

# 단백질 상호작용 네트워크를 통한 유전체 단위반복변이와 트랜스유전자 발현과의 연관성 분석

박 치 현<sup>†</sup> · 안 재 균<sup>†</sup> · 윤 영 미<sup>††</sup> · 박 상 현<sup>†††</sup>

## 요 약

인간 유전체에 존재하는 유전적 구조 변이(genetic structural variation) 중 하나인 유전체 단위반복변이(Copy Number Variation, CNV)은 유전자의 기능 발현과 밀접한 관련이 있다. 특히 특정 유전 질환이 있는 사람들을 대상으로 CNV와 유전자발현의 관계를 밝히는 연구가 계속 진행되고 있지만, 정상인 유전체에 대한 CNV의 기능적 분석은 아직 활발히 이루어지고 있지 않다. 본 논문에서는 다수의 정상인 샘플에서 찾아낸 공통된 CNV에 대하여 유전자들과의 기능적 관계를 유전자의 분자적 위치와 상관없이 밝힐 수 있는 분석 방법을 제시한다. 이를 위해 서로 다른 이질적인 생물학데이터를 통합하는 방법을 제시하고 공통된 CNV와 유전자와의 연관성을 분자적 위치와 상관없이 계산할 수 있는 새로운 방법을 제시한다. 제안된 방법의 유의성을 보이기 위해서 유전자 온톨로지(Gene Ontology) 데이터베이스를 이용한 다양한 검증 실험들을 수행하였다. 실험결과 새롭게 제안된 연관성 측정방법은 유의성이 있으며 공통된 CNV와 강한 연관성을 갖는 유전적 기능의 후보들을 시스템적으로 제시할 수 있는 것으로 나타났다.

키워드 : 유전체 단위반복변이, 트랜스유전자, 전장유전체연관분석연구, 단백질상호작용네트워크

## Genome-Wide Association Study between Copy Number Variation and Trans-Gene Expression by Protein-Protein Interaction-Network

Chihyun Park<sup>†</sup> · Jaegyeon Ahn<sup>†</sup> · Youngmi Yoon<sup>††</sup> · Sanghyun Park<sup>†††</sup>

## ABSTRACT

The CNV (Copy Number Variation) which is one of the genetic structural variations in human genome is closely related with the function of gene. In particular, the genome-wide association studies for genetic diseased persons have been researched. However, there have been few studies which infer the genetic function of CNV with normal human. In this paper, we propose the analysis method to reveal the functional relationship between common CNV and genes without considering their genomic loci. To achieve that, we propose the data integration method for heterogeneity biological data and novel measurement which can calculate the correlation between common CNV and genes. To verify the significance of proposed method, we has experimented several verification tests with GO database. The result showed that the novel measurement had enough significance compared with random test and the proposed method could systematically produce the candidates of genetic function which have strong correlation with common CNV.

Keywords : Copy Number Variation, Trans-gene, Genome-Wide Association Study, Protein-Protein Interaction Network

## 1. 서 론

유전체 단위반복변이는 가장 대표적인 유전체 구조적인 변이 중 하나로, 최근 유전체학 연구 분야에서 가장 많은

관심의 대상이 되고 있는 연구 분야이다. 2003년 종료된 휴먼 게놈 프로젝트[1]를 통해서 인간의 전체 DNA 염기서열이 결정되었고, 이후 유전체 내에 존재하는 유전자, 유전 조절 부위, 구조적 변이 등과 같은 모듈들에 대한 분석을 통하여 그 기능을 밝히려는 연구의 필요성이 증가하고 있다. 대표적으로 서로 다른 인간 유전체 내에 존재하는 다형성(Polymorphism)의 하나인 단일염기변이(Single Nucleotide Polymorphism, SNP)를 통하여 병리적 혹은 형질적 특징을 분석하는 연구가 많이 진행되어 왔다[2][3]. 기존의 연구들은 찾아낸 SNP의 존재를 통하여 특정 질병을 갖는 그룹에 대한 유전적 표식(Genetic Marker)으로 활용해왔다. 실제로 다

\* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2010-0008639).

\*\* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2010-0010688).

† 준 회 원 : 연세대학교 컴퓨터과학과 박사과정

†† 종 심 회 원 : 가천의과학대학교 정보공학과 교수

††† 종 심 회 원 : 연세대학교 컴퓨터과학과 교수(교신저자)

논문접수 : 2010년 11월 26일

수정일 : 1차 2011년 2월 14일, 2차 2011년 2월 15일

심사완료 : 2011년 2월 16일