Bio-Informatics

BIO-INFORMATICS is the professional field in which computer and information technology is used to collect, store, share, analyse, interpret and disseminate data from biomedical and/or biological research.

The main subcategories of this specialisation are: genomics, transcriptomics, proteomics (including protein modelling, structures and functions), metabolomics and the integration of data from these areas. Bio-informaticians are employed to conduct biological and biomedical research in scientific institutions and in companies in the pharmaceutical, biotechnology, food and plant-breeding industries.

Institutions that offer the programme

- HAN University of Applied Sciences, Nijmegen
- Hanze University of Applied Sciences, Groningen
- University of Applied Sciences Leiden

ILLUSTRATION OF PROFESSIONAL FIELD

Occupations, jobs and roles for graduates are mostly to be found in the following professional domains (for a full description of the professional domains, see Section 2). A few examples are given for each domain.

Research and development
- Analysing large data sets from high-throughput laboratory research
- Scientific programmer

Application and production
- Managing gene and protein databases
- Analysing gene-sequencing data flow

Commerce and customer service
- Biotechnology data consultant

National programme profile

	Competence							
	Research	Experimentation	Development	Management	Advice	Instruction	Leadership	Self-management
Minimum national attainment target adopted for the programme	III	III	_**	II	I	I	I	II

** Students can choose to raise the level of these competences by making certain choices in their range of subjects, internship and graduation project during the last two years of their course.

TYPICAL TEXTBOOKS

- *Campbell Biology,* L.A. Urry, M.L. Cain e.a.
- *Essentials of Genetics / Concepts of Genetics,* W.S. Klug, M.R. Cummings e.a.
- *General, Organic and Biochemistry,* K. J. Denniston, J.J. Topping e.a.
- *Starting out with Python,* T. Gaddis
- *Data Structures and Algorithms using Python,* R.D. Necaise
- *Bioinformatics and Functional Genomics,* J. Pevsner
- *Using R for Introductory Statistics,* J. Verzani
- *Data Mining, I. Witten,* E. Frank e.a.
- *Statistics for the Life Sciences,* M. L. Samuels, J.A. Witmer e.a.

The list of typical textbooks serves as an illustration to give an impression of the level at which the subject is taught in the study programme.

Bio-Informatics

KNOWLEDGE

Biology
- **General Biology:** organisms, tissues, evolution
- **Cell biology:** cell structure, energy supply, transport, cell-cell communication, mitosis, meiosis, gene regulation, signal transduction routes, metabolic routes
- **Genetics:** DNA structure, replication, code, transcription, translation, epigenetics, mutations, SNPs, structural variations, inheritance, population genetics
- **Microbiology/virology:** construction, diversity, metabolic strategies, evolution

Chemistry
- **General chemistry:** atomic structure, periodic system, molecules, nomenclature, reactions, kinetics
- **Biochemistry:** building blocks, macromolecules, carbohydrates, lipids, proteins, enzymes, metabolism

SKILLS

Informatics
- **Programming in Python:** data structures, control structures, modular approach, GUIs, Biopython, python database approach, data types, object-oriented programming
- **Programming in Java:** object-oriented programming, application of algorithms, inheritance
- **Programming in R:** scripting, data analysis, statistics, data visualisation
- **Web technology:** HTML, CSS, web services
- **Databases:** relational design, implementation, querying and programmatic interaction (MySQL)
- **Workflow tools:** e.g. Galaxy, Snakemake
- **Linux:** bash scripting, Regex

Bio-informatics
Knowledge and analytical skills of at least the subjects listed below:
- **Sequencing technologies:** NGS technologies, assembly, mapping, NGS application areas (e.g. de-novo & re-sequencing) exome sequencing
- **Algorithmic aspects of sequences:** alignment, mapping, graphs, scoring matrices
- **Sequence annotation:** BLAST and related software
- **Gene expression analysis:** RNA-seq data, Bioconductor
- **Homology and phylogenics**
- **Practical use of bio-informatics tools:** e.g. BLAST, OMIM, Genome Browsers, Genbank, Uniprot, KEGG, MSA tools, topology prediction, PFAM, PROSITE, YASARA PDBe, Gene Expression Omnibus, FASTQ, mappers & aligners & assemblers

Statistics & data analysis
- **Sampling:** data types, population and sample, errors, bias, variation, (un)certainty
- **Descriptive statistics:** e.g. average, median, standard deviation, range, interquartile range
- **Visualization:** e.g. box-plot, histogram, scatterplots, Venn diagram, trees, heatmaps
- **(Hypothesis) tests:** e.g. t-test, ANOVA, chi-squared, Wilcoxon, non-parametric
- **Cluster analysis:** distance measurements, hierarchical clustering, k-means clustering
- **Regression:** linear, non-linear, multivariate, PCA
- **Datamining/machine learning:** e.g. Decision Trees, Naive Bayes, k-Nearest Neighbour, Neural Networks, SVM
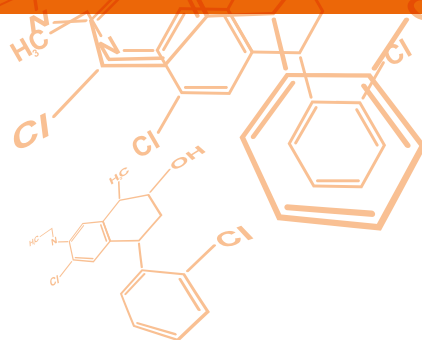
The Body of Knowledge and Skills is a summary of graduates' basic knowledge and basic skills which has been prepared by the HBO-programmes in consultation with the professional field. These are obtained during the first two years of education.

