

Bioinformatics research and service groups in Finland

This document has been automatically generated from the material provided
by the Finnish bioinformatics research groups.

CSC -Scientific Computing Ltd.

<http://www.csc.fi/molbio/bioverkko/>

Contents

1	Bioinformatics research groups at the universities	4
1.1	Helsinki University of Technology	4
1.1.1	Laboratory of Computer and Information Science	4
1.2	Tampere University of Technology	5
1.2.1	Institute of Signal Processing	5
1.3	University of Helsinki	5
1.3.1	Department of Biological and Environmental Sciences, Genetics	5
1.3.2	Department of Computer Science	6
1.3.3	Department of Computer Science, SYSFYS Project	7
1.3.4	Faculty of Medicine	7
1.3.5	Institute of Biotechnology, Bioinformatics Group	8
1.4	University of Kuopio	8
1.4.1	Department of Neurobiology, A.I. Virtanen Institute	8
1.4.2	Department of Pharmaceutical chemistry	9
1.5	University of Oulu	9
1.5.1	The Biocenter and the Dep. of Biochemistry	9
1.6	University of Tampere	10
1.6.1	Institute of medical technology	10
1.7	University of Turku and Åbo Akademi University	12
1.7.1	Computational Biomodelling Laboratory	12
1.7.2	The Structural Bioinformatics Laboratory	12
1.7.3	TUCS Bioinformatics Laboratory	13
2	Bioinformatics service units and research institutes	15
2.1	Biomedicum Helsinki	15
2.1.1	GIU - Genome Informatics Unit	15
2.2	CSC, the Finnish IT center for science	16
2.2.1	Bioinformatics group	16
2.3	Helsinki Institute for Information Technology HIIT	17
2.3.1	Basic Research Unit	17
2.4	Turku Centre for Biotechnology	18
2.4.1	Bioinformatics core facility	18
2.5	VTT Biotechnology, Espoo	19

2.5.1	Quantitative Biology and Bioinformatics group . . .	19
3	Finnish bioinformatics companies	22
3.1	Biocomputing Platforms Ltd Oy	22
3.2	FBD Ltd.	22
3.3	GeneOS Ltd.	23
3.4	Genolyze Ltd.	23
3.5	Jurilab Oy	23
3.6	Medicel Ltd.	24
3.7	Mobidiag Oy	25
3.8	Orionpharma Ltd.	26
3.9	Triacle Biocomputing	26
3.10	Visipoint Oy	27

1 Bioinformatics research groups at the universities

1.1 Helsinki University of Technology

1.1.1 Laboratory of Computer and Information Science

Group leader: **Samuel Kaski**

Home page: <http://www.cis.hut.fi/projects/mi/>

Bioinformatics at the Laboratory of Computer and Information Science, Helsinki University of Technology

Research The Laboratory of Computer and Information Science has strong research activity in developing statistical models and algorithms for analyzing biological processes and mining biological data. Two closely cooperating groups of our laboratory apply modern computational and statistical methods to problems of computational biology: (i) Pattern Discovery group (H. Mannila, J. Hollmen) of the From Data to Knowledge center of excellence, and (ii) Statistical Data Mining and Bioinformatics group of the Finnish Center of Excellence in Adaptive Informatics Research (S. Kaski).

Teaching The lab offers a bioinformatics major based on the active research of the laboratory in both methods and bioinformatics, both to students in the degree programmes in computer science and in bioinformation technology. The lab additionally participates in a new international Master's programme on bioinformatics, starting in 2006 in collaboration with University of Helsinki, and in several graduate schools partly in the field of bioinformatics.

The goal is advanced understanding of the computational, modeling, and data mining methods required for understanding biological and medical systems, analyzing large datasets measured from them, and in solving modern biomedical research problems. Since all advances in the field cannot be predicted, the studies include a strong general-purpose methodological basis which does not leave the students speechless when the field progresses. At the moment, for instance data analysis experts are needed in research institutions and companies in the biotech, medical, and pharmaceutical industries.

The studies consist of advanced studies in at least one of the subfields of bioinformatics or neuroinformatics. Lecture courses and special courses on timely topics are available for that. Additionally, a strong general-purpose understanding of the methods is needed: data mining, statistical modeling or machine learning. Within these limits the studies can be planned freely.

More information about the bioinformatics Master's programme

- <http://www.cs.helsinki.fi/bioinformatiikka/mbi/>

1.2 Tampere University of Technology

1.2.1 Institute of Signal Processing

Group leader: **Olli Yli-Harja**

Home page: <http://www.cs.tut.fi/~yliharja/>

Signal Processing for Systems Biology (SPSB), in Tampere University of Technology

The research team, Signal Processing for Systems Biology (SPSB), involves three local research groups at TUT/ISP: the Computational Systems Biology group led by Prof. Olli Yli-Harja, the Computational Neuroscience group led by Dr. Marja-Leena Linne, and the Data Modelling group led by Prof. Jaakko Astola. The research team has currently 27 members. It will be a part of Academy Centre of Excellence in 2006-2011.

Our main research goal is to develop novel signal processing methods that can be applied in modern systems biology. We mainly concentrate on studying the structure and dynamics of biosystems, but also contribute to modelling of novel biosystem measurement technologies. With the term "biosystem", we refer to interacting cell populations, as well as cell and molecular biological systems, e.g. intracellular molecular pathways.

The research team has four main research directions: (i) theory of nonlinear signal processing applied in modelling of biosystems, (ii) models of biosystem measurement technologies, (iii) discrete large-scale models of biosystems, and (iv) modelling of biological data.

