

제14회 통계유전학워크샵

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea
July 15(Mon) ~ July 20(Sat), 2019



Program

Session 1(July 15 ~ July 16)

01	Application of NGS to Translational Genomics	Murim Choi(Seoul Natl. Univ.)
02	Biostatistics using R	Jinheum Kim(Univ. of Suwon) Ju Hyun Park(Dogguk Univ.)
03	Introduction to Genetics	Sun Shim Choi(Kangwon Natl. Univ.)
04	Epigenome Data Analysis	Inkyung Jung(KAIST)

Session 2(July 17 ~ July 18)

05	Large-scale Genome Data Analysis using KoreanChip	Bong-Jo Kim(Korea Natl. Institute of Health)
06	Analysis of Single Cell RNA-seq Data	Semin Lee(UNIST)
07	Machine Learning for Bio-data Mining	Kyu-Baek Hwang(Soongsil Univ.)
08	Human Microbiome Analysis	Han-Na Kim(Kangbuk Samsung Hosp.)

Session 3(July 19 ~ July 20)

09	Genomic Epidemiology	Ji Wan Park(Hallym Univ.)
10	Cancer Genomics	Sanghyuk Lee(Ewha Womans Univ.)
11	Omics-based Systems Biology	Daehee Hwang(Seoul Natl. Univ.)
12	Multi-omics Data Analysis	Je-Keun Rhee(Pusan Natl. Univ.) SungHwan Kim(Konkuk Univ.)

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 01. NGS를 통한 중개연구(Application of NGS to Translational Genomics)

강사: 최무림, 이영하, 유태경, 조재소, Jana Kneissl, 이정은, 이초룡, 이정하, 전은영
소속: 서울대학교 의과학과

강의개요

NGS를 기반으로 하는 유전체학의 발전에 의하여 질병 원인의 유전적 이해도가 가파르게 상승하고 있다. 이러한 경향에 힘입어 유전체학을 이용한 중개연구도 활발히 이루어지고 있으며 이를 통하여 정밀의학으로 대표되는 기초적, 임상적 발전이 이루어지고 있다. 본 강좌에서는 질병의 유전학적 이해를 통한 최근 동향을 알아본 후 중개연구를 위한 여러 종류의 질환 연구의 예시와 실습을 수행할 것이다. 또한 중개연구의 발전에 힘입은 정밀의학의 예시와 실습, 최근 연구 동향을 학습한다. 이를 위하여 NGS 데이터의 해석과 이를 위한 공공 DB의 사용법, 다양한 유전성 질환(희귀질환, 복합질환, 암을 비롯한 somatic 질환)의 연구 디자인, 분석법, 데이터 해석과 결과의 의생명적 의미 부여 과정을 다룰 것이다. 각 실습 시간은 주강사에 의한 약간의 이론적인 강의와 참여강사에 의한 실제 데이터를 이용한 실습으로 구성되어 있다.

- 준비물: 노트북 지참 실습: 인터넷 접속, linux, R 사용
- 수강생수준: 제한 없음

교육일정

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	질병의 유전학적 이해 및 정밀의학의 구현	최무림	강의
	11:00-12:30	NGS methodology	최무림	강의
	14:00-15:30	중개연구를 위한 public DB, data repository의 종류와 활용법	이영하, 전은영	강의/실습
	15:40-17:10	Whole exome sequencing 분석과 Mendelian 질병 원인 규명법	이정하, 조재소	강의/실습
DAY 2	09:20-09:40	실습을 위한 컴퓨팅 환경 점검	모두	실습
	09:40-11:10	RNA sequencing 데이터 분석	유태경, Jana K.	강의/실습
	11:20-12:30	Single cell RNA sequencing 데이터 분석	이정하/이초룡	강의/실습
	14:00-15:30	Somatic disease의 유전체적 접근과 이해	전은영/조재소	강의/실습
	15:40-17:00	Whole genome sequencing 데이터 분석	최무림	강의/실습

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 02. R을 이용한 기초통계학(Biostatistics using R)

