

cDNA microarray를 이용한 생쥐 배아줄기세포에서의 세포 분화능 조절에 관련된 유전자의 통계적 탐색 분석¹⁾

차경준²⁾, 김철근³⁾, 이영진⁴⁾, 김성주⁵⁾

요약

본 연구는 줄기세포 응용연구의 기반 연구로서 생쥐 배아줄기세포 (Embryonic stem cell)에서의 세포 운명 결정 및 분화에 대한 분자 수준에서의 기작을 이해하기 위해 cDNA microarray 방법을 통하여 대단위적인 유전자 발현 양상을 확인하고자 하였으며, 생쥐 초기 배아 발생 과정 및 배아줄기세포 분화 중에 전능성 유지 및 분화에 관련되리라 추측되는 유전자들에 대해서 K-means clustering 기법과 경로분석 (path analysis)을 통해 유전자 발현 양상을 확인하고자 하였다.

주요용어 : cDNA microarray, 배아줄기세포, K-means clustering, 경로분석

1. 서론

수정란은 발생 과정 동안 각기 다른 구조와 기능을 나타내는 여러 장기나 조직으로 분화하여 성장한 후 정상적인 개체를 형성한다. 이러한 수정란의 분화 과정과 성체 세포의 재생은 여러 가지 줄기세포의 분화능에 의해 조절된다. 줄기세포는 자기 복제 능력을 지니며, 특수한 환경하에서는 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 분화능 (pluripotency)을 지니고 있다 (Evans and Kaufman, 1981). 시험관 내 분화 유도 및 생체내 이식 과정 중에서 다양한 세포로 분화할 수 있는 성체줄기세포 (adult stem cell)는 성체의 각 조직이나 기관에 존재하고 있으며, 위치한 조직이나 기관에 존재하는 세포로의 제한된 분화능을 갖는다. 한편, 최근 성체줄기세포는 이식된 조직이나 장기의 특정 환경에 따라 다른 계열의 세포로 분화할 수 있는 유연성 (plasticity)을 갖는다고 보고되고 있다 (Blau 등, 2001). 줄기세포의 존재가 증명되고 줄기세포를 이용하여 퇴행성 뇌 질환, 심장 질환, 당뇨병 등의 난치병의 세포 치료 가능성도 제시되고 있지만, 줄기세포의 분화 조절 및 유연성에 대한 분자적 기작에 대한 연구는 미진한 편이다.

따라서 본 연구는 이러한 줄기세포 응용연구의 기반 연구로서 세포 운명 결정 및 분화에 대한 분자 수준에서의 기작을 이해하기 위해 생쥐 배아줄기세포를 이용한 cDNA microarray 방법 (Schena 등, 1995)을 통하여 대단위적인 유전자 발현 양상을 확인하고자 하였으며, 생쥐 초기

1) 위 연구는 21세기프론티어연구개발사업단인 세포응용연구사업단의 연구비 지원(M102KL010001-02K1201-00351)에 의해 수행되었음.

2) 한양대학교 수학과 교수

3) 한양대학교 생명과학과 교수

4) 한양대학교 대학원 생명과학과

5) 한양대학교 대학원 수학과

배아 발생 과정 및 배아줄기세포 분화 중에 전능성 유지 및 분화에 관련되리라 추측되는 유전자들에 대해 K-means clustering 기법을 이용하여 군집화하고, 경로분석 (path analysis)으로 인과모형 (causal model)을 제시하고자 하였다.

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 위하여 약 5000여개의 생쥐 cDNA 클론들이 심어져 있는 Mouse cDNA Microarray Membrane (Research Genetics)을 사용하였으며, Pathways 3 software를 이용하여 데이터를 얻었다.

2.2 통계적 분석

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 위하여 분화 후 0 시간의 배아줄기세포의 유전자 발현량과 각 분화단계별 유전자 발현량의 비를 구한 후, 밑이 2인 logarithm을 취했다. 그 후 데이터의 normalization은 Pathways 3 software에서 제공하는 control spot normalization과 데이터의 intensity에 관해서는 median normalization을 사용하였다. 그리고 각 분화단계별 유전자 발현량 차이에 관한 유의성 검정은 비모수 검정의 하나인 Kruskal-Wallis test를 이용하였고, 유전자 발현 양상의 차이에 대해서는 K-means clustering을 사용하여 발현 양상에 따라 그룹화를 하였다. 그리고 K-means clustering의 결과 중 관심 대상이 되는 cluster를 선택하여 유의한 유전자 후보군으로 결정하였다. 유전자 후보군에 포함된 유전자는 인과모형 (causal model)이라 할 수 있는 다변량 분석 기법 중 하나인 경로분석 (path analysis)을 통해 분석하였다. 마지막으로 발현량 차이의 검정에서 통계적인 유의성은 $p\text{-value} < 0.05$ 인 경우로 처리하였다.

