

cDNA microarray 를 이용한
대장암에서의 간 전이와 연관된
유전자군 탐색

연세대학교 대학원
의 과 학 과
기 동 혁

cDNA microarray 를 이용한
대장암에서의 간 전이와 연관된
유전자군 탐색

연세대학교 대학원
의 과 학 과
기 동 혁

cDNA microarray 를 이용한
대장암에서의 간 전이와 연관된
유전자군 탐색

지도교수 라 선 영

이 논문을 석사학위 논문으로 제출함

2005 년 12 월 일

연세대학교 대학원

의 과 학 과

기 동 혁

기동혁의 석사 학위논문을 인준함

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

연세대학교 대학원

2005 년 12 월 일

차 례

국문요약	1
I. 서론	4
II. 재료 및 방법	7
1. 대상 환자 및 검체 채취	7
2. RNA 추출 및 증폭	7
가. Total RNA 추출	7
나. 일차 주형 cDNA 합성	8
다. 이차 주형 cDNA 합성	9
라. <i>In vitro</i> 전사	10
3. cDNA microarray	11
가. 형광표지 및 보합작용	11
나. Image scanning	13
4. 생물 정보학 기법을 이용한 결과 분석	13
가. 결과 전처리 과정	13
나. 유전자 발현 양상 분석	14
다. 유의 유전자군 선정	14
라. 간 전이 관련 후보 유전자 군의 기능 해설	16
III. 결과	18
1. Gene expression profiling 결과 전처리 과정	18
2. Unsupervised hierarchical clustering 을 이용한 유전자 발현 양상 관찰	25
3. 원발성 대장종양과 간 전이 종양조직에서 다르게 발현하는 유전자군 선별	27

가. 정상 대장과 간 조직에서의 장기 특이 유전자의 선별 및 학습군에서의 제거	27
나. 학습군의 학습과 교차 검증	33
다. 검증군을 이용한 검증 및 최적 유전자군 선별	38
4. 간 전이 관련 후보 유전자 군의 기능해설	48
가. Web DATABASE 를 이용한 분석	48
나. 선별된 유전자 중 전이 관련성이 알려진 유전자의 발현 양상 관찰	51
IV. 고찰	54
V. 결론	60
참고문헌	61
영문요약	68

그림 차례

Figure 1. Study scheme	17
Figure 2. Representative scanned image of cDNA microarray	22
Figure 3. M-A plot before and after normalization	23
Figure 4. Box plot before and after normalization	24
Figure 5. Two-way hierarchical clustering of all experiments	26
Figure 6. Hierarchical clustering of normal colon and liver tissues with 4,616 differentially expressed genes	32
Figure 7. Training with filtered 8,207 genes in 14 train set using PAM	35

Figure 8.	
A. Cross-validated probability at each threshold with 14 train set	36
B. Supervised hierarchical clustering at each threshold with 14 train set	37
Figure 9.	
A. Validation of selected genes with test probability at each threshold	40
B. Supervised hierarchical clustering of selected genes with test probability at each threshold	41
Figure 10. Expression of top 10 up-regulated genes in liver metastasis tumor	52
Figure 11. Expression of top 10 down-regulated genes in liver metastasis Tumor	53

표 차례

Table 1. Patients characteristics	20
Table 2. Components of patient samples	21
Table 3. List of top 20 up-regulated genes in normal liver tissue compared to normal colon tissue	28
Table 4. List of top 20 down -regulated genes in normal liver tissue compared to normal colon tissue	30
Table 5. List of 126 liver metastasis related genes	42
Table 6. Ontological information of 126 liver metastasis related genes	50

국문 요약

cDNA microarray를 이용한 대장암에서의 간 전이와 연관된 유전자군 탐색

대장암의 주된 사인은 간 전이에 의한 것으로 알려져 있다. 이러한 간 전이 기전을 이해하기 위하여 cDNA microarray를 이용한 gene expression profiling 기법을 사용하여 대장암에서의 간 전이에 관련되어 있는 유전자를 선별하였다.

대장암 환자 중 진단 당시 수술적 절제가 가능한 간 전이를 동반한 22명의 환자들로부터 수술 당시 정상 대장 조직, 원발성 대장 종양 조직, 간 전이 종양 조직을 각각 쌍으로 얻고, 가능한 경우 추가적으로 정상 간 조직과 대장 용종을 채취하였다. Total RNA를 추출하여 mRNA T7 linear amplification 방법을 이용해 mRNA만을 선택적으로 증폭한 후, 17K cDNA microarray에 reference RNA를 이용한 indirect 방법으로 microarray 실험을 수행하였다. 결과들을 평균화와 전처리 하고 unsupervised hierarchical clustering을 통하여 전체 유전자의 발현 양상을 살핀 후에, 14개의 학습군과 8개의 검증군으로 나누어 분석을 진행하였다. 학습군에서 Significant Analysis of Microarrays (SAM) 을 이용하여 정상 대장 조직과 간 조직에서 유의하게 발현이 차이가 나는 장기 특이 유전자를 선별하고 추후 분석에서 제외하였다. 장기 특이 유전자군이 제거된 유전자를 이용하여 학습군에서 원발성 대장 종양과 간 전이 종양에서의 특이 유전자를 Prediction Analysis of Microarrays (PAM)을 이용하여 선별 및 검정한 후, NIH-DAVID database

(<http://apps1.niaid.nih.gov/david/>)를 이용하여 유전자의 기능을 살펴보았다.

모든 실험 결과를 100 % non-missing proportion에서 선별된 12,823 개의 유전자를 이용하여 unsupervised hierarchical clustering 을 수행하였을 때, unsupervised clustering 임에도 불구하고 크게는 정상 대장, 정상 간, 원발성 대장 종양, 간 전이 종양 조직으로 구분되는 경향을 보였다. 정상 대장과 간 조직으로 SAM 을 이용하여 false discovery rate 0.062 % 에서 4,616 개의 장기 특이 유전자를 선별하고 이 유전자 군을 제거한 8,207 개 유전자를 이용하여 분석을 진행하였다. 14 쌍의 원발성 대장 종양과 간 전이 종양 조직 학습군을 PAM을 이용하여 학습하고 training error와 교차 검정의 misclassification error가 공통으로 0인 구간에서 81개, 126개, 252개의 유의 유전자군을 먼저 선별했다. 이 세 개의 유전자군을 각각 8 개의 독립된 검증군을 이용한 검정과 supervised hierarchical clustering을 통하여 원발성 대장 종양과 간 전이 종양을 적절히 구분할 수 있는 126 개의 최적 유전자군을 선별하였다.

이렇게 선별된 126 개의 유전자 중 간 전이 종양에서 38 개는 up regulation 되었고 88 개는 down regulation 되었다. 선별된 유전자에는 중에는 WNT5A, lipocalin 2, E-cadherin, deiodinase 등 기존에 알려진 oncogene 이나 cell adhesion 관련 유전자들과 16 개의 EST 가 포함되어 있었다. 특히 MMP-1, MMP-2, WNT5A 는 원발성 대장 종양에 비해 간 전이 종양에서 발현 양이 감소하였고, 이들은 간으로의 전이하는 초기단계에 관여하는 것으로 여겨진다. 반대로 TIMP-1은 발현 양이 증가함을 관찰하였다 ($p < 0.01$). 결론적으로,

대장암에서 whole genome의 유전 발현 양상을 살펴보고, 간 전이 과정에서 중요한 역할을 하리라 사료되는 126개의 유전자를 확인하였다.

핵심되는 말 : 대장암, 간 전이, Gene expression profiling, cDNA microarray

cDNA microarray 를 이용한 대장암에서의 간 전이와 연관된 유전자군 탐색

<지도교수 라 선 영>

연세대학교 대학원 의과학과

기 동 혁

I. 서론

전 세계적으로 2 번째로 빈번하게 발생하는 대장암은 식생활의 변화로 우리나라에서도 발생 빈도가 매우 빠르게 증가하고 있다¹. 대장암의 사인 중 가장 큰 요인은 간 전이에 의한 것으로 알려져 있으며, 수술 당시 간 전이를 지닌 환자의 5 년 생존율은 5~10 %정도로 낮다. 만일 환자를 진단하는 시점에서 종양의

잠재적인 높은 전이 능력이나 나쁜 예후를 미리 예측할 수 있게 되면 수술 후 추적 프로그램이나 항암보조요법을 수행할 환자 군을 구분할 수 있고 결과적으로 생존율을 높이는데 매우 유용할 것으로 기대할 수 있지만²⁻³ 최근 대장암의 진단과 치료에 있어서 많은 발전이 있음에도 불구하고, 진단 당시에 early stage 환자가 advanced stage 로 발전하거나 혹은 간으로의 원격 전이가 있을지에 대한 예측 가능성은 거의 희박하다.

대장암을 포함한 많은 종양들은 하나의 clonal origin을 갖지만 분열과 성장을 계속 하면서 다양한 유전적 변이가 축적되어 임상적으로 진단 당시에는 종양 내의 세포성장을, 핵형, antigenicity, immunogenicity, marker enzyme, 각종 약물에 대한 민감성 정도뿐만 아니라 침윤과 전이능력 등 생물학적 특성이 다양한 여러 세포들로 구성되게 된다. 이와 같은 종양 이형성(tumor heterogeneity)은 다양한 임상 양상과 치료 반응으로 현재 암 치료의 가장 중요한 제한점으로 사료된다. 특히, 이렇게 이형성을 이룬 세포들 중 proteinase 등의 효소와 세포부착에 관여하는 분자들을 조절할 수 있는 침윤과 전이 능력이 뛰어난 clone들에 의해 원격 전이가 일어나기 때문에⁴⁻⁵ 전이가 일어난 clone과 그렇지 않은 clone은 유전적으로 차이가 있으며 이러한 유전적 차이를 생물학적으로 이해하려는 노력들이 시도되고 있다. 현재까지 이러한 연구에 의해 알려진 생물학적 표지자로는 *KRAS2*, *p53*, *p21/WAF/CEP1*, *CD44*, *COX-2*, cytokeratin-19, proliferating cell nuclear antigen, MMP-9, prolactin, b-hCC, cyclin D1⁶, VEGF-C⁷, E-cadherin⁸ 등이 있지만, 대부분의 연구에서 한 개 혹은 제한된 수의 분자에 대해 진행되고 있어 암의 발생과 전이의 전 과정을 이해하는 데는 한계가 있

다.

cDNA microarray 기법은 한 번의 실험으로 수 천에서 수 만 개의 유전자 발현 정도를 동시에 확인할 수 있는 high-throughput 방법으로 생물학적이거나 임상적인 연구에 있어 복잡하고 조합된 종양세포 내의 유전적 변화를 감지할 수 있는 유용한 방법이다⁹⁻¹⁴. 많은 논문들에서 cDNA microarray 기법을 이용하여 tumor의 특성을 밝히고 subgroup을 임상적으로나 생물학적으로 밝히고자 하는 연구들을 보고하였다¹⁵⁻²⁰.

실제로 본 연구실에서도 위암이나²¹ Thymoma의 subgroup을 분류하거나 국소 진행성 직장암 환자들의 항암 약물 방사선 병행요법 결과를 바탕으로 치료에 대한 효과와 부작용을 예측할 수 있는 생물학적 표지자를 분자 유전학 수준에서 탐색, 선정하여 그 가능성을 입증했다. (GY Lee et al. submitted)

본 연구에서는 이러한 cDNA microarray 기법을 이용하여 간전이 가능성이 있는 대장암 환자의 원발성 대장 종양 (colorectal primary tumor)과 간 전이 종양 (liver metastasis tumor) 검체를 이용해 gene expression profiling 을 시행하여 간 전이와 관련된 유전자군을 탐색하여 간 전이의 세포병리학적 기전 연구과 그에 따른 개인 맞춤 치료의 근간을 제공하고자 한다.

II. 재료 및 방법

1. 대상 환자 및 검체 채취

2003년부터 2005년까지 연세대학교 의과대학 신촌 세브란스병원, 연세 암센터에 내원하여 대장암으로 진단받은 환자들 중 진단 당시

간으로의 원격 전이를 지니고 있는 22 명을 대상으로 하였다. 이들로부터 수술 당시 대장의 종양병소와 병소에서 멀리 떨어진 정상부위, 간으로 전이된 종양병소를 쌍으로 채취하였고, 이때 대장의 용종 (3 개)이나 간의 정상조직 (10 개)을 채취할 수 있는 경우 함께 채취하였다. 채취된 조직은 채취 직후 액체질소를 이용하여 급속 냉동하고 실험 시까지 $-80\text{ }^{\circ}\text{C}$ 에서 보관하였다. 모든 연구는 연세대학교 세브란스병원 임상연구 심의위원회 (Institutional Review Board)가 승인한 동의서에 자의로 동의한 환자를 대상으로 진행하였다.

2. RNA 추출 및 증폭

가. Total RNA 추출

보관된 100 mg 정도의 조직은 액체질소 내에서 막자와 막자사발을 이용하여 균질화 한 후, 1 mL의 TRIzol (Invitrogen, Carlsbad, CA, USA)을 이용하여 용해했다. 200 μl 의 chloroform을 넣고 $4\text{ }^{\circ}\text{C}$ 에서 10 분간 반응시킨 후, $4\text{ }^{\circ}\text{C}$ 에서 14,000 rpm으로 25 분간 원심분리를 하여 상층액만을 분리시켰다. 분리된 상층액과 Isopropanol을 $-20\text{ }^{\circ}\text{C}$ 에서 30 분간 반응시킨 후, $4\text{ }^{\circ}\text{C}$ 에서 14,000 rpm으로 25 분간 원심분리 하여 RNA를 침전시켰다. 침전된 RNA는 70 % 에틸 알콜로 수세하고, 상온에서 남아있는 알코올을 건조시키고, ultra-pure water에 녹였다. Total RNA의 양과 질은 Nanodrop (Nanodrop technology, Rockland, USA), 1 % agarose gel 전기 영동을 이용하여 확인했다. 대조군으로 사용할 Yonsei reference RNA (연세의대

암전이연구센터, Seoul, Korea)는 YCC-3, YCC-B1, HCT-116, SK-Hep-1, A549, HL-60, MOLT-4, HeLa, HT-1080, Caki-2, T98G의 11개 인간 종양 세포주로부터 추출한 같은 양의 total RNA를 섞어 준비했다²².