강사: 박주현, 김진흠
소속: 동국대학교 통계학과, 수원대학교 데이터과학부

강의개요

본 강좌에서는 기초통계학의 기본 개념을 학습하고 상용 소프트웨어인 R을 이용한 실습을 병행하고자 한다. 본 강좌에서 사용하는 R 패키지는 SAS, SPSS, MINITAB, STATA 등과 달리 무료로 제공되며 저장 공간이 많이 필요하지 않기 때문에 현재 가장 널리 사용되고 있는 소프트웨어이다. 또한 R 패키지는 다른 통계패키지와 달리 개발이 용이하고 최근 연구 결과들이 패키지에 탑재될 때까지 소요되는 시간이 짧아서 통계 관련 연구에 매우 유용하게 사용되고 있다. 본 강좌에서는 R 설치와 기본 사용법, R을 활용한 그래프, 일변량 자료의 정리와 요약 방법 등 R을 활용한 자료 처리 방법과 추정과 가설검정, 두 집단의 비교 및 분산분석, 상관분석 및 회귀분석, 범주형 자료의 독립성검정을 포함하는 이변량 자료의 분석 방법 등 기초 통계 이론에 대하여 강의할 것이다.

교육일정

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	R 설치 및 소개	박주현	인터넷접속필요
	11:00-12:30	R을 활용한 그래프	박주현	인터넷접속필요
	14:00-15:30	일변량 자료의 정리 및 요약	박주현	인터넷접속필요
	15:40-17:10	확률과 확률분포, 표본분포	박주현	인터넷접속필요
DAY 2	09:20-10:50	추정	김진흠	인터넷접속필요
	11:00-12:30	가설검정	김진흠	인터넷접속필요
	14:00-15:30	두 집단의 비교 및 분산분석	김진흠	인터넷접속필요
	15:40-17:10	이변량 자료의 분석: 상관분석, 회귀분석, 독립성검정	김진흠	인터넷접속필요

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 03. 기초유전학(Introduction to Genetics)

강사: 최선심
소속: 강원대학교

강의개요

본 강의에서는 통계유전학 연구를 수행하고자 하는 유전학 비전공자를 위하여, 멘델 유전학 기초 및 확장, 유전자 연관 및 유전자 매핑의 개념을 이해시키고, 유전자와 유전자의 변형된 형태들 (alleles)과 형질의 상호관계를 이해시키는데 초점을 두고자 한다. 이에, 수강생들은 기본적인 유전학에 대한 이해를 기반으로, 멘델법칙에 의해 설명되는 단순하지만 매우 치명적인 유전병의 원인이 되는 변이가 가계도와 유전자연관 (linkage)를 통하여 어떻게 찾아지는 지 이해한다. 뿐만 아니라, 매우 복잡한 유전변이와 환경적 영향까지 결부되어 있는 복잡한 만성유전병을 정의하고, 집단 내 개별자들이 가진 유전체 내에 존재하는 수많은 변이들 중에 어떻게 질병의 원인이 되는 변이들을 찾게 되는지 기초적인 집단유전학의 개념을 이해한다. 나아가, 유전학적 연구 방법론에 기초하여 연구를 설계하고 연구 결과를 해석할 수 있는 기초를 다진다.

- 준비물: 강의교재 외 특별한 준비물 필요없음
- 실습: 없음
- 수강생수준: 기초 유전학 지식이 필요한 사람으로, 일반생물학, 기초분자생물학 지식이 있는 사람

교육일정

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	기초유전학 overview	최선심	강의
	11:00-12:30	Cell division: mitosis/meiosis	최선심	강의
	14:00-15:30	Mendelian inheritance	최선심	강의
	15:40-17:10	Expansion of Mendelian inheritance	최선심	강의
DAY 2	09:20-10:50	Genetic linkage and mapping	최선심	강의
	11:00-12:30	Heritable diseases	최선심	강의
	14:00-15:30	Gene vs. DNA vs 발현	최선심	강의
	15:40-17:10	Genetic variation/GWAS	최선심	강의

제14회 통계유전학워크샵

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 04. ChIP-seq 및 염색질 3차구조 데이터 분석(Epigenome Data Analysis)

