

ORIGINAL ARTICLE

Array CGH를 이용한 DNA 복제 수 변화에 대한
분석에 따른 유방암의 예후 인자에 대한 후보
유전자로서의 COL18A1황기태 · 정중기 · 정인목 · 허승철 · 안영준 · 안혜성 · 장미수¹ · 김정아² · 한원식^{2,3} · 노동영^{2,3}서울대학교 보라매병원 외과 · ¹병리과, ²서울대학교 의과대학 암연구소 · ³외과학교실COL18A1 as the Candidate Gene for the Prognostic Marker of Breast Cancer
According to the Analysis of the DNA Copy Number Variation by Array CGHKi-Tae Hwang, Jung Kee Chung, In Mok Jung, Seung Chul Heo, Young Joon Ahn, Hye Seong Ahn, Mee Soo Chang¹,
Jeong-Ah Kim², Wonshik Han^{2,3}, Dong-Young Noh^{2,3}Departments of Surgery and ¹Pathology, Seoul National University Boramae Hospital, Seoul; ²Cancer Research Institute and ³Department
of Surgery, College of Medicine, Seoul National University, Seoul, Korea

Purpose: We tried to select and validate the candidate gene for the prognostic marker of breast cancer by comparing the analysis of copy number variation (CNV) between normal breast tissues and breast cancer tissues by performing array comparative genomic hybridization (CGH). **Methods:** Array CGH was performed with using the fresh frozen tissues of 77 breast cancer patients. We selected the clones with more than a 20% frequency of gain or loss, and the clones with gain or loss in more than 2 consecutive clones. We finally selected the clones that were statistically significant on the survival analysis. We searched for the candidate gene that belonged to the candidate clones and we selected the final candidate gene that is assumed to be most related to the carcinogenesis of breast cancer by searching for information of the individual gene. We performed RT-PCR to validate the RNA expression of the final candidate gene with using the breast tissues of another 20 breast cancer patients. **Results:**

Eleven (10 in the gain group and 1 in the loss group) clones were finally selected as candidate clones. The significant CNVs with gain were found in the regions of 1q23.1, 1q41, 1q44, 5p15.33, 8q21.3, 15q26.3, 17q12 and 21q22.3 and the significant CNV with loss was found in 14q32.33. COL18A1 (21q22.3) was selected as the final candidate gene and the RT-PCR results revealed that the expression of COL18A1 was up-regulated in the cancer tissues of 18 of the other 20 (90%) breast cancer patients. **Conclusion:** We selected COL18A1 (21q22.3) as the candidate gene for the prognostic marker of breast cancer by comparing the analysis of CNVs from the array CGH. The RNA of COL18A1 was over-expressed in breast cancer tissue, as determined by RT-PCR.

Key Words: Breast neoplasms, Collagen type XVIII, Gene amplification, Microarray analysis

중심단어: 유방 신생물, 콜라겐 type XVIII, 유전자 증폭, 미세배열 분석

책임저자: 노동영

110-744 서울시 종로구 연건동 28, 서울대학교병원 외과

Tel: 02-2072-2921, Fax: 02-766-3975

E-mail: dynoh@plaza.snu.ac.kr

접수일: 2009년 7월 29일 게재승인일: 2009년 11월 4일

본 논문은 서울대학교 보라매병원 연구비(06-2007-5)와 보건복지부의 Korea Health 21 R&D 프로젝트 연구비(01-PJ3-PG6-01GN07-0004)의 지원을 받아 이루어졌음.

본 논문은 2009년 제35차 대한암학회에서 포스터 발표되었음.

서론

유방암의 예후 인자와 관련된 유전자에 대한 연구가 많이 진행되어 왔으나, 아직도 이들 유전자로 전체적인 유전적 기전을 설명하기에는 많은 한계가 있으며, 향후 많은 추가적인 연구를 필요로 하고 있다.

최근에 미세배열(microarray)기술을 이용한 연구를 통해 많은 수의 유전자들의 발현을 동시에 연구할 수 있는 방법이 개발되면서, 각종 질환과 연관된 유전자의 연구에 대한 속도가 빨라지고 있다. 이를 바탕으로 전통적인 병리학적 특성 외에 분자 유전학적 특성에 대한 정보 규명이 빠른 속도로 증가하고 있으며, 이에 따라 분자 유전학적 특성에 따른 특화된 치료가 실제 임상에서 적용되는 사례가 증가되고 있으며, (1) 이의 대표적인 경우가 현재 3상 임상 연구로 진행되고 있는 Microarray In Node negative Disease may Avoid ChemoTherapy trial이라고 할 수 있다.

특히 미세배열을 기반으로 하는 array comparative genomic hybridization (array CGH)를 이용하여 고형암에 대한 분자적인 특성을 규명하여 임상적인 유용성을 증명하는 연구가 활발해지고 있는데, (2,3) array CGH는 전체 유전체(whole genome)의 DNA를 대상으로 하는 고효율성의 연구방법으로, 한 번의 실험으로 동시에 얻어지는 전체 유전자의 복제 수 변화(copy number variation)에 대한 분석을 통해 다양한 연구 분야에 적용되고 있다. (4) Array CGH의 경우 칩에 탑재되는 유전 정보의 종류에 따라 complementary DNA (cDNA) array, bacterial artificial chromosomes (BAC) array, oligonucleotide array, molecular inversion probe array, Solexa sequencing technology array 등의 여러 가지 종류로 세분할 수 있는데, (2) 이번 연구에서는 BAC array 방식을 사용하였다.

저자들은 이전 연구에서 31쌍의 재발 및 비재발 유방암 환자들의 유방암 조직을 이용한 array CGH 실험 분석을 통해 전신 재발과 관련된 유전자에 대한 연구를 보고한 바 있다. (5) 이 연구 결과를 바탕으로 하여, 이번 연구에서는 새로운 15명의 유방암 환자들에 대한 array CGH 실험 결과를 추가하여 모두 77명의 유방암 환자들에 대한 array CGH 결과를 재분석하여 유방암의 예후 인자와 가장 관련이 있는 유전자에 대한 연구를 진행하고자 하였다. 즉, 이미 임상병리학적 정보를 알고 있는 77명의 유방암 환자를 대상으로 시행한 array CGH 실험 결과를 이용하여 정상 조직의 DNA와 대비되는 유방암 조직의 DNA에 대한 복제 수 변화를 분석하고, 또한 생존 분석을 시행함으로써 유방암의 예후 인자와 관련 있는 유전자를 발굴하고, 이 유전자에 대해 새로운 환자들의 유방암 조직을 이용하여 역전사효소 중합연쇄반응(reverse transcriptase-polymerase chain reaction, RT-PCR) 실험을 시행함으로써 후보 유전자의 RNA 발현 양상을 검증하고자 하였다.