1.3 University of Helsinki

1.3.1 Department of Biological and Environmental Sciences, Genetics

Group leader: **Päivi Onkamo**

Home page: Not available

As started in the beginning of this year, I do not yet have a group of my own. However, I'm planning to start it during the next year. The scientific profile I intend to have, or actually maintain, is an interplay of direct biological applications and methodological development: the deep knowledge of the application area gives the best understanding of the new methods needed - or the further development needs of the existing ones. On the other hand, strong background of close collaboration with methodological scientists gives an invaluable advantage: "common language" with methodological side, which makes successful collaboration a whole lot easier.

To date, I have mostly been working with gene mapping methods - first in the invention and creation of Haplotype Pattern Mining, which is a genetic pattern search algorithm based on data mining methods - and its application on various human disease gene mapping studies. Presently, I concentrate on multi-phenotype genetic analysis on human disease, both with existing statistical approaches, and on the other hand emerging new technologies, which are being developed in e.g. collaborative groups of Hannu Toivonen and Liisa Holm.

1.3.2 Department of Computer Science

Group leader: **Esko Ukkonen**

Home page: <http://www.cs.helsinki.fi/research/fdk/>

The group of Professor Ukkonen has worked in the field of bioinformatics since the 1980s when they started to develop algorithms for DNA sequence assembly. These algorithms have influenced to the sequence assembly algorithms that were later on used, for example, in the genome projects. Later on, Ukkonens' group studied theory of string methods, in order to develop effective algorithms for string analysis tasks, like string comparison or similarity searches. The algorithms that resulted from these studies, have gained a significant status and they have, for example, been used in several computer science books (D. Gusfield: Algorithms on strings, trees, and sequences - Computer Science and Computational Biology. Cambridge University Press 1997; B. Smyth: Computing patterns in strings. Pearson/Addison-Wesley 2003).

Collaboration with biological research groups strengthened again at the end of 1990 when the group started to study analysis of gene expression data. Both Alvis Brazma and Jaak Vilo have been working in Ukkonens' laboratory on this topic. One of the first computational methods to predict transcription factor binding sites in the DNA were developed in Ukkonens' group. Currently the group is studying the following bioinformatics related topics:

- Modeling metabolic networks and flux analysis
- Development of data management tools for VTT-TRAC gene expression measuring technology

- New algorithms for haplotype analysis
- Algorithms for detecting gene regulatory factor binding sites in mammalian genomes
- Reconstruction of macromolecular complexes based on electron microscopy images

1.3.3 Department of Computer Science, SYSFYS Project

Group leader: **Juho Rousu**

Home page: <http://www.cs.helsinki.fi/group/sysfys/>

The group led by Docent Juho Rousu concentrates in computational systems biology, in particular computational methods for analysing metabolic networks. The current research topics include:

- Metabolic flux estimation using isotopic tracers
- Methods for analysing metabolite data generated by tandem mass spectrometers
- Machine learning methods for metabolic reconstruction and protein function prediction.

The group consists of two senior researches, one post-doctoral researcher, two PhD students and three MSc students.

1.3.4 Faculty of Medicine

Group leader: **Sampsa Hautaniemi**

Home page: <http://www.ltdk.helsinki.fi/sysbio/csb/>

The focus of research in the Hautaniemi lab is to analyze complex biological systems using systems biology approach. Systems biology research is an interdisciplinary effort to gain understanding of the function and control of biological processes using mathematical methods and statistical experimental design principles. The central theme of our research is to produce reliable, experimentally testable biomedical predictions and suggest novel directions for future experiments.

We are interested in developing methods to integrate data from various sources (genetic, gene expression, protein activations, phenotypes, etc.). We are also developing bioinformatics tools to link the results from our analyses

to biomedical databases and result in robust predictions that can be tested experimentally.

Another main thrust is to construct and analyze dynamic mathematical models for signaling transduction pathways and suggest targets for effective therapeutic interventions, study the effects of perturbations on the pathways, and eventually translate the acquired information to diagnostics and therapeutics.

1.3.5 Institute of Biotechnology, Bioinformatics Group

Group leader: **Liisa Holm**

Home page: <http://ekhidna.biocenter.helsinki.fi/frontpage>

The group's main interests are protein sequence-structure-function relationships and computational analysis of gene regulation. The recent research projects of the group include development of tools for protein interactome data analysis, programs for transcription factor binding site prediction and verification (POCO and POBO) and a protein domain database, based on automatic protein domain decomposition algorithm (ADDA).

At the moment the research group consists of four post doc researchers, three doctoral students and three other researchers.

1.4 University of Kuopio

1.4.1 Department of Neurobiology, A.I. Virtanen Institute

Group leader: **Gary Wong**

Home page: http://www.uku.fi/aivi/neuro/index_genomics.shtml

Research Interests:

Microarray Gene Expression Analysis - project for analyzing and extracting the vital information from the massive amounts of data produced by microarrays. Applications from the neurobiology of disease, pharmacology, and model systems are used as input data. Various clustering algorithms are used including hierarchical clustering and self-organizing-maps, a neuron based machine learning algorithm.

Production of Bioinformatic Tools for Processing Gene Lists - gene lists are produced from microarray, RNA interference, proteomics and other genomics studies. The gene lists are analyzed for over representation of thematic groups based on gene ontologies. Web-enabled implementations are developed for users that arise from algorithms developed in the laboratory.

Integrated Visualization Platform - publically available and our own laboratory written bioinformatic data analysis tools are gathered and a data analysis pipeline is constructed from these tools. Data formatting and visualization

applications are produced in order to provide users ability to gather, analyze, and interpret biological data from diverse (e.g. sequence, microarray, proteomic, chromosome localization) sources.

