

2012 제 7 회 통계유전학워크숍 강좌 소개

- 개최 일시: 2012년 8월 6(월) - 11일(토)
- 개최 장소: 숭실대학교 진리관
- 주 최: 한국유전체학회 & 숭실대학교 의생명시스템학부 및 기초과학융합연구소

■ 프로그램

일 정	강 좌 명	주 강 사	소 속
8 월 6(월)- 7(화)	1. R 을 이용한 기초통계학 (Biostatistics using R)	원성호	중앙대학교
	2. 유전학 기초 (Introduction to Genetics)	오범석	경희대학교
	3. 암유전체학 (The Cancer Genome Atlas)	Peter Park	Harvard Medical School
	4. 전장유전체분석 (GWAS: SNP & CNV)	정연준 이종영	가톨릭대학교 국립보건연구원
8 월 8(수)- 9(목)	5. 유전체 역학 (Genome Epidemiology)	박지완	한림대학교
	6. 차세대 염기서열 데이터 분석 (NGS Data Analysis)	이환석	(주)마크로젠
	7. 네트워크 및 기능분석 (Network Analysis & In silico Functional Analysis)	황태현 김동섭	Minnesota Univ. KAIST
	8. 후성유전체 분석 기법 및 기능분석 (Epigenomics)	엄수중	세종대학교
8 월 10(금)- 11(토)	9. 초중급 생명정보학 (Bioinformatics for Beginners and Intermediate Users)	한윤수 손태권	중앙대학교 숭실대학교
	10. NGS 집단유전체 자료의 분석 (Analysis of NGS Population Genomic Data)	김희발	서울대학교
	11. 질병유전체 분석법 (Genetic Variation and Diseases)	이종극	울산대학교
	12. 전사체 분석 (Transcriptome Analysis)	추인선	한국생명공학연구원

■ 강좌 소개

8월 6(월) - 7(화)

1. R을 이용한 기초통계학 (Biostatistics using R)

강사: 김진흠, 원성호

소속: 수원대학교 통계정보학과, 중앙대학교 응용통계학과

◆ 강의 개요

본 강좌에서는 기초통계학의 기본 개념을 학습과 상용 소프트웨어인 R을 이용한 실습을 병행하고자 한다. 본 강좌에서 사용하는 R 패키지는 SAS, SPSS, MINITAB, STATA 등과 달리 무료로 제공되고 저장 공간이 많이 필요하지 않아, 현재 가장 널리 사용되고 있는 소프트웨어 중의 하나이다. 또한 다른 패키지와 달리 개발이 용이하여 최근 연구 결과들이 패키지에 탑재되는 주기가 짧아 통계 관련 연구에 매우 유용하게 사용되고 있다. 강좌에서 다룰 주요 주제는 R의 설치 및 기본 사용법, 일변량 자료를 정리하고 요약하는 방법, 상관분석 및 회귀분석, 범주형 자료의 독립성 검정을 포함하는 이변량 자료의 분석 방법, 확률 및 확률분포, 표본분포 이론, 모수 추정과 통계적 가설검정, 두 집단의 비교를 위한 통계적 방법, 분산분석 등이다.

- * 교육생 준비물: 개인용 노트북
- * 참여 조교: 임성민 (수원대학교 통계정보학과 석사과정)
박수연 (중앙대학교 응용통계학과 석사과정)
- * 교육일정: 추후공지 (강의/실습 병행, 인터넷 접속 필요)

2. 유전학 기초 (Introduction to Genetics)

강사: 오범석, 임지은

소속: 경희대학교

◆ 강의 개요

유전학 비전공자를 위하여 분자유전학, 전달유전학, 집단유전학, 질병유전학, 통계유전학 등의 기본 개념을 이해시키고 관련된 데이터베이스를 소개한다. 유전자의 구조 및 발현, Human Genome Project 및 유전변이, 멘델의 유전법칙, 가계도, 유전질환의 기전 및 Genetic counseling, 복합질환의 유전학적 이해 등의 주제를 다룬다. 이들 주제를 통하여, 유전학적인 기초 지식을 가르치고 분석 결과에 대한 생물학적인 이해를 높여 적절한 실험디자인을 하도록 도와주고자 한다. 그 외에도 최신 지견으로 Personalized Medicine, Personal Genome Sequencing, Epigenetics 등에 대한 간단한 설명도 포함된다.