강사: 정인경, 박대찬
소속: 카이스트 생명과학과, 아주대학교 생명과학과

강의개요

본 강좌에서는 후성유전학의 기본 개념 소개와 함께 최근 후성유전학 연구에서 중요하게 활용되고 있는 ChIP-seq 및 염색질 3차구조에 관한 관련 이론, 실험 방법, 그리고 기본 데이터 분석을 실습과 함께 숙지하고자 한다. 이에 본 강좌의 첫날은 후성유전학 및 ChIP-seq 실험, 그리고 기초 데이터 분석 이론을 소개한 후 이어서 ChIP-seq 데이터 분석 기초/고급 실습을 통해 MACS peak calling, peak-to-gene assignment by bedtools, UCSC genome browser visualization with bigwig, peak overlap analysis, sample clustering with protein-binding 및 TF motif analysis 분석법을 숙지하려 한다. 둘째 날은 최근 유전자 조절의 신규 기전으로 제시되고 있는 염색질 3차구조에 대한 전반적인 소개와 관련 데이터 분석 방법을 소개한 후 2018년도에 개발된 3DIV를 활용하여 웹기반 염색질 3차구조 데이터 분석과 R을 활용한 Hi-C 데이터 기본 데이터 분석 방법을 익히려 한다.

- 준비물: 노트북 지참
- 실습: 서버 접속

교육일정

참여교과: 김규광(카이스트 생명과학과)

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	후성유전체 및 ChIP-seq 개요	정인경	강의
	11:00-12:30	ChIP-seq 데이터 분석 개요	정인경	강의
	14:00-15:30	ChIP-seq 분석 실습 (기초)	박대찬	강의/실습
	15:40-17:10	ChIP-seq 분석 실습 (고급)	박대찬	강의/실습
DAY 2	09:20-10:50	염색질 3차구조 개요	정인경	강의
	11:00-12:30	염색질 3차구조 데이터 분석 방법	정인경	강의
	14:00-15:30	3DIV 기반 Hi-C 데이터 분석	정인경	강의/실습
	15:40-17:10	R을 활용한 Hi-C 데이터 분석 실습	정인경	강의/실습

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 05. 한국인칩을 이용한 대규모 유전체정보 분석(Large-scale Genome Data Analysis using KoreanChip)

강사: 김봉조, 김영진, 문상훈, 황미영, 한소희, 공진화
 소속: 국립보건연구원

강의개요

본 강좌에서는 국립보건연구원에서 제작한 Korea Biobank Array (이하 한국인칩)를 소개하고, 이를 이용한 유전체정보 분석 방법을 설명한다. 한국인칩은 한국인 질병유전체 연구를 위해 한국인 전유전체분석 정보 및 마이크로어레이 정보를 기반으로 제작된 SNP칩으로써 한국인 유전체 대표성 확보 (genomic coverage)를 위해 tagging SNP 및 기능 유전 변이 등 약 83만개의 프로브로 구성되어있다. 본 강좌에서는 한국인칩사업 소개, 한국인칩 콘텐츠 특징 및 성능분석 결과, 수십만 명 규모의 유전체정보 분석 등 한국인칩을 이용한 전장유전체 연관성 분석의 모든 과정을 습득할 수 있도록 실습을 병행하여 진행한다.

- 준비물: 노트북 지참
- 실습: 인터넷 접속

교육일정

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	한국인칩사업 소개 (현황 및 분양 방법 등)	김봉조	강의
	11:00-12:30	한국인칩 콘텐츠 특징 및 성능 소개	문상훈	강의
	14:00-15:30	한국인칩 유전체정보 정도관리 (1)	김영진	강의
	15:40-17:10	한국인칩 유전체정보 정도관리 (2)	황미영	실습
DAY 2	09:20-10:50	한국인칩 유전체정보 분석 (연관성, 메타, Conditional, GRS 분석 등)	김영진	강의
	11:00-12:30	한국인칩 유전체정보 분석 (연관성, 메타, Conditional, GRS 분석 등)	한소희	실습
	14:00-15:30	한국인칩 유전체정보 분석결과 해석 (1)	문상훈	강의
	15:40-17:10	한국인칩 유전체정보 분석결과 해석 (2)	공진화	실습

제14회 통계유전학워크샵

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 06. 단세포 전사체 데이터 분석(Analysis of Single Cell RNA-seq Data)

강사: 이세민

소속: 울산과학기술원 생명과학부

강의개요

본 강좌에서는 최근 각광 받고 있는 단세포 전사체 데이터 분석 기술에 대한 소개와 실제 데이터에 대한 분석 실습을 병행하고자 한다. 단세포 전사체 분석 기술은 세포의 분화, 암의 진화, 면역세포 프로파일링 및 종양 내 이질성 분석 등에 활용되고 있으며, 관련 기술과 응용 사례에 대한 소개 및 현재 가장 널리 사용되고 있는 10x Genomics사의 Chromium Single Cell Gene Expression Solution을 사용하여 생산된 단세포 전사체 데이터를 위주로 다양한 분석 방법에 대한 실습을 진행하고자 한다.