C. Elegans as a model system for studying human neurodegenerative diseases - transgenic worm models are produced to understand the molecular and cellular events leading to neurodegeneration. Various putative or candidate neurotoxicity causing genes are over expressed in worms using coexpression of green-fluorescent protein (GFP) in neurons as a marker, followed by fluorescent microscopy. Various worm phenotypes are scored and characterized

1.4.2 Department of Pharmaceutical chemistry

Group leader: **Antti Poso**

Home page: <http://www.uku.fi/farmasia/fake/modelling/index.shtml>

The PMC-group, headed by Antti Poso uses molecular modeling and other bioinformatics methods to discover and optimize drug molecules. The critical factor in the activities is integration of bioinformatics and molecular modeling together with synthesis chemistry, biology, pharmacology, neural sciences, and information technologies within fixed research projects. In addition to the drug discovery projects the group aims to develop and commercialize new software for the needs of drug discovery.

At the moment the group has eight people working in the field of bioinformatics. The goal is that on average one Ph.D. student will graduate from the group each year.. Further, the group is actively supporting the new graduates to find their way to industry or other research groups after their graduation. The group is also actively recruiting post-doc level researchers from abroad. For the masters degree students, the group tries to provide an opportunity for a 3-6 month research period in foreign research groups. The most important foreign contacts at the moment are the Heinrich-Heine University in Dusseldorf, Martin-Luther University in Halle, University of Perugia (Italy), University of Fuda (Shanghai) and two companies: SIRC/TCM (Shanghai) and SIRTRIS-pharmaceutical (Boston USA).

1.5 University of Oulu

1.5.1 The Biocenter and the Dep. of Biochemistry

Group leader: **Andre Juffer**

Home page: <http://www.biochem.oulu.fi/Biocomputing>

The Biocomputing group in Oulu is predominantly concerned with computer simulations of biomolecular systems. The general research interests are

concerned with (i) the understanding of biomolecular structure, dynamics and function, and (ii) the development and application of computer models and simulation methods for molecular systems. An important simulation tool is molecular dynamics, which relies on the classical equations of motion to simulate the dynamics of for instance a solvated protein in water or even a membrane environment. Also, the technique of Monte Carlo simulation is employed. For instance, in combination with continuum electrostatics and the boundary element method, this simulation technique is used to calculate acid dissociation constants of titrating sites in proteins. More recently, a stochastic event-based model has been developed that is capable of simulating ribosome traffic along mRNA sequences and mimics in fact recombinant protein synthesis as it happens inside the cell.

An important challenge for the future is to reach larger length and longer time scales. For this, we follow two strategies. One is concerned with the implementation of Grid aware simulation tools and the second is to develop alternative simulation models that operate at a coarse-grained or mesoscopic level to cover substantially longer time and length scales. Most likely, it is the combination of both that will be successful. For the first strategy, we have successfully build a Linux-based computing cluster that is scheduled to become part of the new national material science Grid in 2005. For the second strategy, we are currently developing, among other, a method that is based upon dissipative particle dynamics (DPD), where special attention is given to long ranged electrostatic interactions. Its application to protein simulation would involve selecting suitable degrees of freedoms, a set of DPD particles and their interactions.

1.6 University of Tampere

1.6.1 Institute of medical technology

Group leader: **Mauno Vihinen**

Home page: <http://bioinf.uta.fi>

Bioinformatics group started in Tampere in the beginning of the 1998. IMT Bioinformatics group consists of research scientists, technical staff and undergraduate students. The team is truly multidisciplinary including bioinformaticians, physicians, chemists, computer scientists, statisticians, biologists and biochemists. Currently the group consists of close to 20 members.

Research:

The main research topics of the group are bioinformatics mainly related to the analysis of genotype-phenotype and protein structure-function correlations usually in relation to human hereditary diseases, especially immunodeficiencies. Another major research area is related to gene and protein expression (microarrays, proteomics) data analysis. The group has studied for

many years bioinformatics related to immunology and developed several data and knowledge bases for immunodeficiencies as well as analysed structure-function and genotype-phenotype correlations of immunodeficiency-related proteins, genes and mutations.

Recently the Bioinformatics team has moved to systems biology with the aim of modelling mathematically certain immunological processes, which are important in health and disease.

The group is in charge of over 100 databases and information sources, which are freely available in Internet. The goal of the research is to understand biological processes in detail in atomic, structural and genetic levels.

The group has been active in many different fields of bioinformatics including e.g. sequence analysis, analysis of gene variations (SNPs, mutations), database and knowledge base development, protein structure prediction, studies related to structure-function relationships, protein engineering, microarray and proteomics data analysis, systems biology, database management, algorithm and program development, large studies of "-omics wide" features.

Education:

The group has provided education in bioinformatics throughout the years not only to students in Tampere, but also nationally and internationally. Courses have been in different fields of bioinformatics. Two larger course packages (25 and 12 credit units) have been provided together with the Center for Extension Studies, University of Tampere year 2002 and 2005. Bioinformatics has been one of the four specialization lines in the Biotechnology curriculum at IMT since 2001.

Our distance learning courses obtained year 2002 the National Quality Award in Web-based Teaching from the Ministry of Education and the National Board of Education. Currently three courses, Introduction to Bioinformatics, Bioinformatics in Functional Genomics, and Structural Bioinformatics are available. Fourth course in protein modelling will be released later this year.

Hundreds of students have participated on the different lecture and distance courses. As a new initiative an International Master of Science (MSc) Programme in Bioinformatics will start in September 2006. This programme will be jointly arranged with the Department of Information Technology, University of Turku. 15 places will be available in both universities.

