ20073	NM_017654	1	0	2	0	NaN	1.42e-36
-	ESTs	AV757530	1	0	2	0	NaN	1.42e-36
LUM	lumican	NM_002345	2	0	2	0	NaN	1.42e-36
AMY2B	amylase, alpha 2B	NM_020978	2	0	2	0	NaN	1.42e-36
NCALD	neurocalcin delta	NM_020122	2	0	2	0	NaN	1.42e-36
DKFZP434L1021	potassium channel modulatory factor	NM_020122	1	0	1	0	NaN	4.21e-19
PRH1	proline-rich protein HaeIII subfamily 1	NM_006250	1	0	1	0	NaN	4.21e-19
CA6	carbonic anhydrase VI	NM_001215	1	0	1	0	NaN	4.21e-19
AMY2A	amylase, alpha 2A	NM_000699	1	0	1	0	NaN	4.21e-19
BAT1	HLA-B associated transcript 1	NM_004640	1	0	1	0	NaN	4.21e-19
JTV1	JTV1 gene	NM_006303	1	1	1	1	80.47	4.47e-10
HNRPC	heterogeneous nuclear ribonucleoprotein C	NM_004500 NM_031314	1	1	2	17	9.54	2.47e-04
RBBP7	retinoblastoma-binding protein 7	NM_002893	0	1	0	244	0	7.62e-02
CALM2	calmodulin 2	NM_001743	0	1	0	223	0	9.04-02

결과, 이중에서 70례가 이미 알려진 유전자이었으며, 그 종류는 모두 27종이었고, 나머지 130례의 유전자는 non-redundant gene으로 밝혀졌다 (Table 6.).

6. *in situ* hybridization을 통하여 풍부하게 발현되는 새로운 유전자의 검색

위에서 얻은 130례의 nonredundant gene을 사람의 타액선 조직에 직접 *in situ* hybridization을 한 결과 이중에서 23례의 유전자가 사람의 타액선에 풍부하게 발현되면서 각기 특징적인 타액선 세포 내 발현 양상을 보였다 (Table 7).

Genes with unknown function

알려진 유전자들 중에서 아직 그 기능이 밝혀지지 않은 것들에 대한 결과는 다음과 같다

1) C76-007 : Homo sapiens chromosome 19,

Table 6. Molecular cloning using the subtracted cDNA library of human adult submandibular gland

Genes (numbers)	clones
*histatin 3 (14)	C75-015, C75-023, C75-024, C75-031, C75-044, C75-050, C75-052, C76-016, C76-021, C76-034, C76-35, C76-042, C76-045, C76-056
*histatin 1 (9)	C75-003, C75-006, C75-017, C75-042, C75-043, C75-047, C75-011, C76-040, C76-053
*proline-rich protein HaeIII subfamily 2 (4)	C75-004, C75-051, C76-009, C76-046, C76-049
*proline-rich protein BstNI subfamily 2 (2)	C75-012, C75-046
*proline-rich protein HaeIII subfamily 1 (1)	C76-027
lacrimal proline-rich protein (1)	C76-052
myosin heavy chain (4)	C77-040, C77-052, C77-081, C77-086
*statherin (3)	C75-001, C76-033, C76-037
cystatin S (3)	C76-004, C76-010, C76-039
highly basic protein (3)	C76-003, C76-023, C76-044
collagen type 1, α2 chain (2)	C77-075, C77-122
cystatin SA (1)	C75-018
mucin (MG2) (1)	C76-002
prolactin induced protein (1)	C75-002
RIG like mRNA (1)	C75-016
prostate binding protein(1)	C75-019
*carbonic anhydrase VI (1)	C75-020
lambda-immunoglobulin constant region complex (1)	C75-048
immunoglobulin lambda chain variable region (1)	C76-036
proteasome (1)	C76-008
ubiquitin (1)	C77-039
heat shock protein (1)	C77-133
c-myc oncogene (1)	C76-015
PTD017 mRNA (pituitary tumor gene) (1)	C76-014
synaptopodin (1)	C76-034
Homo sapiens chromosome 19, BAC CIT-B-393i15 (BC301323) (1)	C76-007
Homo sapiens chromosome 17, clone hRPC.1050__D__4 (1)	C76-012
non-redundant genes 130 cases	
Total	200 cases

