

# Netherlands Conference on BioInformatics 2004

## “Images of Life”



**Book of Abstracts  
7 & 8 October 2004  
“De Appel”  
Groningen**



**RuG**



**Hanzehogeschool Groningen**

## Programme Booklet

### Netherlands Conference on BioInformatics 2004

# “Images of Life”

October 7-8, 2004  
Conference Centre De Appel  
Groningen, The Netherlands

**Websites:**            [www.cmbi.ru.nl/wgbioinf/](http://www.cmbi.ru.nl/wgbioinf/)  
                              [www.cmbi.ru.nl/wgbioinf/symp2004/](http://www.cmbi.ru.nl/wgbioinf/symp2004/)  
                              [www.kncv.nl](http://www.kncv.nl)  
                              [www.hanze.nl/](http://www.hanze.nl/)  
                              [www.rug.nl/](http://www.rug.nl/)  
                              [www.rug.nl/rc/hpcv/visualisation/](http://www.rug.nl/rc/hpcv/visualisation/)  
                              [portalgroningen.medialab.nl/](http://portalgroningen.medialab.nl/)

### Organizing Committee

- Jan-Peter Nap, on behalf of the local Groningen committee
- Frank van Enkevort, on behalf of the Working Group BioInformatics

**Contact:**  
Netherlands Working Group BioInformatics  
p/a CMBI  
Radboud University  
PO Box 9010  
6500 GL Nijmegen

## Table of Contents

	Page
Welcome	4
Scientific programme	5
How to come to the CAVE	7
Abstract of Lectures	8
Abstracts of Posters	18
Sponsors	31
Netherlands Working Group BioInformatics	32
Participants	33

**Netherlands Conference on Bioinformatics 2004**  
**"Images of Life"**

Presented by  
the Workgroup BioInformatics

**Topics**  
Genome Annotation  
Networks and Modeling  
Visualization and Animation  
Text Analysis / Data Mining  
Expression Data / Statistics  
DEMO : 3D Visualization in the CAVE

October 7 - 8, 2004  
Hanzehogeschool Groningen, "De Appel"

For online registration and more information on program visit:  
<http://www.cmbi.ru.nl/wjbioinf/symp2004>

Hanzehogeschool Groningen

KNCV  
NVBMB

Rijksuniversiteit Groningen

**Welcome** to the Netherlands Conference on Bioinformatics 2004 **Images of Life**

Dear participant,

Welcome to Groningen! And welcome to the Netherlands Conference on Bioinformatics 2004. On behalf of the organizing committee, we are very pleased to celebrate our first lustrum with you. It all started in 1999 at the CMBI in Nijmegen with the formation of a national committee of bioinformaticians from university and industry. Since then, we have visited Amsterdam, Wageningen and Utrecht for our annual symposium. To celebrate this year, it is the first 2-day meeting showing that Dutch bioinformatics is going strong. This lustrum bioinformatics meeting aims to inform, meet and discuss about your activities and progress in bioinformatics and to establish further collaborations in national and transboundary interdisciplinary networks.

As KNCV BioInformatics symposium, it is the joint effort of the Working Group BioInformatics of the Dutch Society of Biochemistry and Molecular Biology (NVBMB), section of the Dutch Chemical Society (KNCV) and the local Groningen organizing committee in which both the *Rijksuniversiteit Groningen* (University of Groningen) and the Hanzehogeschool Groningen (Hanze University Groningen) participate. We set as our goal to put together an attractive scientific program with emphasis on the topic "Images of Life", images and life in terms of visualization, annotation and data analysis. We are especially proud of the CAVE demonstrations that you will be able to attend. We are grateful to the various sponsors that helped to make this meeting possible.

We are also especially pleased that in this meeting also the Hanzehogeschool (Hanze University for Professional Education) is involved and with it, Dutch HBO bioinformatics. You are here in the conference centre "De Appel" of the Hanzehogeschool. Parallel to this meeting the students of the Hanzehogeschool have organized the first HBO Bin-day for fulltime students in bioinformatics from Groningen, Leiden and Arnhem/Nijmegen. These students will follow our morning program and have their own program in the afternoon. We hope to integrate these students in our bioinformatics activities in the future. Soon these students are looking for study stays and jobs as specialized bioinformatics technicians. That's where you come in again.

At the end of this meeting, you should have experienced the saying 'There goes NOTHING above Groningen'. We urge you to "Image your Life" in Groningen also in the evening between the two meeting days. We trust that you will enjoy this meeting both scientifically and socially.

Best regards from the organizers,

Jan-Peter Nap

&

Frank van Enkevort

Groningen organization committee  
Peter Terpstra  
Jos Roerdink  
Ritsert Jansen

Netherlands Working Group BioInformatics  
Arno Siebes  
Jaap Heringa  
Jack Leunissen  
Ton Rullmann

## Scientific programme

The plenary sessions take place in “De Appel”, Hanzehogeschool. Groningen. The 3D demonstrations on Day 2 take place in “De Zernikeborg”, Rijksuniversiteit Groningen

### Thursday, October 7.

**09.30** Registration

**10.30** Welcome and opening (Jan-Peter Nap, Frank van Enckevort)

**10.40** Session I **Genome Annotation and network modeling** (chair: Roland Siezen)

**10.45** Martijn Huynen (NCMLS / CMBI, Nijmegen, NL)

Reliable protein function prediction from genomics data: application to the proto-mitochondrial proteome

**11.30** Lars Juhl Jensen (EMBL Heidelberg, DE)

STRING: prediction of functional relations, modules, and networks from heterogeneous genome-scale data

**12.00** Alexander Goesmann (CeBiTec, Bielefeld, DE)

BRIDGE: building a bioinformatics software platform for the integration of heterogeneous data from genomic explorations.

**12.30** Poster flashes (odd posters 1 min. each) (chair: Ivo van Stokkum)

**12.45** Poster session odd posters, in combination with lunch

**14.00** Sef Heijnen (TU Delft, NL)

Modeling of metabolic networks

**14.30** Session II **Expression Data I** (chair: Peter Terpstra)

**14.35** Cor Verweij (VUMC, Amsterdam, NL)

Towards personalized medicine: molecular dissection of a rheumatoid arthritis using genomics technology

**15.10** Session III **Visualization I** (chair: Frank van Enckevort)

**15.15** Sasha Panfilov (Dundee, UK / Utrecht University, NL)

Study of reentrant arrhythmias in anatomically based models of the human heart

**16.00** Coffee break

**16.30** Anton Koning (Erasmus MC, Rotterdam, NL)

Visualization and virtual reality for genomics

**17.00** Jos Roerdink (IWI, University of Groningen, NL)

Visualization of large data sets and biological networks in interactive environments

**17.30** Poster flashes (even posters 1 min. each) (chair: Ivo van Stokkum)

**17.45** Poster session even posters in combination with 1<sup>st</sup> lustrum drinks

**19.00** Closure

**Friday, October 8**

Plenary sessions in “De Appel”, Hanzehogeschool Groningen

Parallel 3D demonstrations in “De Zernikeborg”, Rijksuniversiteit Groningen (see below)

**09.00** Registration

**10.00** Opening (Frank van Enckevort)

**10.10** Session IV **Data Mining** (chair: Jaap Heringa)

**10.15** Christian Blaschke (BioAlma & CNB, Madrid, ES)

Text mining in biology

**11.00** Erik van Mulligen (Collexis & EUR, NL)

Associative conceptual spaces

**11.30** Daniela Wieser (EBI / EMBL, Hinxton, UK)

Data mining in large scientific databases: predicting protein annotation and filtering erroneous protein annotation

**12.00** Poster session (all posters) in combination with lunch

**13.10** Session V **Expression Data and Statistics II** (chair: Jan-Peter Nap)

**13.15** Ritsert Jansen (GBIC, Groningen, NL)

Expression arrays in genetic analyses

**13.45** Chris Evelo (BiGCaT and NUGO Maastricht, NL)

Integration of expert counseling and text-mining in microarray analysis through improved pathway maps

**14.15** Session VI **Visualization II** (chair: Jan-Peter Nap)

**14.15** Wim de Leeuw (CWI, Amsterdam, NL)

Interactive analysis of confocal image data

**15.00** Closure of the Symposium (Arno Siebes)

**15.10** Coffee and Farewell

Parallel to plenary sessions (chair: Jos Roerdink) in the “Zernikeborg”

Small groups will be put together based on registration and demand

**10.10** Cave demonstrations

**11.10** Cave demonstrations

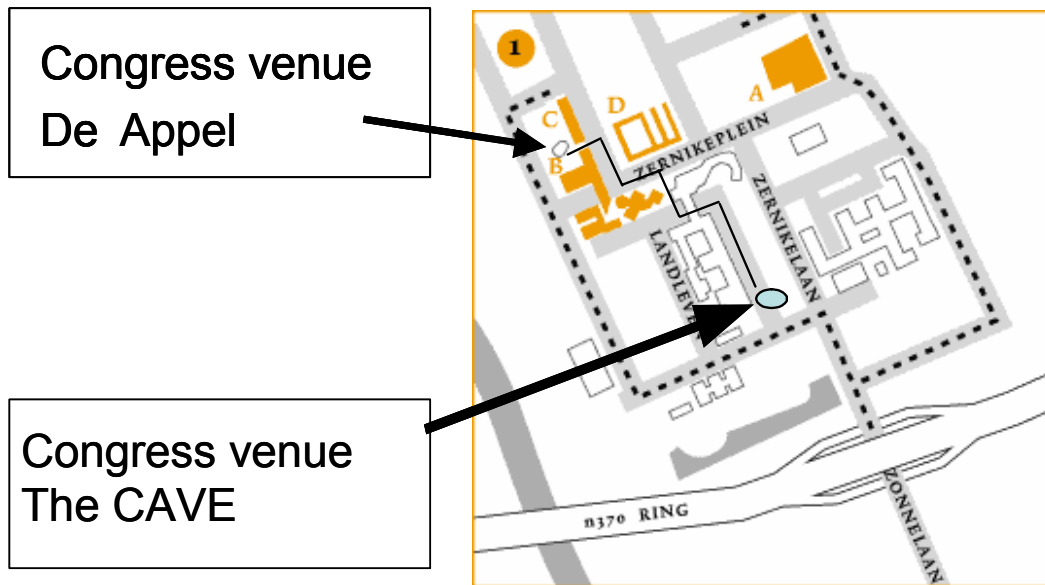
**12.10** Cave demonstrations

**14.00** Cave demonstrations

**15.20** Cave demonstrations (optional)

## How to come to the CAVE

The CAVE is the venue for the visualisation demonstrations The CAVE is at about 5-10 min walking distance from Conference Centre “De Appel”.