Programmer Freerk van Dijk:
'A lot of data requires a lot of computing power'

Name: Freerk van Dijk
Age: 25
Course of study:
Bio-Informatics
Place of employment:
Genomics Coordination
Centre, Genetics Depart-
ment, University Medical
Centre Groningen
Job: Scientific programmer

'I chose this programme because I was interested in biology. One of my hobbies is computers/gaming and I quickly made the link with this course of study. As I had a pre-university diploma, I was able to follow the fast-track bio-informatics programme. I more or less started in the second year and then had to take a number of subjects from the first year. I made up the lost ground in programming languages by putting in extra hours. A lot of attention was also focused on biology and laboratory work during the programme.

My work currently involves setting up pipelines/systems for analysing next-generation sequencing (NGS) data. This technology is used to replicate human DNA (mainly from the exons) and analyse it, which generates massive amounts of data. The analysis part requires a lot of computing power, which means that these analyses have to be performed on computing clusters. Consideration has to be given not only to the open source software required but also to ways of distributing these analyses in parallel. The ultimate goal is to detect variations in the genome and validate any causal variations in the laboratory. We are also responsible for supplying servers and storage capacity for software which the genetics department uses in different DNA analyses and assisting the researchers by providing scripts, software, etc.

Genome
I am currently working on the "Genome of the Netherlands" project, in which we have mapped out the complete genome of 250 father-mother-child trios. The aim is to provide a clear picture of all natural variation in the Dutch population. This data can be used to rule out frequently occurring variations in our population as being the cause of illness in sick people.

The programme fitted in really well with my job. Initially, I did notice that I had followed the fast-track programme: I was lacking a number of hours of programming experience, although you catch up in the course of your work. Lab technologies are also important because they are used to prepare samples for NGS. The analysis of NGS data was not included in the programme, because that technology was still new at the time.

Experimentation and research are the most important competences in my job. The professional field is still fairly new, which means that a lot of software and analyses have not been standardised. The result is that a lot of new software and analysis methods have to be designed and implemented. To me, UMC Groningen is a challenging environment where the focus is on research. This ensures that there is enough variety to keep the job interesting and there are all kinds of opportunities for advancement in the area of research.' ◼
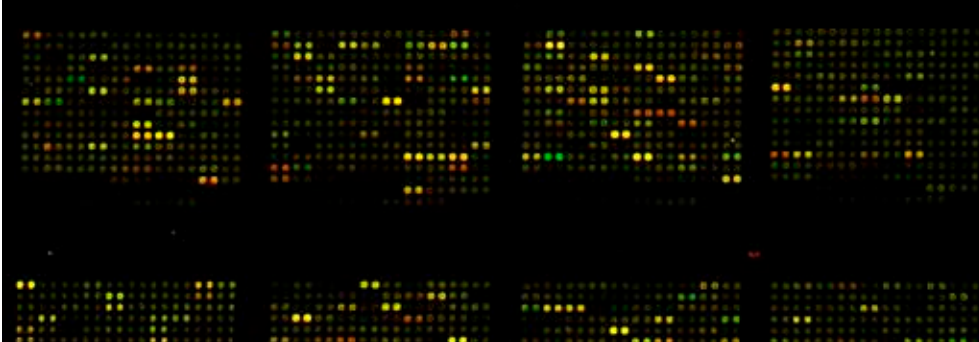
Bio-informatician Varshna Goelela:

'You always have to be eager to learn'

'I chose the bio-informatics programme because the combination of programming and biology appealed to me, partly on account of my previous ICT manager programme in senior secondary education (MBO). In addition, bio-informatics emerged from a course evaluation test as the best match for me. It was a fairly new course at the time – a lot of challenges and a lot of opportunities on the jobs market. And what I also regard as a major advantage: as a bio-informatician, you can progress in a very broad area or a very specialised area.

The programme was everything I expected it to be: a good combination of biology and programming. Now, during my day-to-day work, this combination seems to have been an ideal match. My choice of subject for my graduation project was mainly a practical and technical one. At the time, Next-Generation Sequencing was a relatively new technology that was making great advances and did not yet figure very much in the programme. I deliberately chose a graduate internship that did provide this opportunity and was able to learn how to analyse Next-Gen data. Looking back, this seems to have been a good choice as I am still making full use of the knowledge that I acquired at that time.

As a bio-informatician, I am responsible for carrying out data analysis and visualising various type of experimental data, such as transcriptomics, microbial data and metabolomics. I am also working on the development of bio-informatics pipelines and tools which are used both inside and outside our group. An example of this is a pipeline I have developed for automating quality control, preprocessing and normalisation of Illumina microarray data.

As far as I'm concerned, the programme fitted in really well with my job. During my studies, I also did various projects that provided me with the knowledge that I still use every day in my job. What the programme didn't cover so much was how important the design of a study is for data analysis. That's a pity, because the design of a study has a major impact on the performance of the analyses.

Communication
Research is an important part of a bio-informatician's job, so it's an essential competence for me. Sometimes, something will come your way that you have to learn about yourself, so I think you always have to be eager to learn. Good communication is another requirement. As a bio-informatician, you are surrounded by statisticians, scientists and ICT specialists, so it's useful to be able to communicate with each of them to streamline projects and analyses. Self-management is also important, for example, for enabling you to meet your deadlines.

I would still like to do a Masters degree in bio-informatics so that I can specialise further. In the professional field, I can see the way bio-informatics is playing an increasingly important part in science and advances in science ensure that tools are becoming increasingly efficient and analyses can be carried out faster and more accurately. I hope that I will be able to be part of these exciting advances in future.' ■

Naam: Varshna S. Goelela
Age: 26
Course of study:
Bio-Informatics
Place of employment: TNO Zeist, Microbiology and Systems Biology department
Job: Bio-informatician