

Bio-informatica



BIO-INFORMATICA is het werkveld waarbij computer- en informatietechnologie wordt gebruikt om gegevens uit biomedisch en/of biologisch onderzoek te verzamelen, op te slaan, toegankelijk te maken, te analyseren, te interpreteren en weer te verspreiden.

Belangrijke deelgebieden van dit vakgebied zijn: genomics, transcriptomics, proteomics (inclusief eiwitmodellering, -structuur en -functie), metabolomics en de integratie van data van deze gebieden. Bio-informatici zijn werkzaam in het biologisch en biomedisch onderzoek bij wetenschappelijke instituten, en bij bedrijven in de farmaceutische, biotechnologische, voedingsmiddelen- en plantveredelingsindustrie.

Instellingen die de opleiding aanbieden

- Hanzehogeschool Groningen
- Hogeschool Leiden
- Hogeschool van Arnhem en Nijmegen, Nijmegen

Landelijk opleidingsprofiel

| | Competentie | | | | | | | |
|--|-------------|----------------|-------------|---------|-----------|------------|--------------|-------------|
| | onderzoeken | experimenteren | ontwikkelen | beheren | adviseren | instrueren | leidinggeven | zelfsturing |
| Minimum landelijk vastgesteld eindniveau van de opleiding | III | III | - ** | II | I | I | I | II |

** Studenten kunnen ervoor kiezen om het niveau van deze competentie op te hogen door bepaalde keuzes te maken in hun vakkenpakket, stage en afstuderen tijdens de laatste twee jaar van hun studie



WERKVELDILLUSTRATIE

Beroepen, functies en rollen van de bachelor zijn met name in de volgende beroepsdomeinen te vinden (zie voor een volledige omschrijving van de beroepsdomeinen Hoofdstuk 2). Per domein zijn enkele voorbeelden gegeven.

Research en development

- Analyseren van grote datasets uit high throughput laboratoriumonderzoek
- Wetenschappelijk programmeur

Toepassing en productie

- Beheer gen- en eiwitdatabanken
- Analyseren van gen-sequencing datastroom

Commercie, service en dienstverlening

- Biotechnologisch dataconsultant

KENMERKENDE STUDIEBOEKEN

- *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.

De lijst met kenmerkende studieboeken dient ter illustratie om een indruk te geven van het niveau waarop binnen de opleiding wordt gewerkt.

KNOWLEDGE

Biologie

- **Algemene Biologie:** organismen, weefsels, evolutie
- **Celbiologie:** celstructuur, energievoorziening, transport, cel-celcommunicatie, mitose, meiose, genregulatie, signaaltransductieroutes, metabole routes
- **Genetica:** DNA-structuur, replicatie, code, transcriptie, translatie, epigenetica, mutaties, SNP's, structurele variaties, overerving, populatiegenetica
- **Microbiologie/virologie:** bouw, diversiteit, metabole strategieën, evolutie

Chemie

- **Algemene chemie:** atoombouw, periodiek systeem, moleculen, nomenclatuur, reacties, kinetiek
- **Biochemie:** bouwstenen, macromoleculen, koolhydraten, lipiden, eiwitten, enzymen, metabolisme

SKILLS

Informatica

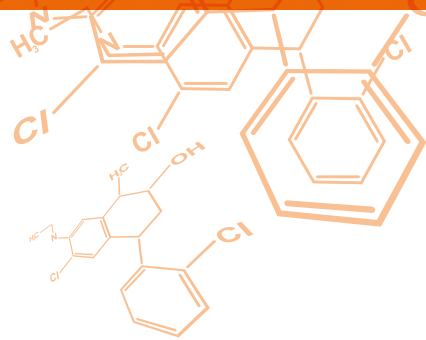
- **Programmeren in Python:** datastructuren, controlestructuren, modulaire aanpak, GUI's, Biopython, python-databasebenadering, datatypen, object-georiënteerd programmeren
- **Programmeren in Java:** object-georiënteerd programmeren, toepassen van algoritmen, overerving
- **Programmeren in R:** scripting, data-analyse, statistiek, datavisualisatie
- **Webtechnologie:** HTML, CSS, web services
- **Databases:** relationeel ontwerp, implementatie, querying en programmatische interactie (MySQL)
- **Workflow tools:** bijvoorbeeld Galaxy, Snakemake
- **Linux:** bash scripting, Regex