- 준비물: 노트북 지참
- 실습: 인터넷 접속, linux, R 사용
- 수강생수준: Linux 기본 명령문 사용 가능자 또는 생물정보학 입문자

교육일정

참여조교: 정형오(울산과학기술원 생명과학부)

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	Introduction	이세민	강의
	11:00-12:30	Data processing	이세민	강의
	14:00-15:30	Data analysis	이세민	강의
	15:40-17:10	Data visualization	이세민	강의
DAY 2	09:20-10:50	Processing scRNA-seq data	이세민, 정형오	실습
	11:00-12:30	Quality check	이세민, 정형오	실습
	14:00-15:30	Biological Analysis	이세민, 정형오	실습
	15:40-17:10	Clustering visualization	이세민, 정형오	실습

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 07. 바이오데이터마이닝을 위한 기계학습(Machine Learning for Bio-data Mining)

강사: 황규백
소속: 송실대학교 컴퓨터학부

강의개요

본 과정은 대용량 데이터에 숨겨져 있는 패턴(규칙)을 추출하는 데 유용한 기계학습의 이론 및 구체적인 방법과 이를 실제로 적용하기 위한 R 프로그래밍에 대한 것이다. 기계학습 기술은 크게 예측모델을 구성하기 위한 지도기계학습(supervised learning), 설명모델을 구성하기 위한 비지도기계학습(unsupervised learning)으로 구분하며, 이와 관련된 용어와 개념, 원리와 알고리즘을 다룰 것이다. 지도기계학습 방법으로 선형회귀(linear regression)를 비롯하여 logistic regression, linear discriminant analysis, decision tree를 집중적으로 강의한다. 또한, 바이오데이터마이닝에서 흔히 접하게 되는 희박 데이터(sparse data) 문제의 완화를 위한 방법으로 ridge 및 lasso regularization과 bagging, random forest, boosting 등의 방법을 다룬다. 강의에서 다룬 기계학습 방법을 실제 문제에 적용하는 것은 R 프로그래밍 환경에서 실습한다

- 준비물: 노트북 지참(실습 시 데이터 다운로드를 위한 인터넷 접속 필요)
- 실습: R 및 RStudio
- 수강생수준: 간단한 프로그래밍 및 확률/통계에 대한 기초 지식이 있는 자

교육일정

참여조교: 이다빈(송실대학교 컴퓨터학부)

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	강의 소개 Bias-Variance Trade-off 선형회귀	황규백	강의
	11:00-12:30	로지스틱 회귀 Linear Discriminant Analysis	황규백	강의
	14:00-15:30	분류 성능 평가 기법 R 설치 및 기초 실습	황규백	강의/실습
	15:40-17:10	선형회귀 및 분류 실습	황규백	실습
DAY 2	09:20-10:50	교차검증 및 Bootstrap 변수 선택 Ridge Regression	황규백	강의
	11:00-12:30	Lasso 및 PCR Decision Tree	황규백	강의
	14:00-15:30	Bagging, Random Forest, Boosting Shrinkage 실습	황규백	강의/실습
	15:40-17:10	Decision Tree 실습	황규백	실습

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 08. 인체 마이크로바이옴 분석(Human Microbiome Analysis)