* 교육생 준비물: 분석실습을 위해 개인용 노트북 지참 (Day 2)

* 교육일정: 추후공지 (강의/실습 병행, 인터넷 접속 필요)

3. 암유전체학 (The Cancer Genome Atlas)

강사: Peter J. Park, Hye-Jung Elizabeth Chun

소속: Center for Biomedical Informatics,

Harvard Medical School, USA; Canada's Michael Smith Genome Science Center, British Columbia
Cancer Agency, Canada

◆ 강의 개요

The Cancer Genome Atlas (TCGA) is a flagship project of the National Institutes of Health (NIH) in the US. The aim of this \$275 million dollar project is to apply the latest genomic technologies for comprehensive characterization of cancer genomes, with the data to be freely available to all. The pilot phase of the project consisted of several data types (e.g., gene expression, DNA copy number, DNA methylation, microRNA expression, targeted sequencing) for glioblastoma multiforme, ovarian, and lung cancers. In the current phase, 20 tumors are being characterized in greater detail, using high-throughput sequencing technologies. These data have become the standard datasets for the tumor types profiled and have resulted in dozens of follow-up papers, both in methodological and experimental aspects. In this workshop, we will teach the participants how to take advantage of the TCGA data. We will describe the various data types, how to access raw data, how to access the results of interim analyses that are being made available to the community, a set of core bioinformatics tools in the standard analysis pipeline, caveats on data quality, and current plans for data production and publications. Hands-on sessions will allow the users to explore these resources. Depending on the interests of the participants, additional discussions on specific analyses (e.g., copy number, RNA-seq or whole-genome analysis) may take place. There will also be a brainstorming session on potential analyses using data, with ideas solicited from the participants.

Requirements: The lectures will be given in English; some discussions may be in Korean. Participants are expected to be familiar with next-generation sequencing technologies and have read at least one of the main consortium papers (*Nature* 455:1061, 2008; *Nature* 474:609, 2011)

About the lecturers: Dr. Park has been an investigator on the TCGA project since its inception. He is currently a co-investigator of the Genomic Data Analysis Center at the Broad Institute/Harvard Medical School, which serves as the main analysis hub; he also leads the analysis efforts for whole-genome sequencing data being produced at the Harvard Genome Characterization Center. Ms. Elizabeth Chun is a lead analyst at the British Columbia Cancer Agency, which produces RNA-seq and miRNA-seq data for the project.

* 교육생 준비물: 개인용 노트북

* 교육일정: 추후공지 (인터넷 접속 필요)

4. 전장유전체분석 (GWAS: SNP & CNV)

강사: 정연준, 이종영, 고민진, 김영진, 허혜진,
정승현, 김지홍

소속: 가톨릭 의과대학교 인간유전체다형성연구소, 국립보건원 유전체센터

◆ 강의 개요

최근 질환 또는 특정형질의 유전적요인 발굴을 위한 주요 방법으로 대두되고 있는 전장유전체연관분석을 실질적으로 수행해 나가는데 필요한 각 단계의 방법을 강의 및 실습을 통하여 습득할 수 있도록 한다. SNP 와 아울러 대표적 유전다형성 지표로 알려진 인간 유전체 복제수 변이 (copy number variation, CNV)는 개개인의 여러 가지 생물학적 특성을 결정짓는데 기여하는 것으로 알려져 있는데 한 예로 CNV 내부에 질병관련 유전자들이 포함되는 경우 다양한 질병에 대한 감수성에 있어 개인차를 유발하는 데 기여한다. 본 강좌에서는 CNV 기반 임상연구 설계 의 기본, 다양한 칩 플랫폼을 이용한 whole-genome copy number 분석, CNVRuler 를 이용하여 실제 임상데이터를 접목한 CNV-GWAS 분석 실무 및 genetic distance 계산을 실습 위주로 학습한다.

