인간 질병 네트워크로부터 얻은 질병 단백체의 특성 분석

Characterization of Diseasomal Proteins from Human Disease Network

이윤경, 구자을, 여명호*, 강태호*, 송민동**, 유재수*, 김학용
충북대학교, 충북대학교*, 건국대학교**

Lee Yoon Kyeong, Ku Jaeul, Yeo Myeong Ho*, Kang Tae Ho*, Song MinDong**, Yoo Jae-Soo*, Kim Hak Yong
Chungbuk National Univ., Chungbuk National Univ.*, Konkuk Univ**

요약

본 연구는 질병과 관련이 있는 단백질들은 질병네트워크를 형성함에 있어서 매우 중요한 인자로 작용할 가능성이 있다는 아이디어에서 출발한다. 우리는 Online Mendelian Inheritance in Man(OMIM)과 SWISS-PROT으로부터 인간의 단백질 데이터와 질병 정보를 확보하고 질병관련 단백질의 단백질 상호작용 네트워크를 구축 한 후, 이를 바탕으로 질병네트워크를 구축했다. 그 결과 단백질 상호작용 네트워크에는 CALM1, ACTB 및 ABL2와 같은 40개의 허브 단백질이 존재하는 것을 확인했다. 단백질 상호작용 네트워크와 질병 네트워크를 통해서 우리는 질병들간의 상관관계와 각 질병에 작용하는 단백질들의 상관관계를 파악할 수 있었다. 구축된 질병네트워크로부터 APP, ABL1 및 STAT1과 같은 38개의 질병단백체를 찾아냈다. 우리는 이전 연구에서 허브 단백질들이 서브 질병네트워크에서 질병 단백체의 경향이 있다는 것을 증명했다. 하지만, 본 연구에서 전체 질병 네트워크를 분석한 결과 전체 40개의 허브 단백질 중 단 18% 허브 단백질만이 질병단백체임이 확인되었다. 현시점에서 허브 단백질-질병단백체 경향성이 전체 질병네트워크와 서브 질병네트워크간의 차이를 설명할 수 없다. 비록 우리가 이러한 풀리지 않은 문제를 안고 있지만, 단백질-질병 네트워크의 구조 및 기능 분석은 복잡한 인간 질병 시스템에서 분자 수준의 기작과 생물학적 과정을 이해하는데 중요한 정보를 제공할 것이다.

Abstract

We initially obtained human diseases-related proteins dataset from the OMIM and the SWISS PROT and then constructed disease-related protein-protein interaction network. The protein network contains 40 hub proteins such as CALM1, ACTB and ABL2. The protein network can be derived the map of the relationship between different disease proteins, denoted disease interaction network. We demonstrate that the associations between diseases are directly correlated to their underlying protein-protein interaction networks. From constructed the disease-protein bipartite network, we derived 38 diseasomal proteins, including APP, ABL1 and STAT1. We previously demonstrated that hub proteins in the network tend to be diseasomal proteins in the disease-related protein sub-networks. However, we found that 18% hubs are only diseasomal proteins in the whole disease network. At this point, we could not elucidate difference in the hub-diseasomal proteins tendency between sub0network and whole network. In spite of we still have unsolved problems, our results elucidate that the discovery of protein interaction networks assigned by diseases will provide insight into the underlying molecular mechanisms and biological processes in complex human disease system.

Ⅰ. 서 론

질병에 대한 연구를 하기 위해 과학자들이 전통적으로 사용하던 방법은, 연구하고자 하는 각각의 질병에

* 본 논문은 2009년 교육과학기술부와 한국산업기술재단의 “지역혁신인력양성사업”과 교육과학기술부의 “지역거점연구단육성사업/충북BIT연구중심대학육성사업단”의 지원을 받아 수행된 연구임