강사: 김한나

소속: 강북삼성병원 연구지원실, 성균관대학교 삼성융합의과학원 임상연구설계평가학과

강의개요

최근 위장관, 피부, 구강 등의 인체 미생물과 인간의 건강 및 질병간의 연관성이 밝혀지면서, 미생물을 질병 예방이나 진단 및 치료 등의 맞춤의료에 활용하고자 메타지놈연구가 활발히 진행되고 있다. 대용량 시퀀싱 기술의 발달은 기존 배양 기반의 미생물연구에 혁신적인 발전을 가져왔으며, 이렇게 얻어진 메타지놈 자료는 인간 유전체자료 분석과 마찬가지로 생물정보학기술이 바탕이 된다. 그러나 인간 유전체자료와 달리, 한가지 종이 아닌 다양한 미생물 군집 데이터를 분석하고 해석 해야하므로, 이를 위해 반드시 알아야 할 기초 지식과 데이터 분석, 그리고 결과의 해석 능력이 필요하다. 본 강좌에서는 메타지놈을 처음 접하는 초보자에게, 관련 논문을 읽고 이해할 수 있을 정도의 기초이론에 대한 상세한 개념 설명 뿐만 아니라, 16S rRNA gene의 amplicon sequencing 에 기반-집중하여, 실제 연구디자인 및 시퀀싱 raw 데이터부터 미생물 군집분석까지 QIIME2 프로그램으로 직접 실습으로 진행해보므로써 실제 연구에 바로 활용 가능하도록 한다. 본 강의를 통하여 수강생들은 다음과 같은 기술을 익힐 수 있다.

- 메타지놈 연구 디자인
- Marker gene (16S rRNA gene amplicon) 시퀀스 데이터의 퀄리티 해석
- 16S rRNA gene 시퀀스 raw 데이터(fastq)로부터 미생물의 feature table 생성
- 미생물의 다양성 (α -diversity, β -diversity) 및 taxonomic profile 분석 및 결과 해석
- QIIME2 외 다양한 프로그램 활용을 위한 메타지놈 데이터의 importing/exporting

- **준비물:** 노트북 지참(노트북 메모리 RAM 4GB 이상, Window 7/ 64bit 이상 필요)
- **실습:** 인터넷 접속, Linux 사용, Window/Linux/Mac OS 사용가능
- **수강생수준:** 마이크로바이옴 분석을 처음 접하는 초보자

제14회 통계유전학워크샵

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 08. 인체 마이크로바이옴 분석(Human Microbiome Analysis)

강사: 김한나

소속: 강북삼성병원 연구지원실, 성균관대학교 삼성융합의과학원 임상연구설계평가학과

교육일정

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	인체 마이크로바이옴 연구 소개 및 메타지놈연구 디자인	김한나	강의
	11:00-12:30	16S rRNA gene 시퀀싱 Raw 데이터의 이해와 정도관리	김한나	강의
	14:00-15:30	다양성 (α-diversity, β-diversity) 및 Taxonomic Profiles 의 이해	김한나	강의
	15:40-17:10	QIIME2 프로그램의 개요 및 설치	김한나 (박은교,김신성)	강의/실습
DAY 2	09:20-10:50	Fastq 파일로부터 Feature Table 만들기 (Importing, Demultiplexing, Denoising)	김한나 (박은교,김신성)	강의/실습
	11:00-12:30	Phylogenetic tree 와 Taxonomy Classification	김한나 (박은교,김신성)	강의/실습
	14:00-15:30	미생물 다양성 분석 및 군집 구조 비교 분석 (α-diversity, β-diversity, Taxonomic analysis)	김한나 (박은교,김신성)	강의/실습
	15:40-17:10	QIIME2 외 다양한 프로그램 활용을 위한 메타지놈 데이터의 importing/exporting	김한나 (박은교,김신성)	강의/실습

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 09. 유전체 역학(Genomic Epidemiology)

강사: 박지완, 조윤신, 지선하

소속: 한림대학교 의학과, 한림대학교 바이오메디컬학과, 연세대학교 보건대학원 역학건강증진학과

강의개요

유전체역학은 유전과 환경의 상호작용이 질병의 발생과 분포에 미치는 영향을 밝히기 위한 학문이다. 본 강좌에서는 질병유전체학의 기본개념을 설명하고 최신 유전체역학연구에서 보편적으로 사용되는 연구방법론과 대표적인 분석 tool을 이용하여 임상-역학-유전변이 데이터를 분석하고 결론을 추론하는 방법을 습득한다. 질병위험요인을 밝히기 위한 Study design(가족, 환자-대조군, 코호트 연구)과 연구유형(유전적 연관성, 시계열분석, 유전-환경 상호작용, 멘델 무작위 분석법)에 적합한 분석 소프트웨어 사용법, 표본 수 산정, 통계분석법을 중심으로 중급 수준의 데이터 분석능력 습득을 목표로 한다.