Werkgroep



## **Abstracts of Lectures**

BioInformatica

## (L1) Comparative genomics for reliable protein function prediction

M. A. Huynen

Nijmegen Center for Molecular Life Sciences (NCMLS), Center for Molecular and Biomolecular Informatics (CMBI) & University Medical Center, Radboud University, Nijmegen.

The accumulating wealth of genomes and other types of genomics data gives us the opportunity both to predict function as well as to trace its evolution at a genomic scale. On the one hand we can reconstruct the evolution of complete genomes and pathways, on the other hand we can predict new pathways and the functions of the proteins within them based on their correlated behavior in genomics data. I will show both quantitative as well as qualitative analyses of either aspect and will thereby focus on the (proto)mitochondrial proteome.

## (L2) STRING – Prediction of functional relations, modules, and networks from heterogeneous genome-scale data

L.J. Jensen

EMBL, Heidelberg, Germany

STRING is a web resource (<http://string.embl.de>) that contains predicted functional associations between proteins from 100+ fully sequenced organisms. Several different genomic context methods are integrated with analysis of microarray expression data, other large-scale experimental data, and PubMed abstracts to infer relations, which are subsequently transferred across species. The resulting association network allows us to place many uncharacterized proteins into a functional context, to predict functional modules, and to provide a network overview of large biological systems such as the mitochondrion or the mitotic cell cycle. In my presentation, I will also briefly discuss novel genomic context methods to be included in future versions STRING.

## (L3) BRIDGE - Building a bioinformatics software platform for the integration of heterogeneous data from genomic explorations

A. Goesmann

CeBiTec, University of Bielefeld, Germany

The flood of data acquired from the increasing number of publicly available prokaryote genomes has led to new demands for bioinformatics software. With the growing amount of information resulting from high throughput experiments new questions arise that often focus on the comparison of genes, genomes and their expression profiles. Inferring new knowledge by combining different kinds of "post genomics" data obviously necessitates the development of new approaches that allow the integration of variable data sources into a flexible framework. Based on the GenDB-2 genome annotation system we are currently working on the integration of transcriptomic and proteomic data into a platform for systems biology. We have developed several specialized components (GenDB for genome annotation, EMMA for transcriptome data analysis, ProDB for proteome analysis and GOPArc as a Gene Ontology and pathway architecture) and established a project management system for combining different data sources. In addition to an application programmer's interface, all components can be used via a web front-end and/or a graphical user interface. I will describe our concept for the integration of heterogeneous data and focus on recent developments of the GenDB genome annotation system.

## (L4) Modeling of metabolic networks

J.J. (Sef) Heijnen

Technical University, Delft, The Netherlands

In recent years significant experimental and theoretical progress has been achieved for transcriptomics and proteomics, but much less for metabolomics. However metabolomics is the essential link from genes and proteins following genetic and/or environmental perturbations towards cellular (dys)functioning aimed at high microbial productivity, metabolic disorders and balanced nutrition. Recently several innovations for metabolomics have been achieved. (1) Mathematical modeling of metabolomic measurements for in-vivo kinetics using lin-log kinetics. This non-linear kinetic formulation can accurately handle large steady state and dynamic perturbations in metabolite, flux and enzyme levels, allows analytical solutions of metabolic reaction networks and the kinetic parameters are easily obtained by linear regression. Examples will be shown for glycolysis and Penicillin pathway. (2) LC-MSMS based high performance analysis of intracellular metabolites which includes the novel use of labeled whole cell extracts as internal standards. (3)  $^{13}\text{C}$ -tracer studies to unravel metabolic pathways using direct LC-MSMS based analysis of mass isotopomers of intracellular metabolites. (4) The development of the Bioscope, a novel microreactor (2ml vol.), which is used to obtain high quality metabolome transients during metabolic perturbations of cellular systems in a short (0-100 sec.) time frame to enable the formulation of in-vivo kinetics of cellular systems. (5) A novel strategy for functional genomics in silent phenotypes has been developed and tested *in silico* using lin-log kinetic modeling and metabolome data. These tools contribute significantly to the use of metabolomic information in the context of System Biology.

## (L5) Towards personalized medicine: molecular dissection of a rheumatoid arthritis using genomics technology

C. Verweij

VUMC, Amsterdam.

Rheumatoid Arthritis (RA) is a complex chronic inflammatory disease, which is recognized as a heterogeneous disease based on variability in clinical presentation and responsiveness to treatment between patients. Unfortunately, good criteria to classify the different forms of RA are not yet available, which poses major problems in prognosis, drug development and in predicting responsiveness to a given drug. We have successfully applied DNA-microarray technology for large-scale gene expression profiling in the synovial tissues and fibroblast-like synoviocytes (FLS) of inflamed joints, and peripheral blood cells of patients with RA. These studies revealed significant heterogeneity in gene expression profiles between patients with RA that may define a molecular basis for the well-recognized but as yet poorly understood heterogeneity in RA. The observed differences in expression profiles provide opportunities to stratify patients based on molecular criteria for clinical studies and evaluation of targeted therapies, and increase our insight in the biological pathways that are concomitant with either disease subtype.

## (L6) Study of reentrant arrhythmias in anatomically-based models of the human heart

A.V. Panfilov\* and K. H. W. J. ten Tusscher

Department of Theoretical Biology, Padualaan 8, Utrecht University, Utrecht, 3584 CH, The Netherlands \*Division of Mathematics, University of Dundee, Dundee, United Kingdom

We report on the development of models for the human ventricles. The model is based on anatomical data on the geometry and fiber orientation in the human heart which was gathered and computerized by Hren (1996). It uses the monodomain description of cardiac tissue and employs the 'weighting functions' numerical approach, which allows integration of the equations in domains of complex shape with explicit numerical integration schemes. The main framework of the model can be coupled with any type of description of cardiac cells. We discuss different types of such description: low dimensional models, reduced, gamma ionic models and the full ionic models for human ventricular cells. We discuss in details the ten Tusscher et al. (2003) model which is based on recent experimental data on the major ionic currents for human ventricular cells: the fast sodium, L-type calcium, transient outward, rapid and slow delayed rectifier, and inward rectifier current. The model includes intracellular sodium, potassium and calcium dynamics, allowing for the realistic modeling of calcium transients, frequency dependence of the intracellular sodium concentration, and the positive contraction staircase typical for human ventricular myocardium. Using this model we study normal and abnormal wave propagation. We show that for normal parameter values the reentrant arrhythmia (scroll wave) persists in the course of time and does not breakup into fibrillation. The dynamics of the

scroll wave is meandering, however the meandering of the scroll wave rotating in the right ventricle is less pronounced than that in the left ventricle of the heart. By modifying the properties of the cardiac tissue, we were able to find parameters for which the scroll wave becomes unstable and breaks down into a multiple wavelet pattern that is considered to be a model of ventricular fibrillation. We compare the results of these simulations with the recent clinical data by Nanthakumar et al.(2004), who performed epicardial recordings of excitation patterns during fibrillation in the human heart using a 4x5cm electrode plaque. First we show that the ECG pattern occurring in our model is similar to that during ventricular fibrillation and that its dominant frequency is close to that recorded by Nanthakumar et al.(2004) and others. We also compare the wavelet dynamics occurring in the 4x5cm left ventricular epicardial patch in our model to those observed by Nantakumar et al. We find that surface reentry can be observed during approximately 10% of the time, and find large activation fronts following repeatable pathways, smaller fractionated wave fronts and epicardial breakthrough waves, similar to the observations by Nanthakumar et al. Finally, we study the role of reentry in the three-dimensional organization of fibrillation in the ventricles by counting the number of filaments, the organizing centers of scroll waves and hence a measure for the number of reentrant sources present. We determine the number of filaments as a function of time, compute their individual and total lengths and compare them with previous simulations using a canine ventricular geometry. The bottom row of the figure shows an typical activation pattern during fibrillation (C) and the scroll wave filaments present at that same moment (D). Our study supports the results by Nanthakumar et al.(2004) that ventricular fibrillation in the human heart may be organized by only a small number of sources.

References: (1) Hren, R. 1996 A realistic model of the human ventricular myocardium: application to the study of ectopic activation, PhD. thesis, Dalhousie University, Halifax, Canada (1996); (2) Ten Tusscher K.H, Noble D, Noble PJ, Panfilov AV., A model for human ventricular tissue. *Am.J.Physiol.*,v.286,p.H1573-1589 (2004); (3) Nanthakumar, K., Walcott, G. P., Melnick, S., Rogers, J. M., Kay, M. W., Smith, W. M., Ideker, R. E., Holman, W., Epicardial organization of human ventricular fibrillation, *Heart Rhythm*,v.1,14-23 (2004)

## (L7) Visualization and virtual reality for genomics.

A. Koning

Erasmus Medical Center, Rotterdam.

The huge amount of data, both experimental and annotational, that available to researchers in the various "-omics" fields today, calls for advanced visualization techniques to be able to gather knowledge from it in an efficient manner. However, even these techniques, which often reduce the amount of data visible at a given moment, are limited by the finite space of a computer screen. Using the infinite space of a virtual world, a virtual reality system like the CAVE™, offers new possibilities to present, combine and investigate large datasets.