나. 일차 주형 cDNA 합성

RNA 증폭은 연세의대 암전이연구센터의 표준 protocol에 따라 진행하였다. 먼저 4 μ l의 total RNA를 주형으로 하여 일차 증폭을 수행했다. RNA 주형에 2 μ g의 oligo-dT/T7 primer (5'-GGCCAGTAATTGTAATACGACTCACTATAGGGAGG CGG-3', Genotech, Daejeon, Korea)와 9 μ l의 RNase-free water를 넣고 65 °C에서 10 분간 반응시켰다. 반응 후 2분 간 얼음에서 식힌 후, 4 μ l의 5 X first strand buffer (Invitrogen, Carlsbad, CA, USA), 2 μ l의 100 mM DTT (Invitrogen, Carlsbad, CA, USA), 2 μ l의 10 mM dNTP mix (Invitrogen, Carlsbad, CA, USA), 2 μ l의 RNAsin (Promega, Madison, WI, USA), 2 μ l의 SuperScript II (Invitrogen, Carlsbad, CA, USA)를 넣고 1시간 동안 42 °C에서 역전사 반응을 시켰다.

다. 이차 주형 cDNA 합성

이차 주형을 합성하기 위하여 일차 주형 cDNA가 만들어진 용액에 91 μ l의 RNase-free water, 30 μ l의 5 X second strand buffer (Invitrogen, Carlsbad, CA, USA), 3 μ l의 10 mM dNTP mix, 10 U의 DNA ligase, 4 U의 DNA polymerase I, 2 U의 RNase H를 넣어 16°C

에서 2 시간 동안 반응시켰다. 반응이 끝난 후에 20 U의 T4 DNA polymerase (Invitrogen, Carlsbad, CA, USA)를 첨가한 후 16 °C에서 5분간 반응시켰고 1 M NaOH와 0.5 M EDTA 각각 10 μ l씩을 첨가하여 65 °C에서 10분간 두어 반응을 비활성화했다. 그 후에 25 μ l의 Tris-HCl을 이용하여 중화하고 phenol : chloroform : isoamyl alcohol (25:24:1) 97 μ l를 첨가하여 잘 섞어준 후 Phase Lock Gel™ (Eppendorf, Westbury, NY, USA)에 넣어 13,000 RPM으로 2분간 원심 분리하여 cDNA만을 정제했다. cDNA가 존재하는 상층액을 따로 분리하여 상층액 부피의 0.5 배 만큼의 7.5M ammonium acetate와 2.5 배만큼의 95% EtOH을 넣은 후, 4 °C에서 13,000 RPM 이상으로 30 분간 원심 분리하여 cDNA pellet을 얻고, 80 % EtOH로 수세한 후 상온에서 건조시켰다. 건조된 cDNA pellet은 RNase-free water 8 μ l를 넣어 용해했다.

라. *In vitro* 전사

In vitro 전사는 합성된 두 가닥의 cDNA 를 이용하여 T7 MEGAscript kit (Ambion, Austin, TX, USA)를 이용하였다. 간략히 설명하면, 위에서 만들어진 8 μ l의 cDNA 용액에 75 mM rATP, rUTP, rCTP, rGTP, enzyme mix, 10X reaction buffer 를 각각 2 μ l씩 첨가하여 37 °C에서 5 시간 동안 반응시켰다. 이렇게 증폭된 mRNA 는 RNeasy mini kit (Qiagen, Valencia, CA, USA)를 이용하여 제조사의 설명에 따라 정제하여 얻었으며, 증폭된 mRNA 의 양과 질은 Nanodrop (Nanodrop technology, Rockland, USA)과 1 % agarose gel 전기 영동을 이용하여 평가하였다.

3. cDNA microarray

cDNA microarray는 알려진 유전자들과 EST를 포함하고 있는 17,104개의 유전자가 집적된 human 17K cDNA chip (CMRC-GT, Seoul, Korea)을 사용하였고 실험은 test sample과 reference RNA를 서로 비교하는 indirect 방법을 사용하였다. 이 때, test sample은 Cy5-dUTP를 이용하여 labeling 하고 reference RNA는 Cy3-dUTP로 labeling 하였다.

가. 형광표지 및 보합작용

이 과정은 CMRC standard protocol에 따라서 수행하였다. 간단히 설명하면, 증폭된 mRNA 2 μg 에 6 μg 의 random primer (Invitrogen, Carlsbad, CA, USA)를 넣어 RNase-free water로 전체 양이 21 μl 가 되도록 맞춘 후 65 $^{\circ}\text{C}$ 에서 10 분간 반응시켰다. 그리고 증폭된 mRNA와 random primer 혼합 용액에 8 μl 의 5 X first strand buffer, 4 μl 의 100 mM DTT, 2 μl 의 SuperScript II RT, 2 μl 의 20 X low dT/dNTP mix, 1 μl 의 RNAsin, 4 μl 의 Cy3 혹은 Cy5-dUTP (Genomic Tree, Daejeon, Korea)를 첨가하여 42 $^{\circ}\text{C}$ 에서 1시간 30분간 반응을 시켰다. 0.1 M NaOH 15 μl 를 첨가하여 RNA 주형을 분해시키고 65 $^{\circ}\text{C}$ 에서 30분간 두어 반응을 비활성화 하였다. 0.1 M NaOH와 동량의 0.1 M HCl로 중화한 후, QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA)를 이용하여 Cy3 혹은 Cy5로 labeling 된 probe들만을 정제하고 200 μl 의 ultra-pure water로

두 번씩 녹여냈다. 이렇게 얻어진 용액에 각각 20 μg 의 Human COT-1 DNA (Invitrogen, Carlsbad, CA, USA), yeast tRNA (Invitrogen, Carlsbad, CA, USA), poly(A) RNA (Sigma, Saint Louis, MO, USA)를 넣고 Microcon YM-30 column (Millipore, Bedford, MA, USA)를 이용하여 최종 양이 40 μl 가 되도록 농축시킨 후 100 $^{\circ}\text{C}$ 에서 2분간 변성 한 후 사용했다.

cDNA microarray chip은 사용하기 전에 3.5 X Sodium chloride/Sodium Citrate buffer (SSC), 0.1 % Sodium Dodesyl Sulfate (SDS), 10 mg/ml Bovine Serum Albumin (BSA)의 용액에서 42 $^{\circ}\text{C}$ 로 1시간 동안 pre-hybridization 한 후 3차 증류수로 2번 수세하고 isopropanol을 이용하여 완전히 건조시킨 후 사용하였다. 농축된 probe은 최종적으로 80 μl 의 25 % formamide, 5 X SSC, 0.1 % SDS의 상태로 맞추어 준비된 cDNA microarray chip에 덮개유리를 얹고 그 틈으로 주입한 후, 42 $^{\circ}\text{C}$ 에서 16시간 동안 빛이 차단된 상태로 반응시켰다. 반응이 끝난 microarray는 순서대로 2 X SSC, 0.1 % SDS 용액, 1 X SSC, 0.1 % SDS 용액, 0.2 X SSC 용액, 0.05 X SSC 용액에서 각각 2분씩 수세하고 회전 건조기를 이용하여 건조시켰다.

나. Image scanning

Hybridization이 끝난 microarray는 GenePix 4000B scanner (Axon Instruments, Foster City, CA, USA)를 이용하여 형광 신호를 감지하고 GenePix pro 4.0 software (Axon Instruments, Foster City, CA, USA)를 이용하여 이미지화 하였다. 이때 non-specific

형광 신호에 의한 간섭을 최소화하기 위하여 local background subtraction method에 따라 background 신호를 제거하고 수치화된 결과를 얻었다.

4. 생물 정보학 기법을 이용한 결과 분석

가. 결과 전처리 과정

얻어진 유전자 발현 결과는 실험 간의 variation을 줄이기 위하여 intensity-dependent within pin-tip group Lowess normalization²³ 방법을 이용하여 표준화 하였다. 모든 실험 간에서 100 % 존재하는 유전자만을 선별하고 (Non-missing proportion (NMP) 100 %), 반복 실험 간의 상관관계는 S-Plus 2000 software (Insightful, Seattle, WA, USA)를 이용하여 확인했다. 후에, 반복 실험 간의 log ratio수치는 평균하여 하나의 데이터화 하여 사용했다. microarray 내에 존재하는 중복 유전자는 Genebank ID를 이용하여 찾아낸 뒤 평균값을 취하여 분석을 시행하였다.

나. 유전자 발현 양상 분석

이렇게 얻어진 전체 자료들을 이용하여 unsupervised hierarchical clustering을 시행하여 전체 환자군의 유전자 발현 경향을 파악하였다 (Eisen software, Stanford Univ.)²⁴. TreeView (Eisen software, Stanford Univ.)²⁴ 에서 보여지는 결과에서 녹색, 빨간색, 검정색은 각각 유전자의 down-regulation, up-regulation 및 변화 없음을 나타낸다.

다. 유의 유전자군 선정

총 22명의 환자 중 간으로의 전이가 병리학적 소견으로 명확했던 18명의 환자를 random number generator를 이용하여 무작위로 학습군과 검증군을 나누었다. 학습군은 각각 쌍으로 14 예의 정상 대장, 대장 종양, 간 전이 종양 조직과 이들 중 정상 간 조직을 얻을 수 있었던 8예로 구성되어 있으며, 검증군은 각각 쌍으로 4예의 정상 대장, 대장 종양, 간 전이 종양 조직과, 실험은 진행하였으나 병리학적 소견에서 GIST나 SCC 형태의 종양 혹은, 간으로의 전이가 불분명한 4예로 구성되었다.

정상적인 대장과 간의 각 장기 간에 유전자 발현 양상이 다른 장기 특히 유전자 군을 선별하기 위하여 *t*-test 알고리즘을 바탕으로 분산의 영향을 최소화하고, 두 그룹 이상을 비교할 수 있는 알고리즘이 추가되어 microarray 결과를 분석할 수 있도록 개발된 Significance Analysis of Microarray (SAM)²⁵ 을 사용하여 학습군의 정상 대장과 간 조직을 이용하여 two-class SAM을 수행하였다. 유의 유전자는 False Discovery Rate (FDR) 을 조절하여 결정했다.

원발성 대장 종양과 간 전이 종양에서 다르게 발현하는 간 전이 관련 후보 유전자 군의 선별을 위해 Prediction Analysis of Microarray (PAM)²⁶ 을 이용했다. 유전자 선별은 학습군을 이용하여 가장 작은 training error rate 과 최대의 cross-validation probability, 그리고 확연히 두 군집으로 구분 될 수 있는 hierarchical clustering 결과를 토대로 결정했다. 이렇게 선별된 유전자는 독립된 검증군을 이용하여 검증하였다 (Figure 1).

라. 간 전이 관련 후보 유전자 군의 기능 해설

통계적 방법과 생물 정보학 기법을 통하여 선별된, 대장 종양과 간 전이 종양에서 다르게 발현되는 최적의 후보 유전자들의 기능적인 부분은 University of Stanford와 National Institutes of Health (NIH) 에서 제공하는 웹사이트인 <http://source.stanford.edu/cgi-bin/source/sourceSearch> (SOURCE) 와 <http://apps1.niaid.nih.gov/david> (DAVID) DATABASE 를 이용하여 검색하였으며, 각 사이트에서 제공되는 유전자의 기능 정보가 다를 경우, 그 두 가지의 정보를 모두 참고하였다.

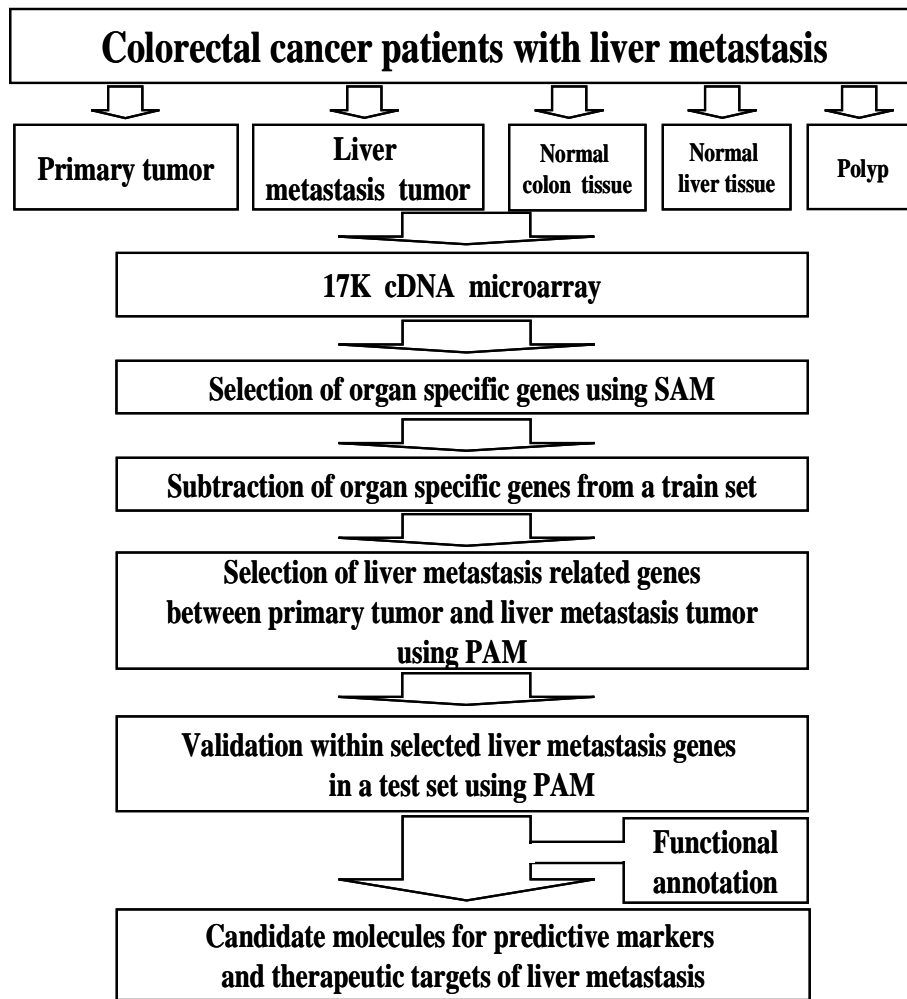


Figure 1. Study scheme

SAM; Significance Analysis of Microarray, PAM; Prediction Analysis of Microarray.

