

특정변화패턴 식별을 위한 염기서열 집단간의 다형성 분석 및 시각화 도구

이일섭¹, 이건명^{2*}

¹충북대학교 소프트웨어학과 학생, ²충북대학교 소프트웨어학과 교수

A Polymorphism Analysis and Visualization Tool for Specific Variation Pattern Identification in Groups of Nucleotide Sequences

Il Seop Lee¹, Keon Myung Lee^{2*}

¹Student, Department of Computer Science, Chungbuk National University

²Professor, Department of Computer Science, Chungbuk National University

요 약 유전체는 생명체가 가지고 있는 모든 유전적 정보를 담고 있다. 특정 종 내에서는 개체별로 고유의 특성이 나타나며, 이 특성은 유전체의 염기서열 분석을 통해 확인할 수 있다. 종내 개체들 사이에 조금씩 다른 염기에 대해 유전적 연관성을 규명 짓고, 더 나아가 질병과의 연관성을 찾는 전장유전체 연관분석 연구가 많이 진행되고 있다. 종 내의 조금씩 발생하는 염기변이를 파악하는 것은 개체의 다형성을 파악하기 위해 중요하다. 이 논문에서는 종 내 여러 개체의 염기서열에서 대립형질 빈도의 특정변화패턴을 쉽게 파악할 수 있는 분석 및 시각화 도구를 제안한다. 그리고 수두 대상포진 바이러스의 계대 배양한 pOka strain 염기서열 데이터를 이용해 실험하여 분석과 시각화의 실용성을 보인다. 본 제안도구를 통해 종 내의 대립형질 빈도의 변화를 탐색하고 유전적 요인을 찾는 연구효율의 증진을 기대할 수 있다.

주제어 : 시각화, 데이터 분석, 자동화, 단일염기다형성, 전장 유전체 연관 분석

Abstract A genome contains all genetic information of an organism. Within a specific species, unique traits appear for each individual, which can be identified by analyzing nucleotide sequences. Many Genome-Wide Associations Studies have been carried out to find genetic associations and cause of diseases from slightly different base among the individuals. It is important to identify occurrence of slight variations for polymorphisms of individuals. In this paper, we introduce an analysis and visualization tool for specific variation pattern identification of polymorphisms in nucleotide sequences and show the validity of the tool by applying it to analyzing nucleotide sequences of subcultured pOka strain of varicella-zoster virus. The tool is expected to help efficiently explore allele frequency variations and genetic factors within a species.

Key Words : Visualization, Data analysis, Automation, Single Nucleotide Polymorphism(SNP), Genome-Wide Association Study(GWAS)

*This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Seoul Accord Vitalization Program(IITP-2018-2012-1-00598) supervised by the IITP(Institute for Information & communications Technology Promotion)

*Corresponding Author : Keon Myung Lee(kmlee@cbnu.ac.kr)

Received October 16, 2018

Revised November 12, 2018

Accepted December 20, 2018

Published December 31, 2018

1. 서론

최근 유전체 연구의 성장과 더불어 염기서열 분석 기술의 발달에 따라 모든 생명체의 복잡한 염기서열을 빠르고 저비용으로 얻을 수 있다[1]. 한 개체의 전체 염기서열은 완전한 유전 정보의 총합이며, 모든 유전자와 유전자가 아닌 부분으로 구성된다. 유전자는 DNA서열 가운데 정보를 갖는 부분을 뜻하며, 유전자가 아닌 부분은 정보가 없는 무작위 서열로 비부호화 DNA서열 또는 쓰레기 DNA(Junk DNA)이라 하며 인간의 게놈 가운데 99%에 해당된다[2]. 쓰레기 DNA의 존재 이유와 역할이 무엇인지는 아직 밝혀지지 않았으나, 최근 그 중요성과 관련된 연구가 나타나고 있다[3,4].