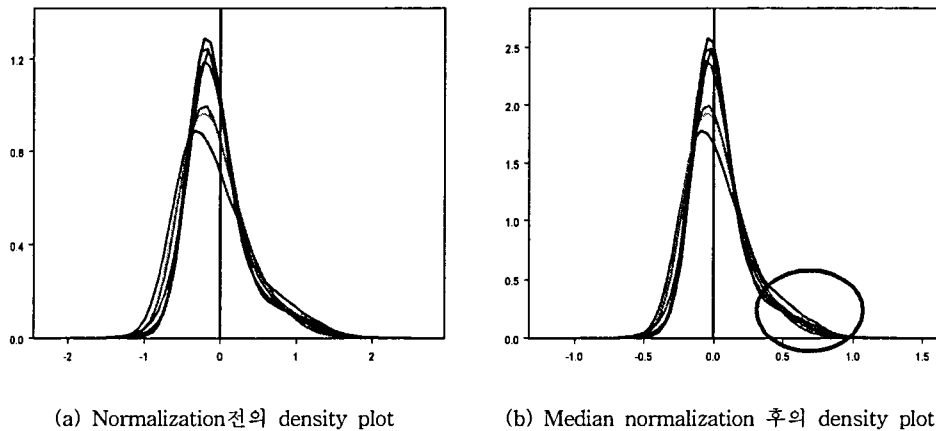
3. 연구결과

3.1 분화단계별 유전자 발현량의 분석과 Normalization

생쥐 배아줄기세포의 각 분화단계별 유전자의 발현량을 비교, 분석하기 위해 밑이 2인 logarithm을 사용하여 살펴본 데이터의 분포는 [그림 1(a)]와 같다. 그림에서 살펴볼 수 있듯이 0을 중심으로 데이터의 대부분이 약간 왼쪽으로 치우쳐 있고, 오른쪽 꼬리가 길게 나타났다. 데이터의 정규성을 살펴보기 위해 통계프로그램인 SAS를 이용하여 정규성 검정 및 QQ-plot을 그려본 결과 정규성을 만족하지 않았다. 따라서 분화단계별 유전자 발현량의 차이를 살펴보기 위해 비모수적 방법 중 하나인 Kruskal-Wallis test를 사용하였고, 그 결과는 [표 1]에서 제시된 바와 같이 각 분화단계별 유전자의 발현량은 유의한 차이를 보였다 ($p\text{-value} < 0.0001$).

cDNA microarray의 각 유전자에 관한 intensity의 normalization은 median normalization을 사용하였는데, 그 결과는 [그림 1(b)]와 같다. Normalization을 적용하기 전의 그래프와 비교해보면, 왼쪽으로 치우쳐 있던 분포는 0을 중심으로 이동하였고 분산이 전에 비해 많이 작아졌음을

알 수 있다. 하지만 오른쪽 꼬리 부분은 여전히 길게 나타나고 있는데, 이는 median normalization의 단점이라 할 수 있다. 분포의 양쪽 끝에 위치한 유전자들은 유의한 유전자 후보군인데 이번 연구에서는 오른쪽 꼬리 부분에 위치한 유전자들, 즉 [그림 1(b)]의 타원에 속하는 부분의 유전자를 유의한 유전자 후보군으로 가정하였다.



[그림 1] Normalization 전, 후의 density plot

[표 1] 배아줄기세포에서의 시간대별 발현량에 관한 평균과 표준오차

배아줄기세포	평균과 표준오차
12 hr	0.034±0.0026
24 hr	0.040±0.0028
36 hr	0.044±0.0034
48 hr	0.034±0.0031
72 hr	0.036±0.0030
96 hr	0.041±0.0027
120 hr	0.038±0.0030