## (L8) Visualization of large data sets and biological networks in interactive environments

J. Roerdink

Institute for Mathematics and Computing Science, University of Groningen

Visualization is becoming ever more important in the area of genomics, as it generically contributes to the interpretation of data which are of high dimension and/or large size. Well known are two-dimensional dot plots, sequence logos, DNA atlases, or metabolic pathway maps. In addition, techniques from graph visualization are increasingly applied to represent, retrieve, display, and explore biological networks, such as phylogenetic trees, metabolic pathways or regulatory genetic networks, either as traditional two-dimensional images or in three dimensions, using interactive displays and virtual environments. Emphasis is put on interactive manipulation of visualized structures by providing users with tools to search, reorganize, control the level of detail (pan and zoom), interrogate, and derive new useful information. I will discuss some general requirements for such systems, briefly mention a number of systems which have recently appeared, and emphasize the need for evaluation studies to assess user performance and suitability of proposed techniques (e.g. in which cases a three-dimensional representation should be chosen instead of a two-dimensional one). Almost invariably current systems are limited in terms of interactivity, adaptiveness of views, possibilities for collaborative work, knowledge representation, allowable model dynamics, and literature coverage. Other important issues are the need for intuitive and unified user interfaces for presenting the visualizations to (remote) users, the use of open software standards to enhance portability, and the potential for dissemination, e.g. by developing tools as internet based applications. I also will briefly look at plans to extend the current tools to techniques for simultaneous visualization of many cellular processes (metabolic pathways, signaling and information pathways, gene regulatory networks, transport routes, replication machinery, protein dynamics) as modules which can be combined in whole "virtual cell" visualizations. Visualization tools should be developed in parallel with modeling and simulation tools, which can complement experimental data and allow the analysis of network behavior under controllable conditions. This development should be carried out in close cooperation with experimental groups. Some recent approaches to the mathematical characterization of biological networks will be discussed, which have wide applicability, ranging from gene regulation to brain connectivity networks.

## (L9) Text mining in biology

C. Blaschke

BioAlma & CNB, Campus Universidad Autonoma, Madrid, Spain

Data access is, in general, an easy task because data is produced in a defined process and stored in a structured way that allows the application of automatic analysis methods like standard statistical packages. Information is much less structured and contained mostly in text documents written by humans to communicate to other humans. Human language is, compared to formal computer languages, highly flexible and ambiguous and its interpretation is domain specific, i.e. the same word can have different meanings depending on where it is used. This makes text documents extremely difficult to interpret by computers because they lack the world knowledge that humans use to understand the content of a document. Therefore, the development of systems for the automatic processing and analysis of scientific literature is absolutely needed, saving valuable time and resources and providing a way for not missing important data. Some advances have been made over the years and the first useful information extraction systems specifically applied to the field of biomedicine and pharmaceutical research are becoming available. The presentation will focus on

- \* how the very specialized language applied in this field and the extremely rich and complex terminology that on top of that changes continually have to be treated and processed to create useful text analysis systems (the problem of "bio-entity" detection in text)
- \* how searches for the "right" document can be shortened
- \* how the information network extracted from the literature can be visualized and explored to make higher order patterns accessible that at the level of simple documents are not evident.
- \* A report in the BioCreative evaluation of text mining systems will be given.

## (L10) Associative conceptual spaces

E. van Mulligen

Institute of Medical Informatics, Erasmus Medical Center, Rotterdam and Collexis BV, Geldermalsen

The approach of our group to assist in managing the data and information deluge is based on the premises that the flood of information will lead to two major trends in scientific discovery: 'From Reading to consulting, and From Reading to meta-analysis'. We therefore concentrate on tools and techniques to assist both these processes. A scientist venturing into a new area during the discovery process should be able to find experts to consult as an integral part of the workflow and at any moment, meta-analysis of all relevant information relevant to the research question under study should be an option. The guiding principle in the research is, that texts and other data sources should be pre-mined for unambiguous concepts, rather than for words or terms. In addition, the semantic relationships between concepts should be used to build semantic knowledge representations of information resources. To assist this scientific process, the group

aims at developing and validating advanced techniques for the processing and analysis of large, complex, and heterogeneous medical and biological data sets, and attempts to link information to people (experts) The good news: Information is not a scarce commodity anymore. Sharing the wealth of data sources world-wide is therefore rapidly becoming the norm. The Biosemantics Group, in close collaboration with its many partners, will develop highly innovative tools for massive concept mining, enrichment of thesauri and ontologies and meta-analysis of large amounts of distributed resources. The philosophy behind our approach is that fundamentally different technologies are needed when handling established knowledge versus new or tacit knowledge. Natural Language Processing (NLP) tools to mine concepts and semantics from texts have their strengths, but they are relatively slow. Complex, domain-specific knowledge is difficult to feed into the rules of these systems. A purely statistical approach can also recognize frequent combinations of words and symbols as potential concepts, but has as a major advantage that it needs multiple documents to make its statistical recognition possible and that it is not human-validated. Thesaurus based technology, as exploited by one of our main research partners Collexis, is extremely fast, but will by definition only find established concepts and concept combinations. The only way out of the "catch 22" situation of the massive knowledge production on the one hand and the inability to meaningfully interpret and combine data on the other hand, is the intelligent combination of all available text mining and meta-analysis technologies where they fit best. A special emphasis of our recent research was on solving ambiguity problems (mainly synonymy and homonymy) during information mining in large text corpora. This is obviously not a goal in itself and the final aim is to generate validated material for massive meta-analysis of information resources. The first route for this development has been the associative concept space (ACS) approach for the meta-analysis of large numbers of papers. Scientists have been meta-analyzing diverse literature resources for many decades and have come up with new insights through intelligent combination of concepts and their interrelationships in the whole set of papers under study. Computational tools that assist the researcher in this combination process have already resulted in testable (and proven) hypotheses. ACS and other technologies will be further developed in the consortium to allow for massive meta-analysis of hundreds of thousands of database records at a time, after prior disambiguation. The approach is expected to cause a quantum leap in our ability to handle and mine massive amounts of information.

References: (1) Weeber M, Schijvenaars BJA, Van Mulligen, EM, Mons, B, Jelier R, Van der Eijk CC, Kors, JA. Ambiguity of human gene symbols in LocusLink and MEDLINE: Creating an inventory and a disambiguation test collection. Accepted for publication, AMIA 2003; (2) C.C. van der Eijk, E.M. van Mulligen, J.A. Kors, B. Mons, J. van den Berg, "Constructing an associative concept space for literature-based discovery", in *Journal of the American Society for Information Science and Technology*, Submitted to JASIST

## (L11) Data mining in large scientific databases: predicting protein annotation and filtering erroneous protein annotation

D. Wieser

EBI / EMBL Outstation, Hinxton, United Kingdom.

High-throughput protein sequencing projects in recent years provide the scientific world with a wealth of decoded protein sequences. A great deal of these sequences is stored in the UniProt knowledgebase, which unifies the former Swiss-Prot, TrEMBL and PIR-PSD protein databases into a single resource. With the Swiss-Prot section of UniProt users obtain a manually curated dataset of high qualitative value. Unlike the Swiss-Prot proteins, the 1 500 000 protein sequences in the TrEMBL section of UniProt are often functionally uncharacterized, as human curators can not keep up with the speed in which newly sequences are submitted. Several automated annotation systems were developed at the EBI to provide information about these TrEMBL proteins until they undergo a literature curation process by a human expert before it is integrated into Swiss-Prot. All systems learn annotation rules from the well characterized Swiss-Prot proteins, based on their sequence patterns and assign annotation to the proteins in TrEMBL after applying these rules. One of the systems, Spearmin, suggests about 4 million annotation entities, i.e. keywords, protein names and description lines, for the TrEMBL proteins with reliability between 90 - 95%. Even though Spearmin is an elaborate system, its accuracy is limited, as it uses a machine learning approach. For efficiency reasons Spearmin occasionally has to ignore important information in the training-set. A second automated annotation system, Xanthippe, was developed to avoid most of the questionable predictions. It post-processes the annotation of the automatic annotation system Spearmin and marks it, if necessary, as possibly erroneous. Xanthippe proved to be a successful approach that currently is able to detect about 50 % of the prediction errors. In this talk, the Spearmin and Xanthippe system will be introduced. It will be explained in detail how the combination of the two annotation systems can help to improve both quality and quantity of large-scale protein annotation in the TrEMBL section of UniProt.

## (L12) Expression arrays in genetic analysis

R.C. Jansen

GBiC, University of Groningen

The recent successes of genome-wide expression profiling in biology tend to overlook the power of genetics. I here propose a merger of genomics and genetics into 'genetical genomics'. This involves expression profiling and marker-based fingerprinting of each individual of a segregating population, and exploits all the statistical tools used in the analysis of quantitative trait loci. Genetical genomics will combine the power of two different worlds in a way that is likely to become instrumental in the further unraveling of metabolic, regulatory and developmental pathways. Results of the first mouse and yeast studies will be presented.

### (L13) Integration of expert counseling and text-mining in microarray analysis through improved pathway maps.

Ch. Evelo

BiGCaT Bioinformatics, University of Maastricht

Microarray analysis can substantially be improved by analysis and visualization of expression results at the pathway level. Several pathway mapping tools have been developed, of which GenMapp is the best known. Such mapping tools typically take the expression changes for the annotated array reporter sequences and look for the number of reporters for which corresponding proteins are present in a pathway map. Such a pathway map can consist of a textbook pathway with a biologically logical layout. In such a map proteins will typically be present as nodes and interactions as edges. A map can also consist of nothing more than a table of proteins that are annotated to share a common property. In this way mapping tools can for instance be used to analyze expression with respect to levels in the gene ontology (GO) database. There are several practical limitations to the use of pathway tools. On one hand array reporters need to be correctly and uniquely annotated to existing proteins, and the protein ID used needs to be available as backpage information in the map. On the other hand pathways need to be logically organized and to be as complete as possible. Bioinformaticians and nutrigenomics researchers within the European network of excellence NUGO realized that current metabolite maps are often incomplete and need to be improved. For this purpose we developed a process workflow in which first available information from GenMapp, GO and Reactome will be combined. Next a selected group of NUGO experts (including those that participate in the IOP gut health) will be asked to comment on the general layout of the pathways and to identify crucial entities (proteins, reactions, metabolites) on the maps. These crucial entities will then be used as starting points for text mining in Pubmed. For this purpose we wish to collaborate with the Biorange textmining initiative. The available data from databases and literature will then be combined and expert input will be asked in two more rounds. The resulting adapted maps will be entered in Biopax format (we will likely create GenMapp to Biopax converters and vice versa and maybe a simple Biopax editor) and contributed to Genmapp and Reactome.