게놈은 종마다 다를뿐 아니라 집단, 개체 마다 조금씩 차이를 보인다. 이러한 차이는 염기서열의 차이에서 비롯되며, 염기서열의 차이는 개체의 대립형질을 결정해 특정 종 내의 개체의 다형성을 결정짓는데 중요한 역할을 한다. 종 내의 염기서열에서 단일 염기의 차이를 단일 염기다형성(Single Nucleotide Polymorphism, SNP)라 하며, 인류전체중 적어도 1% 이상이 갖고 있는 공통된 SNP를 분석하는 해맵 프로젝트(Hapmap Project)과 같이 SNP 분석을 통해 개체간 어떤 유전적 차이로 이어지는지 분석하는 많은 연구가 이루어졌다[5,6].

이 논문에서는 염기서열의 염기분포를 분석하고, 염기서열간의 염기분포의 차이를 쉽게 파악할 수 있는 분석 및 시각화 어플리케이션을 제안한다. 그리고 수두대상포진바이러스(VZV)의 염기서열 데이터를 통해 실험한 결과를 소개한다.

2. 관련연구

2.1 단일염기다형성(SNP)

개체군에서 개체간의 염기서열 차이는 단일염기의 차이가 많으며, 염기의 결손(deletion), 삽입(insertion) 및 치환(substitutions)를 통해 나타난다. 한 개체군의 염기서열에서 대립형질을 결정할 때, 가장 많은 염기를 다수염기(major), 두 번째로 많은 염기를 소수염기(minor)로 분류한다.

2.1.1 MAF(Minor Allele Frequency)

MAF는 두 번째로 가장 많은 염기인 소수염기의 빈도

를 의미하며, 집단 유전학(Population Genetics) 연구에서는 개체군 내의 공통적 변이와 희귀 변이를 구분할 수 있는 정보를 얻을 수 있어 중요하다. MAF가 5% 이상인 경우를 공통변이, 5% 미만은 희귀 변이로 간주하며 희귀 변이는 염기서열 분석 방법에 따라 염기서열 분석 에러(Sequencing Error)와 구분이 필요하다[7]. 일반적으로 비부호화 DNA 영역이 부호화 DNA영역보다 공통적 변이가 더욱 빈번하게 발생한다[8].

2.1.2 공통변이 (Common Variants)

개체군 내에서 흔히 발생하는 변이로, 일반적으로 MAF가 5% 이상인 경우를 공통변이로 간주하며, 하나의 개체군에서 SNP로써 변이를 구분하기 위해서는 MAF가 적어도 1% 이상이어야 한다. 지난 약 10년간 인간의 질병과 관련한 유전체 및 GWAS연구는 공통변이에 초점이 맞추어 연구되었다[9,10]. 공통변이는 한 개체군에서 흔히 발생하는 질병과 연관되어있다고 알려져 있다[11].

2.1.3 희귀변이 (Rare Variants)

개체군 내에서 드물게 발생하는 변이로, 일반적으로 MAF가 5% 미만인 경우를 대상으로 한다. 희귀변이는 개체군 내에서 흔히 발생하는 공통변이만으로는 설명하기 힘든 질병에 대한 잃어버린 유전성(missing heritability)과 일부 연관성이 있다고 알려져 있으며[12], 빠른 개체 증가와 음성 선택이 약한 환경에서 공통변이보다 더욱 해로운 질병을 유발할 수 있다고 알려져 있다[13].

2.2 전장 유전체 연관 분석 (GWAS)

전장 유전체 연관 분석(Genome-Wide Association Study, GWAS)은 모든 유전체의 유전적 변이가 어떤 형질과 연관성을 갖는지 탐색하는 연구이다. 한 개체군내에서 흔히 발생하는 질병(common disease)들은 개체들이 흔히 가지고 있는 공통변이(common variants)에 의해 발생된다는 가설을 기반으로 질병 및 약물 반응성에 대한 유전적 요인을 총체적으로 탐색하는 연구로 진행되고 있다[14].

2.2.1 맨하탄 플롯(Manhattan Plot) 시각화

GWAS 연구에서 흔히 사용되는 산점도이며, 거리가 가까운 유전자 위치끼리는 유전자 재조합으로부터 영향을 적게 받아 연관성을 가진다는 사실을 기반으로 유의

확률(p-value)을 이용한 시각화이다. Fig. 1과 같이 GWAS 분석 결과 의미 있는 유전 위치에 있는 SNP들의 p-value가 고층빌딩과 같이 나타나는 특징을 가진다.