III. 결과

1. Gene expression profiling 결과 전처리 과정

원발성 대장 종양과 간으로의 원격전이 된 종양에서 다르게 발현되는 유전자 군의 선별을 위해 진단 당시 간으로의 원격 전이가 있는 22 명의 대장암 환자로부터 수술 시 얻어진 원발성 종양과 전이 종양조직, 정상 대장과 간 조직, 양성종양 조직의 총 79 예의 검체를 이용하여 gene expression profiling 을 수행하였다 (Table 1, 2). 추가적으로, 무작위적인 14 번의 반복실험을 수행하여 상관계수 0.96 ± 0.013 이상의 높은 재현성을 확인하였다. 전체 실험에서 flag %는 4.20 ± 1.67 %로 실험 간에 나타나는 variation 은 차이가 없음을 확인했다 (Figure 2).

Gene expression profiling의 수치화 된 결과는 실험 간의 variation을 줄이기 위하여 표준화 하였으며, 표준화 수행하기 전과 후의 변화 정도를 M-A plot과 Box plot을 이용하여 관찰 하였다 (Figure 3, 4). M-A plot에서 검정색은 표준화 전에 각 spot이 갖는 log ratio와 intensity를 나타내는 것이고 주황색은 표준화 수행 후에 보정되어 나타난 값을 의미한다. 표준화 전과 후에 큰 변화가 많지 않아 양질의 실험과 실험의 재현성을 확인하였다. 표준화 수행 후, NMP 100%를 취해 반복실험의 결과와 중복 유전자는 평균하여, 최종적으로 12,823 개의 유전자를 분석에 사용하였다.

Gender	Male : Female	15:7
---------------	---------------	------

Table 1. Patients characteristics (n=22)

Age (years)	Median (Range)	63 (33-77)
Primary tumor location		
Colon & Sigmoid	8	
Rectal	13	
Anorectal	1	
Tumor size (cm)		
Primary tumor	Median (Range)	4.5 (3-9)
Liver metastasis tumor	Median (Range)	4 (2-8)
Histology		
Adenocarcinoma	20	
SCC[†]	1	
GIST	1	
Duke's stage of primary tumor[‡]		
B	4	
C	17	

[†] Metachronous

‡ Evaluated in 21 carcinoma tissues except GIST

Patient	Normal colorectal tissue	Polyp	Primary colorectal tumor	Liver metastasis tumor	Normal liver tissue
---------	--------------------------	-------	--------------------------	------------------------	---------------------

Table 2. Components of patient samples

CPF001 [†]	+	+	+	+	+
CPF002	+	+	+	+	
CPF003	+	+	+	+	+
CPL001	+		+	+	
CPL002	+		+	+	
CPL003	+		+	+	
CPL004	+		+	+	
CPL005	+		+	+	+
CPL006	+		+	+	
CPL007 [*]	+		+	+	+
CPL008	+		+	+	
CPL009	+		+	+	
CPL010 [‡]	+		+	+	
CPL011	+		+	+	
CPL012	+		+	+	+
CPL013	+		+	+	+
CPL014	+		+	+	+
CPL015	+		+	+	+
CPL016	+		+	+	+
CPL017	+		+	+	+

CPL018 [*]	+	+	+
CPL019	+	+	+

[†] SCC, [‡] GIST, ^{*} uncertain liver metastasis, the (+) represent completed experience.

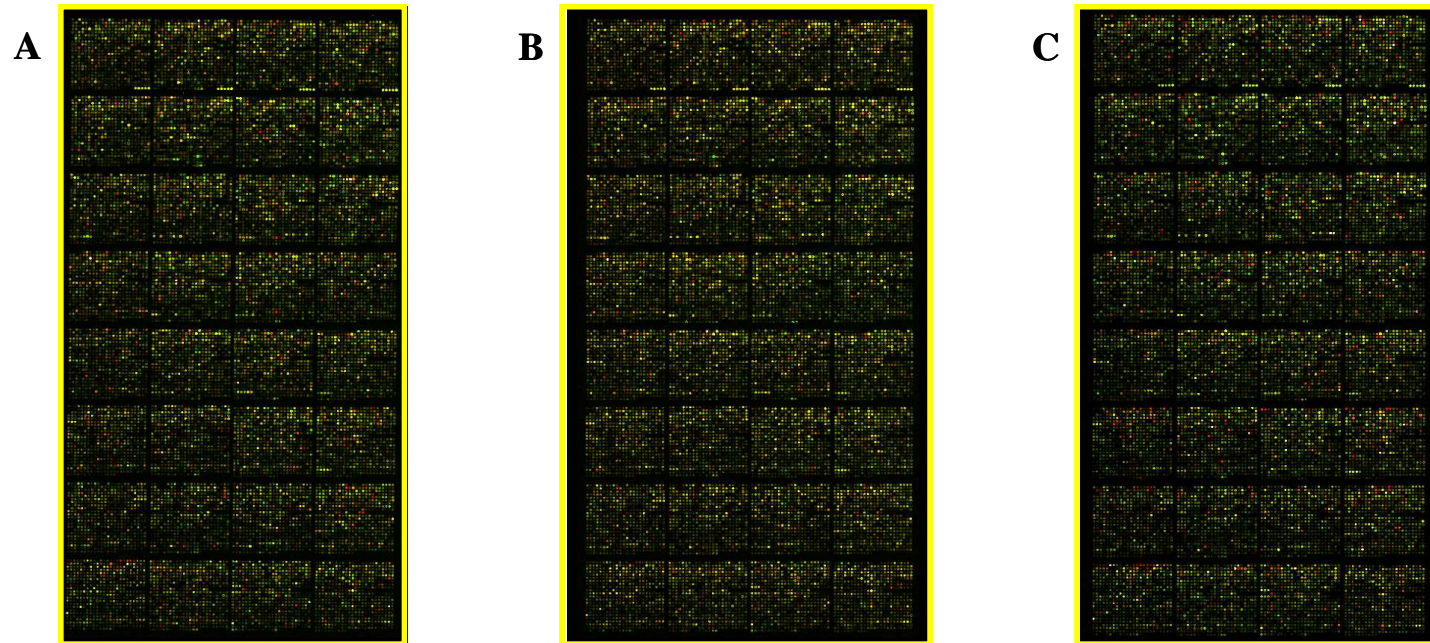


Figure 2. Representative scanned image of cDNA microarray
A, Normal colorectal tissue. B, Primary colorectal tissue. C, Liver metastasis tumor.

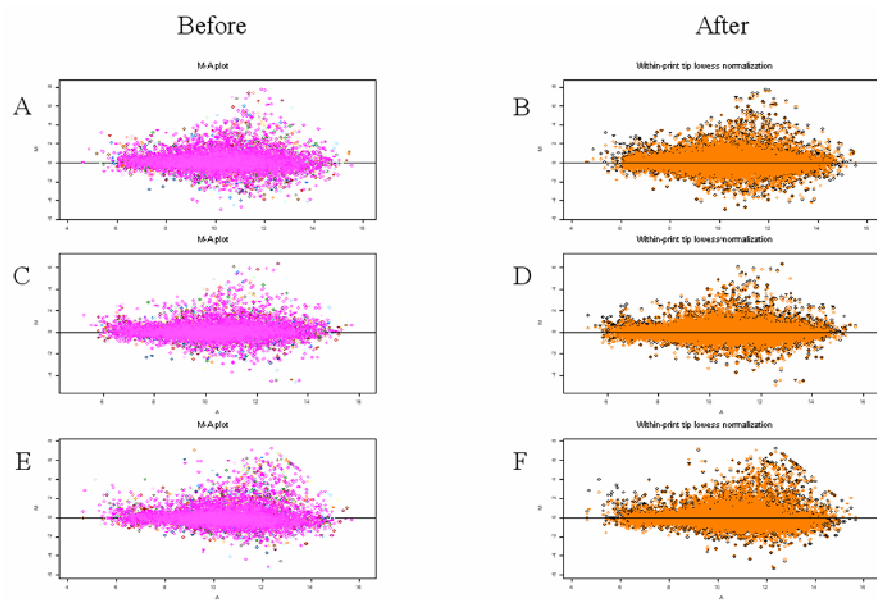


Figure 3. M-A plot before and after normalization

A and B, Normal colorectal tissue. C and D, Primary colorectal tumor. E and F, Liver metastasis tumor. The log ratios ($M = \log_2 R/G$) are plotted on the y-axis against the log of the geometric mean of the signal intensities ($A = \log_2 \sqrt{R/G}$) for each spot on the slides.

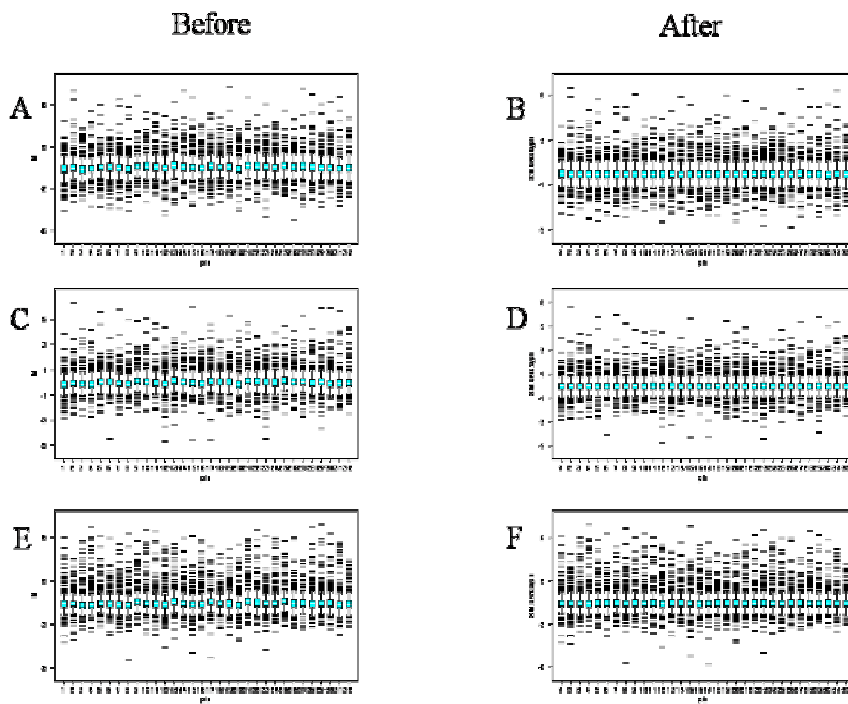


Figure 4. Box plot before and after normalization

A and B, Normal colorectal tissue. C and D, Primary colorectal tumor. E and F, Liver metastasis tumor. This plot displays a statistical summary consisting of the median, upper and lower quartiles as well as the range. The central box in a plot represents the Inter Quartile Range which is defined as the difference between 75th and 25th percentile or the upper and lower quartiles.

2. Unsupervised hierarchical clustering을 이용한 유전자 발현 양상 관찰

Gene expression profiling 실험을 수행 한, 전체 79개의 RNA 수

준에서의 유전적 변화 양상을 2차원 계층 군집 분석을 수행하여 관찰하였다 (Figure 5). Unsupervised hierarchical clustering 임에도 불구하고 크게는 정상 대장조직, 정상 간 조직, 종양 조직의 세 부분으로 구분되는 양상을 보였으며 종양 조직군에서도 원발성 종양과 간 전이 종양의 구분이 어느 정도 가능한 것을 관찰하였다. 본 결과에서는 같은 종양이라 할지라도 원발성 종양과 전이 종양의 유전자 발현이 다르다는 것을 제시하며 각 장기 별로도 유전자의 발현 양상이 다름을 관찰할 수 있었다.

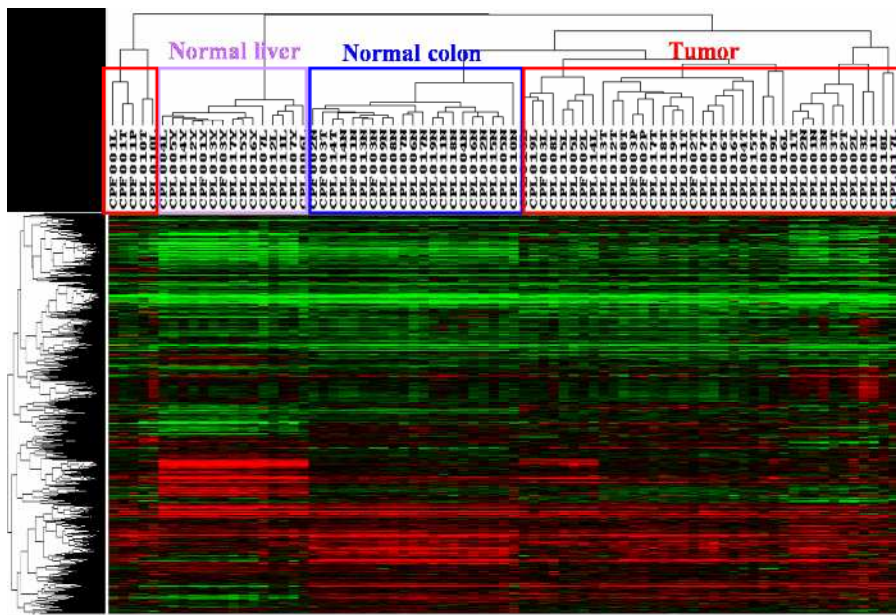


Figure 5. Two-way hierarchical clustering of all experiments
 The N, T, L, Y and P represent normal colorectal tissue, primary colorectal tumor, liver metastasis tumor, normal liver tissue and polyp, respectively. A number represents each patient's sample.

