Aalto University

UNIVERSITY OF HELSINKI

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Helsinki Institute for Information Technology HIIT

# Annual Report 2016

Helsinki Institute for Information Technology HIIT
Tietotekniikan tutkimuslaitos HIIT
Forskningsinstitutet för Informationsteknologi HIIT

Helsinki Institute for Information Technology HIIT is a joint research institute of Aalto University and University of Helsinki for basic and applied research in information technologies. HIIT is a strategic partnership of the two universities with a mission to support top-level research, seamlessly moving between fundamental methods and technologies to novel applications and their impact on people and society, and combining top ICT research of the two universities into a positive collaborating critical mass that supports a strong Helsinki ICT brand.

HIIT groups belong to several national Centers of Excellence and co-operate with the information industry and with sciences applying information technology. HIIT works in a multidisciplinary way, with scientists from computer, natural, behavioural and social sciences, as well as from humanities and design. HIIT partners with several international and Finnish companies as well as with universities and research institutions in Europe, North America and Asia. The current foci of research are computational modelling and data analysis, and ubiquitous ICT in the modern networked world.

HIIT is located in two different sites in the Helsinki Metropolitan area: at the Otaniemi campus of Aalto University and at the Kumpula campus of University of Helsinki. HIIT is currently led by Professor Petri Myllymäki, who started his 5-year term as the Director of HIIT in August 2015. HIIT was founded in 1999 and has nowadays an annual budget of approximately 10 million euros. HIIT's research is funded by Aalto University, the University of Helsinki, Tekes, the Academy of Finland, the European Union, private companies, and foundations financing Finnish hi-tech research. The research by HIIT has been assessed and deemed excellent by its Scientific Advisory Board in 2008, 2012 and 2016; in the Aalto University Research Assessment Exercise in 2009; in the Evaluation of Research and Doctoral Training at the University of Helsinki in 2011; and by the Scientific Advisory Board of the Aalto School of Science in 2016.

www.hiit.fi

Krisztina Cziner and Noora Suominen de Rios (eds.)

Helsinki Institute for Information Technology HIIT

# Annual Report 2016

*Krisztina Cziner and Noora Suominen de Rios (eds.)*

**Contact Information**

Helsinki Institute for Information Technology HIIT
Tietotekniikan tutkimuslaitos HIIT (in Finnish)
Forskningsinstitutet för Informationsteknologi HIIT (in Swedish)

*hiit-info@hiit.fi*
*www.hiit.fi*

**Otaniemi Site**

Postal address at Computer Science Building:
Helsinki Institute for Information Technology HIIT
PO Box 15400, FI-00076 Aalto, Finland

Street address:
Computer Science Building, Konemiehentie 2, Espoo
Telephone: +358 9 47001

**Kumpula Site**

Postal address:
Helsinki Institute for Information Technology HIIT
PO Box 68, FI-00014 University of Helsinki, Finland

Street address:
University of Helsinki, Department of Computer Science, Exactum
Gustaf Hällströmin katu 2b, Helsinki
Telephone: +358 9 1911
Fax: +358 9 191 51120

Unigrafia, Espoo 2017

Cover image by Teemu Pitkänen, 2016
Lay-out: Anna Bergman, Noora Suominen de Rios

# Table of Contents

## HIIT in brief

Helsinki Institute for Information Technology HIIT is a joint research institute of Aalto University and University of Helsinki for basic and applied research in information technologies. HIIT is a strategic partnership of the two universities with a mission to support top-level research, seamlessly moving between fundamental methods and technologies to novel applications and their impact on people and society, and combining top ICT research of the two universities into a positive collaborating critical mass that supports a strong Helsinki ICT brand.

HIIT groups belong to several national Centers of Excellence and co-operate with the information industry and with sciences applying information technology. HIIT works in a multidisciplinary way, with scientists from computer, natural, behavioural and social sciences, as well as from humanities and design. HIIT partners with several international and Finnish companies as well as with universities and research institutions in Europe, North America and Asia. The current foci of research are computational modelling and data analysis, and ubiquitous ICT in the modern networked world.