방 법

대상 환자

이번 연구에서 이용된 array CGH 실험 및 RT-PCR 실험은

모두, 서울대학교 보라매병원 또는 서울대학교병원에서 수술 받은 환자들 중에서 실험에 필요한 조직이 확보되어 있으면서, 필요한 임상 정보를 얻을 수 있는 77명의 여성 유방암 환자를 대상으로 하였다. 모든 환자들의 조직학적 진단명은 침윤성 유관암이었으며, 진단 당시 원격 전이가 있거나, 신보조강화화학요법(neoadjuvant chemotherapy)을 시행한 경우는 대상에서 제외하였다. 환자의 나이는 진단 당시의 나이로 정의하였고, 수술 방법은 크게 유방전체절제술과 유방부분절제술의 두 군으로 분류하였다. 유방암의 병기는 2002년에 출판된 American Joint Committee on Cancer 6번째 판을 따랐으며, 조직학적 등급 및 핵 등급은 Nottingham combined histologic grade를 따랐다. 에스트로겐수용체 및 프로게스테론수용체는 면역조직화학 검사 결과를 이용하였고, 종양의 10% 이상에서 양성 결과가 나온 경우 양성으로 판정하였다. (6) 호르몬수용체는 에스트로겐수용체 또는 프로게스테론수용체가 양성 결과가 나온 경우에 양성으로 정의하였고 두 수용체가 모두 음성 결과가 나온 경우에 음성으로 정의하였다. HER2의 경우에는 형광제자리부합법(fluorescence in situ hybridization, FISH) 검사결과가 있는 경우 이를 준용하였고, FISH 검사결과가 없는 경우에는 면역조직화학 검사 결과가 0 또는 1(+)인 경우는 음성, 2(+) 또는 3(+)인 경우는 양성으로 판정하였다.

조직 채취

Array CGH 실험을 위해 사용된 유방암 조직들은 모두 유방암의 수술 시에 수술장에서 채취되었으며, 조직 표본이 환자로부터 분리되는 즉시 암의 중심부에서 암조직이 채취되었다. 이 조직들은 채취된 직후 질소 탱크에 보관되었으며, 이후 -80°C의 냉동고에 이동 보관되었다. RT-PCR 실험에 사용된 암조직 역시 같은 방법으로 서울대학교 보라매병원 수술장에서 채취되었으며, 또한 암조직에서 충분히 떨어진 부위의 정상 조직을 추가로 채취하여, 질소 탱크에 보관해 오던 조직을 본 실험에 이용하였다.

Array CGH 실험

이번 연구는 MacArray™ Karyo 4000 칩(Macrogen Inc., Seoul, Korea)을 이용하였으며, 이 칩에 대한 정보는 <http://www.macrogen.co.kr>에 기술되어 있다. 이 칩은 75×25 mm 크기이며, 4,096 (2¹²)개의 spot이 2열로 배열되어 있어 1회의 실험으로 1쌍의 실험 결과를 얻을 수 있다. 실험 방법은 이전에 보고된 방식으로 진행하였으며, 간략히 요약하면 다음과 같다. 이 칩에는 4,044개의 BAC이 포함되어 있으며 전체 유전체에 대해 평균적으로 약 1 Mbp의 해상도를 지닌다. 환자의 유방암 조직에서 PureGene kit (Gentra Systems, Minneapolis, USA)를 이용하여 DNA를 추출하였으며, 실험의 대조군으로 사용되는 DNA

는 상업용으로 판매되는 정상인의 DNA를 이용하였다(Human Genomic DNA: Female; Promega Corporation, Madison, USA). 유방암 환자의 DNA에 Cy3를 결합시키고 대조군 DNA에는 Cy5를 결합시킨 뒤, hybridization을 시행하였다. 실험 결과로 얻어진 영상에 대해 GenePix 4200 A (Axon Instruments, Inc., Union City, USA)를 이용하여 스캔한 뒤, MacViewer 1.6 프로그램 (Macrogen Inc., Seoul, Korea)을 이용하여 영상을 분석하였다. 이 분석에 따라 각각의 clone에 대해 환자의 DAN와 대조군의 DNA의 Cy3/Cy5 형광 비율에 따라 Log_2 비율로 변환된 수치로 전환되며, 최종 분석을 위해 LOWESS normalization을 시행하였다. Log_2 비율은 Cy-3의 Log_2 값에서 Cy-5의 Log_2 값을 뺀 값으로 정의하였다($\text{Log}_2 \text{ ratio} = \text{Log}_2 \text{ Cy3} - \text{Log}_2 \text{ Cy5}$).

BAC 클론

MacArray™ Karyo 4000 칩에 포함된 클론들 중에서 Y 염색체 부분을 제외하고, 22개의 상염색체와 X 염색체에 해당하는 클론들의 개수는 모두 4,044개였다. Array CGH 실험 결과로 생성된 spot의 영상이 좋지 못하여 MacViewer 1.6 프로그램으로 그 영상을 분석할 수 없는 경우, 최종적인 Log_2 수치를 표시할 수 없었으며, 이런 경우 최종 수치는 결측치로 처리하였다. 전체 77명의 유방암 조직에 대해 20% 이상(16명 이상)의 조직에서 결측치로 처리된 클론들의 경우는 분석에서 제외하였다. 이런 이유로 전체 4,044개의 클론들 중에서 38개의 클론은 이번 연구의 분석에서 제외하였으며, 최종적으로 나머지 4,006개의 클론들에 대한 결과를 분석하였다. 환자의 DAN와 대조군의 DNA에 각각 표지된 Cy3 및 Cy5의 형광 정도에 대한 normalized log_2 비율의 값이 0.25보다 큰 경우를 gain으로 정의하였고, -0.25보다 작은 경우를 loss라고 정의하였다. 이러한 gain/loss의 기준은 이전의 실험을 바탕으로 산출된 경험적인 값이며, 표준편차의 3배에 해당하는 값이다.

후보 클론의 선정

분석의 대상이 된 4,006개의 클론들 중에서 전체 조직의 20% 이상(16명 이상)에서 gain/loss가 관찰되는 클론들을 먼저 선정하였으며, 이 클론들 중에서 다시 3개 이상의 연속된 클론에서 동시에 gain/loss가 있는 클론들을 선정하였다.(5) 이렇게 선정된 클론들 중에서 무병 생존 분석 및 전체 생존 분석 등의 생존 분석을 시행하여 유의한 클론들을 선정하였다.