3. 원발성 대장종양과 간 전이 종양조직에서 다르게 발현하는 유전자군 선별

가. 정상 대장과 간 조직에서의 장기 특이 유전자의 선별 및 학습군에서의 제거

장기 특이 유전자가 결과분석에 있어 영향을 미치는 것을 방지하기 위하여 이들 유전자를 전체분석에서 제외했다. 즉, 대장과 간의 장기 특이 유전자를 제거하고자 학습군의 14개의 정상 대장 조직과 8개의 정상 간 조직의 microarray 결과를 supervised two-class SAM을 수행한 결과, FDR 0.062 %에서 통계학적으로 유의한 4,616 개의 장기 특이 유전자 군을 선별하였다 (Table 3, 4). 이 유전자들을 hierarchical clustering하여 TreeView를 통하여 시각화하였고 정상 대장과 간 조직은 이들 유전자만으로 확연히 구분되어짐을 확인하였다 (Figure 6). 이렇게 분류된 유전자 군을 학습군의 원발성 대장 종양과 전이 종양 유전자에서 제거하여 8,207 개의 유전자를 선별하여 분석을 수행하였다.

Table 3. List of top 20 up-regulated genes in normal liver tissue compared to normal colon tissue

No	Accession No.	Name	Normal Colon	Normal Liver	Ratio
1	H50500	cytochrome P450, subfamily IIE (ethanol-inducible)	-0.11	6.94	7.1
2	AA233185	insulin-like growth factor binding protein 1	-1.93	5.86	7.8
3	AA702487	G5C protein	-0.18	5.89	6.1
4	AA011414	fibrinogen, A alpha polypeptide	-2.16	5.88	8.0
5	AW001034	group-specific component (vitamin D binding protein)	0.50	5.66	5.2
6	AA961402	apolipoprotein A-II	-0.94	6.44	7.4
7	T73090	plasminogen	0.25	6.42	6.2
8	AI985788	haptoglobin	0.06	6.87	6.8
9	AI343711	aldehyde oxidase 1	-0.16	6.29	6.5
10	T62036	complement component 4-binding protein, alpha	-0.37	4.91	5.3

Normal colon and liver denotes the mean log R/G of all normal colon and liver samples respectively in training set. Ratio denotes the log R/G expression ratio between normal colon and liver samples.

Table 3. Continued

No	Accession No.	Name	Normal Colon	Normal Liver	Ratio
11	AA677287	fibrinogen-like 1	-2.88	4.76	7.6
12	AA700876	orosomuroid 1	-0.58	4.85	5.4
13	AI283497	hemopexin	-0.11	5.91	6.0

14	AA931102	hypothetical protein FLJ21935	-1.63	4.28	5.9
15	AI829222	transthyretin (prealbumin, amyloidosis type I)	0.47	7.35	6.9
16	H05140	regucalcin (senescence marker protein-30)	-0.79	4.58	5.4
17	AI001134	apolipoprotein A-I	-0.13	5.60	5.7
18	N66008	Putative prostate cancer tumor suppressor	0.25	7.07	6.8
19	AA918982	Homo sapiens cDNA: FLJ22704 fis, clone HSI12602	-1.46	4.76	6.2
20	AA699812	methionine adenosyltransferase I, alpha	-0.39	5.35	5.7

Table 4. List of top 20 down -regulated genes in normal liver tissue compared to normal colon tissue

No	Accession No.	Name	Normal Colon	Normal Liver	Ratio
1	AW075441	tumor-associated calcium signal transducer 1	3.46	-3.76	-7.2
2	AI340883	tumor-associated calcium signal transducer 1	3.41	-2.86	-6.3
3	AW009764	chloride channel, calcium activated, family member 1	7.27	0.51	-6.8
4	AI339538	solute carrier family 26, member 3	7.69	0.13	-7.6
5	AW082097	protease inhibitor 3, skin-derived (SKALP)	1.95	-4.45	-6.4
6	AA130579	lectin, galactoside-binding, soluble, 4 (galectin 4)	3.97	-1.07	-5.0
7	AA630328	lectin, galactoside-binding, soluble, 3 (galectin 3)	2.55	-1.90	-4.4
8	AA259115	hypothetical protein FLJ20160	2.02	-2.01	-4.0
9	H59916	CD24 antigen (small cell lung carcinoma cluster 4	2.61	-2.71	-5.3

10	AI318604	antigen) hypothetical protein FLJ20063	2.84	-2.15	-5.0
----	----------	---	------	-------	------

Table 4. Continued

No	Accession No.	Name	Normal Colon	Normal Liver	Ratio
11	AI364374	Ste-20 related kinase	0.30	-2.70	-3.0
12	H23978	general transcription factor IIB	7.02	0.87	-6.2
13	N49284	v-myb avian myeloblastosis viral oncogene homolog	-0.84	-4.33	-3.5
14	AI955582	glucosaminyl (N-acetyl) transferase 3, mucin type	2.32	-1.65	-4.0
15	AA088861	cadherin 17, LI cadherin (liver-intestine)	1.89	-1.17	-3.1
16	AA454597	golgi membrane protein GP73	2.25	-1.21	-3.5
17	AI675714	anterior gradient 2 (<i>Xenopus laevis</i>) homolog	2.11	-3.12	-5.2
18	H84871	ESTs	0.27	-2.66	-2.9
19	AA976699	chromogranin A (parathyroid secretory protein 1)	3.39	-1.39	-4.8
20	AA284492	tetraspan 3	1.04	-2.19	-3.2

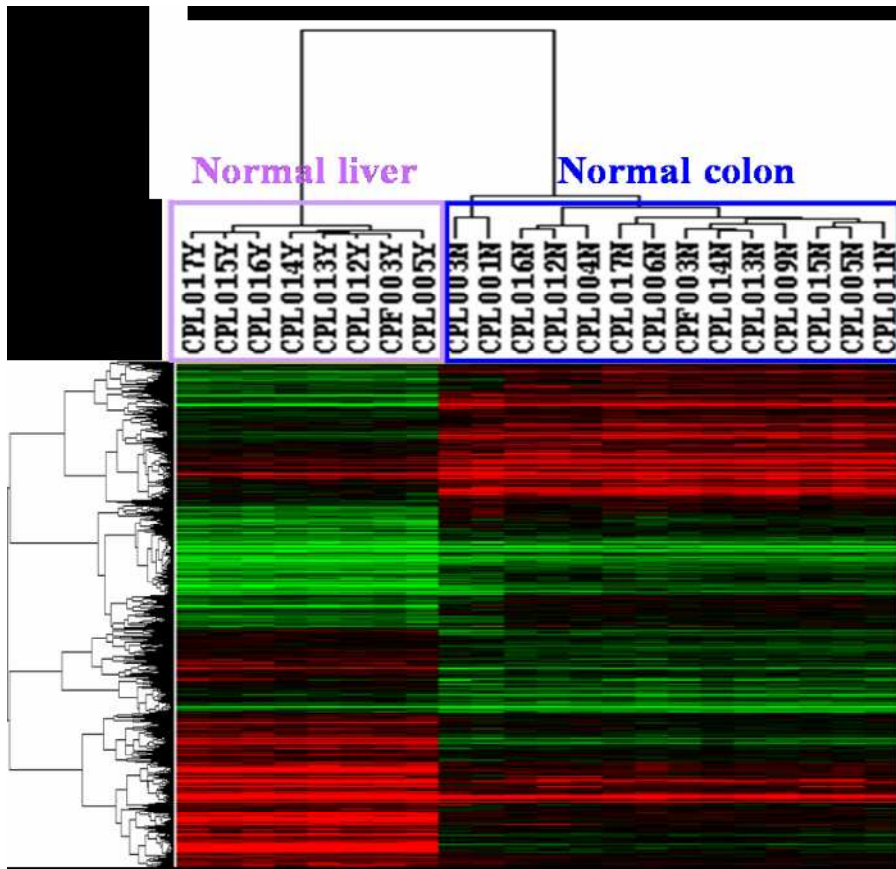


Figure 6. Hierarchical clustering of normal colon and liver tissues with 4,616 differentially expressed genes

나. 학습군의 학습과 교차 검증

장기 특이 유전자군을 제외한 8,207 개의 유전자로 학습군에서 원발성 대장 종양과 간 전이 종양조직을 판별 할 수 있는 유전자군을 찾고자, PAM을 시행하여 두 조직 간의 발현 양이 통계학적으로 유의한 유전자 군을 선별하였다. 유의한 유전자 수를 줄여가면서 확인하여 training error rate이 0 %인 구간을 확인하였고 이 구간에서 38 개부터 1899 개까지의 유전자군이 존재하였다 (Figure 7-A). 유의한 유전자군의 범위를 줄이기 위하여 학습군 내의 원발성 대장 종양과 간 전이 종양을 각각의 군으로 나누어 교차 검정을 실시하여 두 조직 모두에서 misclassification error rate이 0 %인 구간을 확인했고, 그 구간에서 각각 81, 126, 252 개의 유전자를 선별하였다 (Figure 7-B). 이렇게 선별된 유전자를 이용하여 환자 개인별로 교차검정을 이용한 예측확률을 산출했을 때 모든 결과에서 원발성 대장종양과 간 전이 종양 조직을 100 % 구분은 가능하였으나, threshold 가 1.7751에서 선별된 81개의 유전자로 검증한 결과에서는 하나의 예에서 예측확률이 61.8 %로 다른 유전자 군으로 검증한 것보다 낮은 양상을 보였다 (Figure 8-A).

또한 선별된 세 개의 유전자군 별로 유전자 발현양을 확인하여 hierarchical clustering하고 각각의 경향을 확인한 결과, 모든 결과에서 완전하게 원발성 대장 종양과 간 전이 종양으로 나뉘는 것을 확인했다 (Figure 8-B).

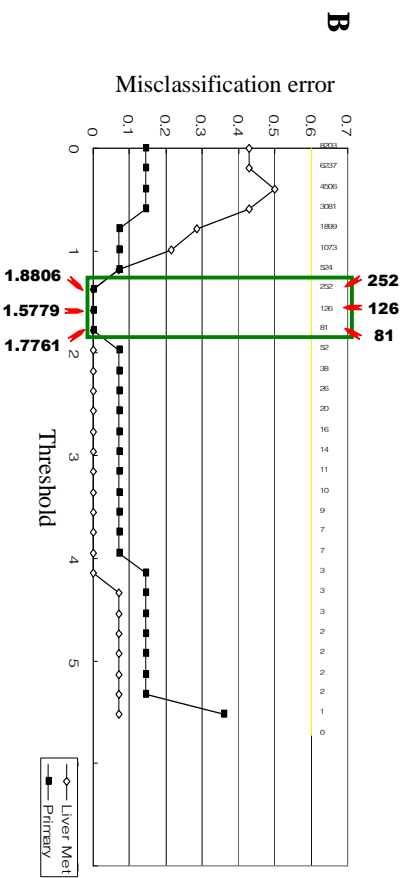
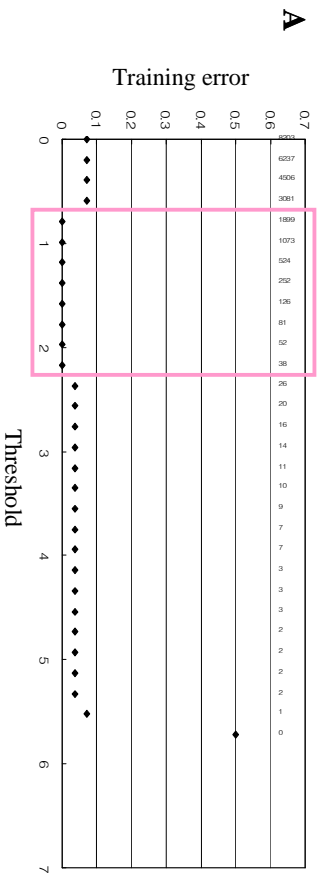


Figure 7. Training with filtered 8,207 genes in 14 train set using PAM
A, Training error plot. B, Individual cross-validation plot.

