

Statistical Analysis about Ability to Mouse Embryonic Stem Cell Differentiation using cDNA Microarray¹⁾

Hang-Suk Choi²⁾ · Sung-Ju Kim³⁾ · Young-Jin Lee⁴⁾
Kyung-Joon Cha⁵⁾ · Chul-Geun Kim⁶⁾

Abstract

As a foundation study of stem cell applied research, it is necessary to identify the large gene expression through cDNA microarray to understand principles of the level of molecular about cell function. In this paper, we investigated the gene expression through the K-means clustering method and path analysis with genes related to pluripotency and differentiation in an mouse early stage embryonic development process and embryonic stem cell differentiation. We find a few biological phenomenon through this study. Also, we realize that this process provides functional relationship of unknown genes.

Keywords : cDNA microarray, Embryonic Stem Cell, K-means, Path Analysis

-
- 1) This work supported by Cell Applied Research Committee of The 21st Century Frontier Research Development Committee(M102KL010001-02K1201-00351).
 - 2) First Author : Ph.D student, Dept. of Mathematics, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.
E-mail : neuldol@ihanyang.ac.kr
 - 3) Ph.D student, Dept. of Statistics, Korea Univ., Anam-dong Seongbuk-Gu, Seoul, 136-701 Korea.
 - 4) Ph.D student, Dept. of Life Science, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.
 - 5) Corresponding Author : Professor, Dept. of Mathematics, Hanyang Univ., 17, Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.
E-mail : kjcha@hanyang.ac.kr
 - 6) Professor, Dept. of Life Science, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.

1. 서론

수정란은 발생 과정 동안 각기 다른 구조와 기능을 나타내는 여러 장기나 조직으로 분화하며 성장하여 정상적인 개체를 형성한다. 이러한 수정란의 분화 과정과 성체 세포의 재생은 여러 줄기세포의 분화능에 의해 조절된다. 줄기세포는 자기 복제가 가능하며, 특수한 환경에서는 다양한 조직이나 세포로 분화할 수 있는 능력을 지닌다 (Weissman, 2000; Blau 등, 2001).

생쥐 배아줄기세포는 포배(blastocyst)의 내세포괴(inner cell mass: ICM)에서 유래된 미분화 상태의 세포로서 신체의 모든 조직세포로 분화할 수 있는 전분화능(pluripotency)을 지니고 있다(Evans and Kaufman, 1981). 성체줄기세포(adult stem cell)는 성체의 각 조직이나 기관에 존재하고 있으며, 위치한 조직이나 기관에 존재하는 세포로의 제한된 분화능을 갖는다. 최근 성체줄기세포가 이식된 조직이나 장기의 특정 환경에 따라 다른 계열의 세포로 분화할 수 있는 유연성(plasticity)을 갖는다는 연구 결과도 보고되고 있다(Blau 등, 2001). 줄기세포의 존재가 이미 증명되었고 줄기세포를 이용한 퇴행성 뇌 질환, 심장 질환, 당뇨병 등의 난치병에 대한 세포 치료 가능성이 제시되고 있지만, 줄기세포의 분화 조절 및 유연성에 대한 분자적 기작에 대한 연구는 여전히 미진한 편이다.

따라서, 본 연구는 이러한 줄기세포 응용연구의 기반 연구로 세포 운명 결정 및 분화에 대한 분자 수준에서의 기작을 이해하기 위해 생쥐 배아줄기세포를 이용한 cDNA microarray 방법(Schena 등, 1995)으로 대단위적인 유전자 발현 양상을 확인하고자 하였으며, 생쥐 초기 배아 발생 과정 및 배아줄기세포 분화 중에 전능성 유지 및 분화에 관련되리라 추측되는 유전자들을 K-means clustering으로 군집화하고, 경로분석(path analysis)으로 인과모형(causal model)을 제시하였다.

2. 연구 대상 및 분석 방법

2.1 연구 대상

분화 후 0 Hr, 12 Hr, 24 Hr, 36 Hr, 48 Hr, 72 Hr, 96 Hr, 120 Hr (Hr : Hour)의 시간에 따라 생쥐 배아줄기세포에서의 유전자 발현 양상을 대단위적으로 분석하기 위해 생쥐 cDNA 클론들이 실어있는 Mouse cDNA Microarray Membrane(Research Genetics)을 사용하였으며, Pathways 3 software를 이용하여 전체 5344개의 데이터를 얻었다.