The students will have a Bachelor's degree in a related field such as molecular biology, biochemistry, genetics, computer science, or statistics. The entire curriculum will be in English. The curriculum will last for two years and be based on different types of courses including e.g. distance learning, seminars, lectures, and intensive courses where students with different backgrounds work together.

1.7 University of Turku and Åbo Akademi University

1.7.1 Computational Biomodelling Laboratory

Group leader: **Ion Petre**

Home page: <http://www.tucs.fi/research/labs/combio.php>

The research of the laboratory centers on the computational modelling of biochemical systems. The approach is to understand as computations the myriad of biochemical processes that evolve in parallel, influence each other, propagate signals, or cooperate on various tasks. Our goal is to increase the understanding of how entire cells adapt, communicate, and survive in dynamic environments, all in terms of computations. Having sound computational models for a biochemical system allows one to employ formal reasoning about its pathways or regulatory networks, formulating predictions and/or running simulations. Such models are also useful for designing novel sorts of computations based on the principles that underline the functioning of bio-systems.

Our group is leading research on computational bio-processes, including computational processes in living cells, as well as nature-inspired human-designed computations. The general interest of the laboratory is gaining an understanding of fundamental structures behind the functioning of all kinds of bio-systems. We have considerable expertise in building discrete models, based on combinatorics, graph theory, stochastic processes, etc.

Our work is funded by Academy of Finland and the European Union. Our group currently consists of three senior members, two PhD students and five undergraduate students. A list of our graduate course on Bioinformatics can be found at <http://www.abo.fi/ipetre/>

1.7.2 The Structural Bioinformatics Laboratory

Group leader: **Tiina Salminen** and **Mark Johnson**

Home page: <http://www.abo.fi/fak/mnf/bkf/research/johnson/>

The Structural Bioinformatics Laboratory (SBL) consists of two research groups, that of Mark Johnson and Tiina Salminen, located at the Department of Biochemistry and Pharmacy at Åbo Akademi University. We have nearly 20 researchers and staff. Our research is in the field of structural biology, mainly involving proteins - the molecules encoded by genes. Our research is multidisciplinary and made in close cooperation with experimental groups since our research approaches are focuses on computer-based molecular modeling and experimental structure determination using x-ray crystallography. The latter approach requires extensive computational support in order to solve molecular structures. Furthermore, we have an extensive program whereby we analyze the experimental structures deposited within the

protein data bank, as well as the development of our own software for molecular visualization and manipulation.

Our collaborative research projects consider a wide range of proteins and their interactions with other molecules. Cell-surface receptors, signalling proteins within the cell and nucleus; proteins within plants, other animals, but mainly human and having importance for health. Our collaborative partners provide experimental data and test our computationally assisted hypotheses. We, in turn, provide molecular interpretations for their experimental results.

The unit has a 32-processor Intel-based server for parallel applications, 30 other computers, 3 terabytes of disk space, gigabit networking is partially in place, and access to all major commercial and public domain data bases and software needed for structure based modeling. We belong to several of the software and database consortia organized by CSC.

In addition to basic research, teaching remains a major part of our activities. We train M.Sc. and Ph.D. students in structural biology, computer-based methods, etc. We teach courses at all levels, providing all students within the department, those belonging to the Health Bioscience degree program, and some students from the University of Turku with a base level of computer literacy. We provide introductory and advanced courses in molecular modeling, bioinformatics, and computer-based drug discovery.

1.7.3 TUCS Bioinformatics Laboratory

Group leader: **Tapio Salakoski**

Home page: <http://www.tucs.fi/research/labs/bioinf.php>

Tucs Bioinformatics Laboratory is a unit of Turku Centre for Computer Science (TUCS) and Department of Information Technology, University of Turku. TUCS is a joint special unit of the three universities in Turku: University of Turku, Åbo Akademi University, and Turku School of Economics and Business Administration, coordinating all university research and education in the field of information technology in Turku. TUCS Bioinformatics Laboratory was established in 2001 to work jointly with the Bioinformatics group at Turku Centre for Biotechnology, a joint biocentre of University of Turku and Åbo Akademi University, acting as a core facility providing all the researchers with equipment, services, and expertise in biotechnology.

TUCS Bioinformatics Laboratory is headed by Professor Tapio Salakoski. The laboratory currently involves senior researchers Dr. Jorma Boberg, Dr. Jouni Järvinen, and Dr. Alexandr M ylläri, as well as some 15 PhD and MSc students. The main sources of funding include TEKES, the National Technology Agency of Finland, and the TUCS Graduate School.

Professor Salakoski is a board member of SocBIN, Society of Bioinformatics in Nordic Countries, organizing annually the International Bioinformatics Conference. Academic collaboration include groups at University of Turku,

Turku Centre for Biotechnology, Åbo Akademi University, University of Tampere, Universite de Franche-Comte, France, Karolinska Institute, Sweden, Stanford University, USA, National Institute of Telecommunications, Poland, Tokyo Denki University, Japan, and University of Luton, the U.K.

Importantly, the group collaborates actively with several biotechnical, pharmaceutical, bioinformatics, language technology, and software companies. A new spin-off bioinformatics company Genolyze started in 2003 involving members and collaborators of the laboratory.

The group's expertise lies in developing and applying machine learning and knowledge-based analysis and modeling techniques in a multi-disciplinary domain. Current research at the laboratory is focused to the development of bioinformatics methods, resources, and tools for discovering protein-protein interactions in scientific literature (Bio-NLP); gene expression and promoter analysis; and modelling the relationships between protein sequence, structure, and function, as well as cellular signalling and metabolic networks. We are also developing bioinformatics databases and analysis tools.