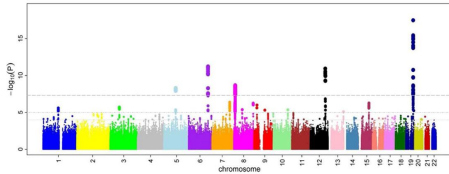


Fig. 1. Manhattan plot for GWAS
Genomic coordinates are displayed along the X-axis, with the negative logarithm of the association p-value for each SNP displayed on the Y-axis

최근 SNP 및 GWAS 연구를 통해 인간 DNA 서열의 변이가 어떤 질병을 일으키고, 병원균, 화학 물질, 약물, 백신 등에 어떻게 반응하는지 파악할 수 있어 중요하며, 개인화된 약물의 개발을 위해 응용된다[15].

3. 단일염기다형성 분석 및 시각화 도구

3.1 시스템 구성도

본 논문에서 제안하는 분석 및 시각화 도구는 Fig. 2와 같은 요소들로 구성된다. 그래픽 유저 인터페이스를 통해 정렬된 염기서열들로 구성된 파일의 입력과 입력 데이터에 해당하는 유전정보를 기입할 수 있다. 분석기에서는 선택한 분석 종류에 따라 분석해 엑셀문서화하며, 시각화는 개체사이의 MAF의 변화량과 사용자 상호작용을 통한 관심영역 필터링으로 구성된다.

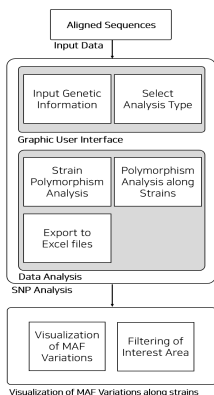


Fig. 2. System configuration for the tool
System configuration consists of input module to read in the sequences aligned, GUI for entering the information of input file, and polymorphisms analyzer and visualization.

3.2 입력 데이터

정렬된 염기서열 데이터를 입력데이터로 받으며, 염기서열의 정렬은 무료로 사용가능한 ClustalW, Kalign, MAFFT와 같은 다중서열정렬(Multiple Sequence Alignment) 프로그램을 통해 가능하다[16]. 또한, 입력데이터는 염기서열의 위치별 염기종류 A, G, C, T의 빈도와 게놈 구조(Genome Structure), 반복되는 서열, ORF (Open Reading Frame), NCR(Non-Coding Region) 정보로 구성된다.

3.3 단일염기다형성 분석기

3.3.1 그래픽 유저 인터페이스

분석기의 그래픽 유저 인터페이스는 Fig. 3과 같이 구성된다. 정렬된 염기서열 데이터를 입력하고 분석에 필요한 컬럼 명 설정 및 유전적 정보를 입력할 수 있는 영역으로 구성된다. 분석은 게놈구조와 반복되는 지역, 유전자 발현지역과 발현되지 않는 지역의 4가지 기준으로 각 염기위치에서 소수염기의 비율 변화를 집게한 엑셀문서를 제공한다. 분석 종류는 개체 다형성 분석과 개체 사이의 다형성 분석이 있으며, 3.3.2절과 3.3.3절에서 설명한다.



Fig. 3. Graphic User Interface of polymorphisms analyzer

A user can input the file of aligned nucleotide sequences and genetic information of it.

3.3.2 개체 다형성 분석

염기서열 분석 데이터의 신뢰성을 위해 동일위치에서 단일염기 빈도의 합계 35 이상인 염기위치만을 분석 및

시각화 대상으로 한다. 각각의 개체에 대해 소수염기 빈도를 분석하며 게놈구조, 반복되는 서열, ORF, NCR 및 염기의 Major/Minor 조합별 MAF 구간에 따라 분석을 한다. MAF 구간은 공통변이에 해당하는 5-15%, 15-25%, 25%이상 및 희귀변이에 해당하는 1.0-2.5%, 2.5-5.0%로 구분한다.

3.3.3 개체사이의 다형성 분석

개체간의 소수염기 빈도의 변화를 분석하며, 각 개체 사이의 MAF의 편차를 계산하여 분석에 사용한다. 또한, Major/Minor 조합의 변화하는 영역은 별도로 추출하여 보여주며, 그 외 분석방법 및 결과 형식은 3.2.1의 개체 다형성 분석과 같다.