A

■ Primary tumor ◇ Liver metastasis tumor

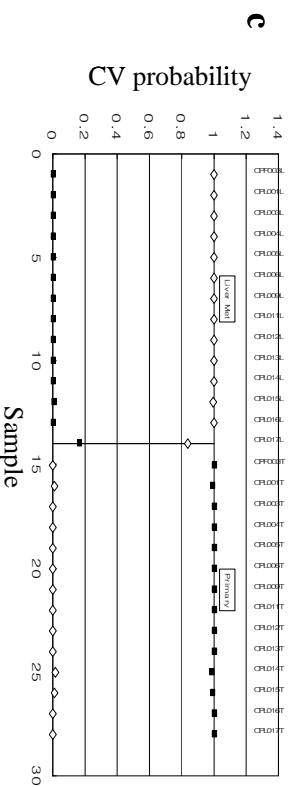
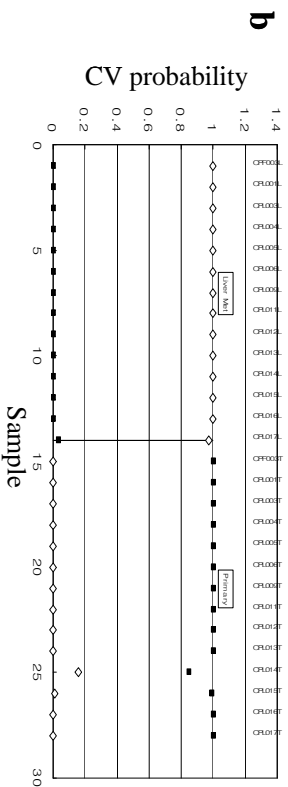
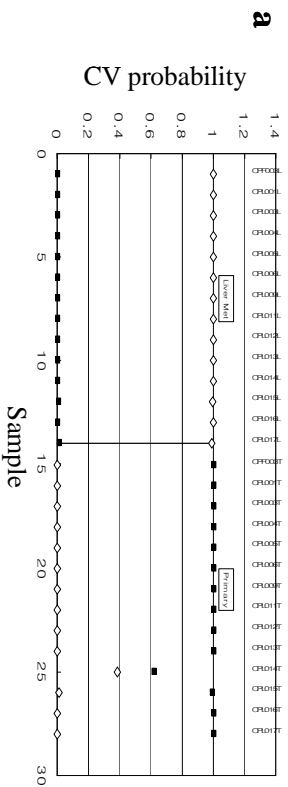
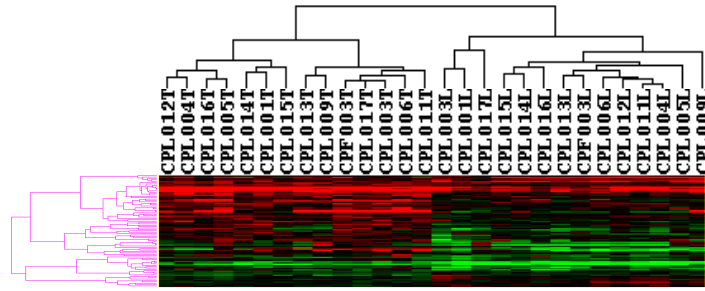


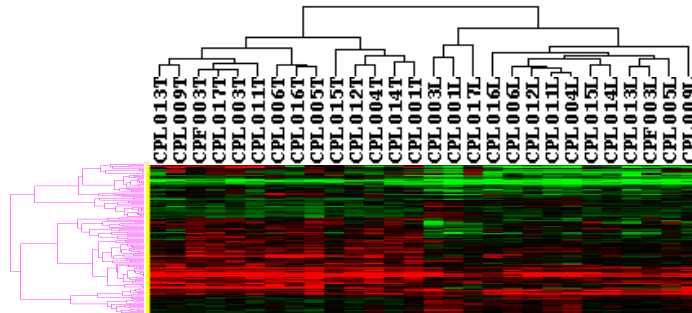
Figure 8-A. Cross-validated probability at each threshold with 14 train set
a, 81 genes at threshold 1.7761. **b**, 126 genes at threshold 1.5779. **c**, 252 genes at threshold 1.8806.

B

a



b



c

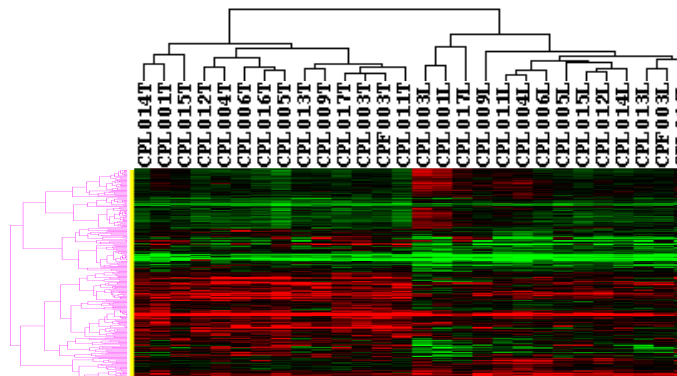


Figure 8-B. Supervised hierarchical clustering at each threshold with 14 train set

a, 81 genes at threshold 1.7761. **b**, 126 genes at threshold 1.5779. **c**, 252 genes at threshold 1.8806. The T and L represent primary colorectal tumor and liver metastasis tumor.

다. 검증군을 이용한 검정 및 최적 유전자군 선별

상기와 같이 학습군에서 선별된 세 군의 유전자들을 이용하여 학습군과는 독립적인 8예의 검증군으로 PAM을 사용하여 검정을 실시하였고, 세 군의 유전자 군에서 모두 원발성 대장 종양과 간 전이 종양을 100% 구분함을 확인하였다 (Figure 9-A). 세 유전자군의 검증 결과 모두에서 원발성 대장암의 구분은 가능했고 간으로의 전이가 불분명 하거나 SCC type인 검증군 예에 대한 특이점은 발견되지 않았다. GIST 예의 경우 81개, 126개, 252개 유전자군에서 각각 간 전이 종양의 predicted test probability가 84.3 %, 87.2 %, 98.5 %였다.

Hierarchical clustering 결과에서는 81 개 유전자 군과 126 개 유전자 군에서는 크게 원발성 대장 종양과 간 전이 종양의 두 가지로 나뉘지며, GIST로 확인된 CPL010 만이 원발성 대장 종양과 간 전이 종양 모두가 간 전이 종양군의 가지로 포함되어 있었다. 이와 달리, 256개 유전자 군의 hierarchical clustering은 크게 원발성 대장 종양과 간 전이 종양의 두 개 가지로 나뉘지기는 하나, GIST로 판명된 CPL010 외에 확실한 원발성 대장 종양과 간 전이 종양을 지닌 CPL002와 CPF002가 섞이는 경향을 나타냈다 (Figure 9-B).

위의 PAM의 결과와 hierarchical clustering 결과를 종합하여 최적의 원발성 대장 종양과 간 전이 종양에서 다르게 발현되는 간 전이 관련 후보 유전자를 아래와 같은 기준으로 선별하였다. 첫째, 학습군의 교차검정 결과에서 높은 probability를 보이며 둘째, 검증군의 probability가 적절한 수준으로 예측되고 셋째, hierarchical clustering으로 원발성 대장 종양과 간 전이 종양의 두 군으로 구분이 가능한 지에 대한 여부를 확인했다. 첫번째 기준

으로는 126개와 252개 유전자군이 적합하였고, 두번째 기준에서는 81개와 126개의 유전자군이 기준과 잘 부합하였다. 세번째 기준에서는 126개 유전자군이 적절한 결과를 보여주어 최종적으로 threshold 1.5779에서 원발성 대장 종양과 간 전이 종양의 두 군에서의 유전자 발현이 유의하게 다른 126 개의 유전자를 선별하였다 (Table 5). 선별된 126개의 유전자는 38개가 간 전이 종양에서 원발성 대장 종양에 비해 up-regulation 되어 있고 88개는 down-regulation 되어 있으며, 기능이 알려진 유전자들 외에 16 개의 EST 가 포함되어 있었다.

A

■ Primary tumor ◇ Liver metastasis tumor

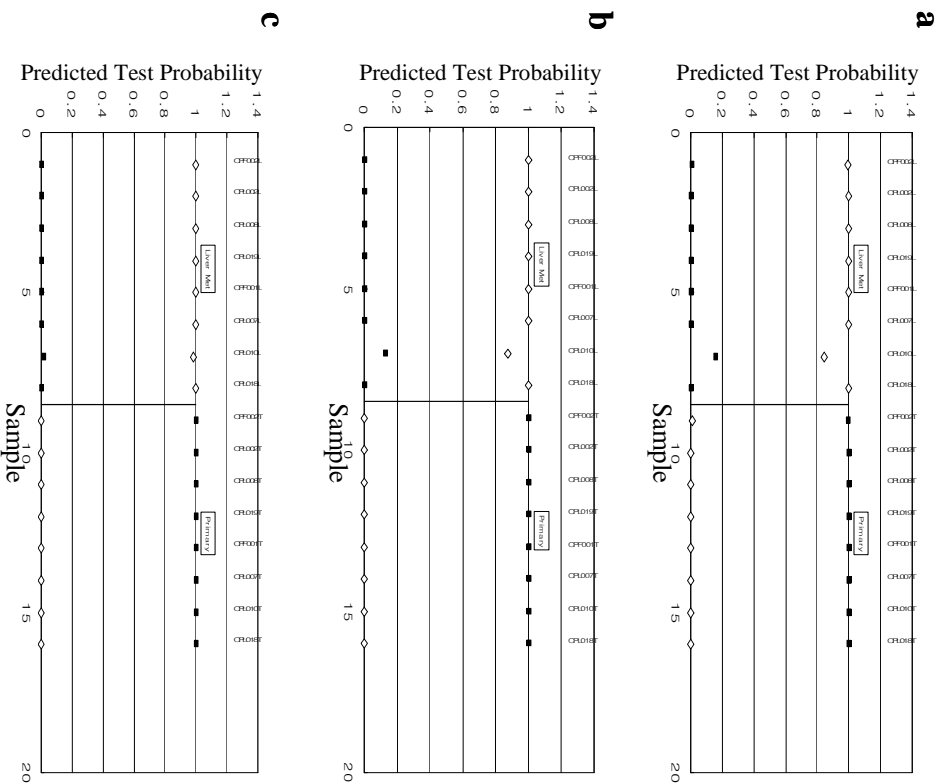


Figure 9-A. Validation of selected genes with test probability at each threshold

a. 81 genes at threshold 1.7761. **b.** 126 genes at threshold 1.5779. **c.** 252 genes at threshold 1.8806.

B

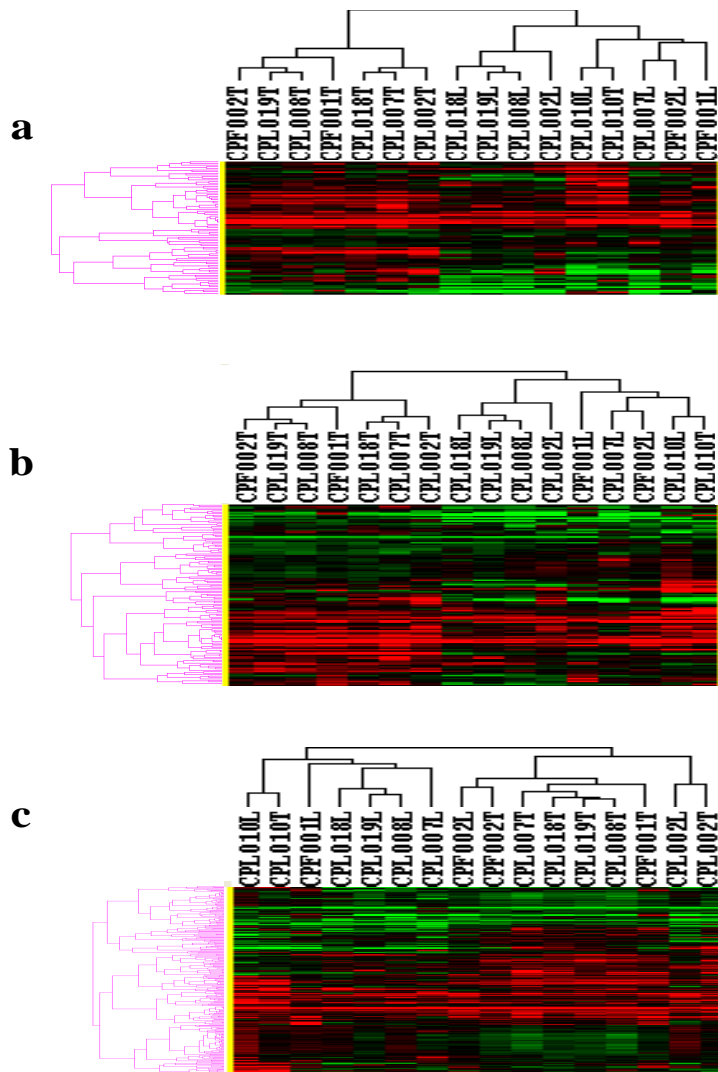


Figure 9-B. Supervised hierarchical clustering of selected genes with test probability at each threshold

a, 81 genes at threshold 1.7761. **b**, 126 genes at threshold 1.5779. **c**, 252 genes at threshold 1.8806.

Table 5. List of 126 liver metastasis related genes

ID	Name	Biological process	Primary tumor	Liver Mets. tumor	Ratio	Fold change
AA844864	regenerating islet-derived 1 beta (pancreatic stone protein, pancreatic thread protein)	cell adhesion	2.52	-0.10	-2.6	-6.15
AI671174	cadherin 1, type 1, E-cadherin (epithelial)	cell adhesion	1.40	0.91	-0.5	-1.41
AA287024	ectonucleoside triphosphate diphosphohydrolase 1	cell adhesion	1.15	0.68	-0.5	-1.38
AA436152	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	cell adhesion	0.82	0.41	-0.4	-1.33
AI888888	melanoma adhesion molecule	cell adhesion	0.96	0.57	-0.4	-1.31
N69335	collagen, type IX, alpha 1	cell adhesion	0.45	0.13	-0.3	-1.25
AA159577	apomucin	cell adhesion	0.32	0.04	-0.3	-1.21
AA666363	ficolin (collagen/fibrinogen domain-containing) 3 (Hakata antigen)	cell adhesion	-0.11	0.22	0.3	1.26
AI652836	T cell activation, increased late expression	cell adhesion	-0.08	0.25	0.3	1.26
AA449715	sushi-repeat-containing protein, X chromosome	cell adhesion	-0.79	-0.45	0.3	1.26
AI655374	stromal cell-derived factor 1	cell adhesion	1.43	2.73	1.3	2.47
AA400973	lipocalin 2 (oncogene 24p3)	cell growth	0.58	-1.41	-2.0	-3.97
AA287828	up-regulated by BCG-CWS	cell growth	0.52	-0.38	-0.9	-1.86
AA521384	up-regulated by BCG-CWS	cell growth	0.32	-0.55	-0.9	-1.82
AI992096	H3 histone family, member K	cell growth	0.71	0.16	-0.6	-1.47
T64134	small inducible cytokine subfamily A (Cys-Cys), member 13	cell growth	0.98	0.44	-0.5	-1.46
AA029299	potassium large conductance calcium-activated channel, subfamily M, beta member 1	cell growth	0.80	0.32	-0.5	-1.40
AI126362	choroideremia (Rab escort protein 1)	cell growth	0.43	0.04	-0.4	-1.31
AA504160	ATPase, H ⁺ transporting, lysosomal (vacuolar proton pump), alpha polypeptide, 70kD, isoform 1	cell growth	0.09	-0.30	-0.4	-1.31
AA709044	ADP-ribosylation factor GTPase activating protein 1	cell growth	0.12	-0.28	-0.4	-1.31
AI076718	H2B histone family, member R	cell growth	0.20	-0.06	-0.3	-1.20
AI342012	macrophage scavenger receptor 1	cell growth	0.96	1.41	0.4	1.36

Primary tumor denoted the mean log R/G of all primary colorectal cancer samples. Liver Mets. tumor denoted log R/G of all liver metastasis tumor samples. Ratio denoted the expression ratio between primary colorectal cancer and liver metastasis tumor samples.

