Universitat Pompeu Fabra Barcelona

COURSE PLAN
Genome Phenome Analysis for Human Health (GPA)
Master in Bioinformatics for Health Sciences

2015-2016 Academic Year
Qualification. Master's Degree

1. Description of the subject

- Subject name: Genome Phenome Analysis for Human Health
- Code: GPA
- Total credits: 5 ECTS
- Workload: 125 hrs
- Year: 1st Year MSc
- Term: 3rd Term
- Type of subject: Optional
- Centre: Department of Experimental and Health Sciences /
- Teaching language(s): English
- Teaching team/Teaching staff:
  - Subject Coordinator: Arcadi Navarro
  - Teaching staff: Ferran Casals, Hafid Laayouni, Juan Antonio Rodriguez, Francesc Calafell, David Hughes

Other references

- Groups: 1 single group
- Timetables: Taught at various times, usually in the mornings
- Building: Dr. Aiguader 80
- Classrooms: TBD
- Comments: Please bring your own laptop for the hands-on sessions

2. Teaching guide (Campus Global)

- **Introduction**

  This subject focuses on the conceptual underpinnings and the bioinformatics tools that are used to learn about the genetic architecture of human interindividual phenotypic differences. The course, in short, is about linking genotypes and phenotypes. Obviously, the most relevant such differences are those relating to disease, such as the presence/absence of a certain condition or differential responses to treatments. However, other complex traits are also covered in the course. An emphasis is made on novel methods used to ascertain the causal genetic variants associated with complex disease, which are currently being applied to study many diseases of great relevance for public health. Recent studies are used as illustrative examples.

- **Associated competences**

  General competences

  Instrumental:
  1. Proficient reading/writing/listening scientific English related to the subject.
  2. Knowledge of software to do quality scientific presentations and reports related to the subject.
  3. Elements of Statistics: concepts of association, correlation and interaction.
  4. Elements of Genomics: genes, alleles, genetic variants, markers, phenotypes, expression, complex traits, genomes, exomes, causal variants, heritability.

  Interpersonal:
  1. Group work
  2. Ability to solve by yourself a given problem

  Systemic:
  1. Analysis and synthesis abilities
  2. Ability to search for information

  Specific competences

  1. To understand the complex nature of the factors generating phenotypes.
  2. To understand the concept of heritability.
  3. To become familiar with the different forms of human genome variation.
  4. To understand the concept of Mendelian Disease.
  5. To understand the concepts of Complex Trait and Complex Disease.
  6. To understand the genealogical and correlational structure of human genome variability.
  7. To master the basics of Linkage Disequilibrium measures and testing.
  8. To master the concept of genotype and haplotype.
  9. To understand and apply the concept of Linkage Mapping
  10. To understand and apply the concept of Association Test.
  11. To master different allele-disease association measures.
  12. To understand what interactions are and how they can be tested
  13. To master and apply the concept of logistic regression.
  14. Ability to use extant genetic epidemiology software.
  15. To understand the concept of multiple testing problems.
  16. To master different methods to correct for multiple testing.
  17. To get the essentials of Genome Wide Association studies
  18. To master information retrieval from disease-related public databases.
  19. To understand and use basic variant calling techniques
  20. To understand how new technologies of sequencing have revolutionized the field
  21 Capacity to decide how to apply NGS technologies to different types of disease
  22 To understand the new challenges and limitations of NGS technologies

Contents

Contents 1: Block 1: Overview of Health, Disease and Genomics.

Concepts	Procedures
Health and disease. Diseases as phenotypes. Disease types: the Mendelian to complex spectrum	To be able to obtain information from disease databases.
Variation in the human genome. Types of variation. Dynamics of variation. Sequencing and Genotyping techniques. Copy number variation and disease.	To be able to obtain information from variation databases.
Linkage disequilibrium: the genealogical basis of association between Disease Susceptibility Loci and Genetic Markers	To be able to compute linkage disequilibrium between markers. To be able to infer haplotypes from genotypes.
Quantitative trait loci (QTLs). Heritability concept.	To be able to use the basic concepts of heritability to perform basic calculations of the contributions of genetic and environmental factors to a given trait.
The common variant/common disease paradigm	To understand the diversity of possible genetic architectures of human traits.