TUCS Bioinformatics Laboratory organizes bioinformatics education as a specialized study line in the Computer Science curriculum at the University of Turku since 2001, currently comprising some 20 courses. In 2006, University of Turku and University of Tampere will launch a joint International Bioinformatics Master's Degree Program.

2 Bioinformatics service units and research institutes

2.1 Biomedicum Helsinki

2.1.1 GIU - Genome Informatics Unit

Group leader: **Juha Saharinen**

Home page: <http://www.giu.fi/>

Genome Informatics Unit (**GIU**) is a research-oriented bioinformatics unit that also functions as a bioinformatics core facility. The unit was established in 2002, as a collaborative project between the University of Helsinki and National Public Health Institute of Finland (KTL), Department of Molecular Medicine. The goals of the unit are two-fold: 1) GIU is a research-oriented unit, conducting its own research in bioinformatics, namely various software and database development projects as well as collaborating with molecular biology groups on projects requiring more advanced bioinformatics skills; and 2) GIU is a core-unit in bioinformatics and thus providing computational resources: bioinformatics software, databases and servers for the researchers.

GIU has its own computer server infrastructure (<http://www.giu.fi/Infra/tabid/64/Default.aspx>), including 100 CPU Linux cluster, Citrix Metaframe/Windows Terminal Server farm, multiple database and www-servers, and 10 TB disk space plus required backup tape robotics. A very large selection of bioinformatics applications (600 software titles) and databases (100, including multiple metazoan genomic databases) are installed and updated (see <http://www.giu.fi/Services/GIUSoftwareandDatabases/tabid/57/Default.aspx>) on GIU servers. The average around clock utilization of the Linux cluster is currently about 80%. The instrumentation is paid by various research grants.

GIU is linked with other core-facilities and research units, like Finnish Genome Center in Biomedicum Helsinki. These units are using BBUs computer hardware for data analysis, storage and distribution. In addition, software and database development projects are in progress between GIU and

these other units.

The personnel in GIU range from post docs to graduate and under-graduate students to database and system administrators. Currently about 8-10 persons work in GIU. Majority of the salaries (80% from various short-term research grants).

The co-localization of the bioinformatics unit in the same building with the wet lab laboratories has proven to be a key issue for a successful collaboration between the two disciplines and emphasized the status of bioinformatics being a cross-disciplinary field of science. GIU is currently involved in about 20 own or collaborative research projects, in areas of genomic sequence analysis, gene expression analysis, transcription factor analysis, genetic linkage and association analysis, comparative genomics hybridization analysis and development of bioinformatics databases and software.

The user base of GIU has grown very rapidly. Currently GIU has over 450 users, gained over the three years of GIUs existence. We estimate the number to be 550 users at the end of 2005, based on the previous growth rate. The user account is free, and GIUs policy is to give the account for any researcher conducting academic (non-profit) science in Meilahti campus area, National Public Health Institute of Finland (KTL), in Biocentrum Helsinki group or any of their collaborators in other research organizations in Finland and abroad. We feel that dual role of GIU has been very beneficial, and the best way to run a bioinformatics core-facility is to be actively involved in the research.

2.2 CSC, the Finnish IT center for science

2.2.1 Bioinformatics group

Group leader: **Tommi Nyrönen**

Home page: <http://www.csc.fi/molbio/>

CSC is the Finnish IT center for science, governed by the Ministry of Education. CSC provides modeling, computing, networking and information services for universities, polytechnics, research institutions and industrial companies. Most of the services of CSC are free for researchers working in the Finnish universities.

The bioinformatics group of CSC maintains a wide selection of software and databases that can be used in the servers of CSC or locally, with the licenses provided by CSC. The tool and database selection cover DNA microarray data analysis, gene mapping, molecular modeling and sequence analysis.

CSC also provides support and training in bioinformatics. The customers can, get on-line assistance from the specialists of CSC, use the guide books and web pages produced by CSC or attend to the courses of CSC. During

2003-2004 over 20 bioinformatics courses and events were arranged. Many of these courses were video recorded and are freely available for viewing on the web.

The bioinformatics group of CSC is doing research and development work to enhance its services. Current R and D projects include:

- Developing a national storage and analysis environment for microarray data (Nami/SYSBIO)
- Assembling scattered (yeast) biological data and providing data storage and distribution service (SyMBOLIC/NEOBIO)
- Developing computational software and data analysis environment for molecular discovery (SOMA/DRUG2000, SASYSBIO)
- Integrating and grid-enabling the major databases and software tools in bioinformatics (EMBRACE/EUNoE)
- Ph.D. project in small molecule nuclear receptor modulators
- Building interface to PairsDB-database by prof. Liisa Holm

2.3 Helsinki Institute for Information Technology HIIT

2.3.1 Basic Research Unit

Group leader: **Heikki Mannila**

Home page: <http://www.cs.helsinki.fi/u/mannila/>

The bioinformatics research of Academy professor Heikki Mannila focuses on computational analysis of genome structure and gene mapping. Prof. Mannila leads research projects on e.g. gene sequences, marker and expression data and the utilization of comparative genomics data. Within a consortium belonging to the SYSBIO program of the Finnish Academy, the group of Prof. Mannila is developing methods for analyzing mammalian genomes. This work involves sequencing genomes, detecting inversions and duplications and linking sequence data to functional data. The research is carried out in close contact with the biological research groups.