관련된 물질들과 질병의 작용 기작을 독립적으로 분석하는 방법이었다. 하지만, 생명체의 시스템이 생체 구성 물질들의 독립적인 존재만으로는 완전한 생체 시스템을 유지할 수 없고, 생체 구성 물질들 간의 유기적인 상호작용 속에서 생체 시스템이 제대로 작용할 수 있다. 생체시스템의 이러한 복잡성과 상호 연관성에서 아이디어를 얻어, 우리는 인간을 이루는 전체 단백질의 상호작용 데이터를 기반으로 아토피, 유방암, 백혈병, 혈관생성과 관련된 질병 네트워크를 구축해 분석했다. 네트워크는 노드와 링크로 구성되어 있는데 단백질 네트워크에서 노드는 단백질을 의미하고 링크는 단백질간의 상호작용을 나타낸다[1]. 마찬가지로 질병 네트워크에서 노드는 질병을 의미하고 링크는 질병들간의 상호작용을 의미하는 것이다. 여기서, 다른 노드들에 비해 상호작용이 많은(연결된 링크 수가 많은) 노드들을 허브(Hub)라고 부른다[2]. 단백질 네트워크에서 허브는 세포 생존에 중요하게 관여하는 필수(essential) 단백질 혹은 치사(lethal) 단백질일 확률이 높다고 보고되었다[3][4].

단백질 네트워크를 구축, 분석하고 질병 네트워크를 구축해 분석한 결과를 비교해 봄으로써 복잡한 인간 질병 체계를 시스템수준에서 이해하는데 많은 도움이 될 것이다[5]. 최근에 인간 질병과 유전자 사이에 상관관계를 규명하는 연구 결과가 보고되었는데, 하나의 유전 질환은 다양한 유전자와 연결되어 있고 이들 유전자는 또 다른 질병과 연결되어있어 결국 비슷한 질병들이 연결고리를 형성하고 있음이 밝혀졌다[6].

따라서 본 연구에서는 아토피, 유방암, 혈관생성, 백혈병을 중심으로 단백질 상호작용 네트워크로부터 질병 상호작용 네트워크를 도출하고, 이 네트워크의 분석을 통해 관련 질병들의 질병 단백체를 찾아내고자 한다. 나아가서 이 질병 단백체들의 3차 구조 및 활성부위 등의 규명을 통해 각 종 질병의 작용을 저해 또는 치료할 수 있는 선도화합물의 단초를 제공할 수 있을 것으로 기대한다.

Ⅱ. 관련연구

인체의 모든 부위는 생체 시스템의 균형에 의해서 유지된다. 그 균형이 무너지면 상황에 따라 여러 가지 질병으로 나타나게 된다. 단백질 상호작용 네트워크도 생체 시스템을 이루고 있는 한 부분이기 때문에 단백질 상호작용 네트워크가 무너진다면 생체 시스템의 균형이 깨져 여러 가지 질병이 생길 수 있다[7]. 단백질 상호작용 네트워크는 결과적으로 생체 분자 네트워크를 형성하며 생명체 기능의 항상성을 유지하고 있기 때문이다. 단백질의 기능에 따라 분류된 단백질 상호작용 네트워크로부터 기능이 밝혀지지 않은 단백질들의 기능을 예측할 수 있듯이, 네트워크 상에서 단백질은 다른 단백질과 매우 밀접하게 연결되어 유사한 질병에 관여하기 때문에 밝혀지지 않은 질병관련 단백질 역시 예측할 수 있을 것이다[8]. 예를 들어, 헌팅톤병(Huntington's disease)관련 단백질 네트워크에는 헌팅톤과 관련된 많은 단백질들이 상호작용하고 있는데, 그 중에서 G protein coupled receptor kinase interaction protein인 GIT1가 이 질병에 관여하는 새로운 단백질임이 확인되었다[9].

최근에는 미생물 병원체의 단백질 상호작용 네트워크 분석을 통한 전염병 연구가 새롭게 주목 받고 있다[10]. 인간 단백질 네트워크와 병원체의 네트워크를 분석함으로써 두 개의 네트워크에서 공통적으로 나타나는 상호작용 단백질을 추출할 수 있으며, 이 단백질은 신약 개발을 위한 자료로 활용할 수 있다. 병원체 뿐만 아니라 바이러스 단백질 네트워크와 사람 단백질 네트워크와의 비교를 통해 전염성 질환에 대한 병인, 기작, 치료법 등의 연구가 점차 확산 되고 있다[11]. 이것은 잠재적으로 인간과 환경과의 상관관계까지도 분석할 수 있다는 것을 의미하기 때문에 이러한 단백질 및 질병 상호작용 네트워크에 대한 연구는 급속히 팽창할 것이다.