3.4 개체간 다형성 변화량 시각화

3.4.1 MAF 변화의 시각화

본 시각화는 염기서열간의 MAF 변화를 표현한다. 기존의 소수염기의 비율만을 표기하던 산점도 시각화를 개선해, 다수염기와 소수염기의 종류와 소수염기비율의 변화를 파악가능하다. 시각화는 Fig. 4와 같은 요소로 구성된다. 염기의 종류를 색상을 통해 구분한다. 또한, 소수염기의 비율변화에 따라 밝기가 진해지는 방향으로 변화함을 표현하며, 다수 염기는 소수 염기의 변화하는 방향의 머리 부분에 표현된다. 소수염기의 변화는 변화방향에 따라 염기 색상의 변화를 통해 표현한다. 위 구성요소를 통해 시각화한 모습은 Fig. 5와 같다.

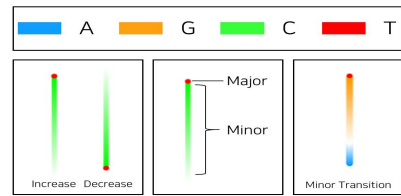


Fig. 4. Snapshots of visualization components

Each base A, G, C, T is identified by colors. Increases and decreases are identified by the direction of brightness changes. Each component is consist of Major placed ends of variation and Minor changes brightness. Minor transition can be identified by changes of base color.

3.4.2 관심영역 필터링

사용자 상호작용 기능을 통해 Fig. 8과 같이 MAF의 변화량과 Major의 염기종류에 따라 데이터를 필터링하여 사용자의 관심영역을 탐색할 수 있다. 염기서열의 위치에 대한 필터링 기능은 Fig. 9의 하단 슬라이더를 통해 가능하며, 게놈구조, 반복서열 및 ORF 영역을 표현할 수 있다. 마우스 클릭을 통해 특정 염기의 상세 정보를 확인 가능하다.

4. 실험

4.1 실험 데이터

실험 데이터는 수두 대상포진 바이러스(Varicella Zoster Virus, VZV)로 pOka strain을 사용하며, 최초의 완벽한 유전체로 보고된 Dumas strain을 레퍼런스로 한다[17]. pOka strain은 1974년 일본에서 수두에 걸린 Oka

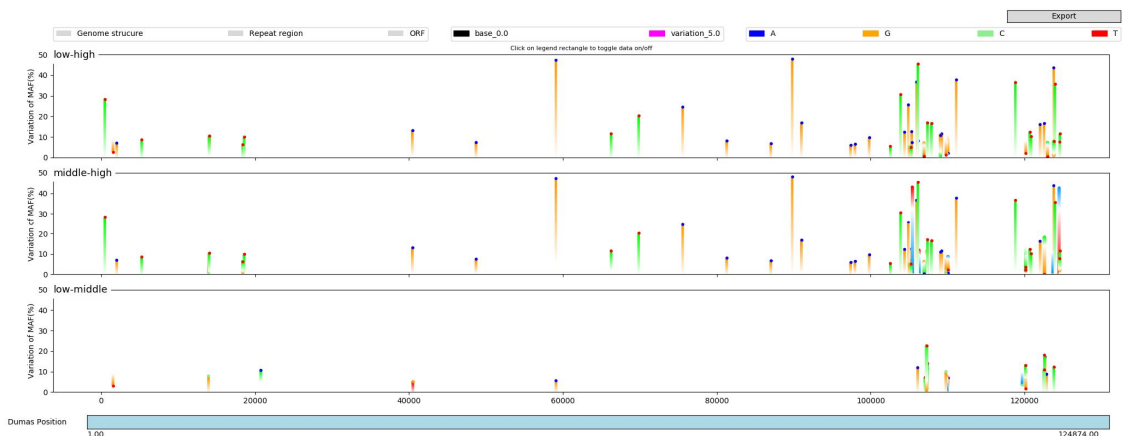


Fig. 5. Visualization snapshots for prominent variations in nucleotide sequences

Each graphs shows variations of MAF aligned with the position of sequences between individuals. X-axis means the position of sequences and Y-axis of each graph means variation of MAF.