2.4 Turku Centre for Biotechnology

2.4.1 Bioinformatics core facility

Group leader: **Stephen Rudd**

Home page: http://www.btk.fi/TCB_Personnel/ID_Cards/id_cards_140.html

The bioinformatics core facility at the Centre for Biotechnology is acting as a support infrastructure for scientists within the BioCity Turku initiative. The bulk of our time and energy is currently utilised in a fruitful cooperation with the Finnish DNA microarray centre that is also located at the Centre for Biotechnology. We are providing critical systems within production, quality and analysis of expression data experiments. These systems may be summarized as:

- Laboratory information management systems (LIMS). This production software is used to track the features that are arrayed on in-house cDNA and oligonucleotide microarrays, The database is used in the process of quality control, production and in the generation of GAL files for the end users of a physical array
- Array reannotation services. Typical annotation associated with a commercial array, or even the features on in-house arrays tend to be minimal and do not form a suitable substrate for meaningful hypothesis based data analysis. Typical GAL files are reannotated to add additional nuance of function, structure and metabolic pathway. Using the openSputnik genome database that has been developed in house we add additional layers of user-specified annotation to our datasets.
- Microarray database. We are developing a MIAME-MAGE compliant microarray database. This is planned as a central repository for all in-house data. The resource will also be scaled up to include large sets of relevant public expression data - this will include all array experimental data from human, mouse and other species that are being actively used within our research sphere.
- Microarray data analysis. We are using the Inforsense Knowledge Discovery Edition to perform in depth analysis and annotation within and across large microarray datasets. We have additionally developed various other in-house statistical modules that we would like to make available to other Finnish researchers. This will require some commitment from CSC to provide access to the required development kit. We provide microarray data analysis as a service to all researchers in Finland.

- **Research and development.** We are actively trying, testing and evaluating other software packages that might provide quantitative or qualitative improvements to user experiences within the field of microarray analysis.

In addition to our roles within microarray analysis the Bioinformatics group has an active research function. We have developed the openSputnik platform for comparative and functional genomics for trait discovery and characterisation. We provide the largest and most comprehensive plant EST database worldwide and perform bioinformatics and genomics research into genome evolution and adaptation. We concentrate on the lower plant clades of hornworts, lycophytes and ferns.

In promoting a holistic view of the contemporary edges of the biological sciences we are developing a database for integrated and holistic access to high-throughput biology. We are developing an automated pipeline for the user-world handling of all data that stem from sequencing machines, real-time PCR machines, microarray scanners, mass-spectrometers and advanced microscopes, cell-sorters and other expensive machinery. We have already developed computational solutions for the handling of sequence data, proteomics and cell imaging data integration will be completed over the summer.

2.5 VTT Biotechnology, Espoo

2.5.1 Quantitative Biology and Bioinformatics group

Group leader: **Matej Oresic**

Home page: <http://sysbio.vtt.fi/qbix/>

The omics revolution and emergence of systems biology is opening new opportunities to characterize biological systems and ask fundamental biological questions. We are exploring the ways to characterize and model biological systems, as well as apply those ways to and pursue our own interests in integrative physiology as related to metabolic diseases, cellular physiology, and more generally to mechanisms controlling the homeostasis of biological systems.

Our facilities include a comprehensive bioinformatics and chemoinformatics system integrating all major databases, as well as analytical facilities for metabolomics.

Main research activities:

(1) **Integrative bioinformatics and conceptual biology**

Overload of information and new technologies in life sciences require new informatics solutions to make sense of available data. More fundamentally, our improved ability to quantitatively characterize biological systems is

challenging the way we formally describe biological systems and how we design experiments to address biological questions. Our aim is to develop new life science knowledge management solutions that can address the bio- and chemoinformatics challenges of systems biology. We have developed software based on three-tier architecture that enables integration and visualization of complex life science data.

(2) Algorithms and methodologies for metabolomics and proteomics

We are primarily interested in developing new approaches for processing and interpreting of mass spectrometry based metabolomics data. For that purpose we have developed software solutions for differential profiling of LC/MS, compound databases containing spectral and pathway information. On the analytical side, we have been developing a lipidomics platform for global screening of lipids, and developed database solutions for their automated identification.

(3) Lipidomic characterization of animal models in studies of lipotoxicity induced insulin resistance (with Antonio Vidal-Puig, Cambridge University)

We are utilizing our lipidomics platform and bioinformatics approaches to characterize the phenotypes of animal models linked to studies of lipotoxicity induced insulin resistance. Specifically, we aim at identifying key relevant endogenous compounds and pathways, aiming at identifying novel targets for interventions.

(4) DIPP Study: Systems Biology Approach to Biomarker Discovery in Type I Diabetes (with Riitta Lahesmaa and Olli Simell)

The overall objective of this study is to identify novel molecular markers that characterize development of diabetes-associated autoimmunity and progression towards overt clinical type I diabetes (T1D). Discovering genetic susceptibility markers and markers reflecting the disease activity is a specific aim of the project. This will be achieved by exploiting and integrating DNA microarray, proteomics, and metabolomics technologies in the analysis of samples obtained from a carefully selected patient population at defined stages of disease development. The data on gene expression, protein and metabolite profiles will be correlated with the Type 1 Diabetes Prediction and Prevention Project in Finland (DIPP) data on HLA gene alleles, autoantibodies, markers of enterovirus infections (antibodies, enterovirus RNA) and demographic and metabolic features of the study children to identify putative associations between a particular biomolecular profile and these key parameters.

(5) Plant systems biology (with Kirsi-Marja Oksman-Caldentey)

We are utilizing methods of metabolomics and bioinformatics to elucidate the pathways involved in production of plant secondary metabolites.

(6) Yeast systems biology (with Merja Penttilä, Esko Ukkonen, Liisa Holm)

Our interest is to develop and apply advanced experimental and computational tools to understand the cell physiology, using yeast as a model system. Specifically, we aim at developing novel computational methods for top-down cellular modelling. We aim at utilizing our methods in metabolic engineering applications.