Table 5. continued

ID	Name	Biological process	Primary tumor	Liver Mets. tumor	Ratio	Fold change
AI923531	deleted in malignant brain tumors 1	protein metabolism	0.50	-1.68	-2.2	-4.55
W58368	eukaryotic translation initiation factor 2B, subunit 3 (gamma, 58kD)	protein metabolism	-1.02	-1.84	-0.8	-1.76
AA453813	sialyltransferase 4C (beta-galactosidase alpha-2,3-sialyltransferase)	protein metabolism	0.36	-0.20	-0.6	-1.47
N72137	FK506 binding protein precursor	protein metabolism	0.39	0.02	-0.4	-1.29
AA039851	protein tyrosine phosphatase type IVA, member 3	protein metabolism	0.52	0.22	-0.3	-1.23
AA903832	KIAA1361 protein	protein metabolism	0.27	0.00	-0.3	-1.21
AI348521	HBS1-like (<i>S. cerevisiae</i>)	protein metabolism	-0.51	-0.15	0.4	1.28
AA913127	glucosaminyl (N-acetyl) transferase 2, I-branching enzyme	protein metabolism	-0.51	-0.15	0.4	1.28
AI184710	toll-like receptor 6	protein metabolism	-0.21	0.23	0.4	1.36
AA668470	regulator of G-protein signalling 5	signal transduction	3.31	2.26	-1.1	-2.07
N78828	wingless-type MMTV integration site family member 2	signal transduction	1.32	0.32	-1.0	-2.00
AA598668	chimerin (chimaerin) 1	signal transduction	0.09	-0.90	-1.0	-1.99
AW071513	MSTP032 protein	signal transduction	2.71	1.90	-0.8	-1.75
R00276	CD38 antigen (p45)	signal transduction	0.61	-0.12	-0.7	-1.66
AI032392	RAB27A, member RAS oncogene family	signal transduction	0.57	-0.15	-0.7	-1.65
AI828190	plasminogen activator, urokinase receptor (uPAR)	signal transduction	0.41	-0.20	-0.6	-1.53
N36174	5-hydroxytryptamine (serotonin) receptor 2B	signal transduction	-0.11	0.33	0.4	1.36
H15574	erythropoietin receptor	signal transduction	-0.24	0.28	0.5	1.43
AA143331	matrix metalloproteinase 1 (interstitial collagenase)	catabolism	2.65	0.10	-2.6	-5.86
AA936799	matrix metalloproteinase 2 (gelatinase A, 72kD gelatinase, 72kD type IV collagenase)	catabolism	0.70	-0.85	-1.5	-2.92
H15155	steroid sulfatase (microsomal), arylsulfatase C, isozyme S	catabolism	0.24	-0.15	-0.4	-1.32
R32410	phospholipase A2, group V	catabolism	-0.21	0.10	0.3	1.24

Table 5. continued

ID	Name	Biological process	Primary tumor	Liver Mets. tumor	Ratio	Fold change
AW075162	tissue inhibitor of metalloproteinase 1	catabolism	1.20	2.14	0.9	1.92
AI081548	cholesterol 25-hydroxylase	lipid metabolism	1.34	0.33	-1.0	-2.01
AA633818	fatty-acid-Coenzyme A ligase, long-chain 4	lipid metabolism	-0.44	-1.09	-0.6	-1.57
AI003792	ethanolamine kinase	lipid metabolism	0.47	-0.03	-0.5	-1.42
AI304790	prostaglandin I2 (prostacyclin) synthase	lipid metabolism	0.81	1.52	0.7	1.64
AA644211	prostaglandin-endoperoxide synthase 2	cell motility	0.04	-0.91	-1.0	-1.94
AA450009	endothelin receptor type A	cell motility	2.28	1.39	-0.9	-1.84
R38717	ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)	cell motility	2.41	1.72	-0.7	-1.61
AA446120	adrenomedullin	cell-cell signaling	-0.09	-0.86	-0.8	-1.71
AA453759	sprouty (Drosophila) homolog 2	cell-cell signaling	1.09	0.67	-0.4	-1.33
AI655747	motilin	cell-cell signaling	-0.21	0.04	0.3	1.19
H08561	Human insulin-like growth factor binding protein 5 (IGFBP5) mRNA	regulation of cell growth	0.42	-0.50	-0.9	-1.89
W46577	endothelial cell-specific molecule 1	regulation of cell growth	1.16	0.49	-0.7	-1.60
T52830	Human insulin-like growth factor binding protein 5 (IGFBP5) mRNA	regulation of cell growth	0.33	-0.04	-0.4	-1.29
AI004169	GDP-mannose pyrophosphorylase B	biosynthesis	0.36	-0.01	-0.4	-1.29
AA126676	prostaglandin I2 (prostacyclin) synthase	biosynthesis	-0.52	-0.08	0.4	1.35
AA504211	tumor necrosis factor (ligand) superfamily, member 11	immune response	0.93	0.32	-0.6	-1.53
AA281635	suppression of tumorigenicity 16 (melanoma differentiation)	immune response	0.28	-0.11	-0.4	-1.31
AA450360	signal sequence receptor, alpha (translocon-associated protein alpha)	regulation of cell proliferation	-0.32	-0.78	-0.5	-1.38
T52674	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)	regulation of cell proliferation	-0.07	-0.36	-0.3	-1.22
AW029226	deiodinase, iodothyronine, type II	amine metabolism	1.00	-0.26	-1.3	-2.39
AI299525	transforming growth factor beta-stimulated protein TSC-22	nuclear metabolism	-0.44	0.28	0.7	1.66

Table 5. continued

ID	Name	Biological process	Primary tumor	Liver Mets. tumor	Ratio	Fold change
AA488893	lipase protein	response to external stimulus	0.11	-0.14	-0.2	-1.18
AA625655	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)	unclassified	3.40	-0.26	-3.7	-12.67
AI972184	wingless-type MMTV integration site family, member 5A	unclassified	-1.11	-3.02	-1.9	-3.76
AA046430	lung type-I cell membrane-associated glycoprotein	unclassified	1.54	0.44	-1.1	-2.14
AW084638	carbonic anhydrase II	unclassified	3.41	2.51	-0.9	-1.86
AI675874	transducin (beta)-like 2	unclassified	-0.01	-0.86	-0.9	-1.81
N98591	interleukin 6 (interferon, beta 2)	unclassified	0.25	-0.60	-0.9	-1.80
AA447761	aminolevulinate, delta-, synthase 1	unclassified	0.29	-0.52	-0.8	-1.75
AI341793	hypothetical protein	unclassified	1.05	0.27	-0.8	-1.72
W69790	downregulated in ovarian cancer 1	unclassified	1.51	0.76	-0.7	-1.68
AA933744	hypothetical protein	unclassified	0.96	0.32	-0.6	-1.55
AA873264	chromosome 4 open reading frame 1	unclassified	-0.15	-0.74	-0.6	-1.50
AA682637	carbohydrate (chondroitin 6/keratan) sulfotransferase 2	unclassified	-0.20	-0.73	-0.5	-1.45
AI970057	secretory leukocyte protease inhibitor (antileukoproteinase)	unclassified	-0.05	-0.58	-0.5	-1.44
AI492358	KIAA1610 protein	unclassified	0.41	-0.10	-0.5	-1.42
AA683520	secretory leukocyte protease inhibitor (antileukoproteinase)	unclassified	-0.20	-0.68	-0.5	-1.40
AA151002	epithelial protein up-regulated in carcinoma, membrane associated protein 17	unclassified	0.72	0.26	-0.5	-1.38
AA505045	Human L2-9 transcript of unrearranged immunoglobulin V(H)5 pseudogene	unclassified	0.50	0.05	-0.5	-1.37
W42849	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)	unclassified	0.38	-0.04	-0.4	-1.34
N53380	protein kinase C, mu	unclassified	0.09	-0.32	-0.4	-1.33
AA455668	Homo sapiens cDNA FLJ13207 fis, clone NT2RP4000023	unclassified	-0.56	-0.96	-0.4	-1.32
AI657191	B-cell receptor-associated protein BAP29	unclassified	0.25	-0.08	-0.3	-1.26

Table 5. continued

ID	Name	Biological process	Primary tumor	Liver Mets. tumor	Ratio	Fold change
AA485214	nucleobindin 2	unclassified	-1.93	-2.27	-0.3	-1.26
T59641	pre-B-cell leukemia transcription factor 2	unclassified	0.31	0.02	-0.3	-1.23
AI391689	Homo sapiens, Similar to hypothetical protein, clone MGC:2992, mRNA, complete cds	unclassified	-0.47	-0.20	0.3	1.20
AA421296	CD68 antigen	unclassified	0.27	0.67	0.4	1.32
AI291437	tetratricopeptide repeat domain 4	unclassified	-0.61	-0.19	0.4	1.34
AA935287	Homo sapiens mRNA; cDNA DKFZp564P1816 (from clone DKFZp564P1816)	unclassified	-0.55	-0.05	0.5	1.42
AA989429	KIAA1505 protein	unclassified	0.47	1.03	0.6	1.48
AA447632	glycoprotein M6A	unclassified	-0.24	0.49	0.7	1.65
AA906961	KIAA0826	unclassified	-0.32	-1.69	-1.4	2.58
AA976712	Transmembrane protein 46	unclassified	-1.62	-2.43	-0.8	1.75
AA948511	Solute carrier family 16 (monocarboxylic acid transporters), member 14	unclassified	0.84	0.27	-0.6	1.48
AA894442	Sine oculis homeobox homolog 4 (Drosophila)	unclassified	-1.11	-1.62	-0.5	1.43
AA973252	ESTs, Weakly similar to ALU6_HUMAN ALU SUBFAMILY SP SEQUENCE CONTAMINATION WARNING ENTRY [H.sapiens]	unclassified	1.03	0.54	-0.5	1.40
AI423516	ESTs	unclassified	0.69	0.21	-0.5	1.39
AI301306	Hypothetical protein BC009264	unclassified	0.40	-0.03	-0.4	1.34
AA931235	ESTs, Weakly similar to CGHU1S collagen alpha 1(I) chain precursor [H.sapiens]	unclassified	0.35	0.04	-0.3	1.24
AA465397	ESTs	unclassified	0.12	-0.18	-0.3	1.23
AI360832	ESTs, Weakly similar to Unknown [H.sapiens]	unclassified	-0.06	-0.35	-0.3	1.22
AI267953	ESTs	unclassified	-0.26	-0.53	-0.3	1.20
AA886739	Histone 1, H2bd	unclassified	0.19	-0.07	-0.3	1.20
AI380488	Shroom	unclassified	0.25	-0.01	-0.3	1.20
AI263120	ESTs	unclassified	-0.34	-0.04	0.3	1.23

Table 5. continued

ID	Name	Biological process	Primary tumor	Liver Mets. tumor	Ratio	Fold change
AI348263	Phosphodiesterase 5A, cGMP-specific	unclassified	-0.58	-0.25	0.3	1.25
AA975949	TatD DNase domain containing 3	unclassified	-0.44	-0.11	0.3	1.26
AA932525	ESTs	unclassified	-0.70	-0.37	0.3	1.26
AI346102	Zinc finger, DHHC-type containing 9	unclassified	-0.07	0.28	0.3	1.27
AI361424	Ribosomal protein L32	unclassified	-0.04	0.35	0.4	1.32
AI536572	EST	unclassified	-0.18	0.25	0.4	1.35
AA910921	ESTs	unclassified	0.12	0.56	0.4	1.36
AI419116	ESTs	unclassified	-0.46	0.00	0.5	1.38
AI298528	ESTs	unclassified	-0.13	0.35	0.5	1.40
AA465727	ESTs, Weakly similar to ALU1_HUMAN ALU SUBFAMILY J SEQUENCE CONTAMINATION WARNING ENTRY [H.sapiens]	unclassified	-0.72	-0.22	0.5	1.41
H82536	Cyclic nucleotide gated channel beta 1	unclassified	1.43	0.72	-0.7	1.63
AA863025	ESTs	unclassified	-0.07	0.25	0.3	1.25
AI380909	Ubiquitin-conjugating enzyme E2E 1 (UBC4/5 homolog, yeast)	unclassified	-0.36	0.01	0.4	1.29
AA894431	ESTs	unclassified	-0.20	0.19	0.4	1.31
AI969860	Hypothetical protein FLJ90396	unclassified	-0.05	0.36	0.4	1.33
AA946953	ESTs	unclassified	-0.26	0.29	0.6	1.47

4. 간 전이 관련 후보 유전자 군의 기능해설

가. Web DATABASE를 이용한 분석

원발성 대장 종양과, 쌍을 이루는 간 전이 종양의 유전자 발현 정도를 비교 분석하여 두 군간에 발현이 차이나는 간 전이 관련 후보 유전자 126개를 선별하였다. 선별된 유전자들은 biological process에 따라 cell adhesion, cell growth, protein metabolism, signal transduction, catabolism, cell motility 등으로 나눌 수 있었다 (Table 6). Cell adhesion 에 속한 관련 유전자 중 regenerating islet-derived 1 beta가 간 전이 대장종양에서 원발성 종양에서 비해 가장 많이 down-regulation 되어 있었으며 6.15 배의 발현저하를 보였다. 반대로 stromal cell-derived factor 1은 가장 많이 up-regulation 되어 두 군간에 2.47 배의 발현 차이를 보였다. Cell growth 관련 유전자 중에서는 lipocalin 2의 유전자 발현 변화량이 3.97 배로 가장 컸으며, protein metabolism에서는 deleted in malignant brain tumors 1 가 4.55 배, signal transduction 에서는 regulator of G-protein signaling 5 가 2.07 배의 변화량을 보였다. Catabolism의 경우 MMP-1과 MMP-2는 각각 5.86배, 2.92배의 down-regulation을 보이고 있으며, TIMP-1 의 경우는 1.92 배 up-regulation 되어 있었다.

