

## Data Intelligence in Sustainable Drug Discovery (J000530)

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

**Credits 6.0**

**Study time 180 h**

**Contact hrs**

60.0h

### Course offerings in academic year 2022-2023

A (semester 1)

English

Gent

### Lecturers in academic year 2022-2023

De Preter, Katleen

GE31

lecturer-in-charge

Bronselaer, Antoon

TW07

co-lecturer

Van Nieuwerburgh, Filip

FW01

co-lecturer

### Offered in the following programmes in 2022-2023

[International Master of Science in Sustainable Drug Discovery](#)

**crdts**

6

**offering**

A

### Teaching languages

English

### Keywords

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

### Position of the course

This course will introduce omics technologies, and bioinformatic concepts and tools as applied in drugging.

### Contents

- 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
- Advanced use of R in drug selection/discovery/repurposing
- Transcriptome data analysis (using R)
- Pipeline for analysis of transcriptome data in context of drug target prioritization/identification, drug repurposing and downstream pathway analysis of drugs
  - Data pre-processing, quality control, visualization/exploration
  - Differential expression analysis
  - Gene ontology and pathway analysis
- Public databases
- Single-cell versus bulk transcriptome data
- 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 three broad categories of models: those based on relations, those based on graphs and those based on trees. We will highlight fundamental operations on those models, 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. Next, we will discuss data quality techniques.
- Some practical skills will be developed by bringing the theoretical principles to live by tackling some use cases. We will work with simple tools to setup a data pipeline and learn how to inspect datasets in a critical way.

#### **Initial competences**

Basics in Python

#### **Final competences**

- 1 Implement simple algorithms (R or Python) in the context of drugging (drug prioritization, drug target identification, etc).
- 2 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 FAIR data principles 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**

Lecture: plenary exercises, Demonstration, Group work, Microteaching, Guided self-study, Lecture, Seminar: practical pc room classes

#### **Extra information on the teaching methods**

Group work for microteaching: in groups of 4/5 students work on a published dataset and run the transcriptome pipeline (data pre-processing, data exploration, differential expression analysis, GSEA, ...). Via microteaching, they present their scripts/approaches and findings to the other students.  
Data management and FAIR data topics will be handled during lectures often combined with demonstrations and with practical sessions.

#### **Learning materials and price**

Laptop/computer  
Recent manuscripts

#### **References**

#### **Course content-related study coaching**

#### **Assessment moments**

end-of-term assessment

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

Written examination, Open book examination, Assignment

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

Written examination, Open book examination, Assignment

#### **Examination methods in case of permanent assessment**

#### **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.