후보 유전자의 선정

상기 방법으로 선정된 후보 클론들에 대해 National Center for Biotechnology Information (NCBI) Map Viewer (www.

ncbi.nlm.nih.gov/mapview; Build 36.3)를 이용하여 유전자를 검색하였다. 즉, 각각의 클론에 대해 해당 염색체를 선택한 이후 각 클론의 시작 및 끝 부분에 대한 정보를 입력하여 구간에 포함된 모든 유전자를 검색하였다. 검색된 각각의 유전자에 대해 PubMed (www.pubmed.gov) 검색을 통해 암과의 연관성 여부를 확인하였다.

RT-PCR

유방암 조직 및 정상 유방 조직의 RNA는 TRIzol® (Invitrogen, Carlsbad, USA) 시약을 이용하여 분리한 후 각각 5 μg 의 전체 RNA를 주형으로 사용하여 10 mM Tris-HCl (pH 8.3), 6.5 mM MgCl_2 , 50 mM KCl, 10 mM DTT, 1 mM dNTPs, random primer 150 ng 및 Moloney Murine Leukemia Virus (M-MLV; Invitrogen, Carlsbad, USA) reverse transcriptase 200 U를 혼합하여 전체 용량을 20 μL 로 정량하였다. 이 혼합물을 37°C에서 1시간 동안 반응시켜 cDNA를 합성한 후 5배로 희석하여, 전체 100 μL 중에서 1 μL 의 cDNA에 0.1 μM primer, 125 μM dNTP 혼합액, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5 mM MgCl_2 및 Taq polymerase (Takara, Shiga, Japan)을 사용하여 최종 25 μL 의 양으로 PCR 반응을 시행하였다. 반응 혼합액은 94°C에서 변성한 후 94°C에서 30초, 55°C에서 30초, 72°C에서 1분 동안 연장하는 3단계의 반응 과정을 시행하였고, 이 조건에서 기준 유전자는 27회, 그리고 COL18A1은 35회의 반응 과정을 거쳤다. RPLP0 (NM_053275)을 기준 유전자로 사용하였으며, 각 primer들의 염기 서열은 다음과 같다. Sense for COL18A1 (NM_130445.2), 5'-ACTTCCCCAGCCTCTTCTTC-3'; antisense for COL18A1, 5'-CCTGCACCTGGTTCTGTGTA-3'; sense for RPLP0, 5'-GGAATGTGGGCTTTGTGTTTC-3'; antisense for RPLP0, 5'-TGCCCTGGAGATTTTAGTGTG-3'.

통계 분석

생존 분석을 위해 Kaplan-Meier 검정을 이용하였으며, 본 연구에서 무병 생존은 수술 이후 전신 재발의 증거가 없는 경우로 정의하였다. 무병 생존 분석에 사용된 기간은, 전신 재발이 있었던 환자의 경우에는 수술 일자로부터 전신 재발이 진단된 일자까지로 정의하였고, 전신 재발이 없었던 환자는 수술 일자로부터 마지막 추적 일자까지로 정의하였다. 전체 생존 분석에 사용된 생존 기간은 사망 환자의 경우에는 수술 일자로부터 사망 일자까지로 정의하였고, 생존 환자의 경우는 수술 일자로부터 마지막 추적 일자까지로 정의하였다. 생존 곡선 사이의 유의성 검정은 log rank test를 사용하였다. RT-PCR 결과의 유의성 판정을 위해서는

단일표본 T 검정 방법을 이용하였다. 통계는 모두 SPSS 16.0 (SPSS Inc., Chicago, USA)을 이용하였으며, 본 연구에 사용된 모든 통계에서 p 값이 0.05 미만인 경우를 통계적으로 유의한 것으로 간주하였다. 생존 분석의 과정에서 다중 비교 분석으로 인한 제1형 오류의 증가를 보정하기 위해 false discovery rate 검정을 사용하였으며, 이 경우 보정된 p 값인 0.05 미만인 경우를 통계적으로 유의한 것으로 간주하였다.

결 과

임상병리학적 특성

전체 77명의 유방암 환자들은 모두 여성이었으며, 평균 연령은 48 (± 10.78 , 31-70)세이었다(Table 1). T2에 해당되는 환자들이 전체의 71.4%로 가장 많은 비율을 차지하였고, T4의 경우는 3.9%로 가장 적은 비율을 보였으며, N3의 경우 32.5%의 비율을 보였으며, N2의 경우 13.0%의 비율을 보였다. 병기는 IIIC인 경우가 32.5%로 가장 많은 부분을 차지하였으며, IIIB인 경우는 1.3%로 가장 적었다. 호르몬 수용체 및 HER2가 양성인 경우가 각각 62.3%, 55.8%였다. 88.3%의 환자가 유방전체절제술을 시행 받았으며, 항암치료, 방사선치료, 호르몬치료를 받은 환자는 각각 89.6%, 54.5%, 55.8%였다. 전체 환자의 평균 추적 관찰 기간은 107 (± 30.62 , 8-115)개월이었고, 이 중에서 34명(44.2%)의 환자에서 전신 재발이 관찰되었으며 전신 재발까지의 평균 기간은 35.6 (± 23.52 , 6-97)개월이었다. 18명(23.4%)의 환자가 유방암으로 인해 사망하였으며, 사망까지의 평균 기간은 41.7 (± 20.10 , 8-86)개월이었다.

후보 클론의 선정

Array CGH 실험 및 분석 결과로 나온 normalized \log_2 비율 값을 기준으로 0.25보다 큰 경우 gain이 있는 것으로 정의하였다. 분석의 대상이 되는 전체 4,006개의 클론들 중에서 gain의 빈도가 전체 77명의 환자 중에서 20% 이상(16명 이상)을 보이는 279개의 클론을 먼저 선정하였으며, 이 중에서 다시 3개 이상의 연속된 클론에서 동시에 gain이 있는 167개의 클론을 선정하였다. 이 167개의 클론에 대하여 각각 생존 분석을 시행하였으며, 그 결과 무병 생존 분석에서 6개의 클론 및 전체 생존 분석에서 5개의 클론이 통계적으로 유의한 클론으로 선정되었다. 이 클론들 중에서 1개의 클론이 무병 생존 분석 결과 및 전체 생존 분석 결과에서 중복되었다. 따라서 gain군에서는, 유방암 조직에서 정상 유방조직에 비해 array CGH 분석 및 생존 분석에서 유의한 차이를 보이는 10개의 클론을 최종 선정할 수 있었다(Table 2).