2.2 통계적 분석 절차

모든 자료의 통계 분석은 SAS(version 8.2, USA)와 LISREL(Linear Structural Relations 8.53, Student Edition)을 사용하였다. 우선 배아줄기세포에서의 분화단계별 유전자 발현량을 비교, 분석하기 위해 분화 직후(0 Hr)의 배아줄기세포의 유전자 발현량과 각 분화단계별 유전자 발현량의 비를 구한 후, 밑이 2인 logarithm을 취하므로

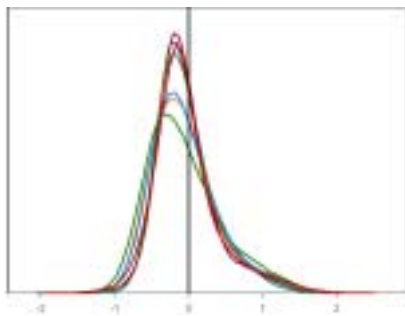
데이터의 scale을 재구성하였다. 그 후 Pathways 3 software에서 제공하는 control spot normalization으로 데이터를 정규화(normalization)한 후, 중심이 0인 위치로 median을 이용하여 데이터 분포를 조정하였다. 그리고 각 분화단계별 유전자 발현량 차이에 대한 통계량을 살펴보고, 유전자 발현 양상의 차이에 대해서는 K-means clustering을 사용하여 발현 양상에 따라 군집화 하였다. K-means clustering 결과 중 관심 대상이 되는 군집을 선택하여 유의한 유전자 후보군을 결정하였고, 선택된 유전자 후보군에 포함된 유전자들을 인과모형(causal model)인 다변량 분석 기법 중 경로 분석(path analysis)으로 최종 모형을 제시하였다.

3. 연구결과

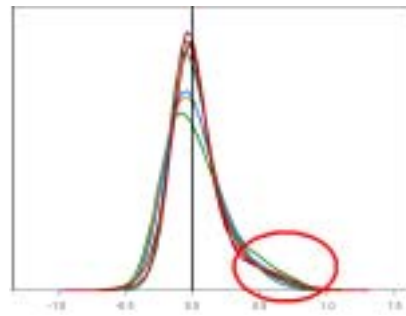
3.1 분화단계별 유전자 발현량의 분석과 Normalization

생쥐 배아줄기세포의 각 분화단계별 유전자의 발현량을 비교, 분석하기 위해 밑이 2인 logarithm을 취한 후 Pathways 3 software에서 제공하는 normalization 한 데이터의 분포는 <그림 1(a)>와 같았다. <그림 1(a)>의 분포는 0을 중심으로 데이터가 왼쪽으로 치우쳤고 오른쪽 꼬리가 긴 것으로 나타났다. 각 분화단계별 유전자 발현량의 평균, 표준편차, 중간값과 범위에 대한 통계량은 <표 1>과 같았다. <표 1>에서 살펴본 결과 평균에 비해 중앙값이 더 작았으며 ($p < 0.001$), 각 유전자별 발현 양상도 다르게 존재하고 있음을 알 수 있었다.

cDNA microarray의 각 시간대별 유전자를 median을 사용하여 분포의 중심이 0이 되도록 한 결과 <그림 1(b)>와 같았다. 결론적으로, 0을 중심으로 왼쪽으로 치우쳤던 분포의 중심이 0쪽으로 이동하였고, 분산은 작아졌다. 그러나 오른쪽 꼬리 부분은 여전히 길게 나타나고 있는데, 이는 median을 사용한 단점이라 할 수 있다(Yang 등, 2001). 분포의 양쪽 끝에 위치한 유전자들은 발현의 차이가 두드러진 유의한 유전자 후보군으로 본 연구에서는 오른쪽 꼬리 부분에 위치한 유전자들, 즉 <그림 1(b)>의 타원에 속하는 부분의 유전자를 유의한 유전자 후보군으로 설정하였다.