Contents 2: Block 2: Family-based methods for disease mapping.

Concepts	Procedures
Family studies, adoption studies, twins. Linkage analysis. Definition. Types. Properties. Examples	To be familiar with family-based inference. To be able to perform basic linkage-mapping calculations.
The Transmission-Disequilibrium test and the Family Based Association Tests.	To be able to perform basic TDTs.

Contents 3: Block 3: Methods and tools for disease gene mapping.

Concepts	Procedures
Association studies. From single polymorphisms in candidate genes to Genome-Wide Association Studies	To be able to perform different association tests.
Population genomics. The 1kGP, P3G projects, WTCCC, Biobanking and other public initiatives.	To master information retrieval from disease-related public sources.
2nd Generation Sequencing. Exomes and Full Genomes.	To be able to understand and decide which NGS approach fits better to a given disease.
Functional annotation of genetic variants. Functional interpretation of results.	To be able to develop functional interpretations of gene mapping studies and to suggest further research avenues.

Contents 4: Block 4: Technical and computational challenges.

Concepts	Procedures
Shortcomings and challenges of large-scale association studies. Replications.	To become aware of the importance of replication
Multiple testing	To be able to understand and implement different strategies for multiple test correction.
Interactions.	To understand the concept of interaction and to be able to deploy different strategies to detect interactions

Study design	Being able to get a synthesis of the different methodologies for genotype-phenotype studies in a critical way

Contents 5: Block 5: Examples of the Genomics of complex disease.

Concepts	Procedures
Examples on the genetics of normal variation: skin color, eye color, personality traits.	To be able to critically assess extant research on the genetic architecture of complex disease.
Recent studies on the genetics of disease. Pharmacogenetics.	To be able to pinpoint specific drawbacks related to different computational methodologies.

Teaching methodology

Approach and general organization of the subject
The course has 5 ECTS credits, comprising 15/20 hours of plenary lectures, 8/10 hours of exercises and hands-on computer classes, 5-7 hours of students' presentations in seminars, 13/18 hours of reading and personal study, and 2 hours performing tests

The subject is based on the understanding of key methodological concepts and tools and on the application of common software found in the genetic epidemiology and statistical genomics labs around the world. As this is a completely incremental subject, the student is advised of the need of strong interaction with the lecturers and of keeping the class material up to date. Thus, the methods used are strongly based on the good understanding of basic principles of computer programming and statistics.

The course includes classical lectures and hands-on exercises.

Training activities*
The subject focuses on practical implementation of different types of tools for simple genotype-phenotype studies. Exercises and tutorials are distributed previous to the corresponding sessions.

Hands-on exercises are based on the use of publicly available resources and software packages, so individual development beyond what is covered in the course can be easily achieved by the students.

Each student will present a short seminar of 20-40 mins on a scientific article of relevance for the field.

Assessment

Assessment system
All the competences that need to be achieved in the course will be evaluated by means of an exam and the end of term. The exam will be individual and based on short questions/answers, some problems and some text questions. In addition, the seminar presentations will be also evaluated.

Grading system
Grades are between 0 and 10 and an overall 5 is needed to pass. The weight that is given to each of the exam sections depends on the performance of the whole group.

3. Programme of activities (Aula Global)*

- Description of the subject
  This subject focuses on the conceptual underpinnings and the bioinformatics tools that are used to learn about the genetic architecture of human interindividual phenotypic differences. The course, in short, is about linking genotypes and phenotypes. Obviously, the most relevant such differences are those relating to disease, such as the presence/absence of a certain condition or differential responses to treatments. However, other complex traits are also covered in the course. An emphasis is made on novel methods used to ascertain the causal genetic variants associated with complex disease, which are currently being applied to study many diseases of great relevance for public health. Recent are used as illustrative examples.