3 Finnish bioinformatics companies

3.1 Biocomputing Platforms Ltd Oy

Bioinformatics contact person: **Pekka Mäkelä**

Home page: <http://www.bcplatforms.com>

Biocomputing Platforms Ltd is a software company that develops and markets data management systems for genetic and clinical studies. At the moment the company has nine employees serving customers from Finland, Sweden, England, and USA.

3.2 FBD Ltd.

Bioinformatics contact person: **Anna-Marja Hoffren** and **Mark Johnson**

Home page: <http://www.fbd.fi>

FBD provides comprehensive computational chemistry consultancy to the Biotechnology, Pharmaceutical and Nutraceutical industries.

Computer aided molecular design (CAMD) has become a keystone within the drug discovery process, bridging the skills of biologists and chemists and helping to promote inter-disciplinary communication. Virtual screening of compound collections (V-HTS) using pharmacophores derived from protein structures can provide meaningful starting points for Lead Generation and Library Design projects. Structure Based Drug Design (SBDD) techniques can also be used within Lead Optimisation studies and in the search for new and novel core structures to accelerate medicinal chemistry progress.

FBD provides expert computer aided drug design services, complemented by protein crystallography, small molecule synthesis and customized software development, to bring value and knowledge through a professional, customer focused approach.

3.3 GeneOS Ltd.

Bioinformatics contact person: **Tarja Laitinen**

Home page: <http://www.geneos.fi>

GeneOS combines the high-quality information resulting from the Finnish healthcare system with the latest scientific achievements in genome research. GeneOS is building patient databases that link the patients' longitudinal health care data with information from their blood samples. GeneOS works with pharmaceutical companies, healthcare providers, molecular diagnostic companies and Finnish healthcare authorities to accelerate testing of new therapies and to offer physicians and patients new diagnostic tools to improve delivery of care.

GeneOS's bioinformatics efforts are focussed on:

- text mining
- clustering of large data sets

3.4 Genolyze Ltd.

Bioinformatics contact person: **Marko Mäenpää**

Home page: <http://www.genolyze.com>

Genolyze is a bioinformatics company founded in Turku 2003. The company provides services related to DNA microarrays and gene expression studies. Genolyze is able to carry through all the phases of microarray studies including experiment design, RNA extraction, as well as all other microarray laboratory services and the demanding mathematical data analysis, which is the core expertise area of the company. Laboratory services are in some cases bought from subcontractors.

Genolyze supports the key activities of a research group and birnings in its own bioinformatics know-how. In addition to gene expression studies, Genolyze's know-how is applicable to other biological and medical research. Combining experimental data from different sources, multidimensional data analyzes, visualization, data-mining and data management are all part of the service concept of the company.

3.5 Jurilab Oy

Bioinformatics contact person: **Pekka Uimari**

Home page: <http://www.jurilab.com/>

Jurilab develops genetic tests for Finnish national diseases, like cardiovascular diseases and also searches for potential drug targets. Jurilab's bioinformatics efforts can be separated into two main phases. The first phase utilizes statistical genetics, i.e., the linkage between DNA markers and a disease is analyzed. In this area, Jurilab has been in collaboration with the HIIT (Prof. Toivonen) within the Tekes-funded Altti project. From one patient 100 000 SNP markers can be analyzed simultaneously (in the future the number of markers will increase to 500 000). This makes data management essential. Jurilab has used the BC Gene database system from Biocomputing Platforms Ltd, which has been developed for effective management of this kind of datasets. Using statistical methods, a set of SNP markers and corresponding genes are selected for more detailed in silico analysis. This is done mainly manually, by, e.g., studying the genome annotation and connection between genes. Automatized tools for this phase are being developed in the Tekes-funded Biomine project (in collaboration with prof. Toivonen from University of Helsinki). In addition Jurilab has done some collaboration with Prof. Carlberg (University of Kuopio) for detecting nuclear receptors from a genome.

3.6 MediceL Ltd.

Bioinformatics contact person: **Christophe Roos**

Home page: <http://www.mediceL.com>

MediceL is a privately owned company founded 2001 in Helsinki. MediceL has two departments: A molecular biology laboratory and a software development laboratory. Each of them occupies about 25 persons including medical doctors, molecular biologists, biochemists, bioinformaticians, statisticians, mathematicians, engineers, and software-engineers.

In the wet lab, operating at Biomedicum, new methodology is developed and applied to drug target discovery and validation projects related to inflammatory diseases and also to protein production in yeast. A special emphasis is put on development of high-throughput measurement technologies using among other highly automated fermentors and sample collection, mass spectroscopy, microarray technology and related wet-lab methods and microscopy.

The software development effort aims at creating an integrated data warehouse platform suitable typically for systems biology research and drug target discovery, but more generally for mostly any molecular biomedical research project.

For this purpose we have created Silvermine, a data management and analysis platform accessible through a set of product suites: Project editor (LIMS) , Workflow editor, Pathway editor (any network/map), Quantitative data workbench (visualisation of high-dimensional data), Microscopy workbench, Text mining workbench, General search and Annotators workbench. The MediceL integration framework is built to allow the plugging in of any tool, from

simple perl scripts to complex software modules.