(a) median을 이용 변환 전 density plot



(b) median을 이용 변환 후 density plot

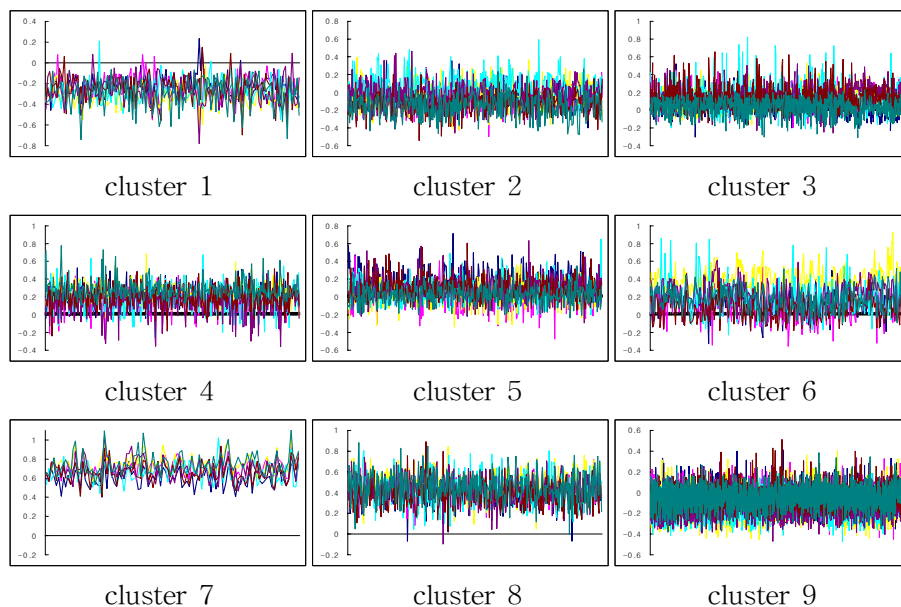
<그림 1> Normalization 전, 후의 density plot

<표 1> 배아줄기세포에서의 시간대별 발현량에 관한 평균, 표준편차, 중간값과 범위

배아줄기세포	평균±표준편차	중간값	범위[최소값, 최대값]
12 hr	-0.0168±0.1917	-0.0513	[-0.5604, 0.7828]
24 hr	-0.0178±0.2013	-0.0575	[-0.6574, 0.8683]
36 hr	-0.0291±0.2519	-0.0728	[-0.7227, 0.9351]
48 hr	-0.0248±0.2276	-0.0588	[-0.6063, 0.9636]
72 hr	-0.0229±0.2216	-0.0593	[-0.8386, 0.9698]
96 hr	-0.0168±0.1974	-0.0577	[-0.7563, 0.8789]
120 hr	-0.0195±0.2183	-0.0573	[-0.8012, 1.0457]

3.2 K-means Clustering

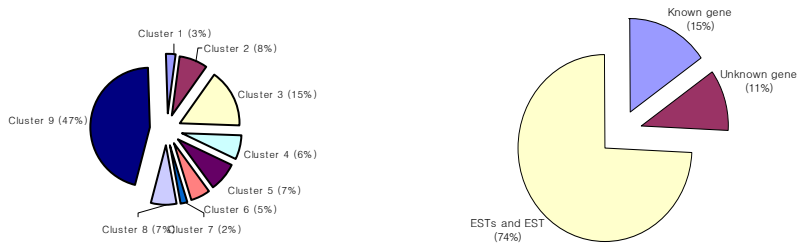
생쥐 배아줄기세포에서의 각 분화단계별 유전자 발현 양상 차이를 이용하여 cDNA microarray의 데이터에 유클리드 거리(Euclidean distance)를 적용한 K-means clustering(K=9)의 결과 <그림 2>와 같았다. 군집 수를 9로 선택한 것은 2~15까지 군집으로 군집화 한 결과 유전자의 발현 패턴이 가장 뚜렷했고, 군집간 거리와 군집내 거리에 의해 적절한 군집 수를 제시해주는 SAS에서의 CCC(Cubic Clustering Criteria) plot에 의해 결정하였다.