라는 3세 남아로부터 분리한 야생주(pOka)로, 실험에는 계대배양 횟수에 따라 p3, p25, p60의 세 개의 염기서열 데이터를 사용한다. pOka strain은 바이러스 특성에 따라 계대배양을 거듭할수록 소수염기의 비율의 변화가 발생한다.

4.2 실험 결과

4.2.1 분석기

각 분석 결과는 엑셀파일로 제공되며 계놈구조, 반복서열, ORF, NCR 및 Major/Minor 염기조합에 따라 각 시트에 MAF 구간별로 작성된다. MAF 구간 5-15%의 계놈구조 및 반복서열별 분석된 염기서열의 결과는 Fig. 6과 같으며, 같은 구간의 Major/Minor 염기조합에 따른 결과는 Fig. 7와 같다. 유전 정보별 해당하는 염기서열의 총 길이와 MAF 구간에 해당되는 염기위치들을 찾아 집계한다. pOka 계놈구조 중 UL영역이 대부분을 차지하며, 배양이 진행됨에 따라 MAF 빈도가 증가함을 확인할 수 있다. 염기조합 결과에서는 주로 A/g와 T/c 조합이 주로 변이가 많이 발생하는 조합임을 확인할 수 있다.

4.2.2 시각화

pOka strain의 시각화 결과는 Fig. 5와 같으며, 계대가 배양됨에 따라 p3-p25, p25-p60, p3-p60사이의 MAF 변화량을 보여준다. 본 시각화에서는 분석기의 결과와 같이 A/g와 T/c가 높은 빈도로 증가하는 경향을 확인할 수 있다. 또한, Fig. 5에서 높은 빈도로 증가하는 경향을 보

이는 영역이 계놈구조 IRS에 해당하며, ORF는 62'에 해당함을 확인할 수 있다.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
28	Range of MAF : 5-15%														
29	Number of GPS							Average MAF at GPS							
30	Region	Dumas Length	N ₁ pO8			N ₂ pO8			Number of GPS / length * 100						
31			pO8 p3	pO8 p25	pO8 p60	average	pO8 p3	pO8 p25	pO8 p60	average	pO8 p3	pO8 p25	pO8 p60	average	
32	TRL	88	0	0	0	0	-	-	-	-	0	0	0	0	
33	UL	104836	5	13	24	14*	6.097%	6.503%	7.738%	7.16%	0.004776	0.012423	0.022901	0.013371	
34	Genome	IRL	88	0	0	0	0	0	0	0	0	0	0	0	
35	Structure	IRS	7320	15	17	16	16*	6.762%	6.934%	9.66%	8.497%	0.201559	0.228556	0.21442	0.214842
36	US	5232	0	0	0	0	-	-	-	-	0	0	0	0	
37	TRS	7320	13	17	18	14.333333*	6.619%	8.456%	9.67%	8.246%	0.176702	0.231459	0.174445	0.194444	
38	Total	124884	33	47	53	44.333333*	6.603%	8.088%	8.702%	8.0%	0.026434	0.037712	0.042379	0.035511	
39	R1	306	0	3	2	1.666667	-	6.53%	10.384%	8.072%	0	1.315789	0.877193	0.730994	
40	R2	326	4	5	5	4.666667	5.586%	7.177%	6.150%	6.35%	1.220994	1.53742	1.53742	1.42189	
41	R3	66	0	2	2	1.333333	-	6.522%	5.089%	5.805%	0	8.333333	6.666667	4.819277	
42	Repeat	R4a	146	9	6	0	5*	6.781%	6.788%	-	6.784%	3.202847	2.135231	0	1.79359
43	Region	R4b	146	10	9	0	6.333333	6.26%	8.206%	-	7.184%	3.558719	3.202847	0	2.213855
44	RS	200	0	1	0	0.333333	-	5.263%	-	5.263%	0	0.379398	0	0.112066	
45	ORI	92	2	0	2	1.333333*	7.571%	-	6.001%	7.086%	2.489136	0	2.179313	1.666667	
46	Total	1282	25	26	11	20.666667	6.441%	7.245%	6.912%	6.844%	1.625488	1.766264	0.709677	1.339649	
47	ORF	111284	9	25	43	25.666667	6.359%	8.555%	9.834%	8.602%	0.002001	0.022486	0.038682	0.022892	
48	NCR	12973	24	22	10	18.666667	6.694%	7.524%	7.538%	7.171%	0.183627	0.170754	0.075279	0.142719	

Fig. 6. Strain polymorphisms analysis results of Genome structure and Repeat region in MAF 5-15%

This is the strain polymorphisms analysis(Section 3.2.1) results in MAF 5-15%. It shows an increasing patterns in UL of genome structure and ORF over the passages.