*표는 Table 5의 DGED 검색에서 나타난 유전자를 표시함

Table 7. Novel genes from human salivary gland cDNA library

clones	positive reaction	positive cells
C75-005	diffusely positive	ductal and acinar cells
C75-007	strong	ductal, acinar cell
C75-008	marked	acinar cell
C75-029	slight	ductal cells*
C76-001	strong	acinar cells*
C76-018	weak	acinar cell
C76-022	moderate	ductal cells*
C76-030	strong	acinar cell
C76-032	slight	acinar cell
C77-046	weak	ductal and acinar cells*
C77-048	strong	mesenchymal cell
C77-049	strong	ductal and acinar cell*
C77-050	weak	ductal and acinar cells*
C77-053	weak	ductal cells*
C77-055	strong	acinar cell
C77-057	marked	mesenchymal cell
C77-074	weak	matrix protein, ductal cell
C77-085	strong	ductal cell
C77-088	weak	acinar cells*
C77-091	moderate	ductal cell
C77-092	strong	ductal, acinar cell
C77-121	weak	ductal cells*
C77-124	s*ng	acinar cells*

※Results form *in situ* hybridization

BAC CIT-B-393i15 (BC301323)이었고 사람 타액선의 도관 및 선상세포에 약한 양성반응을 보였다.

2) C76-012 : Homo sapiens chromosome 17, clone hRPC.1050__D__4, 이었고 사람 타액선의 도관세포에 강한 양성 반응을 보였다.

3) C76-023와 C76-044는 동일한 유전자로서 Homo sapiens 23 kD highly basic protein mRNA이었고 사람 타액선의 도관 및 선상세포에 미만성으로 양성반응을 보였다.

이상과 같이 사람 타액선에 풍부하게 발현되는 새로운 유전자에 대하여 human gene database를 검색한 결과 23 clone들이 새로운 유전자이었고, 3 clone은 사람의 genomic DNA에 위치하는 알려진 유전자이지만 아직 그 기능이 밝혀지지 않은 유전자이다.

총 괄

타액의 구강 내 기능은 구강의 기계적 윤활 기능뿐만 아니라 음식물을 분해하는 생화학적 기능, 음식물의 독성분을 중화하는 기능, 구강 내 세균 감염을 방지하는 기능, 구강 점막을 보호하는 기능 및 전신적 면역 기능을 증진시키는 효과 등으로 매우 다양하다.[3-7] 따라서 최근에는 사람의 타액선 유전자에 대한 관심이 집중되어 지고 있다.[8-12] 여러 학자들에 의하여 다양한 종류의 타액선 유전자들이 발견되어 지고 있는데, 일반적으로 타액 내 단백질은 대부분 신속하게 분해되거나 다른 단백질과 단단하게 cross-linked 되어 변형되기 때문에 타액의 단백질 분석만으로는 제대로 타액 및 타액선의 유전자 검색을 수행하기 어렵다.[13-16] 본 연구자들은 사람의 타액 및 타액선에서 발현되는 새로운 유전자를 검색하기 위하여 일련의 분자 생물학적인 연구를 수행하였으며 발견된 새로운 유전자들이 현재까지 알려진 Genbank database를 통하여 유사성 검색으로 새로운 기능성 유전자의 가능성을 탐색하기 위한 유전 정보 분석을 시행하였다.