* 교육생 준비물: 개인용 노트북

* 교육일정: 추후공지 (강의/실습 병행, 인터넷 접속 필요)

8월 8(수) - 9(목)

5. 유전체 역학 (Genome Epidemiology)

강사: 박지완, 지선하, 서영주

소속: 한림대학교 의과대학 의학유전학교실, 연세대학교 보건대학원 역학건강증진학과,
인하대학교 의과대학 생물통계학연구실

◆ 강의 개요

질병의 유전성 (Familial Aggregation & Correlation), 유전방식 (Segregation analyses), 원인/감수성유전자의 위치추정 (Linkage/Association studies) 등 유전역학의 단계적 연구방법론을 소개하고, 원인/감수성유전자의 인구집단 내 분포특성 (Allele/Genotype frequency, HWE, LD), 질병위험도 (Attributable risk, Odds Ratio, Relative Risk) 측정, 비유전적 요인과 유전-환경상호작용의 영향을 측정하기 위한 통계분석법을 학습한다. 질병위험요인을 밝히기 위한 study design (가족, 환자-대조군, 코호트 연구 등) 방법을 소개하고, 연구유형 (후보유전자, 전장유전체, 유전-환경 상호작용)에 따른 표본수 산정과 통계분석, 원인/연관성 추론방법을 습득한다. 실제 연구사례를 들어 유전체연관성 분석에 보편적으로 사용되는 software (GPC, R, SAS, FBAT, PLINK, GMDR, SDR)을 이용하여 각 주제에 맞는 임상-역학-유전자 자료를 분석하고 결론을 추론하는 실습을 한다. 메타분석, 질병의 위험예측 모델링, NGS 분석을 통한 질병유전체연구, 네트워크 분석, 기능분석 등 최신 유전체연구 분야를 소개하고 문제점 및 해결방법 (혼란변수 통제, 다중가설검정, Quality control 등)을 검토한다.

* 교육생 준비물: 분석실습을 위한 개인용 노트북지참 (Day 2)

* 조교: 홍은표 (한림대학교 의과대학 의학유전학교실)

박만준 (인하대학교 의과대학 통계학과)

* 교육일정: 추후공지 (강의/실습 병행, 인터넷 접속 필요)

6. 차세대 염기서열 데이터 분석 (NGS Data Analysis)

강사: 이환석, 김종일, 김창훈, 박정훈, 주영석, 김지훈

소속: (주)마크로젠

◆ 강의 개요

본 강좌에서는 차세대 염기서열 분석기술로 생성된 유전체 정보에 대한 분석 방법을 소개하고자 함. 최신의 염기서열 분석 장비에 대한 소개 와 Exome Sequencing 등 이를 활용한 다양한 분석 기술 및 이를 활용한 질병 연구 등 최신 연구동향을 소개함. 또한 NGS 데이터 분석 전반에 걸쳐 필요한 데이터 포맷, 소프트웨어 등에 대해 소개하고 생물학적 해석을 위해 필요한 생명지식정보들에 대한 지견을 습득하는 것을 목표로 함.

- * 교육생 준비물: 없음
- * 참여 조교: 김세환, 김지웅, 진행운, 류재면 (마크로젠)
- * 교육일정: 추후공지 (실습 없음, 인터넷접속 필요)

7. 네트워크 및 기능분석

(Network Analysis & In Silico Functional Analysis)

강사: 김동섭, 황태현

소속: 카이스트 바이오 및 뇌공학과

University of Texas Southwestern Medical Center, Dept. Clinical Science

◆ 강의 개요

질병을 이해하고 진단 및 치료법을 개발하기 위한 효과적인 방법으로서, 질병과 관련된 여러 요소들 (diseases, disease genes, mutations, drugs, drug targets, 등등) 사이의 상호작용을 통합하여 네트워크로 표현한 후 이를 시스템 수준에서 분석하는 방법이 발전되고 있다. 본 강의에서는 앞에서 열거한 여러 요소들 사이의 상호작용에 대해 살펴보고, 이를 통합하여 어떻게 통합질병네트워크 구축하는 지에 대해 우선 논의한다. 그 다음 여러 네트워크 분석방법을 소개한 후, 통합질병네트워크를 효과적으로 분석하는 여러 기법을 공부한다. 이를 통해 질병관련 유전자, 질병유발 돌연변이, 질병관련 단백질의 기능변화, 질병네트워크 분석을 통한 신약개발 등에 대해 자세히 논의한다.