### (L14) Interactive analysis of confocal image data.

W. de Leeuw

CWI, Amsterdam

Using confocal microscopy detailed three dimensional scans of fluorescently labeled specimen can be made. Analysis of such data presents a number of challenges due to characteristics of the imaging process and the nature of biological data. The presentation will discuss these challenges and requirements for a system for the quantitative analysis of confocal data. Furthermore, Argos, a system to fulfill these requirements, is presented.

Werkgroep



## **Abstracts of Posters**

BioInformatica

## (P1) Combining experimental data and *in silico* analysis to model the metabolic network of *Lactobacillus plantarum*

B Teusink<sup>1,2,3</sup>, FHJ van Enckevort<sup>2,3</sup>, A Wegkamp<sup>1,2</sup>, D Molenaar<sup>1,2</sup>, J Hugenholtz<sup>1,2</sup>, EJ Smid<sup>1,2</sup> and RJ Siezen<sup>1,2,3</sup>

1. Wageningen Centre for Food Sciences, Wageningen; 2. NIZO Food Research, Ede; 3. CMBI, Radboud University, Nijmegen, The Netherlands

The complete genome of *Lactobacillus plantarum* WCFS1 has recently been sequenced (PNAS USA 2003;100:1990). *Lactobacillus plantarum* is a versatile lactic acid bacterium that is important in many food and feed fermentation processes. Putative biological functions could be assigned to 2,120 (70%) of the 3,052 predicted protein-encoding genes. After prediction of gene function, focus is on the development and improvement of methods and tools to go from genome sequence to gene annotation, to pathway reconstruction and to prediction of phenotype. Important aspects are how and where to incorporate and use experimental (omics) data, and how and to what extent parts of the process can be automated. We have set up different bioinformatics tools and experimental techniques. LacplantCyc was generated from the annotated genome and the MetaCyc database (Karp, Bioinformatics 2002 Suppl 1:S225-32). We also use Simpheny (Genomatica Inc.) for genome scale modeling. Our current network comprises 710 genes (23% of the genome), and a similar number of reactions. The properties of the metabolic network are being investigated within the framework of constraint-based modeling and elementary flux mode analysis. Transcriptome data can be projected on metabolic maps through LacplantCyc and Simpheny, and on genome-maps through the Microbial Genome Viewer, an in-house developed application (Kerkhoven, Bioinformatics 2004). Results of the *in silico* analysis are being compared with experimental data. As an example, biosynthetic capacities of amino acids and vitamins have been compared with their requirement in growth medium. Six out of 28 initial comparisons were inconsistent. This analysis has identified missing enzymes and led to new hypotheses about metabolic pathways.

## (P2) Effect of amplification on gene expression profiles

Rachel I.M. van Haaften, Blanche Schroen, Ben J.A. Janssen, Jos F.M. Smits, Yigal M. Pinto and Chris Evelo  
University of Maastricht

Gene expression microarray technology permits the analysis of global gene expression profiles. However, the range of biological questions to be answered is limited due to the relative large amount of sample needed for a microarray experiment. This amount of sample needed limits the use of small excision biopsies and /or needle biopsies from human or animal tissues. Linear amplification techniques have been developed which can be used to increase the amount of sample derived cDNA. These amplified samples can be hybridised on microarrays. We compared the results of Affymetrix microarray experiments with samples derived from non-amplified rat heart tissue and amplified cardiac biopsies from the same hearts. The biopsies were linearly amplified with T7 polymerase method to obtain enough sample for a microarray experiment. Both kinds of samples were hybridized to the 'Rat Expression Set 230 Array' of Affymetrix. We hereby saw that some specific genes are lost by amplification, and attributed these losses to specific gene characteristics.

### (P3) Combining methods for ortholog prediction

Robert Kerkhoven(1), Richard Notebaart(1), Frank van Enckevort(1,2), Berend Snel(1) and Roland Siezen(1,2,3)

(1). CMBI, Nijmegen; (2). NIZO Food Research, Ede; (3). WCFS, Wageningen.

Discrimination between close alternative homologs in the prediction of orthologous relations is a well-known bioinformatics problem. Here two methods for defining orthologous relations between two genomes were combined, and corrected or confirmed by taking gene context into account. Ortholog relations, in the genomes of *Bacillus subtilis* and *Bacillus cereus*, were predicted with the program INPARANOID and the COG annotation method. In case defined orthologous groups contain many-to-many relations, gene context was used in addition to similarity scores to define the better ortholog relation. It turns out that most of the many-to-many relations found by the COG-method are, as expected, resolved by INPARANOID. In cases where sequence similarity is not discriminating, gene context can be resolving or correcting. The main advantage of the COG method is the detection of 1293 additional homologs which are missed by INPARANOID or difficult to detect with conventional alignment-based methods. A considerable amount of additional orthologs is expected to be derived from this category. In the comparison between *B.subtilis* and *B.cereus* it is shown that genes with a higher sequence similarity are overall more likely to share gene context. Interestingly, this is not the case with genes from highly conserved operons. These occur with equal probability between a 50-100% sequence-identity range. From the INPARANOID analysis it is shown that out of 104 of the one-two orthologous relations (one *B. subtilis* gene has two co-orthologs in *B. cereus*, or vice versa) 65 have completely different gene context. In 10 cases gene context is conserved. In 9 of the 29 remaining situations the highest sequence similarity does not agree with conserved gene context. It can be concluded that gene-context is an additional clue for the identification of orthologs.

### (P4) Wageningen *Arabidopsis thaliana* database (WAtdb): a novel resource for plant genome research

Paulien Adamse and Jan-Peter Nap

Applied Bioinformatics, Plant Research International, Wageningen

As part of a collaborative network of European Plant Databases that PlaNet (see [www.eu-plant-genome.net](http://www.eu-plant-genome.net)) is going to provide to the research community, we are designing and building the WAtDB database. This relational database offers information about mutants, transgenics and natural variants of *Arabidopsis thaliana*. WAtDB includes data about transposon insertion mutants (activation tags and insertion tags; insertion site), profiling (microarray, proteomics and metabolomics), detailed phenotypic descriptions and pictures. Existing guidelines and/or standards to describe experiments, such as MIAME for micro arrays, and MIAPE for proteomics are or will be implemented. Proper ontologies to describe plant phenotypes (Plant Ontology; PO) in sufficient detail are not yet available. Therefore, the development of WAtDB is interlinked with the development of the PO by the international Plant Ontology Consortium. In addition, WAtDB supplies data storage and exploration facilities to improve the exploration and annotation of the *Arabidopsis* genome. WAtDB can also be implemented for other plant species, for example rice. WAtDB will be fully integrated with all other PlaNet databases and can be accessed over the Internet ([www.watdb.nl](http://www.watdb.nl)) and via the BioMoby service of PlaNet. Other partners in PlaNet are MIPS (Neuherberg, Germany), VIB (Gent, Belgium), Genoplante-Info (Evry, France), NASC (Nottingham, Great Britain), John Innes Centre (Norwich, Great Britain) and CNB/CSIC (Madrid, Spain). PlaNet is supported by the European Commission, grant QLRI-CT-2001-00006

### (P5) BioGRID: integrating \*omics data

Morris A. Swertz, E.O. (Bert) de Brock, and Ritsert C. Jansen  
Groningen Bioinformatics Centre (GBiC), University of Groningen, Haren, The Netherlands

Bridging the available information areas in the life sciences domain remains a difficult task with a promising reward. The difficulty lies in semantic and technological diversity among available information sources such as for genomics and proteomics. The ambition of the BioGRID system is to conduct a trial for the introduction of a so-called 'GRID' approach in the biotechnology industry: transparent plug-in integration of data and processing sources. BioGRID integrates three technologies: gene expression statistics in SpaceExplorer, protein interaction maps in PSIMAP, and literature classification in Classification Server. This enables the researcher (i) to prepare gene expression data with methods such as clustering, (ii) compare the expression data with known protein interaction data both simulated (based on 3D structure) and experimental (Y2H) and (iii) to find related documents from sources such as PubMed classified into an ontology. The areas of gene expression and protein interaction are bridged using structural classification of proteins (SCOP). The area of literature is linked via the GeneOntology (GoPubMed). The BioGRID system architecture contains three enabling technologies: (i) a common data model on which the external data is projected such as genes, expressions, proteins, etc.; (ii) a repository of so-called 'access methods' that allows easy management of the connections to external systems such as to query a database or to run statistics; and (iii) a distributed agent framework to allow for parallelisation of computation intensive tasks over multiple worker PCs.

Contact: m.a.swertz@cs.rug.nl

### (P6) MOLGENIS information infrastructure for life science experiments

Morris A. Swertz, E.O. (Bert) de Brock, Alrik L. Lubbers and Ritsert C. Jansen  
Groningen Bioinformatics Centre (GBiC), University of Groningen, Haren, The Netherlands

Experimental genomic research laboratories, such as those working with microarrays, need an adequate information infrastructure to support the management of their data production and research workflow. But what criteria make such an experiment information infrastructure adequate? Experimental genomics databases typically have the following requirements: (i) fast delivery and evolution of the system to keep the system up with the fast developing genomics field; (ii) a custom information structure to support laboratory specific methodologies and research topics; (iii) clear screen organization to survive the flood of data; and (iv) limited total costs of ownership. We have implemented the Molecular Genetics Information System (MOLGENIS) that meets all these requirements by the use of (i) a mass customization software production line that takes (ii) a customized information structure based on community standards as input and results in (iii) a low-maintenance web application (and can be repeated on every change taking only seconds). To strengthen immunity to change in external tools and software (iv) data files are stored as black boxes (instead of decomposing them in database tables) and (v) by loosely linking to other programs (instead of laboriously integrating them). A first version for a group working on microarrays for bacteria was implemented in a period of three man-months, later versions for groups working in animals and medicine only took in the order of man-weeks. We are extending this approach for groups working with proteomics methods like mass spectrometry but also organizations working in clinical trials! MOLGENIS can provide valuable insights and lessons to both software developers and a user community embarking on large-scale genomic projects.