- Total credits: 5 Total number of hours: 35h (plus 2h exam)
- Estimated time spent on the subject: 125h

  - In the classroom: 35h
  - Outside the classroom: 90h

Weekly timetable of learning and assessment activities

Week (dates)	Work in the classroom (plenary, seminar, practical, etc.)	Estimated time (hrs)	Activities outside the classroom (time studying, preparing activities, etc.)	Estimated time (hrs)
1st week	Plenary. Introduction. Linking Genotypes and Phenotypes. diseases as a phenotype. Disease types: the Mendelian to complex spectrum.	1	*Study*	3
	Plenary. How do we know that a disease has a herederitary component? QTLs Heritability concept. Quantifying heritability: familiy studies, adoption studies, twins.	1	*Study*	3
	Plenary. An overview of variation in the human genome. Types of variation. Dynamics of variation. eQTLs. Epigenenomics. Examples of links betwen genomic variation and disease.	2	*Study*	4
	Linkage disequilibrium: the genealogical basis of association between Disease Susceptibility Loci and Genetic Markers. TagSNPs. Rare variants	2	*Study*	4
2nd week	Practical. Databases of human variation: OMIM, dbSNP, HapMap, 1kGP, EVA and others	2	Review of databases. Preparation of exercises. *Study*	6
	Plenary. Overview of how to links phenotype and genotype. Families vs. Unrelated individuals. Linkage analysis. Definition, types and properties. Examples. The Transmission-Disequilibrium Tests: definition. Other family-based association tests. Examples.	1	Study	3
	Plenary. Association studies and	1	*Study*	3

	their Whole Genome Extensions. Design and Analysis of Association studies and Genome-Wide Association Studies (GWAS). Candidate-Gene vs. Genome-wide approaches.			
3rd week	Practical. Candidate-based Association Studies with PLINK & relevant R Packages. Genoptyping. Phasing. Testing	2	Review of databases. Preparation of exercises. Study.	6
	Plenary. Challenges of GWASs (Session 1). The multiple testing problem and correction procedures. The problem of lack of replicability. Publication biases. The Common Variant / Common Disease Paradigm and the heritable causes of disease. The mystery of the missing heritability.	2	*Study*	6
4th week	Plenary. Challenges of GWASs (Session 2). Adding biology into the analysis. Gene-by-gene analysis. Pathways. Haplotype analysis and simultaneous analysis of all SNPs. Testing for interactions. High order analysis.	2	*Study*	4
	Practical. Array-based Association Studies. Session 1: PLINK	2	Review of databases. Preparation of exercises. Study.	6
	Practical. Array-based Association Studies. Session 2: PLINK, VEGAS, GCTA	2	Review of databases. Preparation of exercises. Study.	4
	Plenary. Interpreting GWAS results. Genomic Risk Scores. Variation effect annotation. Translation to the clinic. Complex diseases vs. Cancer.	1	Study.	3
5th week	Practical. Interpretation of Genotype-Phenotype links. Functional annotation of Variants in Cancer	2	Review of databases. Preparation of exercises. Study.	6
	Seminars. Students presentations	2	*Paper preparation. Study*	4
	Update on population genomics, Decode Genetics. The GA4GH. The 1000 Genomes Project. UK100K and other large consortia.	1	*Study*	3
	Rare variants and human disease	1	*Study*	2
	Practical- Analysis of NGS data.	2	Review of formats & software. Preparation of exercises. Study.	6
6th week	Seminars. Students presentations	2	*Paper preparation. Study*	2
	Genome-Phenome analysis with exomes & full genomes. The epigenome roadmap.	2	*Study*	4
	Seminars. Students presentations	2	*Paper preparation. Study*	6
Other	Exam	2	Individual Exam	2

Total hours	37h	90

Assessment of all specific competences is by an exam at the end of term.

Bibliography

There is no text-book bibliography as such, but a wide selection of papers will be distributed during the course.

Also, these books are available in the library

- Evolution in health and disease / edited by Stephen C. Stearns and Jacob C. Koella
- The Genetic basis of common diseases / edited by Richard A. King, Jerome I. Rotter, Arno G. Motulsky
- Genetics and analysis of quantitative traits / Michael Lynch, Bruce Walsh
- Handbook of medical informatics / J.H. van Bemmel, M.A. Musen, editors ; J.C. Helder, managing editor
- Handbook of statistical genetics / editors, D.J. Balding, M. Bishop, C. Cannings
- Human genome epidemiology : a scientific foundation for using genetic information to improve health and prevent disease / edited by Muin J. Khoury, Julian Little, Wylie Burke
- Statistical methods in genetic epidemiology / Duncan C. Thomas