* 교육생 준비물: 추후공지

* 교육일정: 추후공지

8. 후성유전체 분석 기법 및 기능분석

(Epigenomics)

강사: 엄수중, 권영수, 박의현

소속: 세종대학교 생명공학과 유전자제어연구실, 기능유전체학 연구실

◆ 강의 개요

후성유전학의 개념을 설명하고 DNA methylation, histone modification, ncRNA 와 관련된 통계적 분석방법을 강의한다. 후성유전학 (Epigenetics)은 DNA 염기서열의 변화가 아닌 주변 환경에 의한 유전자의 발현조절을 연구하는 분야로서 genome-wide 수준에서 연구하는 것을 후성유전체학 (Epigenomics)이라 한다. 대표적인 후성유전을 결정하는 대표적인 인자로 DNA methylation, histone modification, non-coding RNA 등이 있다. DNA methylation 은 1965 년 처음으로 발견된 이후, DNA 상의 methylation 의 증감으로 유전자의 발현조절이 생기고 이에 따라 세포의 기능저해뿐만 아니라 암 및 기타 질병 등 다양한 형태의 이상 현상들이 형성될 수 있음이 밝혀지고 있다. Histone modification 은 1996 년 histone acetyltransferase (HAT)와 histone deacetylase (HDAC)의 발견된 이후 methylation, phosphorylation, ubiquitination 의 modification 이 발견되었다. 이들 histone 상에 다양한 변형을 통해 유전자의 발현이 정교히 조절되며, 최근 이들 연구로 논문, 약물개발, 특허 등이 급속한 추세로 발표되고 있다. 이 두 가지 후성유전학적인 변형은 non-coding RNA (ncRNA)에 의해서도 영향을 받는다. 이들 ncRNA 중에서 비교적 최근에 발견된 miRNA 의 경우도 다양한 질병과 연관되어 있음이 발표되고 있다. 그러므로, DNA methylation, histone modification, ncRNA 의 패턴을 genome-wide 수준에서 밝히는 과정은 생명현상의 이해 및 질병발생 등을 연구하는 데 필수적인 과정으로 자리 잡고 있다. 현재 후성유전체 사업과 관련하여 HEP, NoE, ANE, AHEAD 과 같은 국제 공동연구소들이 설치되어 운영되고 있으며 미주, 유럽, 아시아의 선진국가들도 연구 및 투자를 증가시키는 추세이다. 최근 Roche, Agilent, Affimatrix, Illumina, 마크로젠, 지노믹트리와 같은 국내외 산업체에서도 기술의 발전을 주도 또는 효과적으로 활용하기 위해 노력 중이다. 이번 강의에서는 세계적인 발전추세에 따른 후성유전체학의 위치를 재확인하고, 후성유전체(Epigenome)의 분석을 위한 첨단 기법의 이론과 실제 사용법 (methylome 분석을 위한 NoME-Seq, oxBS-

Seq , BisChip-seq/ ChIP BS-Seq ; histone 변형 분석을 위한 ChIP-seq; ncRNA 의 기능 분석을 위한 RIP-Chip 등)에 대한 소개와 구체적 강의 및 실습을 진행할 예정이다.