On a low level, we have designed tools for

- importing data from external databases or published results such as Uniprot, PDB, EMBL, Ensembl, Bind, Kegg (pathways, ligand, glycans), Refseq, SGD, Interpro, GO-Database, Pubmed, GEO, transcription factor binding data, GFP-localisation data etc.
- integrating any ontology in standard format like GO, ChEBI, Human development, tissue, cell, phenotype, IntAct etc.
- analysing and calculating sugar structures based on MS-data, for statistical analysis of e.g. microarray measurements or for sequence analysis using a.o. hidden Markov models.
- We have also integrated external tools such as the Emboss suite tools, Clustalw, Meme and Blast.

Altogether, the use of Silvermine offers an integrated environment where very disparate measurements and data types can be used together in a flexible environment allowing high automation and auditing.

3.7 Mobidiag Oy

Bioinformatics contact person: **Jaakko Pellosniemi**

Home page: <http://www.mobidiag.com>

Founded in the year 2000, Mobidiag is specialized in the molecular diagnostics of infectious diseases. The focus area of the company is primarily to develop high-quality content for DNA biochip-based diagnostic assays. Mobidiag has selected clinically relevant panels of bacteria and other infectious disease markers and is developing efficient DNA detection technology for their analysis.

The main product under development is the Prove It diagnostic system, including proprietary biochips with patented DNA sequences for diagnosis of bacterial pathogens and antibiotic resistance markers, a reader instrument, as well as advanced software for the analysis of the diagnostic results.

The Prove It Advisor software uses intelligent algorithms to produce fully automated analysis of the assays. After a patient sample has been tested with the Prove It biochip, the software communicates the potential presence and identity of infection-causing microbes, using a raw image of the biochip as an input. In addition to the Prove It Advisor analysis software, Mobidiag also develops advanced LIMS (Laboratory Information Management System) solutions to support the clinical validation and use of the diagnostic products.

3.8 Orionpharma Ltd.

Bioinformatics contact person: **Stephen Rudd**

Home page: <http://www.orion.fi/english/>

The strategic goal of Orion Pharma is to develop new innovative drugs for global markets. Structural bioinformatics is an essential method in drug discovery process. Molecular modeling is used to model the structure of target protein and to design and optimize new drug candidates, affecting to these proteins.

Orion is currently studying the possibilities to utilize modern genomics, proteomics and bioinformatics methods in characterizing the action profiles of drug candidates and target proteins, to better evaluate the efficiency and security. Data managements, data analysis, visualization and integration with different biological and chemical results are salient bioinformatics application areas. Orion is currently investigating the possibilities for collaboration with Finnish bioinformatics service providers.

3.9 Triacle Biocomputing

Bioinformatics contact person: **Andre Juffer**

Home page: <http://www.triacle-bc.com/>

Products and Services

Triacle Biocomputing is a software development company serving the life science, biomedical and the biotechnology sector. The firm is physically located in Oulu, Finland, and is a spin-off of the Biocomputing group at the University of Oulu. In a nutshell, the firm offers services for:

- Software development
- Consulting
- Web design

Triacle Biocomputing was founded by Dr. A.H. Juffer in 2001. In 2002-2003, funding from the Tekes organization was obtained to carry out a Feasibility Study on Biocomputing with the objective to identify in greater detail industrial software requirements and to establish business relations with life science companies and organizations. Since then, the company has had and still has clients in the life science, biomedical and biotechnology sector. A number of products are under development, including **Molex**, a simulation and optimization tool to model the synthesis of proteins.

Technologies

Life Science and biocomputing Triacle Biocomputing offers a considerable expertise in particular areas of biocomputing, as exemplified in a number of scientific publications in peer reviewed well established journals, in association with the Biocomputing group at the Biocenter and Department of Biochemistry of the University of Oulu.

Software development The development of life science software is based upon sound object-oriented analysis and design. The firm employs, among other, C++, Java and XML. Life science application software can be either developed for the Unix/Linux or the Windows environment, while expertise also exist to build web applications, for instance, with XSLT, Cocoon and native XML-databases.

Website design Triacle Biocomputing implements websites using web standards, that is HTML, XHTML and XML for structure, CSS for presentation, and ECMAScript (javascript) and DOM for behaviour. A great deal of effort is spent to ensure that a website is accessible to every device (e.g. Palm Pilots, web-enabled cell phones, screen readers, text browsers, alternative browsers and devices).

3.10 Visipoint Oy

Bioinformatics contact person: **Olavi Raatikainen**

Home page: <http://www.visipoint.fi/>

Visipoint was founded in February 1999 as a spin-off from the University of Kuopio. The company has its office in Kuopio, with close contacts to the University. Visipoint develops and markets software products for data exploration of large data sets. The software products are used in various applications of life sciences (pharmaceutical, environmental, medical). Examples of the applications of bioinformatics are gene expression and 2-D electrophoresis data clustering and visualisation.

Product Description

Visipoint software is based on tree structured self organizing maps. Visipoint products are developed for personal data mining. They are tools that enable the users, such as biochemists, microbiologists, those who work within -omics and many others, to perform data mining in an interactive way.

Visual Data is the main product of Visipoint. It is a general purpose software that has been used to evaluate the potential of different application areas. Among other fields, we have explored the analysis of scientific research data, process analysis in process industry and also business informatics.

Bioinformatics

Big need of data exploration is in bioinformatics, where huge masses of data are resulting from genetic research, related to various medical or natural science applications. Examples of Visual Data applications are clustering and visualization of 2-D protein gel electrophoresis data and gene expression data.

Drug discovery

Visual Data is used for drug discovery purposes as a part of the total discovery process for clustering the data obtained for molecule libraries.

Further information and testing

Visipoint kindly asks You to contact us by e-mail or phone. Our contact person in the fields of this meeting is Dr Olavi Raatikainen, PhD (biochem). Interested in testing Visual Data? Contact Olavi and get free testing period of full version of the software.