HIIT is located in two different sites in the Helsinki Metropolitan area: at the Otaniemi campus of Aalto University and at the Kumpula campus of University of Helsinki. HIIT is currently led by Professor Petri Myllymäki, who started his 5-year term as the Director of HIIT in August 2015. HIIT was founded in 1999 and has nowadays an annual budget of approximately 10 million euros. HIIT's research is funded by Aalto University, the University of Helsinki, Tekes, the Academy of Finland, the European Union, private companies, and foundations financing Finnish hi-tech research. The research by HIIT has been assessed and deemed excellent by its Scientific Advisory Board in 2008, 2012 and 2016; in the Aalto University Research Assessment Exercise in 2009; in the Evaluation of Research and Doctoral Training at the University of Helsinki in 2011; and by the Scientific Advisory Board of the Aalto School of Science in 2016

Aalto University

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

# Review of Year 2016

## *Overview*

Helsinki Institute for Information Technology HIIT is a joint research institute of Aalto University and the University of Helsinki for basic and applied research in information technology. Our vision is that by 2030, the Helsinki metropolitan area is recognized as one of the top technology hubs in Europe, known in particular of the top level research conducted within ICT, but also of the productive collaboration networks bringing together top ICT researchers with researchers in other disciplines, and with companies and other organizations. HIIT wishes to be a key player in this context, combining top ICT research of the two universities into a positive collaborating critical mass that creates a strong Helsinki ICT brand, and provides a high-impact collaboration platform linking ICT research to other sciences and other areas of the society. HIIT is led by Professor Petri Myllymäki, who started his 5-year term as the Director of HIIT in August 2015.

Two main goals of HIIT are to (1) conduct high level strategic research bridging the two universities, and (2) coordinate big common ICT issues across departments to serve the Helsinki ICT community as a whole. In year 2016 we were once more successful along both lines. We continued our contribution to three Centres of Excellence of the Academy of Finland: Computational Inference (COIN, overlapping with HIIT's CI programme), Inverse Problems, and Cancer Genetics, and coordinated Re:Know, one of Tekes's strategic research openings that are visionary and challenging projects aiming to make serious breakthroughs that will lay the foundation for entirely new areas of business in the future. HIIT was evaluated twice in 2016, and got high remarks for scientific excellence, innovativeness and impact (see page 24).

Re:Know

HIIT can be considered as a great success story, both Aalto University and University of Helsinki see HIIT as a strategically important activity: the two rectors have signed in October 2015 an updated contract establishing the status of HIIT for the future, and HIIT is also highlighted in the more

COIN Finnish Centre of Excellence in Computational Inference Research

Finnish Centre of Excellence in Inverse Problems Research

The Finnish Center of Excellence in Cancer Genetics Research

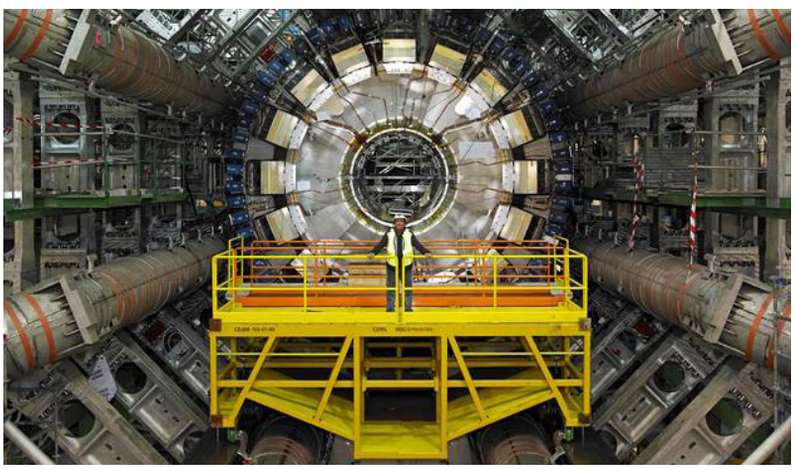
recently signed strategic partnership agreement concerning Aalto and University of Helsinki collaboration on a more general level. Hence it is evident that HIIT has a strong mandate from the both hosting universities to continue its work for combining top ICT research of Aalto University and University of Helsinki into a positive collaborating critical mass that supports an internationally strong Helsinki ICT brand, so that the capital area will be recognized as one of the leading ICT hubs in Europe. During its 15-year history HIIT has tried different organisational models to support this mission, but one thing has become clear: the basic funding of HIIT is not sufficient to sustain a separate department living independently of the relevant ICT departments, but HIIT needs to act more as a collaboration platform facilitating co-operation of the best researchers. One successful model for this type of collaboration is provided by the national Centers of Excellence programme coordinated by the Academy of Finland; The AoF CoE programme model represents a most ambitious scientific yardstick for Finnish science. Inspired by this line of reasoning, we launched in Spring 2016 a call for new HIIT research programmes, aligning it with the Academy CoE Call opened at the same time. As a result of this process, the organizational structure of HIIT's research programmes was renewed completely in 2016, and the new programmes started in August 2016.