현재까지 GenBank database에 수록된 다양한 종류의 cDNA library들을 서로 비교 검색하는 XProfiler와 DGED 방법은 각 조직별로 unique 또는 non-unique하게 발현되는 정도를 나누어 판별할 수 있는데, XProfiler에서는 대략의 유전인자 발현에 대한 빈도 및 유전자를 찾아 낼 수 있으며, 특히 DGED에서는 특이하게 발현되는 유전자를 개별적으로 찾아 낼 수 있었다. Genbank database에는 사람의 타액선 cDNA library가 정상 타액선 및 타액선 암조직의 library와 미확인 타액선 조직 (uncharacterized salivary tissue) library를 포함하여 모두 6 개 있으며, 이들에서 모두 2,585종의 유전자가 기록되어 있다 (known ; 2179, unknown ; 406). 이에 비해 사람 전립선 cDNA library는 261개의 library에서 모두 16,620종의 유전자 (known ; 9,533, unknown ; 7,087)를, 사람 유선 cDNA library는 906개의 library에서 모두 14,068종의 유전자 (known ;

9,220, unknown ; 4,848)를 기록하고 있다. XProfiler 분석에 따르면, 사람의 타액선 cDNA library에서 발견한 2,174종의 알려진 유전자 가운데에 2,011종의 유전자가 사람의 전립선 cDNA library에서도 발견되며, 1,978종의 유전자는 사람의 유선 cDNA library에서도 발견되는 유사성을 보이고 있다

한편, 사람의 타액선과 전립선, 유선의 정상 조직에서 얻은 cDNA library를 분석한 결과, 타액선이나 전립선 및 유선에 특이하게 발현되는 유전자 가운데에는 서로 동일하게 발현되는 유전자의 수보다 각 조직에 각각 특이하게 발현되는 유전자 수가 훨씬 많은 것으로 나타났다. 이는 Genbank에 기록된 유전자들이 대체로 새로운 유전자를 찾기 위하여 설정되어진 연구 방법으로 찾은 유전자들이 기록되어 있기 때문이기도 하지만, 아직도 이들 전립선이나 유선에는 밝혀지지 않은 유전자들이 상당히 많이 있다는 것을 의미한다.

또한, 타액선의 알려진 유전자 20종 가운데 11종의 유전자 (alpha 2A amylase, alpha 2B amylase, proline-rich protein HaeIII subfamily1 포함)가 전립선의 알려진 유전자 총 8,377종에서 발견되고 있으며, 유선의 알려진 유전자 총 6,317종 가운데에서는 타액선의 알려진 유전자 6종이 포함되어 있다. 이러한 수치는 사람의 정상 타액선 cDNA library의 data가 쌓여가면서 계속적으로 증가할 것으로 예상된다. 한편, 정상인 사람 타액선 cDNA library에서 알려진 유전자 20종 가운데 7종은 필자의 실험실에서 사람의 정상 악하선 조직으로 만든 타액선 cDNA library의 cloning 결과를 Genbank 검색하여 찾아낸 27종의 알려진 유전자에 포함되며, 필자의 실험실에서 제작한 타액선 cDNA library에서도 histatin 3가 14개로 가장 많이 존재하였으며, histatin 1, 다양한 type의 proline-rich protein이 그 다음으로 많이 존재함으로써 DGED 분석과 유사한 분포를 보여 주었다 (Table 5, 6).

필자의 실험실에서 제작한 사람의 타액선 cDNA library의 cloning 결과, 유사성이 없는 것으로 확인된 유전자(non-redundant gene)들을 사람의 타액선 조직에 직접 in situ hybridization을 하여 사람의 타액선에 풍부하게 발현되는 23종의 유전자들을 얻었으며, 이 non-redundnat 유전자들의 염기서열을 본 연구에 기술한 방법과 같이 EST blast search, XProfiler 및 DGED 방법 등으로 검색한 결과 본 연구에서 얻은 23종의 유전자들은 아직 Genbank나 world wide gene database에 기록되지 않은 새로운 유전자임을 확인하였다.[17]