Ⅲ. 연구 방법 및 연구 결과

1. 연구 방법

아토피, 백혈병, 유방암, 혈관생성과 관련된 92개의 단백질 리스트를 OMIM(Online Mendelian Inheritance in Man) 데이터베이스로부터 얻었다. 아토피와 관련된 단백질 리스트는 43개, 백혈병과 관련된 단백질은 22개, 유방암과 관련된 단백질은 17개, 혈관생성 관련 단백질은 11개로 총 93개의 리스트 중에서 중복되는 단백질은 1개가 있었다. 92개의 단백질은 [표1]에 요약하였

다.

인간 단백질-단백질 상호작용 데이터는 HPRD (Human Protein Reference Database)로부터 얻어 각 질병 네트워크를 구축하고, 통합된 질병 네트워크도 추가로 구축하였다. InterViewer 프로그램을 이용하여 구축한 네트워크를 시각화 하였으며, OMIM과 Morbid Map, Swiss-Prot/ TreEMBL을 이용하여 네트워크에 포함된 단백질들의 질병 정보를 얻었으며 이를 바탕으로 단백질 정보를 질병 정보로 치환함으로써 단백질 네트워크로부터 질병 네트워크를 구축하였다.

단백질 및 질병 네트워크는 노드(단백질 혹은 질병)와 링크를 이용해 나타낸 것으로 노드간의 상호작용을 링크로 나타내지만, 이때 방향성은 고려하지 않았다.

표 1. 92개 단백질 리스트

번호	단백질명
1	Abelson murine leukemia viral oncogene homolog 1
2	Angiotensin-converting enzyme
3	ALRH
4	Amypoid beta A4 protein
5	Androgen receptor
6	Rho GTPase-activating protein 26
7	ADP-ribosylation factor-like protein 11
8	ASTHMA, susceptibility to
9	ASTHMA-related Traits, susceptibility to, 1
10	ASTHMA-related Traits, susceptibility to, 2
11	ASTHMA-related Traits, susceptibility to, 3
12	ASTHMA-related Traits, susceptibility to, 4
13	ASTHMA-related Traits, susceptibility to, 5
14	ASTHMA-related Traits, susceptibility to, 6
15	Serine-protein kinase ATM
16	Dermatitis, Atopic
17	Dermatitis, Atopic 6
18	BRCA1-associated RING domain protein1
19	Apoptosis regulator BAX
20	B-cell lymphoma/leukemia 10
21	Breast cancer type 1 susceptibility protein
22	Breast cancer type 2 sysceptibility protein
23	Fanconi anemia group J protein
24	Caspase recruitment domain containing protein11
25	Calcium-binding atopy-related autoantigen1
26	Eotaxin
27	C-C motif chemokine 24
28	C-C chemokine receptor type 5
29	Epithelia cadherin
30	CCAA T/enhancer-binding protein alpha
31	Serine/threonine-protein kinase Chk2
32	Chymase
33	Transcription factor ETV6
34	Filaggrin
35	FL cytokine receptor
36	Vascular endothelial growth factor receptor 3
37	Erythroid transcription factor
38	Gap junction beta-2 protein
39	Hepatitis A virus cellular receptor 1
40	HIV-1
41	HLA class 2 histocompatibility antigen, DRB1-1 beta chain
42	Homeobox protein Hox-D4
43	Interferon gamma
44	IGAD1
45	IGER
46	Ig gamma-2 chain C region IGHG2
47	DNA-binding protein Ikaros
48	Interleukin-10
49	Interleukin-12 subunit beta
50	Interleukin-13
51	Interleukin-13 receptor alpha-1 chain
52	Interleukin-4 receptor alpha chain
53	Interleukin-9
54	IPEX
55	Interferon-beta Promoter stimulator protein
56	Tyrosine-protein kinase JAK2
57	Vascular endothelial growth factor receptor 2
58	Mast./stem cell growth factor receptor
59	GTPase KRas
60	KTCN1
61	Menin
62	High affinity immunoglobulin epsilon receptor subunit beta
63	DNA mismatch repair protein Msh2
64	Nijmegen breakage syndrome protein 1
65	Neurofibromin
66	Nucleophosmin
67	P2X purinoceptor
68	Cellular tumor antigen p53
69	Prohibitin
70	PHD finger protein 11
71	Peptidase inhibitor
72	Phosphatidylinositol 3-kinase
73	Platelet-activating factor acetylhydrolase
74	Tyrosine-protein phosphatase non-receptor type11
75	DNA repair protein RAD51 homolog 1
76	DNA repair and recombination protein RAD54-like
77	RB1-inducible coiled-coil protein 1
78	U4/U6.U5 tri-snRNP-associated protein 1
79	SV40 early leader protein
80	Plasma protease C1 inhibitor
81	Solute carrier family 22 member 18
82	Serine protease inhibitor Kazal-type 5
83	Signal transducer and activator of transcription 6