References: MOLGENIS homepage: <http://www.molgenis.nl>; Swertz, M.A. et al. (2004) *Bioinformatics* 20: 2075-2083. Contact: m.a.swertz@cs.rug.nl

### (P7) Chromosomal co-expression domains in the genome of *Arabidopsis thaliana*

Xin-Ying Ren, Mark W.E.J. Fiers, Williem J. Stiekema and Jan-Peter Nap  
Applied Bioinformatics, Plant Research International, Wageningen

Plant growth and development is influenced by higher-order chromatin configurations. We aim to dissect a plant's whole genome organization and to define its functional characteristics. Chromosomal domains are thought to be the basic unit of genome organization. With the help of both sequence and expression data, such domains are being identified and analyzed. Chromosomal co expression domains are here defined as genomic regions where neighboring genes have high correlated expression under different conditions. Using all Arabidopsis MPSS expression data and the MIPS genome annotation, we have identified ~5% of the gene pairs that co express with a high correlation coefficient ( $R > 0.7$ ). This is significantly higher than expected on the basis of random gene combinations. Shared promoters (divergent pairs) did not explain the high number of co expressed genes. The correlated gene pairs are distributed evenly over the whole genome. Co expression appears independent of the distance between the neighboring gene pairs. A small but significant set of gene pairs tend to co express when using all expression data. Their genome locations will be analyzed in more detail. Biological subdivision of expression data (in different tissues or under different experimental conditions) will give additional insight in the existence and importance of co-expression domains. The sequence characteristics of such domains will be analyzed and compared to known boundary elements.

### (P8) Motifcontroller: an automated motif-discovery application

Evert-Jan Blom , Dinne W.J. Bosman , Sacha A.F.T. van Hijum , Patrick J. Ogao , Jos B.T.M. Roerdink and Oscar P. Kuipers  
Department of Genetics and Department of Mathematics and Computing Science, University of Groningen, Groningen.

Motifcontroller is an automated motif discovery and retrieval system. Using a distributed approach multiple motif discoveries are processed and it can detect common motifs in sets of genes. The search space of motifcontroller can be extended by importing known motif -information. Affinity values between clusters can be inferred, hence allowing for benchmarking of different clustering techniques. Our application serves a tripartite purpose

- (i) capable of combining predicted information with existing motif-information,
- (ii) using external tools for the discovery of motifs using a distributed approach, and lastly,
- (iii) providing extensive search capabilities to find common motifs in clusters of genes (e.g. clustering data from time series data).

### (P9) A bioinformatics framework for modeling, visualization and simulation of gene regulatory networks

Bosman D.W.J. , Blom E.J. , Ogao P.J. , Kuipers O.P. and Roerdink J.B.T.M.

Department of Genetics and Department of Mathematics and Computing Science, University of Groningen, Groningen

Transcriptome analysis using DNA-microarrays based on time-series of *Bacillus subtilis* provides a wealth of information. To make optimal use of this data source and existing information on regulatory pathways, our key objective is to develop an automated gene regulatory identification and visualization system. Network information is stored in an object oriented database. Our system will enable users to gain insight in the organization and behavior of large regulatory gene networks by extensive data mining capabilities.

### (P10) Prediction of conserved regulatory networks in Lactobacillus species

M.W.W. Wels, M. Kleerebezem, W.M. de Vos and R. J. Siezen

WCFS / CMBI, Radboud University, Nijmegen

Lactobacilli play a substantial role in food biotechnology and influence our quality of life by their fermentative and probiotic properties. To date, (partial) genomic sequences of 6 different Lactobacillus species are available via the ERGO bioinformatics suite ([http://ergo.integrated\\_genomics.com/ERGO/](http://ergo.integrated_genomics.com/ERGO/)). Here, we describe the *in silico* analysis of the different regulatory networks encountered in these 6 related species. Orthologous genes between the different lactobacilli were predicted using the protein family clustering tools of the ERGO bioinformatics suite (Overbeek, PNAS USA, 1999;96:2896). In addition, operon structures for each of these genomes were predicted. The operon border predictions were based on three parameters: (1) a change in the direction of transcription, (2) a large intergenic region (>100 bp) or (3) the occurrence of a rho independent termination signal in an intergenic region. By combining orthology and operon predictions, orthologous operons were predicted. This analysis showed that the different lactobacillus genomes share many of these orthologous operons. The MEME module (<http://meme.sdsc.edu>) was used to screen the upstream regions of a set of orthologous operons for conserved motifs (cis-elements). This screening showed that orthologous operons frequently share (partially) conserved motifs that are anticipated to be involved in transcription regulation and might represent regulatory protein (trans-acting elements) binding sites. Validation of the predicted regulatory motifs by comparison with experimental data from literature, confirmed this assumption for several of these motifs. Subsequently, cis-element predictions were used to predict regulons, by grouping of operons that share a specific cis-element, suggesting their concerted regulation (regulon). These predictions revealed that many operons are part of the same regulon in different species. This *in silico* analysis of Lactobacillus regulatory networks will be used to facilitate interpretation of global transcription profiling results. Moreover, by comparative distribution analysis of Lactobacillus regulatory proteins, it might be possible to predict functional cis-trans regulatory modules for these species.

### (P11) Development of a generic annotation pipeline: gP

Mark Fiers, Joost de Groot, Sander Peters and Roeland van Ham  
Centre for Biosystems Genomics and Applied Bioinformatics, Plant Research International,  
Wageningen

In order to annotate the potato and tomato sequences which will be generated in the CBSG sequencing projects we are in the process of developing and implementing an automated annotation pipeline. The software will be able to take an incoming sequence (i.e. a BAC) and fully automatic annotate this sequence on a linux cluster and store the annotation results in a database. The annotation results will then be dynamically visualized using publically available software such as the Generic Genome Browser or Ensembl. Important features of the software will be:

- \* Flexible. It will be easy to add other annotation tools.
- \* Incremental. New data or databases should be incorporated efficiently, without having to rerun the complete pipeline.
- \* Database independent. Use of another database or multiple databases is easy.
- \* Interchangeable components. By using standard file formats (GFF, XML) for communication between the pipeline components, introducing new components will be easy. Also implementing new components is simplified this way.
- \* Easy to use. The software will have a clear GUI and will be operable by a non expert annotator.

### (P12) Genome-wide analysis of gram-positive bacteria: bioinformatics and transcriptome analysis

Richard Baerends, Girbe Buist, Chris den Hengst, Sacha van Hijum, Siger Holsappel, Anne de Jong, Harma Karsens, Naomi Karsens, Rasmus Larsen, Andrzej Lulko, Aldert Zomer, Jan Kok and Oscar Kuipers  
Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB),  
University of Groningen, Haren, the Netherlands

The recent availability of the complete nucleotide sequences of the chromosomes of many micro-organisms, among which important human pathogens such as *Streptococcus pneumoniae*, and industrially important organisms such as *Bacillus subtilis* and *Lactococcus lactis*, should allow reconstruction and understanding of the physiology of each organism. Among the prokaryotes the Gram-positive bacteria are a very prominent group that is characterized by a strong evolutionary relatedness. Some species are of great importance in the industrial agro-food chain. Recently, DNA microarray analysis was implemented in our research. DNA microarrays on glass slides are being constructed in-house that contain amplified DNA fragments of the complete genomes of *B. subtilis*, *L. lactis* IL1403, *L. lactis* MG1363, *S. pneumoniae* (together with P. Hermans, Pediatrics, Rotterdam) and *B. cereus*. Using genome-wide transcriptome analysis we aim to reconstruct the metabolic pathways and gene regulatory networks of the bacteria under study, with a focus on stress response, protein secretion, competence development, sporulation, flavour formation, virulence, health risk assessment and food safety. The poster will present the validation of the transcriptome analysis procedure using *L. lactis* IL1403 DNA microarrays and software and bioinformatics tools developed for analysis of the transcriptome data.

### (P13) PhenoLink

B.T.F. Alako, P.B.T. Neerincx and J.A.M. Leunissen  
Bioinformatics, Wageningen University & Research Centre, Wageningen

Finding genes responsible for a certain genetic trait is a laborious task. Using linkage analysis studies the inheritance of known markers can be compared to the inheritance of a trait of interest. This enables researchers to pinpoint one or several marker-defined regions on the genome where one or more genes involved in the trait must be located. Even though linkage analysis narrows the search, depending on the marker resolution there might still be dozens of genes left that could be involved. Screening all those candidate genes for involvement is a time and money consuming effort. The goal of the PhenoLink project is to develop a data mining framework to reduce the amount of candidate genes and hence reduce the amount of wet-lab experiments necessary to find the genes involved. We will combine information from mapping studies with gene expression information to rank the candidate genes from the most likely to the least likely genes involved. The ranking will be based on the idea that if a particular trait is expressed at a certain location (tissue) or during a certain time frame (developmental stage), the gene(s) involved must be expressed at the same location or during the same time frame. We will use information from model species to complement information from the target organism. More specifically, we will start with the chicken (*Gallus gallus*) and tomato (*Lycopersicon esculentum*) as the target organisms, but in the end it should be possible to use the same framework to rank candidate genes in other organisms as well. PhenoLink predictions will reduce the amount of genes that have to be checked for involvement in a specific trait and hence save time and money.