Virtual Northern 검색을 통해 정상조직과 암 조직에서 많이 발현되는 사람의 타액선 유전자들을 각 조직에서 발현도를 분류하여, 연구자료로 사용할 수 있게 된 것은, 많은 시간과 수고를 줄여줄 수 있는 유용한 검색 도구로 생각된다. 따라서 앞으로 본 연구에서 얻은 23종의 새로운 유전자의 염기서열 자료와 Northern blot 결과를 Virtual Northern 검색과 같은 도구에 입력할 계획이다. 그러나, 이번 검색과정에서 본 연구자들은 다음과 같은 문제점들을 발견하였다. 첫째, Genbank database에는 사람의 타액선 cDNA library가 전립선 cDNA library나 유선 cDNA library에 비해 매우 적은 정보가 축적되어 있다. 다시 말해서, 사람의 정상 타액선 cDNA library의 기초적인 data가 다른 기관 조직에 비하여 상당히 부족한 상태이며, 질병과 관련된 database는 아직 초보단계인 것으로 나타났다.[18-20] 또한, 사람의 타액선의 SAGE library나, 생쥐와 같이 실험에 많이 쓰이는 동물의 library도 아직 data가 없었다. 둘째, Salivary Gland Library(ID:5022)의 경우, 287 sequences, 31종의 유전자 정보가 들어 있으나, 조직 성분이 밝혀지지 않아서 data로 적용하기에 어려움이 있다. 따라서, databank에 data를 올릴 때에는 모든 정보를 정확히 밝혀서 효율적으로 사용할 수 있게 하여야 할 것이다. 셋째, 모든 salivary gland library가 non-normalized 상태인 경우이므로, data 사용에 제한이 있다. 앞으로 normalized된 data도 축적되면[21-23] 보다 원활하게 사용할 수 있을 것이다.

결론적으로 본 연구를 통하여 얻은 새로운 유전자 23례 중에는 8clone들이 선상세포에 뚜렷한 양성 반응을 보였고, 7례는 도관세포에, 6례는 선상 및 도관

세포에 미만성으로, 그리고 2례는 중배엽성 조직세포에 국한적으로 양성 반응을 보였다. 그리고 사람의 genomic DNA에 locus를 확인한 3례의 clone은 각각 특이성 있는 새로운 유전자들로 생각된다. 따라서, 사람의 타액선에서 발견된 23례의 새로운 유전자들은 각각의 전체 염기서열을 확인하기 위한 연구와 각각의 단백질 구조 및 기능에 관한 분자생물학적 연구를 계속하여야 하고, 사람의 genomic DNA에 locus를 확인한 3례의 유전자들은 각각의 유전자 구조와 open reading frame을 확인하여 마찬가지로 각각의 단백질 구조 및 기능에 관한 분자생물학적 연구를 계속할 예정이다.