Gain군과 같은 방식으로 normalized \log_2 비율값을 기준으

로 -0.25보다 작은 경우 loss가 있는 것으로 정의하였으며, 전체 4,006개의 클론 중에서 loss의 빈도가 20% 이상인 131개의 클론을 일차 선정하였다. 이 중에서 다시 3개 이상의 연속된 클론에서 동시에 loss가 있는 45개의 클론을 선정하였다. 이 클론들에 대하여 무병 생존 분석 및 전체 생존 분석을 시행한 결과, 각각 1개씩의 클론이 선정되었으나, 이 클론들은 서로 동일하였으며, 따라서 최종 1개의 클론이 loss군에서 후보 클론으로 선정되었다

Table 1. Clinicopathological characteristics of the patients

| Characteristics | Number | % |
|--------------------|--------|------|
| T stage | | |
| T1 | 11 | 14.3 |
| T2 | 55 | 71.4 |
| T3 | 8 | 10.4 |
| T4 | 3 | 3.9 |
| N stage | | |
| N0 | 24 | 31.2 |
| N1 | 18 | 23.4 |
| N2 | 10 | 13.0 |
| N3 | 25 | 32.5 |
| Stage | | |
| I | 5 | 6.5 |
| IIA | 21 | 27.3 |
| IIB | 14 | 18.2 |
| IIIA | 11 | 14.3 |
| IIIB | 1 | 1.3 |
| IIIC | 25 | 32.5 |
| Histological grade | | |
| 1, 2 | 32 | 41.6 |
| 3 | 42 | 54.5 |
| Nuclear grade | | |
| 1, 2 | 37 | 48.1 |
| 3 | 38 | 49.4 |
| Hormone receptor | | |
| Positive | 48 | 62.3 |
| Negative | 29 | 37.7 |
| HER2 | | |
| Positive | 43 | 55.8 |
| Negative | 34 | 44.2 |
| Operation | | |
| TM | 68 | 88.3 |
| BCS | 9 | 11.7 |
| Radiation therapy | | |
| Yes | 42 | 54.5 |
| No | 35 | 45.5 |
| Chemotherapy | | |
| Yes | 69 | 89.6 |
| No | 8 | 10.4 |
| Hormonal therapy | | |
| Yes | 43 | 55.8 |
| No | 34 | 44.2 |

TM=total mastectomy; BCS=breast conserving surgery.

(Table 2).

Gain군에서는 염색체의 1q23.1, 1q41, 1q44, 5p15.33, 8q-21.3, 15q26.3, 17q12 및 21q22.3 부위의 발현이 증가되어 있었으며, loss군에서는 14q32.33 부위의 발현이 감소되어 있는 양상이었다. 전체 77명의 환자 중에서 50% (39명) 이상에서 gain 또는 loss가 관찰되는 clone에 대해 분석한 결과, 1p36.33, 8q23.3, 8q24.3, 8q24.12, 8q24.13, 8q24.21 및 20q13.33에서 50% 이상의 gain이 관찰되었고, 8p23.1, 10q11.22 및 14q32.33에서 50% 이상의 loss가 관찰되었다(Table 3).

후보 유전자의 선정

상기 분석에서 선정된 gain군에서 10개의 클론 및 loss군에서 1개의 클론에 대해 NCBI Map Viewer (Build 36.3)를 이용하여 유전자를 검색하였으며, 해당 유전자에 대해 다시 PubMed 검색을 통해 암과의 연관성 유무를 확인하였다. 최종적으로 gain군에서 7개 및 loss군에서 1개의 유전자가 선정되었다(Table 4). 후보로 선정된 8개의 유전자 중에서 1q21-22에 위치하고 있는 NTRK1 및 21q22.3에 위치하는 COL18A1이 다양한 종류의 고형성 암과의 연관성이 보고되어 있었으며, 특히 COL18A1은 유방암과의 관련성에 대해서도 보고된 바가 있었다. 이번 연구의 분석 결과를 바탕으로 최종적으로 COL18A1에 대해 새로운 환자들의

Table 2. Statistically significant candidate clone list in gain and loss group

| Gain/loss | Clone No. | Frequency (%) | SRFS | OS | Cytoband | Start (kb) | End (kb) | Size (kb) |
|-----------|-----------|---------------|-------|-------|----------|------------|-----------|-----------|
| Gain | c2168 | 45.5 | NS | 0.019 | 1q23.1 | 155063058 | 155208815 | 145757 |
| | c4555 | 27.3 | NS | 0.021 | 1q41 | 221038499 | 221109556 | 71057 |
| | c5615 | 44.2 | NS | 0.012 | 1q44 | 246778289 | 246914669 | 136380 |
| | c5262 | 40.3 | 0.014 | NS | 5p15.33 | 473194 | 562560 | 89366 |
| | c301 | 36.4 | NS | 0.048 | 8q21.3 | 91034606 | 91125957 | 91351 |
| | c2776 | 33.8 | 0.027 | NS | 15q26.3 | 100054894 | 100129431 | 74537 |
| | c2096 | 26.0 | 0.010 | NS | 17q12 | 34946411 | 35061427 | 115016 |
| | c664 | 24.7 | 0.007 | NS | 17q12 | 34989583 | 35061419 | 71836 |
| | c2806 | 22.1 | 0.008 | NS | 21q22.3 | 45569815 | 45705137 | 135322 |
| | c2754 | 29.9 | 0.036 | 0.044 | 21q22.3 | 45592803 | 45703201 | 110398 |
| | Loss | c710 | 50.6 | 0.041 | 14q32.33 | 105821330 | 105907464 | 86134 |
| | | | | | | | | |

SRFS=systemic recurrence-free survival; OS=overall survival; kb=kilobase; NS=not significant.