* 교육생 준비물: 추후공지 * 교육일정: 추후공지

8 월 10(금) - 11(토)

9. 초중급 생명정보학

(Bioinformatics for Beginners and Intermediate Users)

강사: 한윤수 (초급), 손태권 (중급)

소속: 중앙대학교 생명과학과, 숭실대학교 의생명시스템학부

◆ 강의 개요

(초급) 분자생물학 전공자와 생물정보학 입문자를 위한 초급 과정의 생물정보학 강좌로 다음과 같은 필수 기초과정을 다룬다. NCBI 의 PubMed, Gene, Nucleotide, Protein 데이터베이스 등의 검색을 통한 문헌 및 서열 정보의 열람 및 획득; DNA 와 단백질 서열 정보를 이용한 BLAST 서열 검색; 단백질 서열의 기능적 도메인 분석(Pfam, SMART); 단백질 서열의 세포 내 위치 예측(SignalP, TargetP, PSORT II, TMHMM, NucPred, PredGPI); 유사 서열의 다중 정렬 및 비교 분석(ClustalX, njplot); EST 서열을 이용한 새로운 유전자 발굴(cap3); UCSC Genome Browser 를 활용한 유전체 정보 열람 및 분석. 생물정보의 검색, 열람, 획득, 가공 및 분석 방법의 기술과 분석 결과의 올바른 해석 및 활용 방법에 대하여 강의와 실습을 수행한다.

(중급) 초급과정의 생물정보학 강좌 이수자를 대상으로 초급 강좌에서 다루지 않은 생물정보 데이터베이스와 분석 tool 을 다룬다. 사용법과 적용 분야를 중심으로 강의와 실습을 병행하며 진행한다. 실험실에서 생산되는 다양한 생물정보 데이터를 유전체, 발현체, 단백질체의 세가지 분야로 나누어서 강의하며, 분야별 세부내용은 아래와 같다. 유전체 정보의 검색과 분석을 위해 UCSC Genome browser 의 중고급 활용법과 Galaxy site 의 개요, 특징, 사용법 및 타 database 와의 연동 활용 방법을 알아본다. Unigene, GEO 등 유전자 발현 정보를 다루는 데이터베이스의 구조와 사용법 및 활용 방법에 관한 강좌가 실습과 함께 진행된다. 그리고, STRING, HPRD 등 공개

프로그램을 이용한 단백질 상호작용 분석 방법을 강의한다.

* 교육생 준비물: 개인용

노트북

* 참여 조교: 김동선(중앙대학교 생명과학과),

김종범, 김현호, 이준남(숭실대학교 의생명시스템학부)

* 교육일정: 추후 공지 (강의/실습 병행, 인터넷접속 필요)

10. NGS 집단유전체 자료의 분석

(Analysis of NGS Population Genomic Data)

강사: 김희발, 조서애, 곽우리, 이태헌, 문선진, 김재민

소속: 서울대학교 동물생명공학과 생물정보 및 집단유전학실/ C&K 지노믹스

◆ 강의개요

최근 대량의 NGS 데이터의 생산은 기존에는 연구하기 어려웠던 집단유전학 및 진화 정보 분석을 통한 양성진화(positive selection) 혹은 선택흔적 (selective sweep) 영역의 발굴이 가능하게 되었다. 기본적으로 NGS 를 통해서 생산되는 집단유전체 자료는 개체별로는 low coverage (10 fold coverage 이하)의 특성을 가지게 되며 또한 다량의 개체에서 자료가 생성되게 된다. 현재 우리나라에서 진행되고 있는 NGS 유전체 자료의 분석은 염기서열의 어셈블과 유전변이 (SNP, indel, 구조변이)를 calling 하는 데에만 중점이 맞추어져 있는 듯하다. 앞으로 집단유전체 자료가 점점 더 많아지면서 집단유전학에 대한 이해가 필요해지고 또한 집단유전체 자료의 유의성을 검정하기 위한 자료의 시뮬레이션에 대한 이해가 필수적이라고 할 수 있다. 본 강좌에서는 이러한 요구에 필요로 되는 핵심적이고 최소의 집단유전학적 이론과 실습을 강의하고자 한다. 실습강의는 워낙 큰 자료를 다루어야 함으로 수강생 개별적으로 실습은 진행되지 않지만 작은 예를 가지고 강사가 보여주는 강의를 이루어 질 것이고 실습강의에서 이론적 부분도 일부 다루어 질 예정이다. 이론 강의는 NGS 집단 유전체 자료의 분석을 위해 필요로 되는 핵심적인 집단유전학 이론을 강의하고자 한다. 이러한 집단유전체 분석은 사람, 동물, 식물, 미생물/바이러스에 이르는 모든 생물종의 분석에 공통적으로 적용이 가능하다고 할 수 있다.