Table 6. Ontological information of 126 liver metastasis related genes

Biological process	Number
cell adhesion	11
cell growth and or maintenance	11
protein metabolism	9
signal transduction	9
catabolism	5
lipid metabolism	4
cell motility	3
cell-cell signaling	3
regulation of cell growth	3
biosynthesis	2
immune response	2
regulation of cell proliferation	2
amine metabolism	1
nuclear metabolism	1
response to external stimulus	1
unclassified (include 16 ESTs)	59

나. 선별된 유전자 중 전이 관련성이 알려진 유전자의 발현 양상 관찰

선별된 유전자들 중 원발성 대장 종양과 간 전이 종양에서의 유전자 발현 변화가 큰 유전자들을 순서대로 나열하여 up 혹은 down-regulation이 된 각각 상위 10 개의 유전자를 선별하여 도식화하였

다 (Figure 10-A, 11-A). 이 유전자 목록에는 proteolytic activity를 지니며 전이와 관련되어 있다고 알려진 TIMP-1, MMP-1, MMP-2가 모두 포함되어 있었다. 이들 유전자의 각 sample 별 유전자의 발현 여부를 확인한 결과 TIMP-1의 경우 정상 대장, 용종, 원발성 대장 종양, 간 전이 종양에서 점차적으로 증가하는 발현양상을 보였고 (Figure 10-B) MMP-1과 MMP-2의 경우는 정상대장과 용종, 원발성 대장 종양은 점차적으로 증가하는 양상을 보이지만 간 전이 종양에서 급격히 감소하는 경향을 보였다 (Figure 11-B, 11-C). 그 외에도 WNT signal pathway에 관련되어 있는 WNT5A의 fold change가 원발성 대장 종양보다 간 전이 종양에서 3.76 배 down-regulation 되어 있었다 (Figure 11-D).

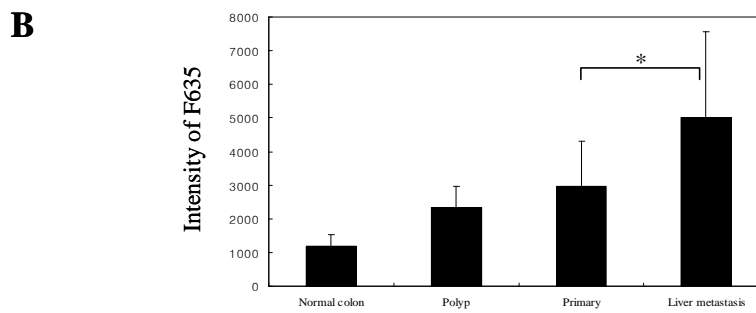
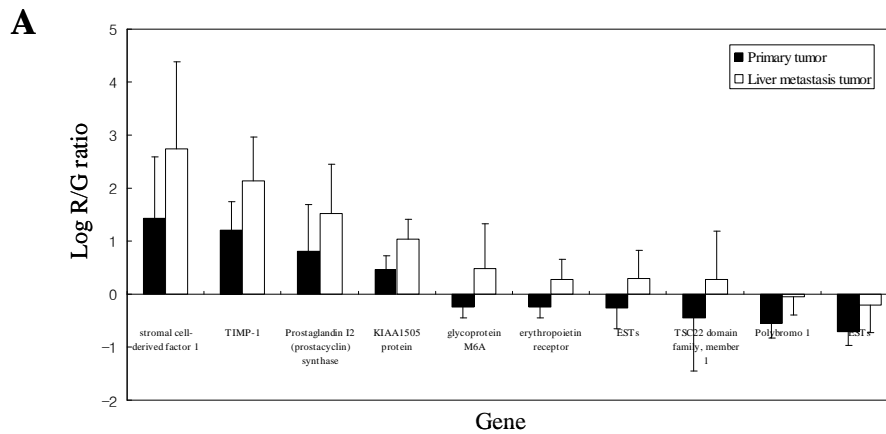


Figure 10. Expression of top 10 up-regulated genes in liver metastasis tumor

A, Comparison of top 10 up-regulated genes expression between primary colorectal tumor and liver metastasis tumor. This graph shows the mean of $\log_2(R/G)$ ratio of colorectal primary tumor and liver metastasis tumor. B, Expression of TIMP-1 according to tumor progression with the mean of raw F635 intensity of TIMP-1. The (*) represents p value below 0.01.

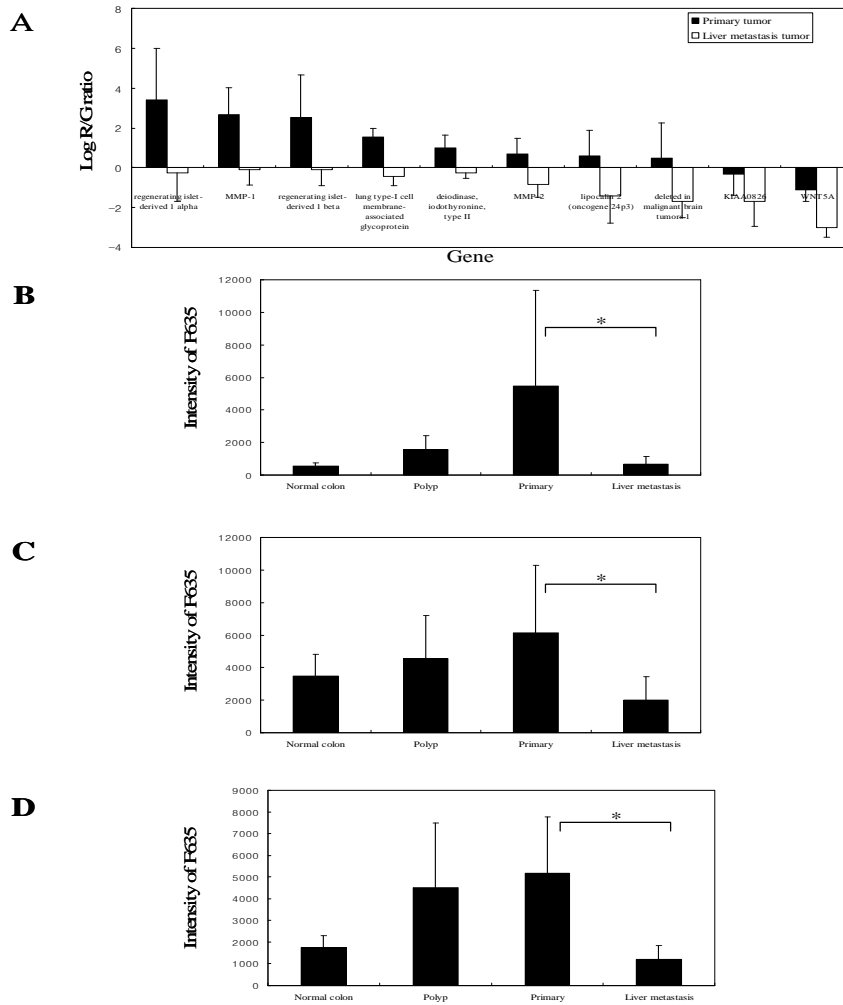


Figure 11. Expression of top 10 down-regulated genes in liver metastasis tumor

A, Comparison of top 10 down-regulated genes expression between primary colorectal tumor and liver metastasis tumor. This graph shows the mean of $\log_2(R/G)$ ratio of colorectal primary tumor and liver metastasis tumor. B, C and D, Expression of MMP-1, MMP-2, WNT5A according to tumor progression with the mean of raw F635 intensity of each gene, respectively. The (*) represents p value below 0.01.

IV. 고찰

대장암은 주로 간으로 전이되며 이 같은 전이가 주된 사인으로 꼽히고 있다²⁷. 그렇기 때문에 대장암 세포가 간으로 전이되는 기전을 분자 수준의 유전자 탐색을 통해 이해하고, 이러한 기전을 적절히 제어할 수 있는 분자를 찾아 종양이 간으로 전이되는 것을 차단하게 되면 환자의 생존율을 증가시킬 것으로 기대한다. 따라서 본 연구에서는 원발성 대장 종양과 간 전이 종양을 한 번의 실험으로 수천 개의 유전자 변화를 관찰할 수 있는 cDNA microarray를 기반으로 한 gene expression profiling을 수행하여 genome-wide screening 하였다.

cDNA microarray는 암 치료에 있어 치료의 유전자 target 이나 새로운 pathway를 찾을 수 있는 고도의 실험기술로써 사용되고 있다²⁸⁻³⁰. cDNA microarray 기법을 이용한 gene expression profiling은 유전자 수준에서 RNA의 변화 양상을 직접적으로 비교할 수 있으며, 실험을 단시간에 수행하고 결과를 관찰 할 수 있는 장점을 지니고 있다. 하지만 한번의 실험에서 50 μ g 이상의 많은 양의 total RNA를 사용하기 때문에 환자에게서 얻은 소량의 검체를 이용하여 실험을 하는 데는 한계가 있다. 이러한 한계를 극복하기 위하여 본 연구에서는 최소 0.5 μ g total RNA만으로도 실험이 가능한 T7 linear mRNA amplification을 이용하여 실험을 수행하였다.

이러한 cDNA microarray의 장점과 RNA 증폭 방법을 이용하여 진단 당시 간 전이를 가지고 있는 대장 종양 환자 22명으로부터 얻은 sample로 원발성 대장 종양과 간 전이 종양의 유전자 발현 변화를 관찰하고 통계학적으로 유의한 유전자 군을 선별하고자 하였다.

먼저, 실험의 유의성을 확인하기 위하여 각 실험에서의 flag %를 계산하였고, 표준화 전과 후의 데이터를 비교하여 양질의 실험임을

확인했다. 또한 14 번의 반복실험 간의 상관계수를 산출하여 실험의 재현성을 확인하고 분석을 진행하였다. 전체 79 예의 unsupervised hierarchical clustering 결과 (Figure 5)에서는 unsupervised 임에도 불구하고, 종양과 정상조직의 유전자 발현이 다른 양상을 보이고 있었으며 같은 정상조직이라 할지라도 각 장기 간에 발현되는 유전자의 양상은 확연히 다르다는 것을 볼 수 있었다. 또한 종양 군으로 집단화 된 부분에서는 동일 환자의 원발성 대장 종양과 간 전이 종양이라 할지라도 *Sridhar Ramaswamy et al*³¹에서 언급한 양상과는 달리 같은 가지로 묶이지 않고 종양의 위치에 따라 분류되는 양상을 보였다. 이것은 동일 환자의 종양이라 할지라도 전이라는 과정을 거치면서 원발성 종양과 전이 종양의 유전자 발현 양상은 달라질 수 있거나, 혹은 종양이 위치한 장기 등의 주변 환경에 의해 이러한 결과에 영향을 줄 수 있다는 것을 제시하고 있다.

정상 대장과 간에서 다르게 발현되어 분석에서 제외된 4,616 개의 유전자에는 실제로 haptoglobin이나 plasminogen 등의 간 기능 관련 유전자 들이 포함되어 있었다. 반면 이 유전자 군에는 insulin-like growth factor binding protein 1이나 cadherin 같은 어느 정도의 이미 알려진 종양 관련 유전자들이 속해 있었지만 이러한 유전자들은 우리가 찾고자 하는 원발성 대장 종양과 간 전이 종양에서만 유전자 발현에 차이를 보이는 유전자를 선별하는데 있어 영향을 줄 수 있는 요인으로 사료되어 분석에서 제외하였다. 따라서 최종 선별된 126 개의 유전자는 장기에 따라 발현 양이 달라질 수 있는 유전자를 제외한 원발성 대장 종양과 간 전이 종양에서만 명백히 발현 양의 차이를 보이는 유전자를 선별할 수 있었다.

결과에는 표기하지 않았으나 장기 특이 유전자로 분리된 4,616 개의 유전자를 제거하지 않고 분석했을 때 실제로 결과에는 상기와 같은 간 기능 관련 유전자들이 원발성 대장 종양 조직과 간 전이 종양 조직에서 다르게 발현되어 포함되었다. 그리고 이 유전자 군을 제외한 후에 원발성 종양과 간 전이 종양만을 hierarchical clustering 했을 때 figure 5의 종양 군 양상처럼 종양 위치 별로 묶이지 않고 환자 각 개인별로 묶이는 양상을 보였다. 이것은 4,616 개의 선별된 유전자 군은 알려진 종양관련 유전자들이 포함되어 있기는 하지만 장기 특이 유전자가 될 수 있는 하나의 증거가 될 수 있다.