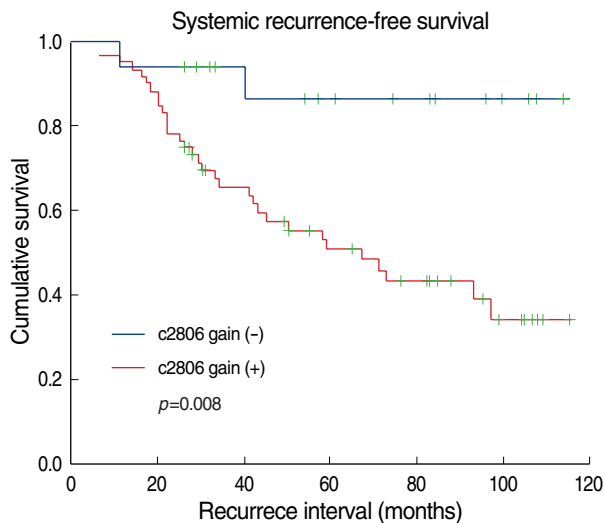
Table 3. Common regions showing gain or loss in more than 50% of all 77 samples

| Region | Clone No. | Cytoband | Start (kb) | End (kb) | Number (%) | Cancer related genes |
|--------|-----------|----------|------------|-----------|------------|--|
| Gain-1 | c5784 | 1p36.33 | 552910 | 563807 | 75 (97.4) | CYHR1, KIFC2, FOXH1, PPP1R16A, GPT, MFSD3, RECQL4, LRRC14, LRRC24, MGC70857, KIAA1688 |
| Gain-2 | c5242 | 8q24.3 | 145647141 | 145761879 | 46 (59.7) | |
| Gain-3 | c4824 | 8q24.13 | 126947484 | 127030285 | 42 (54.5) | SAMD12 |
| Gain-4 | c1394 | 8q23.3 | 116937688 | 117027644 | 41 (53.2) | |
| Gain-5 | c1437 | 8q24.12 | 119396534 | 119465174 | 41 (53.2) | |
| Gain-6 | c1433 | 8q24.21 | 131335147 | 131416013 | 41 (53.2) | |
| Gain-7 | c2733 | 20q13.33 | 61387393 | 61535237 | 39 (50.6) | ARFGAP1, COL20A1, CHRNA4, KCNQ2 |
| Loss-1 | c5256 | 8p23.1 | 7323700 | 7428919 | 57 (74.0) | DEFB106B, DEFB105B, DEFB107B, LOC645489, FAM90A6P, FAM90A7, LOC729339, FAM90A22, FAM90A23 |
| Loss-2 | c4589 | 8p23.1 | 7334384 | 7420885 | 56 (72.7) | DEFB105B, DEFB107B, LOC645489, FAM90A6P, FAM90A7, LOC729339, FAM90A22 |
| Loss-3 | c5126 | 8p23.1 | 7647665 | 7716751 | 55 (71.4) | FAM90A19, LOC729394, FAM90A9, FAM90A10, DEFB107A |
| Loss-4 | c5189 | 10q11.22 | 46320705 | 46408357 | 43 (55.8) | RHEBP1, SYT15 |
| Loss-5 | c710 | 14q32.33 | 105821330 | 105907464 | 39 (50.6) | IGHVIII-25-1, IGHV2-26, IGHVIII-26-1, IGHVII-26-2, IGHV7-27, IGHV4-28, IGHVII-28-1, IGHV3-29 |

kb=kilobase.

Table 4. Candidate gene list from candidate clones in gain and loss group

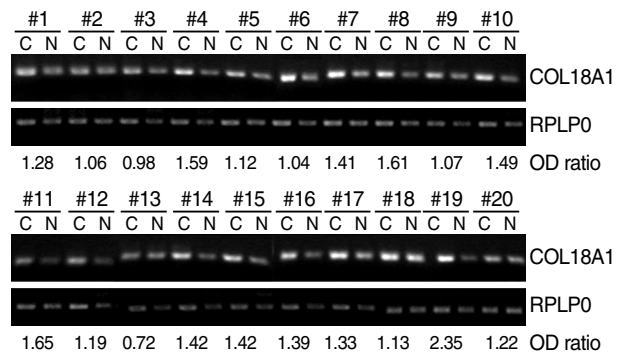
| Gain/loss | Gene | Clone No. | Cytoband | Description |
|-----------|-------------|-----------|----------|---|
| Gain | NTRK1 | c2168 | 1q21-q22 | Neurotrophic tyrosine kinase, receptor, type 1 |
| | ARHGEF11 | c2168 | 1q21 | Rho guanine nucleotide exchange factor (GEF) 11 |
| | AHRR | c5262 | 5p15.3 | Aryl-hydrocarbon receptor repressor |
| | SLC9A3 | c5262 | 5p15.3 | Solute carrier family 9 (sodium/hydrogen exchanger), member 3 |
| | NBN | c301 | 8q21 | Nibrin |
| | PPP1R1B | c2096 | 17q12 | Protein phosphatase 1, regulatory (inhibitor) subunit 1B (dopamine and cAMP regulated phosphoprotein, DARPP-32) |
| | COL18A1 | c2806 | 21q22.3 | Collagen, type XVIII, alpha 1 |
| Loss | IGHV11-25-1 | c710 | 14q32.33 | Immunoglobulin heavy variable (III)-25-1 |

**Figure 1.** Survival curve of systemic recurrence-free survival analysis for the clone which contains COL18A1 gene (c2806) by Kaplan-Meier test. The survival of the group with gain of c2806 clone was better than that without gain of c2806 clone and the difference of survival between two groups was statistically significant ($p=0.008$) by log rank test.

유방암 조직 및 정상 조직을 이용하여 검증하기로 하였다. COL18A1은 2806번의 클론에 포함되어 있는 유전자이며, 이 클론에 대한 무병 생존 분석에서는 통계적으로 유의한 결과를 보였으나 ($p=0.008$), 전체 생존 분석에서는 유의한 결과를 보이지 않았다 (Figure 1).

후보 유전자에 대한 RT-PCR 검증

새로운 20명의 유방암 환자에서 얻은 20쌍의 유방암 조직 및 정상 유방조직에 대해 COL18A1에 대한 RT-PCR을 시행하였다. 그 결과 18명(90%)의 환자에서 정상 유방 조직에 비해 유방암 조직에서 COL18A1의 발현의 증가가 관찰되었으며, 나머지 2명의 환자에서는 차이가 없었으며(Figure 2), 이는 통계적으로 매우 유의한 결과였다($p<0.0001$).

**Figure 2.** Results of validation test by RT-PCR for the RNA expression of COL18A1. The RNA expression of COL18A1 was up-regulated in breast cancer tissues compared to normal breast tissues in 18 of 20 (90%) breast cancer patients. Difference of the RNA expression was statistically significant ($p<0.0001$). RPLP0 was used as reference gene. OD ratio=(OD[COL18A1_Cancer]/OD[RPLP0_Cancer])/(OD[COL18A1_Normal]/OD[RPLP0_Normal]). OD=optical density.