### (P14) GeNeYouS

Genomics Network for Young Scientists (GeNeYous) P.O. box 6, 6500 AA Nijmegen

GeNeYouS is an initiative by young scientists in genomics research. The term genomics is hereby defined as the separate and integral sciences of genetics, studies on the transcriptome, proteome, metabolome and bioinformatics. GeNeYouS aims to bring together PhD students, postdocs and other young scientists. GeNeYouS offers a platform for information exchange between scientists working in the fast evolving field of genomics. Through this platform, young scientists should fully benefit from existing expertise and communicate easily with other people in genomics research. GeNeYouS is supported by the Netherlands Genomics Initiative and Senter IOP Genomics. See also <http://www.geneyous.nl>

### (P15) Usability evaluation of a genomics virtual reality application

Ogao, P.J., Bosman, D.W.J., Blom, E-J., Kuipers, O.P. and Roerdink, J.B.T.M  
University of Groningen

SARAgene is a Virtual Reality framework for genomic visualization that runs in the CAVE. It is developed by SARA, the Dutch High Performance Computing Center in Amsterdam in cooperation with Johnson & Johnson Pharmaceuticals B.V. One of SARAgene's components is intended to help biologists browse and explore a relational database of experimentally verified protein interactions of mammalian organisms. In this poster we explain how we have employed both heuristic and formative evaluation approaches for usability evaluations of virtual environments as a means of attaining a usable and useful virtual reality application for genomic visualization.

### (P16) PSI-PRALINE: a new algorithm for multiple sequence alignment

Simossis V.A. and Heringa J.  
Free University Amsterdam

We introduce a new multiple sequence alignment strategy PSI-PRALINE. The algorithm is an extension to the profile pre-processing technology implemented earlier in the highly successful multiple sequence alignment method PRALINE (Heringa, 1999; Simossis and Heringa, 2003). The alignment of a set of sequences proceeds as follows: First, the algorithm submits each sequence as a query to PSI-BLAST. Secondly, the resulting PSI-BLAST alignment for each sequence is converted to a pre-profile. This way, each sequence is represented as a pre-profile of related sequences and therefore carries an increased amount of evolutionary information about the position-specific characteristics of the sequence. Thirdly, these pre-profiles are aligned using PRALINE's progressive alignment strategy. In addition, similarly to the original PRALINE pre-profile implementation, a user-defined threshold allows quality control of the sequences used to build the PSI-PRALINE pre-profiles. As a result, the PSI-BLAST alignments are filtered before they are used in the pre-profile strategy. The improvement in MSA as well as the quality control in sequence selection allows for several iterative scenarios for database search optimization as well as optimization of the multiple sequence alignment itself.

### (P17) Analysis of transcriptional profiles based on transcriptional modules

André Boorsma, Anna Zakrzewska, Klaas J. Hellingwerf and Frans M. Klis  
UvA SILS - Microbiology

We have analyzed about 700 publicly available expression profiles, using Quontology, which calculates the contribution to gene expression of a particular gene group to the total gene expression using t-test statistics. Gene groups may be defined in various ways. Here we use groups of genes of which the upstream regions are known to physically interact with a specific transcription factor as indicated by chromatin immune precipitation (ChIP-chip) experiments (Lee et al., *Science* 298, 799-804, 2002). A disadvantage of the latter approach is that the ChIP-chip data cannot distinguish between adjacent, oppositely transcribed genes. To establish which gene of each gene pair is regulated by a particular transcription factor, the degree of correlation between the expression of the individual genes of each gene pair and the expression of the corresponding group was determined over a range of 700 expression experiments. Based on the outcome, individual members were either retained or removed from the group. This computational approach reduced the number of gene pairs by about 50%, and also resulted in more homogeneous groups in terms of gene functions.

**(P18) Simulating DNA microarray expression data**

Casper J. Albers, Ritsert C. Jansen, Jan Kok, Oscar P. Kuipers and Sacha A.F.T. van Hijum  
GBiC, Molecular Genetics, GBB, University of Groningen.

Simulation of DNA microarray data allows a researcher to avail of as many datasets as needed to test e.g. statistical methods for differential expression. In experimentally obtained DNA microarray data the actual (differential) gene expressions are either unknown or only estimated. Simulated DNA microarray data contain differentially expressed genes with expressions known beforehand. The latter makes simulated DNA microarray data a valuable addition to experimentally obtained DNA microarray data for the validation of new analysis methods. We introduce a sophisticated and realistic model implemented in a free to use web-based computer program to simulate large amounts of DNA microarray data.

**(P19) Comparison of the complete genomes of *Lactobacillus plantarum* WCFS1 and *Lactobacillus johnsonii* NCC533 reveals extensive differences in chromosome organization and gene content.**

Jos Boekhorst, Roland J. Siezen, Marie-Camille Zwahlen, David Vilanova, Raymond D. Pridmore, Annick Mercenier, Michiel Kleerebezem, Willem M. de Vos, Harald Brüssow and Frank Desiere  
CMBI, Radboud University, Nijmegen

The first comprehensive comparative analysis of lactobacilli was done by comparing the genomes of *Lactobacillus plantarum* (3.3 Mb) and *Lactobacillus johnsonii* (2.0 Mb). *L. johnsonii* is predominantly found in the gastro-intestinal tract, while *L. plantarum* is also found on plants and plant-derived material, and is used in a variety of industrial fermentations. The *L. plantarum* and *L. johnsonii* chromosomes have only 28 regions with conservation of gene synteny (totaling about 0.75 Mb), which are not colinear, indicating major chromosomal rearrangements. Metabolic reconstruction indicates many differences between *L. johnsonii* and *L. plantarum*: numerous enzymes involved in sugar metabolism and in biosynthesis of amino acids, nucleotides, fatty acids and cofactors are lacking in *L. johnsonii*. Major differences were seen in the number and types of putative extracellular proteins, which are of interest because of their possible role in host-microbe interactions. The differences between *L. plantarum* and *L. johnsonii*, both in genome organization and gene content, are exceptionally large for two bacteria of the same genus, emphasizing the difficulty in taxonomic classification of lactobacilli.

## (P20) Discovering possible co-relations and control-regulations between gene pairs in time series microarray data using salient dynamic features

M. Egmont-Petersen

Institute of Information and Computing Sciences, University of Utrecht

Microarray data obtained from time series experiments may reveal possible interactions between the protein coded by one gene and the subsequent transcription of another gene. We regard the subsequent expression ratios as a time series. The features used to analyze the time series data are the local minima and local maxima of the expression ratios of the mRNA. With our representation, relations may be discovered between the most likely time points at which a gene (eventually the associated protein) is active (local maximum) and inactive (local minimum). Our approach makes it feasible to establish a possible regulatory relation between a transcription factor with small absolute changes in expression ratio, and a target gene. We evaluated the robustness of our approach on the dynamic microarray data obtained from Spellman (1998). K-means clustering was used to find 20 cluster centers in the data. The probabilities of the corresponding local extrema between each pair of mean vectors were computed. Subsequently, noise was added to the cluster centers in order to study the degradation of the matching probability distribution.

## (P21) Modeling mass spectra as mixtures of probability distributions

Martijn Dijkstra and Ritsert C. Jansen

GBiC, University of Groningen

Motivation: Surface Enhanced Laser Desorption and Ionization (SELDI) Time Of Flight (TOF) is an example of Mass Spectrometry (MS) that measures protein/peptide-composition of small biological samples by outputting a mass spectrum. A spectrum contains peaks corresponding to proteins in the sample, but overlapping peaks and spurious peaks hamper assessment of the true protein composition. Results: We propose a new procedure to simultaneously fit a baseline and estimate the detected protein masses and their abundances using mixture models. Major advantages of the new approach are (i) the measured protein's abundance is quantified properly by estimating its area under the peak curve, (ii) the areas under the curve can be estimated even if adjacent peaks overlap considerably, (iii) mass-resolution and intensity estimation of a protein are therefore improved and (iv) the underlying peak positions show that spectra over multiple samples are shifted with a constant. Examples are shown on diabetes data. Contact: [m.dijkstra@cs.rug.nl](mailto:m.dijkstra@cs.rug.nl)

## (P22) Identifying a microarray reporter sequence with a well-described gene

Stan Gaj, Joris Korbeeck, Willem Ligtenberg, Rachel van Haaften, Edwin ter Voert and Chris Evelo

BiGCaT BioInformatics, University of Maastricht

Microarray reporter sequences are usually annotated through automated BLAST searches against nucleotide databases such as GenBank/EMBL or UniGene. Often minimum criteria to validate the high quality of the alignments are not given. As a result hits can point out to something similar to the actual sequence. Also hits can often point to relatively non-informative sequences like EST's or sequences annotated being similar to something else. This leads to two basic problems: the quality of the alignment hit and the quality of the annotation of the hit. The aim of this study was to annotate as much reporter sequences as possible with known genes for which protein products are known. Two different approaches have been developed in Perl: 1) a high quality alignment to a custom built database containing only full coding sequences for a well which a well described protein product is known. 2) a data mining procedure using database crosslinks (both direct and SRS-based) to locate well-described proteins starting from finding high quality alignments towards sequences in UniGene clusters. Both approaches were successful in updating the annotation for both commercial (primarily Incyte, Agilent and in lesser extent Affymetrix) and home-spotted arrays. They also proved to be useful for data derived from suppression hybridization studies.

## (P23) Bio-informatics Education @ Hanze University Groningen

Grietinus Plat, Michiel Noback, Omid Givi, Rineke Daatselaar and Jan-Peter Nap

Expertise Centre Bioinformatics, Institute for Life Science & Technology, Hanze University, Groningen, NL

The Institute for Life Science & Technology of the Hanze University for Professional Education (Hanzehogeschool) Groningen offers a fulltime education bioinformatics for bioinformatics technicians. It was the first in The Netherlands to do so. Spring 2004, the team involved won the Zilveren Zandloper from the Dutch Biotechnological Society for their educational efforts. However, how a bioinformatics curriculum for technicians should look like is still being discussed, both nationally and internationally. Current set-up of the curriculum in Groningen will be presented, mainly aiming at stimulating the discussion with Dutch academic and industrial bioinformatics to define what they expect from a bioinformatics technician in the future, compared to for example a laboratory technician. Notably the views on the balance between biology and computer science are of interest. In addition, bioinformatics education works the best with real-life cases: participants are challenged to come up with data suitable for demonstrating and teaching bioinformatics concepts and applications in an educational context.