- 준비물: 노트북 지참(Window 7 이상)
- 실습: 인터넷 접속

교육일정

참여조교: 전하라, 김도윤(한림대학교)

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	유전체 역학의 개요 Overview: Genomic Epidemiology	박지완	강의
	11:00-12:30	질환 유전자 발굴 전략 Disease Gene Identification Strategies	박지완	강의
	14:00-15:30	전장 유전체 연관성 분석 Genome-wide Association Analysis	조윤신	강의/실습
	15:40-17:10	전장 유전체 염기서열 분석 NGS Data Analysis	조윤신	강의/실습
DAY 2	09:20-10:50	생존 분석 Survival Analysis	박지완	강의/실습
	11:00-12:30	유전-환경 상호작용 분석 Gene-Environment Interaction Analysis	박지완	강의/실습
	14:00-15:30	멘델 무작위 분석법 Mendelian Randomization	지선하	강의/실습
	15:40-17:10	유전체 역학 최신 경향/ 연구 발표 New Trends in Genomic Epidemiology/ Research Presentation	박지완	강의

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 10. 암유전체학(Cancer Genomics)

강사: 이상혁, 김종경

소속: 이화여자대학교 생명과학과, 대구경북과학기술원

강의개요

본 강좌에서는 암유전체학의 최근 발전 동향 및 주요 주제를 데이터 분석의 측면에서 살펴보고자 한다. 차세대 시퀀싱(NGS) 방법의 도입에 따라 암유전체학은 비약적인 발전을 거듭하여 맞춤의료, 정밀의료의 첨단 분야로 확립되고 있다. 1일차 강의에서는 차세대 시퀀싱 기반의 암유전체학에 대한 소개와 NGS 데이터의 기본적인 분석 방법을 다룰 것이다. 모든 분석의 근간이 되는 whole exome sequencing과 transcriptome sequencing 데이터 분석을 통하여 somatic variants를 찾고 발현 데이터로부터 DEGs, signature genes, gene fusion 등을 구하는 방법을 실습과 함께 익힐 것이다. 2일차 강의는 최근 암 연구의 패러다임을 바꾸고 있는 단일세포 시퀀싱, 암 면역치료를 중심으로 실제 임상에서 정밀의료 구현을 위한 환자 상세분류 등의 다양한 주제를 다루고자 한다. 실습은 교육생들이 추후 실제 연구에 활용할 수 있도록 분석 프로토콜을 제공하고 설명할 것이다.

- 준비물: 노트북 지참
- 실습: 인터넷 접속

교육일정

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	차세대 시퀀싱과 암유전체학 개요	이상혁	강의
	11:00-12:30	Variant Discovery (Mutations, Indels, CNAs)	이상혁	강의
	14:00-15:30	Transcriptome Analysis (Expression, Signatures, Gene fusions)	이상혁	강의
	15:40-17:10	NGS 데이터 분석 실습	김상욱, 장예은	강의/실습
DAY 2	09:20-10:50	Single Cell Genomics 1	김종경	강의
	11:00-12:30	Single Cell Genomics 2	김종경	강의
	14:00-15:30	Cancer Immunotherapy	이상혁	강의
	15:40-17:10	Cancer Precision Medicine	이상혁	강의

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 11. 오믹스 기반 시스템생물학(Omics-based Systems Biology)

강사: 황대희
소속: 서울대학교 생명과학부

강의개요

This course teaches fundamental methodologies in omics-based systems biology including (1) Type, structure, and characteristics of biological systems (single molecules, molecular complexes, organelles, cells, tissues, organs, circulation systems, etc.), (2) omics (genomics, transcriptomics, epigenomics, proteomics, metabolomics, etc.) technologies, (3) standard data formats for exchange and storage and data normalization and visualization, (4) Identification of genes/proteins/metabolites predominantly expressed in a particular condition (e.g., diseases, stresses, and infections), clustering, and functional enrichment analysis, (5) Types of biological networks and network modeling and analysis, and (6) Identification of key regulators, pathways, or network modules governing a given phenotype. Case studies where these methodologies are applied to real biological questions will be introduced.