고찰

본 연구는 77명의 유방암 환자에 대한 array CGH 실험을 통해 정상 조직에 대한 유방암 조직에서 발현이 증가 또는 감소한 클론들 중에서, 생존 분석에서 유의한 클론을 선정하였으며, 이 클론들 중에서 암과 관련된 유전자를 검색한 뒤 새로운 20명의 환자의 조직을 이용하여 RT-PCR 실험을 통해 그 유의성을 검증하였다. 이 연구를 통해 COL18A1이 정상 유방조직에 비해 유방암 조직에서 그 발현이 유의하게 증가되어 있으며, 또한 생존 분석에서도 유의한 결과를 보이고 있음을 증명하고 있다.

본 연구에 포함된 환자들의 경우 일반적인 유방암 환자에 비해 유방암의 병기가 높은 경향이 있었다. 이는 array CGH 실험에 사용되는 유방암 조직을 수술장에서 채취하기 위해서는 유방암 조직의 크기가 어느 정도 크기 이상이 되는 환자들만이 연구 대상으로 선정되었기 때문으로 사료된다. 즉, T2의 경우가 전체의 71.4%를 차지하고 있으며, N3의 경우는 32.5%, 그리고 IIIC 병

기의 환자는 전체의 32.5%나 차지하고 있다(Table 1). 뿐만 아니라, HER 양성인 경우가 전체의 55.8%를 차지하여 일반적인 유방암 환자의 HER 양성률보다 높은 비율을 보이고 있다. 유방전체절제술을 88.3%의 환자에서 시행한 것도 전체적인 병기가 높은 이유로 생각되며, 이렇게 유방전체절제술을 시행 받은 환자가 많음에도 불구하고 수술 후 보조적으로 방사선 치료를 시행한 경우가 54.5%의 높은 비율을 차지하고 있는 것도 동일한 이유에서 비롯된 것으로 사료된다. 하지만 이러한 환자의 임상병리학적 분포 양상은 정상 유방 조직과 유방암 조직의 유전적 차이를 확인하려는 이번 연구에서는, 분석에 있어서 중요한 장애 요인이 되지 않을 것으로 사료된다.

Array CGH 방법은 미세배열 실험의 한 종류로서, 인간의 전체 유전체를 일정한 숫자의 클론으로 나누어 각각의 클론에 대한 정보를 미리 지도화(mapping)해 놓은 후에, 전체 클론들을 한 장의 칩에 제작을 하고 그 위에 증폭을 원하는 시료를 반응시킴으로써, 한 번의 실험으로 전체 유전체의 변화를 클론 단위로 파악할 수 있는 고효율성의 실험 방법이다. 하지만 실험 비용이 상대적으로 고가이고, 전통적인 유전 연구 방법에 비해 익숙하지 않은 실험 방법이며, 실험 결과로 제시되는 방대한 양의 정보에 대한 분석 방법이 쉽지 않은 등의 이유로 국내 논문에서는 array CGH를 이용한 논문의 보고가 매우 드문 실정이며, 특히 유방암에 대한 보고는 이번 연구가 처음인 것으로 사료된다. 하지만 국내 논문의 경우 유방암과 관련하여 cDNA를 이용한 expression microarray 실험에 대한 논문은 드물게 보고되어 있다.(7,8)

본 연구에서 후보 클론을 선정하기 위해 몇 가지의 선정 기준을 사용하고 있다. 먼저 분석의 대상이 되는 전체 4,006개의 클론 중에서 이미 보고된 논문의 내용에 따라 전체 검체 수의 20% 이상에서 gain이나 loss의 변화가 있으면서, 3개 이상의 연속된 clone에서 동시에 변화가 관찰되는 클론의 경우를 의미 있는 클론으로 간주하였다.(5,9) 이상에서 선정된 clone들 중에서 임상적으로 의미가 있는 clone을 선정하기 위해 본 연구에서는 생존 분석의 방법을 사용하였다. 생존 분석의 방법은 임상적으로 가장 의미 있는 정보 중의 하나이며, 본 연구의 대상이 된 77명의 환자들은 모두 생존 분석에 필요한 임상 정보를 추적할 수 있었고, 그 평균 추적 관찰 기간도 107개월로 생존 분석에 필요한 비교적 충분한 추적 관찰 기간을 가지고 있었다는 점 등이 임상적으로 의미가 있는 클론을 선정하기 위한 기준으로 생존 분석의 방법을 이용한 주요 사유가 되었다. 상기 과정을 통해 선정된 클론들 중에서 무병 생존 분석이나 전체 생존 분석에서 의미 있는 결과가 나온 클론들을 최종 선정하였다(Table 2). 특히 이번 연구에서 최종 후보 유전자로 선정된 COL18A1이 포함되어 있는 클론(c2806)의 경우에는 전체 생존 분석에서는 유의한 차이를 보이지 않았으나, 무병 생존 분석

에서는 유의한 차이를 보여주고 있다(Figure 2).

최종 선정된 클론들은 gain 군에서는 10개의 클론이 선정되었으며, 이들은 1q23.1, 1q41, 1q44, 5p15.33, 8q21.3, 15q26.3, 17q12 및 21q22.3의 부위에 위치하고 있었고, loss군에서는 14q32.33 부위에서 1개의 clone이 선정되었다. 이 중에서 1q41 및 17q12 부위는 전이 유방암에서 그 발현이 증가해 있다는 보고가 있으며, (10) 1q44는 에스트로겐수용체 음성 유방암에서 나쁜 임상 예후와 관련이 있다는 보고가 있고, (11) 5p15.33는 유방암의 전신 재발과 관련이 있으며, (5) 21q22.3의 경우 정상이나 양성 유방 조직에 비해 유방암 조직에서 과발현되어 있다는 보고도 있다. (12)

전체 검체 수의 50% 이상에서 gain이나 loss가 있는 클론들을 살펴보면, 1p36.33, 8q23.3, 8q24.3, 8q24.12, 8q24.13, 8q24.21 및 20q13.33 부위에서 50% 이상의 gain이 관찰되었으며, 8p23.1, 10q11.22 및 14q32.33 부위에서 50% 이상의 loss가 관찰되었다(Table 3). 이처럼 정상 유방 조직에 비해 유방암에서 반복적으로 gain 또는 loss를 보이는 염색체의 부위들은 array CGH 실험 방법 또는 전통적인 CGH 실험 방법을 이용한 이전의 논문에서의 결과와 유사한 경향을 보이고 있으며, (13-16) 향후 많은 논문에서 공통적인 결과를 보이는 염색체의 부위에 대해서는 추가적인 연구의 필요성이 있을 것으로 사료된다.