참 고 문 헌

1. 이영준, 고택수, 한형욱, 이상신, 김은철, 김연숙, 이석근, 지제근. 사람 배아의 두개안면 발생에 관련되는 새로운 유전인자를 찾기 위한 검색. 대한병리학회지 2000; 34:961-71
2. 이석근, 김연숙, 송인선, 이상신, 이영준, 김우호, 지제근. Digoxigenin을 표지한 RNA in situ hybridization의 개선된 연구 방법. 대한병리학회지 2001; 35:98-110
3. Kratzschmar J, Haendler B, Eberspaecher U, Roosterman D, Donner P and Schleuning WD. The human cysteine-rich secretory protein (CRISP) family. Primary structure and tissue distribution of CRISP-1, CRISP-2 and CRISP-3. Eur J Biochem 1996; 236:827-36
4. Lin HH and Ann DK. Molecular characterization of rat multigene family encoding proline-rich proteins. Genomics 1991; 10:102-13
5. Liu B, Rayment S, Oppenheim FG and Troxler RF. Isolation of human salivary mucin MG2 by a novel method and characterization of its interactions with oral bacteria. Arch Biochem Biophys 1999; 364:286-93
6. Miao YJ, Subramaniam N and Carlson DM. cDNA cloning and characterization of rat salivary glycoproteins. Novel members of the proline-rich-protein multigene families. Eur J Biochem 1995; 228:343-50
7. Troxler RF, Iontcheva I, Oppenheim FG, Nunes DP and Offner GD. Molecular characterization of a major high molecular weight mucin from human sublingual gland. Glycobiology 1997; 7:965-73
8. Albone EF, Hagen FK, Szpirer C and Tabak LA. Molecular cloning and characterization of the gene encoding rat submandibular gland apomucin. Mucsmg, Glycoconj J 1996; 13:709-16
9. Albone EF, Hagen FK, VanWuyckhuyse BC and Tabak LA. Molecular cloning of a rat submandibular gland apomucin. J Biol Chem 1994; 269:16845-52
10. Dickinson DP and Thiesse M. A major human lacrimal gland mRNA encodes a new proline-rich protein family member. Invest Ophthalmol Vis Sci 1995; 36:2020-31.
11. Dickinson DP and Thiesse M. cDNA cloning of an abundant human lacrimal gland mRNA encoding a novel tear protein. Curr Eye Res 1996; 15:377-86
12. Weston WM, LeClair EE, Trzyna W, McHugh KM, Nugent P, Lafferty CM, Ma L, Tuan RS and Greene RM. Differential display identification of plunc, a novel gene expressed in embryonic palate, nasal epithelium, and adult lung. J Biol Chem 1999; 274:13698-703
13. Blaker M, Kock K, Ahlers C, Buck F and Schmale H. Molecular cloning of human von Ebner's gland protein, a member of the lipocalin superfamily highly expressed in lingual salivary glands. Biochim Biophys Acta 1993; 1172:131-7
14. Bobek LA, Tsai H, Biesbrock AR and Levine MJ. Molecular cloning, sequence, and specificity of expression of the gene encoding the low molecular weight human salivary mucin (MUC7). J Biol Chem 1993; 268:20563-9
15. Nutt E, Dunwiddie C, Jacobs JW and Simpson E. Identification of novel proteins synthesized in bone cells by antibody screening of a cDNA

참 고 문 헌

expression library. Biochem Biophys Res Commun 1988; 151:382-7

16. Sanghi S, Kumar R, Lumsden A, Dickinson D, Klepeis V, Trinkaus-Randall V, Frierson HF, Jr. and Laurie GW. cDNA and genomic cloning of lacritin, a novel secretion enhancing factor from the human lacrimal gland. J Mol Biol 2001; 310:127-39
17. Shaw P and Schibler U. Structure and expression of the parotid secretory protein gene of mouse. J Mol Biol 1986; 192:567-76
18. O' Connell BC, Xu T, Walsh TJ, Sein T, Mastrangeli A, Crystal RG, Oppenheim, FG and Baum, BJ. Transfer of a gene encoding the anticandidal protein histatin 3 to salivary glands. Hum Gene Ther 1996; 7:2255-61
19. Sabatini LM, He YZ and Azen EA. Structure and sequence determination of the gene encoding human salivary statherin. Gene 1990; 89:245-51
20. Sabatini LM, Ota T and Azen EA. Nucleotide sequence analysis of the human salivary protein genes HIS1 and HIS2, and evolution of the STATH/HIS gene family. Mol Biol Evol 1993; 10:497-511
21. Svendsen P, Laursen J, Krogh-Pedersen H and Hjorth JP. Novel salivary gland specific binding elements located in the PSP proximal enhancer core. Nucleic Acids Res 1998; 26:2761-70
22. Samuelson LC, Phillips RS and Swanberg LJ. Amylase gene structures in primates: retroposon insertions and promoter evolution. Mol Biol Evol 1996; 13:767-79
23. Das S, Marcantonio N, Deponte K, Telford SRr, Anderson JF, Kantor FS and Fikrig E. SALP16, a gene induced in Ixodes scapularis salivary glands during tick feeding. Am J Trop Med Hyg 2000; 62:99-105