GHENT UNIVERSITY

Course Specifications

Valid as from the academic year 2026-2027

Data Intelligence in Sustainable Drug Discovery (J000530)

Course size *(nominal values; actual values may depend on programme)*

Credits 6.0 **Study time 180 h**

Course offerings and teaching methods in academic year 2026-2027

A (semester 1)	English	Gent	lecture
			seminar
			independent work

Lecturers in academic year 2026-2027

Pattyn, Filip	FWO1	lecturer-in-charge
Hertleer, Carla	FWO2	staff member
Van Nieuwerburgh, Filip	FWO1	co-lecturer

Offered in the following programmes in 2026-2027	**crdts**	**offering**
International Master of Science in Sustainable Drug Discovery	6	A

Teaching languages

English

Keywords

Omics technologies, genetic analyses using bioinformatics, data visualization, data management, FAIR data, relational & graph databases, differential expression/pathway/gene ontology analysis, R, Python, SQL & semantic web technologies

Position of the course

This course will introduce omics technologies, data science, programming, knowledge management and bioinformatic concepts and tools as applied in the context of drug development.

Contents

Pharmacogenomics

- Basics in genetics (human genome, epigenome, gene expression regulation, genetic variants)
- Omics techniques such as microarrays, and NGS/MPS
- Application of omics in GWAS, clinical diagnosis, neoantigen cancer treatment, etc.
- Practical sessions on BLAST, NGS read mapping, variant calling and genetic databases such as PharmGKB

Transcriptome data analysis

- Pipeline for analysis of transcriptome data in context of downstream pathway analysis of drugs
  - Data pre-processing, quality control, visualization/exploration
  - Differential expression analysis
  - Gene ontology and pathway analysis
- Public databases

Data science & data management: sustainable storage and usage of data

- Data models and technologies, especially details on the pros and cons of these models/technologies in some existing use cases. Hereby, we will pay particular attention to relational, graph-based and semantic data models. We will highlight fundamental operations on that model, which will lead us to the world of query

languages. In addition, metadata annotation and data standardization in drug discovery research will be handled.
- Usage of data in practical scenarios. This means we will see some best practices to make data available for internal and external sharing and re-use, which brings us to FAIR principles. This is of huge importance in a setting where re-validation and confirmation is becoming more and more important. Also, scenarios with public data initiatives will be handled.
- Some practical skills will be developed by bringing the theoretical principles to live by tackling some use cases. We will work with tools to setup a data extraction and transformation pipeline and learn how to inspect datasets in a critical way.

Initial competences

Basics in statistics, chemistry, biology

Final competences

1 Implement basic algorithms in R for transcriptome analysis
2 Understand and perform a data-mining pipeline on genome and transcriptome data, including visualisation and differential expression analysis
3 Explain the different omics technologies and approaches
4 Extract useful drug discovery information from different datasets
5 Understand and apply data management and data stewardship using FAIR data principles in the context of drug discovery
6 Understand and apply the basics of relational database management, graph databases and semantic web technologies
7 Implement basic scripts in Python for data extraction and transformation.
8 Understand and apply SQL to query relational databases
9 Critically use AI tools in programming, data transformation and analysis in the context of drug discovery

Conditions for credit contract

Access to this course unit via a credit contract is determined after successful competences assessment

Conditions for exam contract

This course unit cannot be taken via an exam contract

Teaching methods

Group work, Seminar, Lecture, Practical, Independent work

Extra information on the teaching methods

Pharmacogenomics, data management, and FAIR data topics will be covered in lectures, often combined with demonstrations and practical sessions.
Pipelines for the analysis of transcriptome data will be explained during lectures and applied in seminars and practical sessions.
During lectures and practical sessions, students will learn the basics of R, Python, SQL, and semantic web technologies.
The critical and responsible use of AI to assist in programming and data science will be addressed during practical sessions, seminars, and group work.
Group work and microteaching: in groups, students will solve practical problems by combining various tools and techniques. Topics relate to omics data analysis, data extraction, transformation and management, and data modeling in the context of drug discovery.

Study material

Type: Syllabus

Name: Recent manuscripts
Indicative price: Free or paid by faculty
Optional: no

Type: Slides

Name: Slides
Indicative price: Free or paid by faculty
Optional: no
Language : English
Available on Ufora : Yes
Online Available : Yes

Type: Software

Name: Python programming language, RStudio, SQL and relational database management software, Integrated Development Environment
Indicative price: Free or paid by faculty
Optional: no
Online Available : Yes

References

Course content-related study coaching

Assessment moments

end-of-term and continuous assessment

Examination methods in case of periodic assessment during the first examination period

Written assessment with multiple-choice questions, Oral assessment open-book, Written assessment with open-ended questions, Written assessment

Examination methods in case of periodic assessment during the second examination period

Examination methods in case of permanent assessment

Presentation, Assignment

Possibilities of retake in case of permanent assessment

not applicable

Calculation of the examination mark

Students who eschew period aligned and/or non-period aligned evaluations for this course unit may be failed by the examiner.