Bio-informatica

- **Sequencing technologieën:** NGS-technologieën, assembly, mapping, NGS-toepassingsgebieden (bijvoorbeeld de-novo & re-sequencing) exome sequencing
- **Algoritmische aspecten van sequences:** alignment, mapping, graphs, scoring matrices
- **Sequence annotatie:** BLAST en gerelateerde software
- **Genexpressie analyse:** RNA-seq data, Bioconductor
- **Homologie en fylogenie**
- **Praktisch gebruik van bio-informaticatools:** bijvoorbeeld BLAST, OMIM, Genome Browsers, Genbank, Uniprot, KEGG, MSA-tools, topologiepredictie, PFAM, PROSITE, YASARA PDBe, Gene Expression Omnibus, FASTQ, mappers & aligners & assemblers

Statistiek & data-analyse

- **Sampling:** typen data, populatie en sample, fouten, bias, variatie, (on)zekerheid
- **Beschrijvende statistiek:** o.a. gemiddelde, mediaan, standaarddeviatie, range, interkwartiel range
- **Visualisatie:** o.a. box-plot, histogram, scatterplots, Venndiagram, bomen, heatmaps
- **(Hypothese) toetsen:** o.a. t-test, ANOVA, chi-kwadraat, Wilcoxon, non-parametrische
- **Clusteranalyse:** afstandsmaten, hierarchical clustering, k-means clustering
- **Regressie:** lineair, niet-lineair, multivariaat, PCA
- **Datamining/machine learning:** o.a. Decision Trees, Naive Bayes, k-Nearest Neighbour, Neural Networks, SVM



De Body of Knowledge and Skills geeft de gemeenschappelijke basis van de opleiding bij de verschillende hogescholen weer. Studenten leren deze kennis en vaardigheden in de eerste twee jaar van hun opleiding. In de laatste twee jaar specialiseren zij zich.

Programmeur Freerk van Dijk: 'Veel data vragen veel rekenkracht'

Naam: Freerk van Dijk
Leeftijd: 25
Studie: Bio-informatica
Werkplek: Genomics
Coordination Center, afde-
ling genetica, University
Medical Center Groningen
Functie: Wetenschappelijk
programmeur

'Ik koos voor deze opleiding, omdat ik geïnteresseerd was in biologie. Een van mijn hobby's is computeren/gamen, en de link met deze studie was toen snel gelegd. Aangezien ik een vwo-diploma had, kon ik de verkorte opleiding Bio-informatica volgen. Ik begon als het ware in het tweede jaar en moest daarnaast een aantal vakken uit het eerste jaar volgen. De achterstand in programmeertalen werd ingelopen door het maken van extra uren. Tijdens de opleiding werd ook veel aandacht besteed aan biologie en laboratoriumwerk.'

Momenteel houd ik me bezig met het opzetten van pijplijnen/systemen om next-generation sequencing (NGS) data te analyseren. Met behulp van deze techniek wordt menselijk DNA (voornamelijk uit de exonen) vermenigvuldigd en geanalyseerd, waarbij enorm veel data gegeneerd worden. Voor het analyseren daarvan is veel rekenkracht nodig, zodat deze analyses uitgevoerd moeten worden op rekenclusters. Hierbij moet dus niet alleen nagedacht worden over de benodigde open source

software, maar ook over hoe deze analyses in parallel gedistribueerd kunnen worden. Het uiteindelijke doel is het in kaart brengen van variaties in het genoom en eventuele causale variaties valideren in het laboratorium. We zijn tevens verantwoordelijk voor het aanbieden van servers en opslagcapaciteit voor software die de afdeling genetica gebruikt voor verschillende DNA-analyses, en het ondersteunen van de onderzoekers door het aanbieden van scripts, software, etcetera.

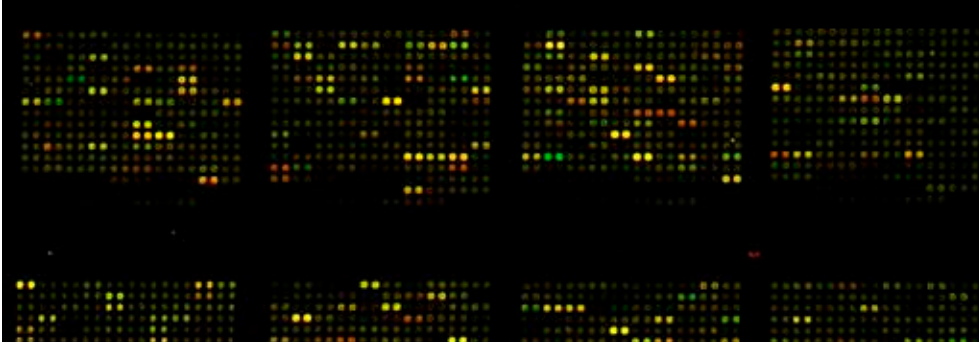
Genoom

Momenteel werk ik aan het project 'Genome of the Netherlands', waarbij we het complete genoom van 250 trio's – vader, moeder en kind – in kaart hebben gebracht. Het doel is om alle natuurlijke variatie in de Nederlandse populatie inzichtelijk te maken. Met behulp van deze data kunnen veelvoorkomende variaties in onze populatie bij zieke mensen als ziekteveroorzakend worden uitgesloten.