원발성 대장 종양과 간 전이 종양에서 유의하게 발현에 차이는 유전자군을 선별하고 검증하기 위해 먼저 전체 sample을 각각 독립적으로 학습군과 검증군으로 나누어 분석을 진행하였다. 학습군을 PAM을 이용하여 학습시켜 training error를 산출하고 이를 다시 한 번 교차 검정해서 misclassification error를 산출했다. 이 두 결과에서 error가 0인 구간에서 선별된 81 개, 126 개, 252 개의 유전자군은 supervised hierarchical clustering 상에서 모두 원발성 대장 종양과 간 전이 종양을 확연히 나누어 지는 것을 확인했지만 81개 유전자군의 교차 검정 probability 가 다른 두 유전자군에 비해 조금 낮은 양상을 보였다. 이 세 개의 유전자군을 검증군을 통하여 검정을 진행한 결과, 81개, 126개 유전자군에서는 검증군에 속해있는 GIST 예의 predicted test probability가 약 0.8 정도로 다른 adenocarcinoma와 작은 차이를 보이고 있으나, 252개 유전자군에서는 1로 나타나 다른 adenocarcinoma와 차이를 구분하기 힘들었다. GIST 예는 분석을 진행하는 동안 항상 다른 양상의

유전자 발현을 보이고 있었기 때문에 다른 SCC 예나 간 전이가 확실하지 않았던 예에 비해 좀 더 의미를 두었다. 이러한 과정을 통해 결과에서 언급한 기준에 따라 선별된 유전자군은 분석에 사용된 환자의 예가 적기 때문에 이것이 절대적이라고 확인할 수는 없지만 이러한 유전자들이 간 전이와 연관되어 있다는 것은 제시할 수 있을 것으로 사료된다.

선별된 126개의 유전자에는 biological process에 따라 cell adhesion, cell growth, protein metabolism, signal transduction, catabolism, cell motility 등으로 분류할 수 있었으며 많은 수의 종양관련 유전자들을 포함하고 있었다. 그 중에는 proteolytic pathway에 관여하는 MMP-1, MMP-2, TIMP-1, uPAR 등의 유전자들이 존재하고 있었으며 TIMP-1을 제외한 유전자들은 모두 간 전이 종양에서 원발성 대장 종양에 비해 발현 양이 down regulation 되어 있는 것을 확인 했다. 정상 대장 조직의 발현 양을 확인한 결과 MMP-1, MMP-2는 발현 양이 작았고 uPAR의 경우는 원발성 대장 종양과 비슷한 정도의 발현을 보이고 있었다. 반대로 TIMP-1의 발현은 정상 조직에서는 낮았지만 악성화 되면서 점점 증가하는 양상을 보였고 이 것은 다른 종류의 종양이기는 하나 *William Hornebeck et al*³²이나 *Valerie Gouyer et al*³³의 결과와 비슷한 양상을 보였다. 또한 몇몇 논문들³⁴⁻³⁵에서 MMP-9 과의 연관성에 대해 언급하고 있었으나 본 실험에서는 MMP-9 유전자들이 모두 flag out 되어있어 확인할 수 없었다.

대장암의 발생에 관여하는 것으로 알려진 WNT pathway³⁶의 WNT5A는 원발성 대장 종양에서 정상조직에 비해 높은 증가율을 보이며 이것은 이미 발표된 논문들과도 맥락을 같이 한다. 특이점은 간으로 전

이가 일어난 후에는 다시 급격히 감소 한다는 점이며 가설이지만 몇몇 보고에서 cell adhesion이나 motility에 영향을 줄 수 있다고 언급되어 이러한 가설을 증명하는데 이 결과가 하나의 증거로 활용될 수 있을 것으로 사료된다.

본 연구에서는 원발성 대장암과 간 전이 대장암에서 발현의 양상이 다른 간 전이 관련 후보 유전자군을 분자적 수준에서 탐색하고자 하였고, 각각을 예측할 수 있는 후보 유전자 군을 high-throughput 기법인 cDNA microarray의 gene expression profiling 기법을 사용하여 다수의 후보 유전자 군을 탐색하였다. 이렇게 선별된 다양한 후보 유전자 군은 종양의 복잡한 생물학적 경로의 여러 측면에서 전이에 관련된 기전을 이해 할 수 있는 척도를 제공해 줄 수 있으리라 생각된다. 이후에, 대상환자의 수를 늘려 실험의 결과를 일반화 하고 간으로의 전이가 없는 원발성 대장 종양과 간으로의 전이가 있는 원발성 대장 종양 조직을 비교 분석하여 선별된 유전자 군과 함께 비교하게 되면 간 전이 관련 예측 혹은 진단 유전자 군을 선별할 수 있을 것이고 이것은 임상적으로 대장암 환자의 생존율을 높이는데 기여할 수 있을 것으로 여겨진다.

V. 결론

본 연구에서는 cDNA microarray의 gene expression profiling 기법을 사용하여 원발성 대장 종양과 간 전이 종양에서 다르게 발현되는 유전자 군을 찾고자 하였다. 정상 대장 검체와 정상 간 검체를 이용하여 장기 특이 유전자 4,616 개를 선별하여 전체 12,823 개의 유전자에서 제거하고 PAM을 이용하여 126개의 간전이 관련 후보 유전자 군을 선별하였다. 선별된 유전자 군들의 유전학적 수준

의 관찰은 대장암의 간 전이 기전을 이해하는데 있어 중요한 기반을 제공할 것으로 생각된다.

참고 문헌

1. Sakamoto J, Ohashi Y, Hamada C, Buyse M, Burzykowski T, Piedbois P. Efficacy of oral adjuvant therapy after resection of colorectal cancer: 5-year results from three randomized trials. *J Clin Oncol* 2004;22:484 - 492.
2. Rempei Y, Yoichi F, Tatsuhiko T, Osamu K, Masao K, Kohei M, et al. Genome-wide screening of genes showing altered expression in liver metastases of human colorectal cancers by cDNA microarray. *Neoplasia* 2001;3:395-401.
3. Meihua L, Yu-min L, Suguru H, Takshi S, Kohei M, Masao K, et al. Gene associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray. *Int J Oncology* 2004;24:305-312.

4. Mordechai G, Isaiah J. Biology of human colon cancer metastasis, *World J Surg* 1995;19:226-234.
5. Francois B, Sebastien S, Severine E, Valery N, Pascal F, Christophe G, et al. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene* 2004;23:1377-1391.
6. Eccles SA, Modjtahedi H, Box G, Court W, Sandle J, Dean CJ. Significance of the c-erbB family of receptor tyrosine kinases in metastatic cancer and their potential as targets for immunotherapy. *Inv Met* 1994-95;14:337-348.
7. Andre T, Kotelevets L, Vaillant JC, Coudray AM, Weber L, Prevot S, et al. Vegf, Vegf-B, Vegf-C and their receptors KDR, FLT-1 and FLT-4 during the neoplastic progression of human colonic mucosa. *Int J Cancer* 2000;86: 174-181.
8. Karube H, Masuda H, Ishii Y, Takayama T. E-cadherin expression is inversely proportional to tumor size in experimental liver metastases. *J Surg Res* 2002;106: 173-178.
9. Naotsugu H, Hiroshi I, Koshi M, Fumiaki T, Tohru U. Analysis of gastric cancer with cDNA microarray. *Cancer Chemother Pharmacol* 2004;54:S21-S24.
10. Lin L, Ranelle CS, Hongqing G, Anton B, Joy KC, Jose EG. Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nature* 1999;5:117-122.

11. Douglas TR, Uwe S, Michael BE, Charles MP, Christian R, Paul S. Systematic variation in gene expression patterns in human cancer cell lines. *Nature* 2000;24:227-235.
12. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531-537.
13. Andrew IS, John BW, Lisa MS, Suzanne GK, Petre D. Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res* 2001;61:7388-7393.
14. James E, Hermes Y, Kevin M, Yanxiang C, Suresh N, Richard F. Analysis of gene expression in single live neurons. *Proc Natl Acad Sci* 1992;89:3010-3014.
15. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503-511.
16. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002;8:816-824.
17. Bertucci F, Nasser V, Granjeaud S, Eisinger F, Adelaide J, Tagett R, et al. Gene expression profiles of poor-prognosis primary breast cancer correlate with survival. *Hum Mol Genet* 2002;11:863-872.

18. Devilard E, Bertucci F, Trempat P, Bouabdallah R, Loriaud B, Giaconia A, et al. Gene expression profiling defines molecular subtypes of classical Hodgkin's disease. *Oncogene* 2002;21:3095-3102.
19. Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci* 2001;98:13784-13789.
20. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, et al. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 2002;1: 203-209.
21. Yang SH, Seo MY, Jeong HJ, Jeung HC, Shin J, Kim SC, et al. Gene copy number change events at chromosome 20 and their association with recurrence in gastric cancer patients. *Clin Cancer Res* 2005;11(2 Pt 1):612-620.
22. Kim TM, Jeong HJ, Seo MY, Kim SC, Cho GB, Park CH, et al. Determination of genes related to gastrointestinal track origin cancer cells using a cDNA microarray. *Clin Cancer Res* 2005;11:79-86.
23. Yee HY, Sandrine D, Percy L, David ML. Normalization for cDNA Microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002;30:e15.

24. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci* 1998;95:14863-14868.
25. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to ionizing radiation response. *Proc Natl Acad Sci* 2001;98:5116-5121.
26. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci* 2002;99:6567-6572.
27. Taylor I. Liver metastases from colorectal cancer: lessons from past and present clinical studies. *Br J Surg* 1996;83:456-460.
28. Jenni CC, Eric CW, Anna T, Susan GH, M Carolina G, Richard E, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003;362:362-369.
29. Gene HM, John ES, Geoffrey KM. Radiation induced histopathologic changes of the breast. *Am J Surg Pathol* 2004;28:47-53.
30. Stephanie SL, Annie NYC, Hextan YSN. Differential gene expression in cervical cancer cell lines before and after ionizing radiation. *Int J Onco* 2003;22:1091-1099.
31. Ramaswamy S, Ross K, Lander E, Golub T. A molecular signature of metastasis in primary solid tumors. *Nat Gen* 2003;33:49-54.

32. William H, Elise L, Emmanuelle P, Philippe B. Beneficial and detrimental influences of tissue inhibitor of metalloproteinase-1 (TIMP-1) in tumor progression. *Biochimie* 2005;87:377-383.
33. Valerie G, Massimo C, Patrick D, Farid Z, Marie-Christine C, Evelyne C, et al. Tissue inhibitor of metalloproteinase 1 is an independent predictor of prognosis in patients with nonsmall cell lung carcinoma who undergo resection with curative intent. *Cancer* 2005;103:1676-1684.
34. Ogata Y, Miura K, Ohkita A, Nagase H, Shirouzu K. Imbalance between matrix metalloproteinase 9 and tissue inhibitor of metalloproteinases 1 expression by tumor cells implicated in liver metastasis from colorectal carcinoma. *Kurume Med J* 2001;48:211-218.
35. Curran S, Dundas SR, Buxton J, Leeman MF, Ramsay R, Murray GI. Matrix metalloproteinase/tissue inhibitors of matrix metalloproteinase phenotype identifies poor prognosis colorectal cancers. *Clin Cancer Res* 2004;10:8229-8234.
36. Behrens J. The role of the Wnt signalling pathway in colorectal tumorigenesis. *Biochem Soc Trans* 2005;33:672-675.

Abstract

Genome-wide screening of genes associated with liver metastasis of colorectal cancers using a cDNA microarray

Dong Hyuk Ki

*Department of Medical Science
The Graduate School, Yonsei University*

(Directed by Professor Sun Young Rha)

Liver metastasis is the major cause of death in colorectal cancer patients. By using cDNA microarray gene expression profiling, we selected genes of metastatic potential from colorectal cancer patients with liver metastasis. We performed cDNA microarray in an indirect design based on the T7 linear mRNA amplification method with paired 22 colorectal normal mucosa, primary tumor, normal liver and metastasis liver tumor tissue RNAs using 17K human cDNA microarray. After the normalization, we evaluated the gene expression profilings of all samples with unsupervised clustering. Next, we selected organ specific genes with normal colon mucosa and liver tissue using Significant Analysis of Microarrays (SAM). After that, we selected and validated liver metastasis related genes from colorectal primary tumor and liver metastasis tumor within genes which subtracted organ specific genes using Prediction Analysis of Microarrays (PAM). The selected genes were functionally annotated using NIH-DAVID database

(<http://apps1.niaid.nih.gov/david/>).

Unsupervised hierarchical clustering with 12,823 genes with 100% non-missing proportion of the 22 paired samples showed different expression profiling between primary tumor and liver metastasis. Next, we selected 4,616 organ specific genes with normal colon mucosa and liver tissue at the false discovery rate of 0.062% using SAM, which were subtracted for not using further analysis. After that, we identified 38 and 88 genes of up and down regulated in metastatic lesions, respectively, with selected 8,207 genes of 14 train set using PAM, and validated the selected genes in 8 independent test set. 126 selected genes contained many known oncogenes and cell adhesion molecules such as wingless-type MMTV integration site family member 5A, lipocalin 2, E-cadherin, deiodinase as well as 16 ESTs. We especially observed that MMP-1, MMP-2 and WNT5A expressions were significantly decreased in liver metastatic tumors ($p < 0.001$), suggesting their roles in early stage colorectal cancer rather than systemic metastasis. However, TIMP-1 expressions were significantly increased in liver metastatic tumors ($p < 0.001$). In conclusion, we scanned whole genome using a cDNA microarray and identified 126 genes which might play a significant role in liver metastasis in colorectal cancer.

Key words : Colorectal cancer, Liver metastasis, Gene expression profiling,
cDNA microarray.