	G	H	I	J	K
2	Range of MAF : 5-15%				
3	Major/Minor	pO8 p3	pO8 p25	pO8 p60	
4	A/G	7	16	14	
5	A/C	5	7	3	
6	A/T	0	0	0	
7	G/A	0	1	8	
8	G/C	1	3	0	
9	G/T	1	1	2	
10	C/A	2	0	2	
11	C/G	1	5	3	
12	C/T	0	1	7	
13	T/A	0	1	0	
14	T/G	10	2	1	
15	T/C	6	10	13	

Fig. 7. Strain polymorphisms analysis results of Major/Minor combinations in MAF 5-15%

This is the strain polymorphisms analysis(Section 3.2.1) results in MAF 5-15%. It shows an increasing patterns in A/g and T/c combinations.

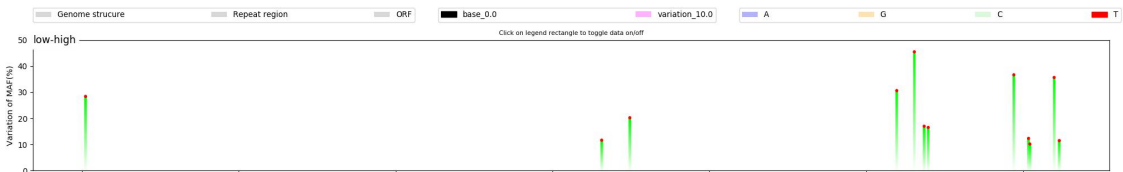


Fig. 8. Bases filtering on major type and MAF variations. This shows all major types are filtered except for the type of Major T and MAF changed more than 10%.

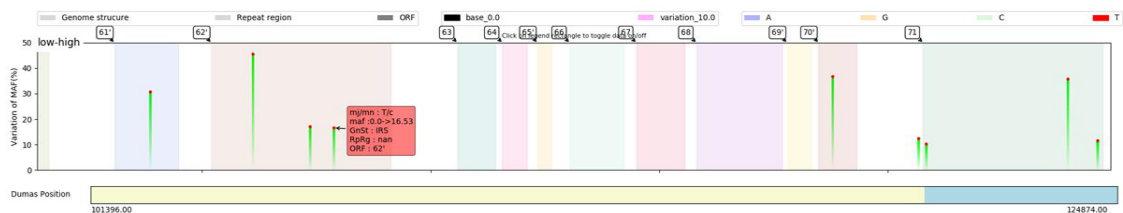


Fig. 9. Bases filtering on the position of sequences with genetic information expression. This shows sequences filtered by the position from 101396 to 124874 with ORF section and details of specific position.

5. 결론

본 논문에서는 염기서열 데이터의 분석 및 시각화 어플리케이션을 제안하고, VZV의 pOka strain 계대배양 염기서열 데이터를 통해 실험한 결과를 소개하였다. 소수염기 빈도의 변화를 파악하기에 용이한 분석과 시각화를 통해 분석의 효율성을 높일 수 있었고, 실험결과 pOka strain의 계대가 지남에 따라 A/g, T/c조합의 MAF가 증가하는 경향과 ORF 62' 영역에서 변이가 흔히 발생된다는 점을 확인할 수 있었다. 향후 연구에서는 각 염기위치간의 MAF 변화의 패턴 또는 연관성을 파악할 수 있는 분석 및 시각화가 이루어진다면 보다 높은 분석효과를 기대할 수 있을 것이다[18].