De opleiding sloot goed bij mijn werk aan. In het begin kon ik wel merken dat ik de verkorte opleiding heb gevolgd: ik miste een aantal uren aan programmeerervaring, al haal je dat tijdens het werk zo weer in. Ook labtechnieken zijn van belang, omdat die bij het prepareren van de samples voor NGS worden gebruikt. Het analyseren van NGS-data was tijdens de opleiding niet aan bod gekomen, omdat die techniek destijds nieuw was.

In mijn werk zijn **experimenteren** en **onderzoeken** de belangrijkste competenties. Het werkveld is nog vrij nieuw, wat er toe leidt dat veel software en analyses niet gestandaardiseerd zijn. Het gevolg is dat er veel nieuwe software en analysemethodes moeten worden bedacht en geïmplementeerd. Het UMC Groningen is voor mij een uitdagende omgeving waar onderzoek voorop staat. Dat zorgt ervoor dat er genoeg afwisseling is om het werk leuk te houden, en er zijn allerlei doorgroei-mogelijkheden richting onderzoek! ■





Bio-informaticus Varshna Goelela: 'Je moet altijd leergierig zijn'

Ik koos voor de opleiding Bio-informatica, omdat de combinatie van programmeren en biologie me aansprak, mede vanwege mijn vooropleiding mbo ICT-beheerder. Bovendien kwam uit een studiewijzertest Bio-informatica als nummer 1-studie voor mij naar voren. Het was destijds een vrij nieuwe studie – veel uitdaging en veel mogelijkheden op de banenmarkt. En wat ik ook als een groot voordeel beschouw: je kunt je als bio-informaticus heel breed of juist heel specialistisch ontwikkelen.

De opleiding was alles wat ik ervan verwachtte: een goede combinatie van biologie en programmeren. Deze combinatie blijkt nu tijdens mijn dagelijkse werkzaamheden een ideale match te zijn. De keuze voor mijn afstudeeronderwerp was vooral een praktische/technische. Next Generation Sequencing was destijds een relatief nieuwe techniek die erg in opmars was en nog niet zoveel aan bod kwam in de opleiding. Ik heb bewust een afstudeerstage gekozen waar ik die mogelijkheid wel had en kon leren Next Gen-data te analyseren. Achteraf bleek dit een goede keuze te zijn geweest, aangezien ik nog steeds uitvoerig gebruikmaak van de kennis die ik toen heb opgedaan.

Als bio-informaticus ben ik verantwoordelijk voor het uitvoeren van data-analyse en de visualisatie van diverse typen experimentele data, zoals transcriptomics, microbiële data en metabolomics. Verder werk ik mee aan de ontwikkeling van bio-informatica pipelines and tools, die zowel binnen als buiten onze groep worden gebruikt. Een voorbeeld hiervan is een pipeline die ik heb

ontwikkeld voor het automatisch uitvoeren van quality control, preprocessing en normalisatie van illumina microarray data.

De opleiding sloot wat mij betreft erg goed aan bij mijn werk. Tijdens mijn studie heb ik verschillende projecten gedaan waarvan ik de kennis nog dagelijks toepas in mijn werk. Wat niet zo aan bod kwam tijdens de opleiding, is hoe belangrijk de opzet van een studie is voor de data-analyse. Dat is jammer, want de opzet van een studie heeft grote gevolgen voor de uitvoering van de analyses.

Communicatie

Onderzoeken hoort echt bij een bio-informaticus, dus dat is voor mij een belangrijke competentie. Soms komt er iets nieuws op je pad dat je jezelf eigen moet maken, dus wat mij betreft moet je altijd leergierig zijn. Goede communicatie is ook een vereiste. Als bio-informaticus zit je tussen statistici, wetenschappers en ict-ers; het is handig als je met iedereen kunt communiceren om projecten en analyses te stroomlijnen. Zelfsturing is ook erg belangrijk, bijvoorbeeld om je aan je deadlines te kunnen houden.

Ik wil nog graag een master in de bio-informatica doen, om mij verder te specialiseren. In het werkveld zie ik dat bio-informatica een steeds belangrijkere rol speelt in de wetenschap en ontwikkelingen zorgen ervoor dat tools steeds efficiënter worden en analyses accurater en sneller kunnen worden uitgevoerd. Ik hoop dat ik in de toekomst deel mag uitmaken van deze spannende ontwikkelingen. ■

Naam: Varshna S. Goelela
Leeftijd: 26
Studie: Bio-informatica
Werkplek: TNO Zeist, afdeling Microbiology and Systems Biology
Functie: Bio-informaticus