- 준비물: 노트북 지참(노트북 메모리 RAM 2GB 이상, Window 7 이상 필요)
- 실습: 인터넷 접속(데이터 및 소프트웨어 다운로드)

교육일정

참여조교: 채세현(서울대학교 생명과학부)

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	Introduction to Systems Biology	황대희	강의
	11:00-12:30	Genomic technology & analysis	황대희	강의
	14:00-15:30	Data analysis of genomic data	황대희	강의
	15:40-17:10	Proteomic technology & analysis	황대희	강의
DAY 2	09:20-10:50	Data analysis of proteomic data	황대희	강의
	11:00-12:30	Network modeling and analysis	황대희	강의
	14:00-15:30	Application case study; Practice	황대희, 채세현	강의/실습
	15:40-17:10	Practice	실습	실습

제14회 통계유전학워크숍

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 12. 빅데이터 분석 기법을 활용한 다중 오믹스 데이터 분석(Multi-omics Data Analysis)

강사: 이제근, 김성환

소속: 부산대학교 치의학전문대학원 치의생명과학과, 건국대학교 응용통계학과

강의개요

본 강좌에서는 빅데이터 통계 분석 기법을 활용하여 다중 유전체 (multiple omics) 데이터를 분석하는 여러 가지 방법들을 학습한다. 주로 통계 패키지인 R프로그램을 이용하여 실제 데이터 및 모의 분석을 병행 실습하고, 현재 많은 관심을 받고 있는 유전체 빅데이터 분석 및 이종 유전체 결합 분석에 대한 그 현황과 방법론을 다루고자 한다. 또한 실제 데이터 분석 실습을 통하여 의생명과학적 의미를 도출하는 방법도 함께 다룬다. 강의는 크게 두가지 파트로 나뉜다.

첫번째 파트는 주로 빅데이터 분석에 사용하는 기본적인 기계학습 기반 방법들에 대해 배운다. 구체적으로는 R을 이용한 기본적인 데이터 분석 방법에서부터, PCA, NMF, t-SNE 등을 이용한 차원 축소 (dimension reduction) 및 시각화 방법을 우선적으로 배운다. 또한 k-means clustering, hierarchical clustering을 비롯하여 heatmap을 이용한 시각화와 consensus clustering에 대해서도 실습을 통해 학습하며, 각종 classification 방법에 대해서도 실습과 함께 살펴본다.

두번째 파트는 주로 다른 종류의 유전체 데이터를 결합하여 분석하는 이종 유전체 결합 분석에 대해 학습한다. 구체적으로는 메타품질관리 (MetaQC), 메타 유전자 탐색 (MetaDE), 메타 군집분석 (Meta clustering), 메타판별분석 (Meta prediction) 및 메타시각화 (Meta visualization) 등의 방법론을 살펴볼 예정이다.

- **준비물:** 노트북 지참 (통계 패키지 R 설치)
- **실습:** 인터넷 접속 (웹서핑 및 CRAN/bioconductor에서 통계 패키지 다운로드 및 설치)
- **수강생수준:** R 프로그램에 대한 경험 및 기초 지식을 권장하나, 미경험자도 강의를 수강할 수 있도록 R 프로그램에 대한 간단하 기본 소개는 강의 초반에 진행할 예정이다.

제14회 통계유전학워크샵

14th Asian Institute in Statistical Genetics and Genomics

Graduate School of Public Health, Seoul National University, Seoul, Korea

July 15(Mon) ~ July 20(Sat), 2019



강좌소개 12. 빅데이터 분석 기법을 활용한 다중 오믹스 데이터 분석(Multi-omics Data Analysis)

강사: 이제근, 김성환

소속: 부산대학교 치의학전문대학원 치의생명과학과, 건국대학교 응용통계학과

교육일정

일자	시간	강의내용	담당강사명	구분
DAY 1	09:20-10:50	R 소개 및 R을 이용한 데이터 분석	이제근	강의/실습
	11:00-12:30	데이터 차원 축소 및 시각화	이제근	강의/실습
	14:00-15:30	데이터 군집화(Clustering)	이제근	강의/실습
	15:40-17:10	데이터 분류(Classification)	이제근	강의/실습
DAY 2	09:20-10:50	Omics통합 분석 개관 및 MetaQC (Quality control)	김성환	강의/실습
	11:00-12:30	MetaDE(differential expressed genes) 분석	김성환	강의/실습
	14:00-15:30	Omics 데이터 통합 시각화(Visualization)	김성환	강의/실습
	15:40-17:10	Omics 데이터 통합 군집 및 분류 (clustering and classification) 분석	김성환	강의/실습