3.2 K-means Clustering

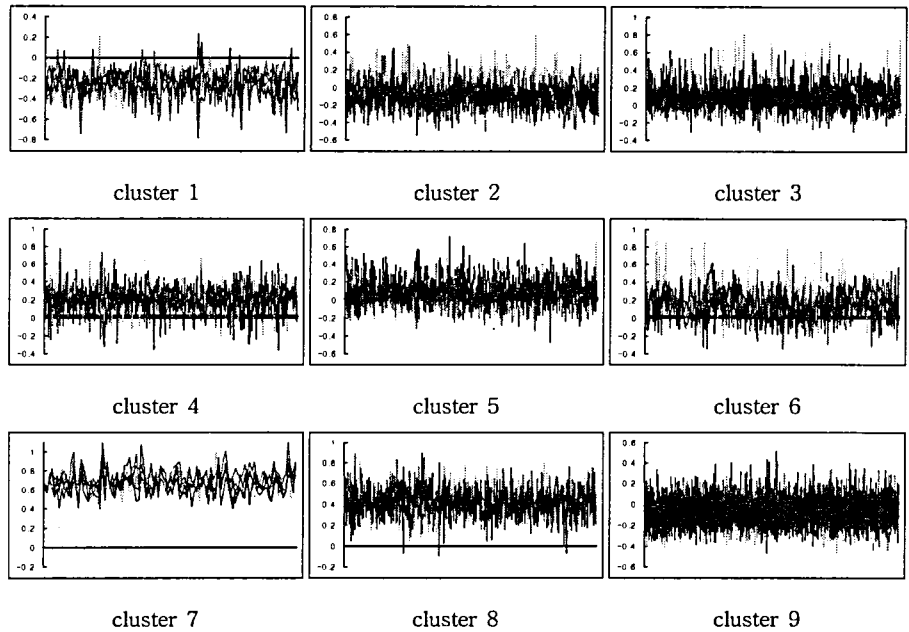
생쥐 배아줄기세포에서의 각 분화단계별 유전자 발현 양상이 다르기 때문에 유전자의 발현 양상에 따른 군집을 나누기 위하여 cDNA microarray의 데이터에 유클리드 거리 (Euclidian distance)를 적용한 K-means clustering (K=9)의 결과는 [그림 2]와 같다.

K-means clustering의 결과를 살펴보면, cluster 9 (47%), cluster 3 (15%), cluster 2 (8%) 등의 순으로 비율이 높음을 [그림 3(a)]를 통해 알 수 있다. 특히 cluster 9의 경우가 대부분을 차지하였는데 이것은 배아줄기세포에서의 각 분화단계별 유전자 발현량은 시간이 지나도 거의 차이가 없음을 보여준다. 즉, cluster 9에 포함된 유전자는 세포의 기본적인 물질대사에 필요한 유전자들로서, 거의 모든 세포 및 조직에서 발현되며 housekeeping gene 혹은 ubiquitous expressed gene들로 사료된다. 이와 비슷한 패턴을 보이고 있는 cluster는 cluster 2, 3, 5였고 분화함에 따라 발현량이 감소하는 유전자군은 cluster 1, 이와는 대조적으로 발현량이 증가하는

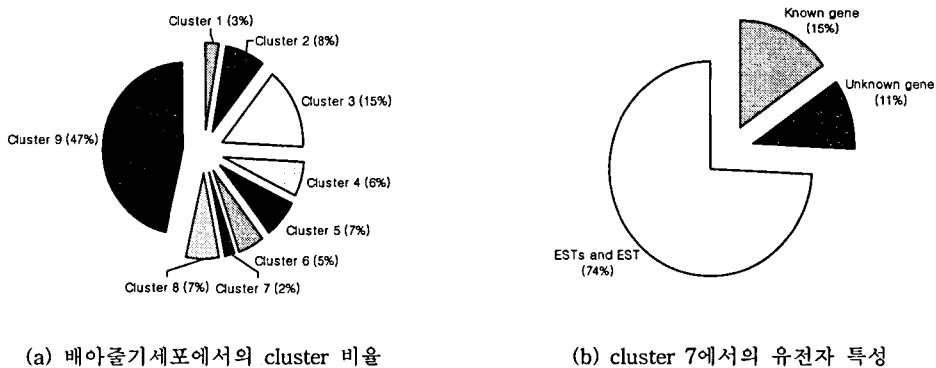
cDNA microarray를 이용한 생쥐 배아줄기세포

유전자군은 cluster 4, 6, 7이었다.

이번 연구에서 유의한 유전자 후보군으로 가정하였던 유전자들은 다른 유전자에 비해 배아줄기세포가 분화함에 따라 발현량이 증가하고 있는 패턴을 보여야 하는데, 그들은 cluster 7에 속해 있는 것을 [그림 2]를 통해 알 수 있었다. 그리고 cluster 7에 포함된 유전자의 특성은 [그림 3(b)]와 같다. Cluster 7에 속한 유전자의 11%는 기능이 알려져 있지 않은 미지의 유전자이고 74%는 기능이 일부만 알려져 있는 EST와 ESTs였다. 반면 기능이 알려져 있는 유전자는 15%에 불과하였다.