<그림 2> 배아줄기세포에서 분화단계별 유전자 발현량의 K-means clustering (K=9)

K-means clustering 결과 <그림 3(a)>와 같이 분석 유전자들은 cluster 9(N=2486), cluster 3(N=819) 그리고 cluster 2(N=429)의 순으로 유전자가 분포하였고, cluster 9에 속하는 유전자는 배아줄기세포에서의 각 분화단계별 유전자 발현량이 시간에 따라 변화가 없었다. 즉, cluster 9에 포함된 유전자는 세포의 기본적인 물질대사에 필요한 유전자들로, 대부분의 세포 및 조직에서 발현되며 housekeeping gene 혹은 ubiquitous expressed gene들로 사료된다. 이와 비슷한 패턴을 보이고 있는 군집은 cluster 2, 3 그리고 5였고, 분화함에 따라 발현량이 감소하는 유전자군은 cluster 1이었으며, 이와는 대조적으로 발현량이 증가하는 유전자군은 cluster 4, 6, 7 그리고 8이었다.

이번 연구에서 유의한 유전자 후보군으로 가정하였던 유전자들은 다른 유전자에 비해 배아줄기세포가 분화함에 따라 발현량이 증가하는 패턴을 보이는 것으로, 그들은 cluster 7(N=106)에 속해 있는 것을 <그림 2>를 통해 알 수 있었다. 그리고 cluster 7에 포함된 유전자의 특성은 <그림 3(b)>와 같았다. 즉, cluster 7에 속한 유전자의 11%는 기능이 알려져 있지 않은 미지의 유전자이고 74%는 유전자 서열의 일부만 알려져 있는 expressed sequence tags(EST & ESTs)였다. 반면 기능이 알려진 유전자는 15%에 불과하였다.



(a) 배아줄기세포에서의 cluster 비율

(b) cluster 7에서의 유전자 특성

<그림 3> 전체 cluster의 비율과 cluster 7의 유전자에 관한 piechart



<그림 4> 경로 분석도에서의 최대우도법에 의한 경로계수

3.3 경로분석 (Path analysis)

유의한 유전자 후보군으로 가정한 cluster 7(N=106)에서 높은 발현량을 나타내는 유전자 중 기능이 알려지지 않은 ESTs와 Unknown2G1228 그리고 기능이 알려진 Veli 2와 Homer 3에 대한 경로분석 path diagram은 <그림 4>와 같았으며, 효과분할표는 <표 2>와 같았다. 경로분석은 총 3개의 모델을 가정하여 분석하였고, 3개의 모델 중 모델의 적합도를 나타내는 root mean squared residual과 good of fit index로 최종 모델을 선택하였다(root mean squared residual=0.02, good of fit index=0.92).

우선, 미지의 유전자인 ESTs와 Unknown2G1228의 기능을 파악하기 위해 기능이 알려진 2개의 유전자와의 상관분석 결과, ESTs와 Homer 3의 상관계수가 0.86으로 가장 높았고, Unknown2G1228과 Veli 2가 0.85, 그리고 ESTs와 Veli2가 0.83이었다. 이로부터 미지의 유전자인 ESTs와 Unknown2G1228는 Homer 3와 Veli 2의 기능과 유사한 기능을 하는 유전자이거나 Homer 3와 Veli 2의 기능적 작용을 촉진 혹은 억제 하는 유전자로 가정할 수 있었다.

이와 같은 가정을 바탕으로 ESTs와 Unknown2G1228이 Homer 3와 Veli2에 작용하는 관계를 파악하기 위해 <표 2>의 효과분할표의 직접효과와 간접효과를 살펴보면, Unknown2G1228의 Veli 2에 대한 직접효과의 경로계수는 0.55이고, 간접효과의 경로계수는 $0.30(=0.61 \times 0.50)$ 이었으며, Unknown2G1228과 Homer 3에 미치는 직접효과의 경로계수는 0.34이고 간접효과의 경로계수는 $0.40(=0.61 \times 0.65)$ 이다. 이와 같은 직접효과와 간접효과로부터 총효과 분석 결과 Unknown2G1228과 Veli2는 0.85인 관계를 갖고, ESTs와 Homer 3는 0.65인 관계로 상호 작용하는 유전자라 할 수 있었다.

<표 2> 유전자의 효과분할표

내적변수	사전변수	상관계수	간접효과	직접효과	총효과
ESTs	UNKNOWN2G1228	0.61		0.61	0.61
Veli 2	ESTs	0.83		0.50	0.50
	UNKNOWN2G1228	0.85	0.30	0.55	0.85
Homer 3	ESTs	0.86		0.65	0.65
	UNKNOWN2G1228	0.74	0.40	0.34	0.74

4. 결론 및 고찰

이번 연구는 줄기세포 응용연구의 기반 연구로 세포 운명 결정 및 분화에 대한 분자 수준에서의 기작을 이해하기 위해 생쥐 배아줄기세포를 이용한 cDNA microarray 방법으로 대단위적인 유전자 발현 양상을 확인하고자 하였고, 생쥐 배아줄기세포에서 분화단계별로 K-means clustering을 통해 군집화하여 관심 대상의 군집에 대한 유전자의 기능과 특성을 살펴보고자 경로분석을 시행하였다.