최종 선정된 클론들을 대상으로 NCBI Map Viewer 및 PubMed 검색을 통해 암과의 연관성이 보고된 바 있는 유전자를 선정하였다(Table 4). 이와 같은 방식으로 선정된 유전자들에 대해서 다양한 고형암과 연관성이 보고되어 있었는데, 특히 NTRK1과 COL18A1은 다양한 암종에서 발암, 치료, 예후 등에 관련이 있다는 것이 보고되어 있었다. NTRK의 경우 췌장암, 갑상선암, 간암, 난소암, 침샘 종양, 폐암 등과의 관련성에 대해 보고되어 있었으며, (17-19) COL18A1의 경우에는 유방암, 위암, 폐암, 방광암, 간암 등과의 관련성에 대해 보고되어 있었다. 특히 COL18A1의 경우 다양한 암종에 대한 관련성이 보고되어 있을 뿐만 아니라, 유방암과의 관련성에 대한 내용도 보고되어 있었으며, 따라서 이 유전자를 이번 분석에서 유방암의 예후 인자와 가장 관련성이 높을 것으로 추정되는 후보 유전자로 최종 선정하였다.

Array CGH 실험 방법은 expression microarray가 RNA를 이용한 실험 방법인 것과는 달리, whole genome DNA를 이용한 방법으로서 DNA 자체의 변화를 측정할 수 있는 대신, RNA의 변화는 직접적으로 측정할 수 없다. 이를 보완하기 위해 본 연구에서는 array CGH 실험 방법으로 선정된 후보 유전자에 대한 RNA 발현을 검증 하기 위해 RT-PCR의 방법을 사용하였다. 이 실험을 위해 새로운 20명의 유방암 환자에 대해 20쌍의 유방암 조직 및 정상 유방 조직을 채취하여 실험을 시행하였는데, 이 결과 18명(90%)의 검체에서 정상 유방 조직에 비해 유방암 조직에

서 COL18A1의 발현이 높게 나왔으며(Figure 2), 이는 통계적으로 매우 유의하였다($p < 0.0001$). 발현의 차이를 보이지 않은 나머지 2명의 임상병리학적 정보상에서는 특이 소견이 관찰되지 않았다. 이를 통해 COL18A1 유전자의 RNA 역시 정상 조직에 비해 유방암에서 유의하게 그 발현이 증가되어 있음을 제시하고 있다.

COL18A1의 공식 명칭은 'collagen, type XVIII, alpha 1'이며 인간 염색체의 21q22.3 부위에 위치하고 있는 108,537 bp 크기의 유전자이다. COL18A1은 인간의 type XVIII 콜라겐의 알파 사슬(alpha chain)을 부호화(coding)하고 있으며 KNO, KNO1, FLJ27325, FLJ34914, MGC74745 등의 명칭으로도 알려져 있다. 이 유전자에 의해 생성되는 type XVIII 콜라겐의 C-말단 분절이 바로 endostatin이며, 이 물질은 인체 내에서 강력한 혈관 생성 억제 기능이 있음이 알려져 있다. COL18A1은 위암, 폐암, 방광암, 간암 및 유방암 등에서 발병, 치료, 예후 등과 관련이 있다는 것이 보고된 바 있다. Iizasa 등(20)은 collagen XVIII의 과발현이 비소세포 폐암에서 나쁜 예후와 연관되어 있으며, 이는 혈액의 endostatin 증가와 관련이 있다는 보고를 하였으며, Swidzińska 등(21)도 폐암 환자에서 혈액 내의 endostatin이 유의하게 증가되어 있음을 보고하였다. Woo 등(22)은 혈액 내 endostatin의 증가가 위암의 나쁜 예후와 연관되어 있음을 보고하기도 하였으며, Kikuchi 등(23)은 lentivirus를 통해 endostatin의 유전자를 주입함으로써, 방광암의 성장을 억제하였다는 보고를 한 바 있다. Hu 등(24)은 endostatin이나 collagen XVIII의 발현의 증가가 간암에서의 나쁜 예후가 있으며, 이는 VEGF의 증가와도 관련이 있음을 보고하였다. 한편 유방암과 관련하여, Balasubramanian 등(25)이 endostatin 4349A allele이 유방암과 관련 있음을 보고한 바 있으며, Beck 등(26)은 endostatin과 prolactin 길항제를 결합한 단백질이 유방암의 새로운 치료제가 될 수 있음을 보고하였으며, Lourenço 등(27)은 COL18A1 유전자의 104NM 다형성을 가진 경우 유방암의 발병 위험이 높음을 보고하기도 하였다.

본 연구는 array CGH 실험을 통해 연구자가 확인하고자 하는 목적에 부합하는 후보 유전자에 대한 선정 과정을 보여주고 있으며, 이를 이용하여 다양한 방법의 후속 연구를 진행할 수 있게 해 준다는 것을 보여주고 있다. 하지만, array CGH 실험 결과에 대한 분석 방법은 아직 공인된 방법이 없어, 연구자에 따라 다양한 분석 방법이 사용되고 있으며, (28) 이러한 분석 방법에 따라 분석 결과가 차이가 날 수 있어 결과의 해석에 신중을 요한다. 또한 array CGH 실험 비용이 비교적 고가인 관계로 많은 수의 시료를 사용하기 어렵다는 점이 통계적인 유의성을 감소시키는 주요 요인이 되고 있으며, 이를 극복하기 위해 메타분석 등의 방법을 사용하기도 한다. (10) Array CGH는 후보 유전자를 찾아내는 검색 기능의 의

미가 많으므로, 여기서 선택된 후보 유전자에 대한 정확한 검증은 또 다른 심도 있는 연구를 필요로 한다는 것에 유의할 필요가 있다. (2) 본 연구를 통해 COL18A1과 유방암의 관련성에 대한 가능성을 제시하고 있지만, 이에 대한 정확한 규명을 위해 다양한 방법의 추가 연구를 필요로 한다.

결론

본 연구는 array CGH 실험 방법을 통하여 정상 유방 조직과 대비되는 유방암 조직의 DNA에 대한 발현 양상을 분석하고, 이어 생존 분석을 시행함으로써, COL18A1을 유방암의 예후 인자와 연관성이 높은 후보 유전자로 선정하였으며, COL18A1에 대한 RNA의 발현 양상에 대해 새로운 유방 조직을 이용하여 확인하였다. COL18A1의 유방암의 예후 인자로서의 유용성에 대해서는 향후 추가적인 연구가 필요할 것으로 사료된다.