[그림 2] 배아줄기세포에서 분화단계별 유전자 발현량의 K-means clustering (K=9)



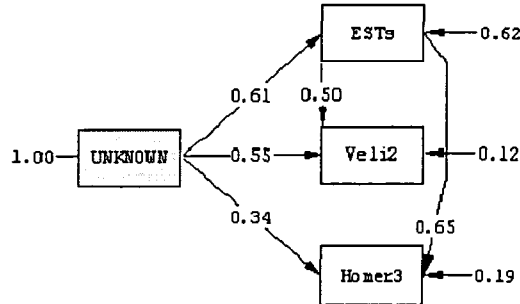
(a) 배아줄기세포에서의 cluster 비율

(b) cluster 7에서의 유전자 특성

[그림 3] 전체 cluster의 비율과 cluster 7의 유전자에 관한 piechart

3.3 경로분석 (Path analysis)

유의한 유전자 후보군으로 가정한 Cluster 7에서 평균적으로 높은 발현량을 나타내는 10개의 유전자를 선택하여 그들 중 1개의 ESTs와 Unknown gene (Unknown2G1128), 그리고 2개의 Known gene (Veli 2, Homer 3)으로 경로분석을 시행한 결과의 path diagram은 [그림 4]이고, 효과분할표는 [표 2]에 주어져 있다.



[그림 4] 경로 분석도에서의 최대우도법에 의한 경로 계수

[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

미지의 유전자인 Unknown2G1128과 ESTs의 기능을 유추해보고자 경로분석을 시행하였는데 그 결과를 살펴보면 ESTs와 Homer 3의 상관계수가 0.86으로 가장 높음을 알 수 있었다. 그리고 Unknown2G1128과 Veli 2, ESTs와 Veli 2 등의 순으로 상관계수가 높았는데 이는 미지의 유전자인 Unknown2G1128과 ESTs가 Homer 3와 Veli 2의 기능인 신경에서 신호전달 기전의 조절자와 같은 비슷한 기능을 가진 유전자라 유추되어진다. 그리고 미지의 유전자가 알려져 있는 유전자에 미치는 영향을 살펴보기 위해 효과분할표에서의 간접효과와 직접효과를 살펴보면, Veli 2는 ESTs의 간접효과 (35%)보다 Unknown2G1128의 직접효과 (65%)가 높았다. 반면, Homer 3의 경우는 ESTs의 간접효과 (54%)가 Unknown2G1128의 직접효과 (46%)보다 높음을 알 수 있었다. 따라서 Unknown2G1128과 Veli 2, ESTs와 Homer 3가 보다 밀접한 관계로서 상호작용을 하는 유전자로 사료되어진다.

4. 결론 및 고찰

이번 연구는 줄기세포 응용연구의 기반 연구로서 세포 운명 결정 및 분화에 대한 분자 수준에서의 기작을 이해하기 위해 생쥐 배아줄기세포를 이용한 cDNA microarray 방법을 통하여 단위적인 유전자 발현 양상을 확인하고자 하였고, 생쥐 배아줄기세포에서 분화단계별로

K-means clustering을 통해 군집화하여 관심 대상의 cluster에 대한 유전자의 기능과 특성을 살펴보고자 경로분석을 시행하였다.

그 결과 4개의 유전자가 선별되었는데 생물학적 의미를 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 알려져 있다 (Kato 등, 1997). 글루타메이트 수용체를 경유한 신경세포간의 신호전달에 있어 Homer 3는 이 수용체에 결합함으로써, 생체주기 (circadian cycle)를 포함한 여러 가지 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 (Simske 등, 1996), *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).

따라서 Homer 3 와 Veli 2는 공통적으로 신경의 시냅스를 경유한 신호전달 기전에 있어 중요한 조절자로서 생쥐 배아줄기세포의 분화시 신경세포 및 여러 가지 세포로의 분화를 조절할 것으로 사료된다. 또한 경로 분석을 토대로, 아직까지 기능이 밝혀지지 않은 유전자들 (ESTs, Unknown2G1128)과의 상호 작용을 통해 각 유전자의 발현 및 기능이 조절될 수 있으리라 판단된다.

참고문헌

- Blau H.M, Brazelton T.R, Weimann J.M (2001), The evolving concept of a stem cell: entity or function?, *Cell*, 105, 829-41.
- Evans M.J, Kaufman M.H (1981), Establishment in culture of pluripotential cells from mouse embryos, *Nature*, 292, 154-6.
- 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.
- 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.
- 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.
- 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.
- 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.
- Weissman I.L (2000), Stem cells: unit of development, units of regeneration, and units in evolution, *Cell*, 100, 157-68.
- Yang Y.H, Dudoit S, Luu P and Speed T.P. (2001), Normalization for cDNA microarray data, *SPIE BiOS*, 2001.
- 강현철 외, SAS Enterprise Miner 이용한 데이터마이닝, (1999)