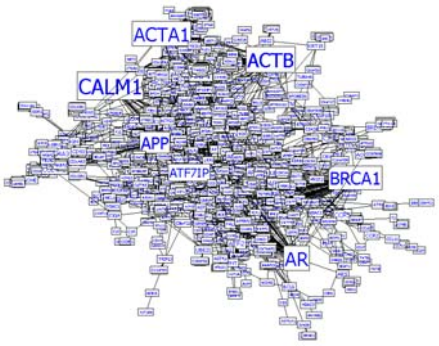
84	tRNA 2'-phosphotransferase
85	Hamartin
86	Tuberous sclerosis 2 protein
87	Tumor susceptibility gene 101 protein
88	Uteroglobin
89	Von Hippel-Lindau disease tumor suppressor
90	Xaa-Pro aminopeptidase 2
91	DNA repair protein XRCC3
92	Zinc finger MYM-type protein 2

2. 연구 결과

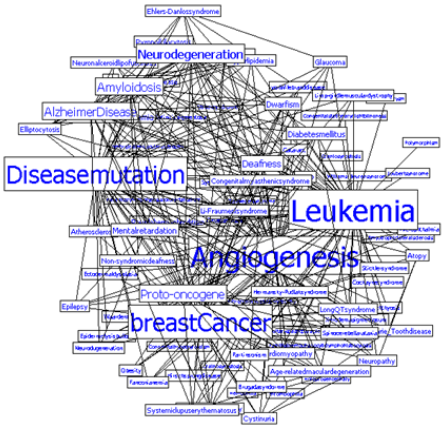
표1에서 보여준 질병관련 단백질과 HPRD의 단백질 상호작용 데이터를 이용하여 4개 질병들의 단백질 상호작용 네트워크를 구축하였다. 구축된 단백질 네트워크는 1,900개의 노드(단백질)과 3,192개의 링크(상호작용)으로 구성되었고[그림1A], ACTA1, ACTB, APP, AR, BRCA1, CALM1, ABL1 등과 같은 허브 단백질도 발견되었다. 네트워크를 구성하는 단백질들을 질병으로 전환하였는데, 단백질 하나가 여러 개의 질병과 연관되어 있을 경우, 하나의 단백질을 다수의 질병 노드로 전환하였으며, 단백질 중에서 질병과 관련이 없는 단백질들은 전환 과정에서 제거되었다.

구축된 질병 네트워크를 구성하는 93개의 질병 리스트 중에서 유방암, 백혈병, Disease Mutation, 신생혈관형성에 과한 질병들이 비교적 다른 질병과의 많은 링크를 가지고 있었다. 아토피 관련 단백질이 43개나 있음에도 불구하고 질병 상호작용 수는 매우 적어 질병 네트워크 상에서 중심 질병으로는 나타나지 않았다. 이는 아토피 질병은 다른 질병과 연계되어 있기 보다는 면역 및 알레르기 관련 몇몇 질병에 한정하여 상호작용하고 있다고 볼 수 있다.