## (P24) How to store an expert's brain, and use it to understand omics?

Rachel van Haafte (1), Marjan van Erk (2), Rob Stierum (2), Susanne Sansone (3), Philippe Rocca-Serra (3) and Chris Evelo(1).

(1) BiGCaT Bioinformatics, Maastricht; (2) TNO, Zeist; (3) EBI, Hinxton, UK.

The European network of excellence NuGO ([www.nugo.org](http://www.nugo.org)) tries to improve understanding of omics results through improved pathway mapping. To enable this; expert information will be collected in a workflow that involves several rounds of expert counseling, text mining and curation. Storage of the expert knowledge turned out to be a larger problem than initially anticipated. The current (v2.0b) version of the well known mapping tool GenMAPP has several drawbacks when it comes to understanding pathways with respect to the reaction types involved, the localization of the reactions and the metabolites formed. In order to be able to store all knowledge collected we first proposed to store the pathways in the new BioPAX (1.0) format and to develop converters to and from BioPAX. Working on these converters we realized that the basics behind the two storage formats are very different. GenMAPP uses genes as the central entity and adds localization information in screen coordinates. Reactions are just arrows with x,y coordinates and so are metabolites and other annotations. BioPAX on the other hand is reaction centered. After discussion with several of the larger players in the field we decided on a fundamentally different solution. Pathway data will be entered in the new EBI Reactions/ Pathway database Reactome. Reactome will be further developed to enable the visualization of omics results using new developed views. Those views will be directly exported to GenMapp 2.0b format. Apart from this we are working with the GenMapp group itself to develop converters to and from the new GenMapp Markup Language (GMML) format for GenMapp 2.0, Reactome and BioPAX. Since GMML is actually a superset of BioPAX we will also propose extensions of the BioPAX format. We think that this workflow will enable us to 1) obtain and understand relevant pathway data from experts and literature; 2) comprehensively store all of the collected information in a larger well-curated database (Reactome); 3) to use the currently most important mapping tool with the new data as soon as we have collected the pathway data and 4) convert all this information to standard pathway format (GMML/BioPAX plus) to make it available for new analytical tools.

## Sponsors

The organizers thank the following sponsors for making this Netherlands Conference on BioInformatics 2004 **Images of Life** possible.

- **Collexis BV**  
[www.collexis.com](http://www.collexis.com)  
Geldermalsen, The Netherlands



- **Dalicon BV**  
[www.dalicon.com](http://www.dalicon.com),  
Renkum  
The Netherlands



- **DSM BV**  
[www.dsm.com](http://www.dsm.com)  
Heerlen, The Netherlands



- **Keygene BV**  
[www.keygene.com](http://www.keygene.com)  
Wageningen  
The Netherlands



- **The MathWorks BV**  
[www.mathworks.nl](http://www.mathworks.nl)  
Gouda The Netherlands



- **Oracle Nederland BV**  
[www.oracle.com](http://www.oracle.com)  
De Meern,  
The Netherlands



- **Organon BV**  
[www.organon.nl](http://www.organon.nl)  
Oss,  
The Netherlands



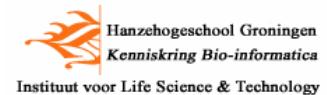
- **Sun Microsystems Nederland BV**  
[nl.sun.com](http://nl.sun.com)  
Amersfoort  
The Netherlands



- **BioASP**  
[www.bioasp.nl](http://www.bioasp.nl)  
Amersfoort,  
The Netherlands



- **Hanzehogeschool Groningen**  
Hanze University, Expertise Centre  
Bioinformatics, Institute for Life Science &  
Technology  
[www.hanze.nl](http://www.hanze.nl)  
[www.bioinf.nl](http://www.bioinf.nl)  
Groningen  
The Netherlands



- **KNCV**  
[www.kncv.nl](http://www.kncv.nl)  
Leidschendam,  
The Netherlands



- **NVBMB**  
[www.kncv.nl/nvbmb/](http://www.kncv.nl/nvbmb/)  
Nijmegen  
The Netherlands



- **Rijksuniversiteit Groningen**  
University of Groningen  
FWN (IWI, GBiC, GBB)  
FMW (GBiC)  
[www.rug.nl](http://www.rug.nl)  
Groningen  
The Netherlands



- **Working group BioInformatics**  
[www.cmbi.ru.nl/wgbioinf/](http://www.cmbi.ru.nl/wgbioinf/)  
Nijmegen,  
The Netherlands



## Netherlands Working Group BioInformatics of the Royal Netherlands Chemical Society (KNCV)

The Working Group BioInformatics (<http://www.cmbi.ru.nl/wgbioinf>) is linked to both the KNCV (<http://www.kncv.nl>) and the NVBMB (<http://www.kncv.nl/nvbmb>).

The goal of this Working Group is to stimulate the bioinformatics research and education in The Netherlands by means of:

- Focus on national strength within the field of bioinformatics
- Set up information programs regarding bioinformatics
- Offer a discussion platform for all information relevant and for all people interested in the field of bioinformatics
- Organize meetings and symposiums
- Improve networks between universities, technical schools, research institutes, and industry
- Improve contacts with international bioinformatics organizations

The annual symposium is a means to achieve these goals. The intended audience is universities, technical schools, research institutes and industry. The Netherlands Working Group BioInformatics wishes to open opportunities in the country in sharing our experience and open new collaborations with experts in this field. Our aim is to create a network of Netherlands and worldwide scientists and technicians working in bioinformatics in the fields of medicine, biology, chemistry, mathematics, physics, informatics, and related fields.

The 2004 board of the Netherlands Working Group BioInformatics consists of:

<b>A. Siebes</b> (Arno)	(chairman)	Utrecht University, Utrecht
<b>F.H.J. van Enkevort</b> (Frank)	(secretary)	Radboud University, Nijmegen
<b>J. Heringa</b> (Jaap)		Free University, Amsterdam
<b>J.A.M. Leunissen</b> (Jack)		Wageningen UR, Wageningen
<b>V. de Jager</b> (Victor)		BioASP, Amersfoort
<b>R.C. Jansen</b> (Ritsert)		University of Groningen, Groningen
<b>H. Mooibroek</b> (Hans)		Wageningen UR, Wageningen
<b>J.A.C. Rullmann</b> (Ton)		Organon NV, Oss
<b>F.H.J. Schuren</b> (Frank)	(for NBV)	TNO Voeding, Zeist
<b>M. Timmers</b> (Marc)	(for NVBMB)	Utrecht University, Utrecht
<b>J.H. de Winde</b> (Han)		DSM & TU Delft

Werkgroep



## Participants

BioInformatica

Dr	Paulien	Adamse	Plant Research International	paulien.adamse@wur.nl
Drs	Blaise	Alako	Laboratory of Bioinformatics WUR	blaise.alako@wur.nl
Dr	Casper	Albers	Groningen Bioinformatics Centre	C.J.Albers@biol.rug.nl
Drs	Rudi	Alberts	Groningen Bioinformatics Centre	r.alberts@biol.rug.nl
Dr	Richard	Baerends	Molecular Genetics, Groningen	baerends@biol.rug.nl
	Hans	Bakker	MathWorks BV The Netherlands	hans.bakker@mathworks.nl
Dr	Mar	Bellido	Hospital del Mar, Aut. Univ. Barcelona, ES	mbellido@imas.imim.es
Dr	Christian	Blaschke	BioAlma & CNB, Campus Universidad Autonoma, Madrid, ES	blaschke@cnb.uam.es
Drs	E.J.	Blom	Molecular Genetics, RuG	e.j.blom@biol.rug.nl
	Jos	Boekhorst	CMBI, Nijmegen	J.Boekhorst@cmbi.ru.nl
Dr	Henk	Bolhuis	Microbial Ecology, University of Groningen	h.bolhuis@biol.rug.nl
Ing	Andre	Boorsma	University of Amsterdam	boorsma@science.uva.nl
	Martin	Borg		
Drs	D.W.J.	Bosman	Computing Science, RuG	dinne@cs.rug.nl
Dr ir	Bernd	Brandt	LUMC	b.w.brandt@lumc.nl
-	Rutger	Brouwer	University of Groningen	r.w.w.brouwer@student.rug.nl
Dr	Girbe	Buist	Molecular Genetics, RuG	g.buist@biol.rug.nl
Dr	Gert-Jan	Caspers	Radboud University Nijmegen	g.caspers@cmbi.ru.nl
Drs	Koen	Cuelenaere	Dalicon BV	K.Cuelenaere@dalicon.com
	Rineke	Daatselaar	Hanzehogeschool Groningen	m.c.c.daatselaar@pl.hanze.nl
Drs	Peter	Dammers	Institute for Life Science & Technology Hanzehogeschool Groningen	p.m.dammers@pl.hanze.nl
	Marinus	Dansen	TNO-Voeding	dansen@voeding.tno.nl
Drs	Mark	de Been	WUR/WCFS/CMBI	mdebeen@cmbi.kun.nl
Dr	Bert	de Brock	Hanzehogeschool/RUG	e.o.debrock@pl.hanze.nl
Drs	Victor	de Jager	BioASP	victor.de.jager@bioasp.nl
Ing	Anne	de Jong	Molecular Genetics, University of Groningen	A.de.Jong@biol.rug.nl
Dr	Wim	de Leeuw	CWI, Amsterdam	Wim.de.Leeuw@cw.nl
Prof dr	Han	de Winde	DSM Life Sciences & TUDelft	j.h.dewinde@tnw.tudelft.nl
	Martijn	Dijkstra	GBiC	m.dijkstra@cs.rug.nl
Drs	Bas E.	Dutilh	CMBI / NCMLS	dutilh@cmbi.ru.nl
Dr	Michael	Egmont-Petersen	Institute of Information and Computing Sciences, Utrecht	michael@cs.uu.nl
Ir	Thijs	Ettema	CMBI, Nijmegen	t.ettema@cmbi.ru.nl
Dr	Chris	Evelo	BiGCaT Bioinformatics Maastricht	chris.evelo@bigcat.unimaas.nl
Ir	Vanessa	Fens	GeNeYouS	vanessa.fens@wur.nl
Ir	Mark	Fiers	Plant Research International	mark.fiers@wur.nl
Drs	Lude	Franke	UMC Utrecht	
Drs	Jingyuan	Fu	Groningen Bioinformatics Center	j.fu@biol.rug.nl
	Maartje	Geerds	Hanzehogeschool	
Drs	Omid	Givi	Hanzehogeschool	omid.givi@gmail.com
	Peter	Gnodde	Hanzehogeschool Groningen, Institute of LST	p.gnodde@pl.hanze.nl
Dr	Alexander	Goesmann	CeBiTec, Universität Bielefeld, DE	Alexander.Goesmann@CeBiTec.Uni-Bielefeld.DE
	Menno	Hardonk	Groningen Bioinformatics Centre, RuG	m.hardonk@med.rug.nl
Drs ing	Tjerk	Harkema	Hanzehogeschool	t.j.harkema@pl.hanze.nl
	Marjan	Heijkamp	BioASP	marjan.heijkamp@bioasp.nl
Prof dr	Sef	Heijnen	TU Delft, Bioprocess Engineering	j.j.heijnen@tnw.tudelft.nl
Drs	Quinta	Helmer	CMBI, Universiteit Nijmegen	qhelmer@cmbi.kun.nl