- * 교육생 준비물: 없음
 - * 교육일정: 추후 공지
- (강의/실습 병행, 인터넷
접속 필요 없음)

11. 질병유전체 분석법(Genetic Variation and Diseases)

강사: 이종극, 김재정

소속: 울산대학교

◆ 강의개요

인간유전체에 존재하는 유전변이형인 SNP(single nucleotide polymorphism)의 기본 개념을 이해하고 유전체 구조(LD; linkage disequilibrium) 및 유전변이형 분석방법(SNP genotyping & haplotyping)의 원리를 배움. 또한 질병 유전자 탐색을 위한 전략, 연관성 연구(association study), 전장 유전체 연관성 분석법 (GWAS), 차세대 염기서열 분석법을 이용한 질병 원인 유전자 발굴(예, exome sequencing)과 유전변이형 기능연구 방법론 (functional study) 등 실제 연구 수행을 위한 다양한 분석 기법을 이해하고 활용분야를 소개함. 아울러 질병 유전자에 대한 자연선택을 통한 진화론적 이해와 유전자 발현조절 기전 연구를 위한 새로운 유전 변이형 분석 기술들을 이해하도록 함. 또한 다양한 종류의 분석 프로그램들의 내용과 사용 방법을 간략한 데모와 더불어 실제로 교육생들이 실습할 수 있는 기회를 제공함. 컴퓨터 분석 데모 및 실습에서는 후보유전자 유전체정보 수집, dbSNP 이용법, PCR primer 선별법, SNP 판별법, HapAnalyzer, Haploview, Plink, PolyPhen, Match 및 기타 SNP 관련 database 이용방법 등을 다루게 됨.

* 참고 강의 교재:

[질병유전체분석법 2]

(월드사이언스, 2010 년

출간)의 내용을 중심으로

강의하고

강의발표용 및 실습용 자료는 강의교재로 인쇄하여

교육 당일에 배포할 예정입니다.

* 교육생 준비물: 개인용 노트북

* 참여 조교: 박영미, 박상용 (울산대학교)

* 교육일정: 추후 공지 (강의/실습 병행, 인터넷접속 필요)

12. 전사체 분석 (Transcriptome Analysis)

강사: 추인선, 김남신, 김선규

소속: 한국생명공학연구원(KRIBB)

◆ 강의개요

전사체(Transcriptome) 데이터 분석은 DNA, RNA 및 단백질 간의 전사(Transcription)와 전이(Translation) 과정에서 특정 조건의 유전자발현 패턴의 변화를 총체적으로 파악하기 위한 것이며, 최근 전체 유전자의 정량화나 발현분석을 목적으로 마이크로어레이(Microarray, DNA chip)와 차세대 시퀀싱(NGS) 기술을 주로 사용한다. 본 강좌에서는 마이크로어레이 유전자발현 프로파일 및 RNA-seq 데이터를 중심으로 유전자발현 패턴 분석, 클러스터링 기법, 유의한 차이를 보이는 유전자 추출, mRNA isoform 수준에서 유전자발현 분석, Fusion 유전자 분석의 기본 분석과 함께 대용량 전사체 데이터를 다룰 때 효율적인 통계패키지 R 을 이용한 다양한 분석법을 다룬다. 또한 유전자 발현패턴을 중심으로 Gene Ontology 와 Pathway 분석, 암연구에서 중요한 예후인자 및 예측인자 발굴, 국제적인 암유전체 프로젝트 데이터 활용을 포함하여 생물정보학적 통합분석 및 개념

브라우저를 이용한 시각화 분석의 내용을 강의와
실습으로 진행한다.

* 교육생 준비물: 개인용
노트북

* 교육일정: 추후 공지 (강의/실습 병행, 인터넷접속 필요)