질병단백체를 발굴하기 위하여 질병과 질병관련 단백질 사이의 상관관계를 이분할 그래프를 만들어 분석해 보았다. 각 질병에 관여하는 단백질과 질병과의 연결고리를 통해서 우리는 각 질병들에서 많이 발견되는 질병 단백체를 발굴할 수 있었다. 그림 1에서 보여준 허브 단백질들 중에서 ACTA1, APP는 질병 네트워크를 분석해 얻은 질병 단백체 리스트에도 포함되어, 질병과 매우 밀접한 질병 단백체임을 확인할 수 있었다. 하지만 질병 네트워크를 통해 얻은 대부분의 질병 단백체들은 단백질 상호작용 네트워크에서 허브로 나타나지 않았던 단백질들이었다.



▶▶ 그림 1. 질병 관련 단백질-상호작용 네트워크



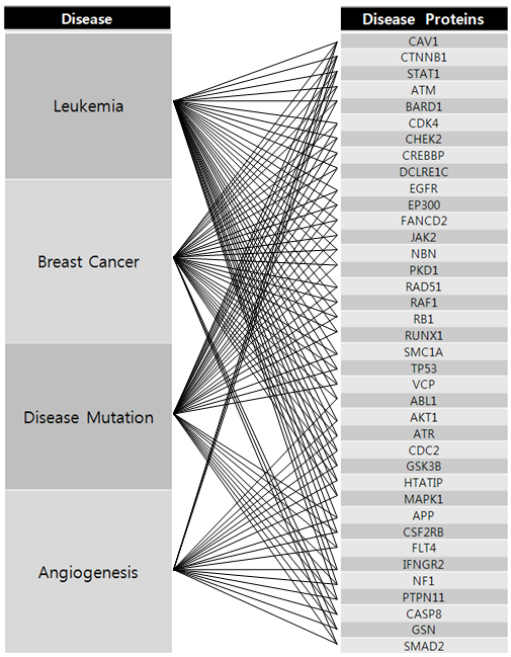
▶▶ 그림 2. 질병-질병 상호작용 네트워크

Ⅳ. 결론

본 연구에서 우리는 생물정보학적 접근 방법을 이용하여 질병 관련 단백질 네트워크와 질병 네트워크와의 상관관계를 규명하였다. 간략하게 요약하면, OMIM 데이터베이스로부터 백혈병, 아토피, 유방암, 혈관생성과 관련된 단백질 리스트를 얻은 다음 단백질 상호작용 네트워크를 구축하였다(그림 1). 네트워크에 포함된 단백질들을 관련 질병으로 전환하여 질병 단백질 네트워크를 구축한 후(그림 2), 이를 바탕으로 질병 단백체를 발굴하였다(그림 3). 이러한 접근 방법은 생체 시스템에서 중요한 역할을 하는 단백질들을 찾아낼 수 있을 뿐만 아니라 질병 네트워크에서 중요하게 작용하는 질병 단백체에 대한 정보도 얻을 수 있다. 그리고 더 나아가 단백질 상호작용 네트워크와 질병 네트워크를 비교, 분석 함으로써 정상적인 시스템에서의 네트워크 특징과

질병 상태에서의 시스템의 차이를 밝히는데 사용될 수 있을 것으로 생각된다.

아토피 관련 질병 질병네트워크에서는 허브 단백질 중 75%가 질병단백체였으나 본 연구 결과에서는 허브 단백질들 중 오직 18%만이 질병단백체임이 확인되었다. 왜 이 두 연구결과에서 차이가 나는지는 현시점에서는 명확하지 않다. 전체 질병네트워크에서는 비교적 상호작용이 적은 소규모 질병들은 사라지기 때문에 허브 단백질들은 사라지고 질병과 질병을 연결하는 정보의 양이 방대하면 허브 단백질 보다는 질병을 서로 연결하는 단백질이 질병단백체로 보여줄 수 있다.