경로 분석에서 선택된 유전자의 생물학적 기능을 살펴보면, Homer 3는 neuronal immediate early gene(IEG)으로서 excitatory synapses에서 많이 발현되며, group I metabotropic glutamate receptors(mGluRs)에 결합하여 신경의 신호전달 기전을 조절하는 것으로 알려졌다(Kato 등, 1997). 글루타메이트 수용체를 경유한 신경세포간의 신호전달에 있어 Homer 3는 이 수용체에 결합해서 생체주기(circadian cycle)를 포함한 여러 신진대사를 조절하는 것으로 알려졌다(Park 등, 1997).

Veli 2(lin 7 homolog b, *C. elegans*)는 카르복실기에 PDZ 도메인이 존재하는 세포 결합-관련(cell junction-associated) 단백질로서, LET-23 수용체 인산화 효소(kinase)에 결합하며, *C. elegans*에서 동정된 본 유전자는 상피 세포에서 다량 발현하고 vulva의 발생을 조절한다(Simske 등, 1996). 한편 신호전달 기전과 관련된 신경세포의 시냅스 결합은 상피 결합과 유사성을 보이는 바, Veli 2는 포유류 신경세포에 존재하여 시냅스를 경유하는 신호전달 기전을 조절할 것으로 알려졌다(Jo 등, 1999).

결론적으로, Homer 3와 관련된 ESTs는 신호전달 기전을 조절하거나 신진대사 조절에 영향을 주는 유전자로 판단할 수 있고, Veli 2와 상호작용하는 Unknow2G1228은 Veli 2와 함께 포유류 신경세포에 존재하며 시냅스를 경유하여 신호전달을 조절하는 유전자로 볼 수 있다.

이와 같이 군집화와 인과 관계를 가진 경로 분석(path analysis)으로 기능이 밝혀지지 않은 다양한 유전자들과 기능이 알려진 유전자와의 관계 및 상호 연관성을 파악함으로써 유전자에 대한 기능 이해와 연구에 토대가 될 것으로 생각된다.

참고문헌

1. Blau H.M, Brazelton T.R, Weimann J.M (2001), The evolving concept of a stem cell: entity or function?, *Cell*, 105, 829-41.
2. Evans M.J, Kaufman M.H (1981), Establishment in culture of pluripotential cells from mouse embryos, *Nature*, 292, 154-6.
3. Jo K, Derin R, Li M and Brecht D.S (1999), Characterization of MALS/Veli-1, -2, and -3: a family of mammalian LIN-7 homologs enriched at brain synapses in association with the postsynaptic density-95/NMDA receptor postsynaptic complex. J., *Neuro.*, 19, 4189-99.
4. Kato A, Ozawa F, Saitoh Y, Hirai K, and Inokuchi K. (1997), Vesl, a gene encoding VASP /Ena family related protein, is upregulated during seizure, long-term potentiation and synaptogenesis, *FEBS Lett.*, 412, 183-9.
5. Park H.T, Kang E.K, and Bae K.W (1997), Light regulates Homer mRNA expression in the rat suprachiasmatic nucleus. *Brain Res. Mol. Brain Res.*, 52, 318-22.
6. Schena M, Shalon D, Davis R.W, Brown P.O (1995), Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science*, 270, 467-70.
7. Simske J.S, Kaech S.M, Harp S.A, Kim S.K (1996), LET-23 receptor

localization by the cell junction protein LIN-7 during *C. elegans* vulval induction. *Cell*, 85, 195-204.

8. Weissman I.L (2000), Stem cells: unit of development, units of regeneration, and units in evolution, *Cell*, 100, 157-68.
9. Yang Y.H, Dudoit S, Luu P and Speed T.P. (2001), Normalization for cDNA microarray data, *SPIE BiOS*.

[2005년 9월 접수, 2005년 11월 채택]