REFERENCES

- [1] S. C. Schuster. (2007). Next-generation sequencing transforms today's biology. *Nature Methods*, 5(1), 16 - 18.
DOI : 10.1038/nmeth1156
- [2] E. Pennisi. (2012). ENCODE Project Writes Eulogy for Junk DNA. *Science*, 337(6099), 1159 - 1161
DOI : 10.1126/science.337.6099.1159
- [3] A. Park. (2012. Sep). Junk DNA - Not So Useless After All. *TIME*
- [4] D. Vendramini. (2004). Noncoding DNA and Teem Theory of Inheritance, Emotions and Innate Behavior. *Medical Hypotheses* 64(3), 512-519.
DOI : 10.1016/j.mehy.2004.08.022
- [5] The International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928-933.
- [6] The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*. 437(7063), 1299 - 1320.
DOI : 10.1038/nature04226
- [7] Q. Chen & F. Sun. A unified approach for allele frequency estimation, SNP detection and association studies based on pooled sequencing data using EM algorithms. *BMC Genomics*, 14(1).
DOI : 10.1186/1471-2164-14-S1-S1
- [8] L. B. Barreiro, G. Laval, H. Quach, E. Patin & L. Quintana-Murci. (2008). Natural selection has driven population differentiation in modern humans. *Nature Genetics*. 40(3), 340 - 345.
DOI : 10.1038/ng.78
- [9] O. Harismendy et al. (2010). Population sequencing of two endocannabinoid metabolic genes identifies rare and common regulatory variants associated with extreme obesity and metabolite level. *Genome Biology*, 11(11), R118.
DOI : 10.1186/gb-2010-11-11-r118
- [10] T. A. Manolio. (2010). Genomewide association studies and assessment of the risk of disease. *The New England Journal of Medicine*. 363(2), 166 - 76.
DOI : 10.1056/NEJMr0905980
- [11] S. Raychaudhuri et al. (2009). Identifying Relationships among Genomic Disease Regions: Predicting Genes at Pathogenic SNP Associations and Rare Deletions. *PLoS Genetics*, 5(6), e1000534.
DOI : 10.1371/journal.pgen.1000534
- [12] B. Bandyopadhyay, V. Chanda & Y. Wang. (2017). Finding the Sources of Missing Heritability within Rare Variants Through Simulation. *Bioinform Biol Insights*.
DOI : 10.1177/1177932217735096
- [13] D. B. Goldstein, A. Allen, J. Keebler, E. H. Margulies, S. Petrou, S. Pertrovski & S. Sunyaev. (2013). Sequencing studies in human genetics: design and interpretation. *Nature Reviews Genetics*, 14(7), 460.
DOI : 10.1038/nrg3455
- [14] C. S. Carlson, M. A. Eberle, L. Kruglyak & D. A. Nickerson. (2004). Mapping complex disease loci in whole-genome association studies. *Nature*, 429(6990), 446-452.
DOI : 10.1038/nature02623
- [15] B. Carlson. (2008. June). SNPs - A Shortcut to Personalized Medicine. *Genetic Engineering & Biotechnology News*, 28(12).
- [16] R. Chenna et al. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, 31(13), 3497 - 500.
DOI : :10.1093/nar/gkg500
- [17] T. A. Manolio. (2010). How to interpret a genome-wide association study. *JAMA*. 363(2), 166-76.
DOI : 10.1001/jama.299.11.1335
- [18] A. J. Davison & J. E. Scott. (1986). The complete DNA sequence of varicella-zoster virus. *Journal of General Virology*, 67(Pt9), 1759-1816.
DOI : 10.1099/0022-1317-67-9-1759

이 일 섭(Lee, Il Seop)

[학생회원]



- 2014년 3월 ~ 현재 : 충북대학교
소프트웨어학과
- 관심분야 : 기계학습, 데이터 분
석, 바이오인포매틱스
- E-Mail : lis123kr@chnu.ac.kr

이 건 명(Lee, Keon Myung)

[정회원]



- 1990년 2월 : KAIST 전산학과 학사
- 1992년 2월 : KAIST 전산학과 석사
- 1995년 8월 : KAIST 전산학과 박사
- 1998년 8월 ~ 현재 : 충북대학교
소프트웨어학과 교수
- 관심분야 : 인공지능, 기계학습,
빅데이터, 바이오인포매틱스
- E-Mail : kmlee@cbnu.ac.kr