▶▶ 그림 3. 질병 네트워크로부터 질병단백체 발굴을 위한 이분할 그래프

이 부분에 대해서는 좀 더 연구하여야겠지만, 특 질병에 대해 중점적으로 분석하기 위해서는 전체 질병 네트워크를 구축하여 분석하기 보다는 특정 질병 네트워크를 만들어 분석하는 것이 바람직할 것이다. 예를 들어 아토피 질병의 경우, 43개의 질병단백질을 가지고 있음에도 불구하고 질병 네트워크에서는 아토피 질병이 명확하게 나타나지 않았다(그림 2).

질병 네트워크 분석을 통해 발굴한 질병 단백체는 분자 수준에서 선도화합물에 대한 실마리를 제공할 수 있을 것이다. 네트워크 분석 결과, 이미 잘 알려져 연구가 많이 이루어진 질병과 그렇지 않은 질병 사이의 상관관계를 통해 분자 수준에서 연결고리를 확인할 수 있었다.

결론적으로, 본 연구에서 제시한 생물정보학적 접근 방법은 좀 더 복잡한 질병 시스템을 분자 수준에서의 기작을 이해하는데 깊은 통찰력을 제공할 것이다. 또한 사람 질병 단백질 전체 네트워크를 구축하여 질병 단백체를 확보하고 아미노산 서열이나 도메인의 특성을 분석하여 신약 개발을 위한 선도화합물을 발굴하는데 정보를 제공할 수 있을 것으로 기대한다.

■ 참 고 문 헌 ■

[1] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A. -L. barabasi, "The large-scale organization of metabolic networks," Naure, Vol.407, No.6804, pp.41-42, 2001.

[2] A. -L. Barabasi, Z. N. Oltvai, "Network biology: understanding the cell's functional organization," Nat. Rev. Genet., Vol.5, No.2, pp.101-113, 2004.

[3] H. Jeong, S. P. Mason, A. L. Barabasi , Z. N. Oltvai. "Lethality and centrality in protein networks," Nature, Vol.411, No.6833, pp.41-42, 2001.

[4] H. Yu, D. Greenbaum, H. X. Lu, X. Zhu, M. Gerstein. "Genomic analysis of essentiality within protein networks," TRENDS in Genomics, Vol.20, No.6, pp.227-231, 2004.

[5] S. Bortoluzzi, C, Romualdi, A. Bisognin, G. A. Danieli, "Disease genes and intracellular protein networks," Physiol. Genom. Vol.15, No.3, pp.223-227, 2003.

[6] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, A. -L. Barabasi, "The human disease network," Proc. Nat'l. Acad. Sci. Vol.104, No.21, pp.8685-8690, 2007.

[7] L. Sam, Y. Liu, J. Li, C. Friedman, Y. A. Lussier, "Discovery of protein interaction networks shared by diseases," Pacific Sym. on Biocomp. Vol.12, No.1 pp.76-87, 2007.

[8] T. Ideker, R. Sharan, "Protein networks in disease," Genome Res. Vol.18, No.4, pp.644-652, 2008.

[9] H. Goehler, M, Lalowski, U. Stelzl, S. Waelter, M. Stroedick, U. Worm, E. E. Wanker, "A protein interaction network links GIT1, an enhancer of huntintin aggregation, to Huntington's disease," Mol. Cell. Vol.15, No.6, pp.853-865, 2004.

[10] J. R. Parrish, J. Yu, G. Liu, J. A. Hines, J. E. Chan, R. L. Finlev Jr., "A proteome-wide protein interaction map for Campylobacter jejuni," Genome Biol. Vol.8, No.7, pp.R130, 2007.

[11] M. A. Calderwood, K, Venkatesan, L. Xing, M. R. Chase, A. Vazquez, A. M. Holthaus, E. Kieff, E. Johannsen, "Epstein-Barr virus and virus human protein interaction maps," Proc. Nat'l. Acad. Sci., Vol.104, No.18, pp.7606-7611, 2007.