Prof dr	Jaap	Heringa	IBIVU, Free University Amsterdam	heringa@cs.vu.nl
Prof dr	Martijn	Huynen	NCMLS / CMBI, Radboud University Nijmegen	huynen@cmbi.ru.nl
Prof dr	Ritsert	Jansen	GBIC, University of Groningen	r.c.jansen@biol.rug.nl
Drs	Antoine	Janssen	Keygene N.V.	gerrie.van-haaster@keygene.com
Dr ir	Luc	Jansz	ID Lelystad	luc.jansz@wur.nl
Ing	Gera L.	Jellema	CMBI	g.jellema@cmbi.ru.nl
Dr	Lars Juhl	Jensen	EMBL Heidelberg, DE	lars.jensen@embl.de
Dr	Eric	Kamst	Erasmus MC, Hematology	e.kamst@erasmusmc.nl
Drs	Robert	Kerkhoven	CMBI	robertke@cmbi.kun.nl
Dr	Frans	Klis	Swammerdam Inst Life Sciences	F.M.Klis@uva.nl
	Sietske	Kloosterman	Hanzehogeschool	
Dr	Anton	Koning	Erasmus MC, Rotterdam	a.koning@erasmusmc.nl
	Lukas	Kortenhorst	Sun Nederland	
Prof dr	Oscar	Kuipers	Molecular Genetics, RuG	o.p.kuipers@biol.rug.nl
Drs	Arnold	Kuzniar	WUR	arnold.kuzniar@wur.nl
Prof dr	Jack	Leunissen	Bioinformatics, WU R	jack.leunissen@wur.nl
Drs	Alrik	Lubbers	Groningen Bioinformatics Centre, University of Groningen	l.a.lubbers@biol.rug.nl
	A.C.M.	Luijf	Bioinformatics Lab., Academic Medical Centre	a.c.luyf@amc.uva.nl
Dr	Scott	Marshall	Swammerdam Inst. Life Sciences	marshall@science.uva.nl
Dr	Arnold	Meijster	HPC&V University of Groningen	A.Meijster@rc.rug.nl
Dr ir	Jan-Peter	Nap	Bioinformatics Hanze University Groningen	j.p.h.nap@pl.hanze.nl
Drs	Pieter	Neerincx	WUR	pieter.neerincx@wur.nl
Dr	Michiel	Noback	Institute for Life Science and Technology	m.a.noback@pl.hanze.nl
	Richard	Notebaart	CMBI Universiteit Nijmegen	R.notebaart@cmbi.kun.nl
Dr	Stephen	Nyangoma	Groningen Bioinformatics Institute	s.nyangoma@cs.rug.nl
Dr	Patrick	Ogao	University of Groningen	ogao@cs.rug.nl
Drs	Martin	Oti	CMBI	m.oti@cmbi.ru.nl
Prof dr	Sasha	Panfilov	University of Dundee, UK / Utrecht University, NL	panfilov@maths.dundee.ac.uk
Dr	Herman	Pel	DSM	herman.pel@dsm.com
	Caspar	Perik	MathWorks BV The Netherlands	Caspar.Perik@mathworks.nl
Drs	Grietinus	Plat	Institute for Life Science & Technology Hanzehogeschool Groningen	g.plat@pl.hanze.nl
Dr ir	Marco	Pool	ID Lelystad	marco.pool@wur.nl
	Pjotr	Prins	WUR Dept. of Nematology	pjotr.prins@wur.nl
Dr	Daphne	Rainey	WUR Dept. of Nematology	daphne.rainey@wur.nl
Drs	Xin-Ying	Ren	Plant Research International – WUR	xinying.ren@wur.nl
Prof dr	Jos	Roerdink	IWI, University of Groningen	j.b.t.m.roerdink@cs.rug.nl
Dr	Marco	Roos	Integrative Bioinformatics Unit (IBU), University of Amsterdam (UvA)	roos@science.uva.nl
Drs	Blanche	Schroen	Universiteit Maastricht	b.schroen@cardio.unimaas.nl
Prof dr	Arno	Siebes	Universiteit Utrecht	arno@cs.uu.nl
Prof dr	Roland	Siezen	CMBI, Radboud University, Nijmegen	siezen@cmbi.ru.nl
Drs	Victor A	Simossis	Vrije Universiteit	vsimoss@cs.vu.nl
	Marcel	Smid	Josephine Nefkens Institute, Erasmus MC	m.smid@erasmusmc.nl
Dr	Berend	Snel	CMBI, Radboud University Nijmegen	snel@cmbi.ru.nl
Ing	Marcel	Sturre	Rijksuniversiteit Groningen MBP	m.j.g.sturre@biol.rug.nl
Drs	Morris	Swertz	Groningen Bioinformatics Centre, University of Groningen	m.a.swertz@cs.rug.nl
Drs	Jifeng	Tang	PRI-WUR	jifeng.tang@wur.nl

Dr	Peter	Terpstra	University Groningen	p.terpstra@med.rug.nl
Dr	Bas	Teusink	WCFS/NIZO food research	bas.teusink@nizo.nl
Ing	Bart	Theelen	CBS	theelen@cbs.knaw.nl
Prof dr	Marc	Timmers	Utrecht University	H.T.M.Timmers@med.uu.nl
	Tilman	Todt	HAN – HLO	tdt@ft.han.nl
Drs	Evertjan	van de Kaa	TU Delft	ejvandekaa@hotmail.com
Drs	Harmen	van de Werken	Wageningen	harmen.vandewerken@wur.nl
	Maarten	van den Bosch	WUR Bioinformatica	maarten.vandenbosch@wur.nl
	Laslo	van den Hoek	UVA	
Ir	Ate	van der Burgt	PRI / WUR	ate.vandenburgt@wur.nl
	Marjolein	van der Glas	MathWorks BV The Netherlands	
Dr	Roald	van der Laan	ErasmusMC, Cell Biology & Genetics	r.vanderlaan@erasmusmc.nl
	Alexander	van der Velden	Sun Nederland	
Dr	Frank	van Enckevort	NIZO Food Res & CMBI Nijmegen	Frank.van.Enckevort@cmbi.ru.nl
	Jan	van Haarst	Plant Research International	jan.vanhaarst@wur.nl
	Karin	van Haren	BioASP	karin.van.haren@bioasp.nl
Ir.	Simon	van Heeringen	Hogeschool Leiden	heeringen.van.s@hsleiden.nl
Dr	Sacha	van Hijum	University of Groningen	s.a.f.t.van.hijum@biol.rug.nl
Drs	Sander	van Hooff	AMC University of Amsterdam	s.vanhooff@amc.uva.nl
Drs	Bart	van Houte	IBI Free University Amsterdam	bvhoute@cs.vu.nl
Drs	Rob	van Linschoten	Institute for Life Science & Technology Hanzehogeschool Groningen	r.j.van.linschoten@pl.hanze.nl
Dr	Erik	van Mulligen	Collexis & EUR Biosemantiek	e.vanmulligen@erasmusmc.nl
Dr	Jan	van Oeveren	Keygene N.V.	gerrie.van-haaster@keygene.com
	Marian	van Os	C.v.B. Hanze University Groningen	
Ing	Barbera	van Schaik	AMC, University of Amsterdam	b.d.vanschaik@amc.uva.nl
Dr ir	Cees	van Sluis	EBT TU Delft	c.vansluis@tnw.tudelft.nl
Dr	Ivo	van Stokkum	Free University Amsterdam	ivo@nat.vu.nl
	Antoine	Veldhoven	Erasmus MC, Dept. of Urology	a.veldhoven@erasmusmc.nl
Ir	Bert	Vermeer		b.vermeer@tiscali.nl
Ir	Harold	Verstegen	Keygene N.V.	gerrie.van-haaster@keygene.com
Prof dr	Cor	Verweij	VUMC, Amsterdam	c.verweij@vumc.nl
Drs	Michiel	Wels	WCFS/CMBI	mwels@cmbi.kun.nl
Dr	Daniela	Wieser	EBI / EMBL Outstation, Hinxton, UK	dwieser@ebi.ac.uk
Dr	Greer	Wilson	CMBI	gwilson@cmbi.ru.nl
Ing	Aldert	Zomer	Groningen - Molecular Genetics	a.l.zomer@biol.rug.nl
	Ria	Zoutman	Hanzehogeschool	
	Eric	Zwanenburg	Oracle Nederland	