참고문헌

1. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer* 2007;7:545-53.
2. Tan DS, Reis-Filho JS. Comparative genomic hybridisation arrays: high-throughput tools to determine targeted therapy in breast cancer. *Pathobiology* 2008;75:63-74.
3. Davies JJ, Wilson IM, Lam WL. Array CGH technologies and their applications to cancer genomes. *Chromosome Res* 2005;13:237-48.
4. van Beers EH, Nederlof PM. Array-CGH and breast cancer. *Breast Cancer Res* 2006;8:210.
5. Hwang KT, Han W, Cho J, Lee JW, Ko E, Kim EK, et al. Genomic copy number alterations as predictive markers of systemic recurrence in breast cancer. *Int J Cancer* 2008;123:1807-15.
6. Chebil G, Bendahl PO, Fernö M; South Sweden Breast Cancer Group; North Sweden Breast Cancer Group. Estrogen and progesterone receptor assay in paraffin-embedded breast cancer--reproducibility of assessment. *Acta Oncol* 2003;42:43-7.
7. Kim HS, Jung JH, Park HY, Lee YH, Chung EJ, Kim MK, et al. Gene expression profile analysis of human breast cancer using cDNA microarrays. *J Korean Breast Cancer Soc* 2003;6:58-67.
8. Han W, Chung KW, Ahn SJ, Noh DY, Youn YK, Oh SK, et al. Gene expression profiles of primary breast cancer tissue using cDNA microarray. *J Korean Breast Cancer Soc* 2002;5:284-90.
9. Kim MY, Yim SH, Kwon MS, Kim TM, Shin SH, Kang HM, et al.

- Recurrent genomic alterations with impact on survival in colorectal cancer identified by genome-wide array comparative genomic hybridization. *Gastroenterology* 2006;131:1913-24.
10. Thomassen M, Tan Q, Kruse TA. Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast Cancer Res Treat* 2009;113:239-49.
 11. Nakatani K, Thompson DA, Barthel A, Sakaue H, Liu W, Weigel RJ, et al. Up-regulation of Akt3 in estrogen receptor-deficient breast cancers and androgen-independent prostate cancer lines. *J Biol Chem* 1999;274:21528-32.
 12. Liu D, Rudland PS, Sibson DR, Barraclough R. Identification of mRNAs differentially-expressed between benign and malignant breast tumour cells. *Br J Cancer* 2002;87:423-31.
 13. Han W, Han MR, Kang JJ, Bae JY, Lee JH, Bae YJ, et al. Genomic alterations identified by array comparative genomic hybridization as prognostic markers in tamoxifen-treated estrogen receptor-positive breast cancer. *BMC Cancer* 2006;6:92.
 14. Naylor TL, Greshock J, Wang Y, Colligon T, Yu QC, Clemmer V, et al. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. *Breast Cancer Res* 2005;7:R1186-98.
 15. Nessling M, Richter K, Schwaenen C, Roerig P, Wrobel G, Wessendorf S, et al. Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer Res* 2005;65:439-47.
 16. Rennstam K, Ahlstedt-Soini M, Baldetorp B, Bendahl PO, Borg A, Karhu R, et al. Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. *Cancer Res* 2003;63:8861-8.
 17. Dang C, Zhang Y, Ma Q, Shimahara Y. Expression of nerve growth factor receptors is correlated with progression and prognosis of human pancreatic cancer. *J Gastroenterol Hepatol* 2006;21:850-8.
 18. Brzezińska E, Karbownik M, Migdalska-Sek M, Pastuszek-Lewandowska D, Wloch J, Lewiński A. Molecular analysis of the RET and NTRK1 gene rearrangements in papillary thyroid carcinoma in the Polish population. Molecular analysis of the RET and NTRK1 gene rearrangements in papillary thyroid carcinoma in the Polish population. *Mutat Res* 2006;599:26-35.
 19. Tokusashi Y, Asai K, Tamakawa S, Yamamoto M, Yoshie M, Yaginuma Y, et al. Expression of NGF in hepatocellular carcinoma cells with its receptors in non-tumor cell components. *Int J Cancer* 2005;114:39-45.
 20. Iizasa T, Chang H, Suzuki M, Otsuji M, Yokoi S, Chiyo M, et al. Overexpression of collagen XVIII is associated with poor outcome and elevated levels of circulating serum endostatin in non-small cell lung cancer. *Clin Cancer Res* 2004;10:5361-6.
 21. Swidzińska E, Ossolińska M, Naumnik W, Trojan S, Chyczewska E. Serum endostatin levels in patients with lung carcinoma. *Rocz Akad Med Białymst* 2005;50:197-200.
 22. Woo IS, Kim KA, Jeon HM, Hong SH, Rho SY, Koh SJ, et al. Pre-treatment serum endostatin as a prognostic indicator in metastatic gastric carcinoma. *Int J Cancer* 2006;119:2901-6.
 23. Kikuchi E, Menendez S, Otori M, Cordon-Cardo C, Kasahara N, Bochner BH. Inhibition of orthotopic human bladder tumor growth by lentiviral gene transfer of endostatin. *Clin Cancer Res* 2004;10:1835-42.
 24. Hu TH, Huang CC, Wu CL, Lin PR, Liu SY, Lin JW, et al. Increased endostatin/collagen XVIII expression correlates with elevated VEGF level and poor prognosis in hepatocellular carcinoma. *Mod Pathol* 2005;18:663-72.
 25. Balasubramanian SP, Cross SS, Globe J, Cox A, Brown NJ, Reed MW. Endostatin gene variation and protein levels in breast cancer susceptibility and severity. *BMC Cancer* 2007;7:107.
 26. Beck MT, Chen NY, Franek KJ, Chen WY. Prolactin antagonist-endostatin fusion protein as a targeted dual-functional therapeutic agent for breast cancer. *Cancer Res* 2003;63:3598-604.
 27. Lourenço GJ, Cardoso-Filho C, Gonçalves NS, Shinzato JY, Zeferino LC, Nascimento H, et al. A high risk of occurrence of sporadic breast cancer in individuals with the 104NN polymorphism of the COL18A1 gene. *Breast Cancer Res Treat* 2006;100:335-8.
 28. Chari R, Lockwood WW, Lam WL. Computational methods for the analysis of array comparative genomic hybridization. *Cancer Inform* 2007;2:48-58.