HIIT has renewed itself also administratively: after 2016 HIIT is no longer a normal university department with its own staff, but it should be seen more as a strategic initiative providing added value for both universities, enabling activities and challenges that otherwise would not be considered. As the themes of HIIT overlap with several departments of both universities, we do not wish HIIT to be seen as a competitor of the research activities of the departments, but instead as a useful collaboration platform. This means that administratively HIIT has become more like a virtual umbrella organization that does not any more maintain its own premises, IT, or staff for administration, but utilizes the existing services of the universities and their departments. Concretely this re-organization manifested itself at the end of 2016, when the last remaining HIIT researchers (and the Director) moved from the Open Innovation House in Otaniemi to the CS Building that hosts the Computer Science Department of Aalto, as well as some departments of the School of Electrical Engineering. This also means that HIIT researchers are from now own always double-affiliated: the primary affiliation is the department hosting the researcher, and the secondary affiliation can be HIIT, if the research participates in one of HIIT's research programmes or other activities.

## *Research highlights*

Later we will describe the goals of HIIT's research programmes in more detail and give examples of high-quality publications produced in these programmes. Some of these results have also been published as highlight news in HIIT's web pages (see www.hiit.fi/highlights). The idea here is that the highlight news are not only of high scientific value, but demonstrate high-impact potential or may be otherwise of more general interest. Below we will list examples of our highlight results in 2016 that demonstrate the quality and impact of our achievements.

## *Computational social science is coming to Finland, says Helsingin Sanomat*



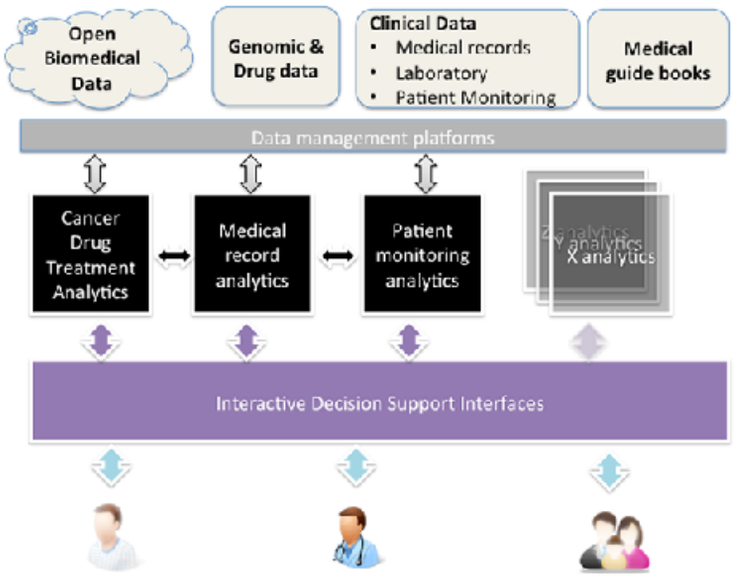
Helsingin Sanomat predicted eight science stories for the year 2016. These included the usual suspects of the Hyperloop, the cousine of the Higgs boson and the Rosetta mission. It also included a highlight about big data and computational methods entering the social sciences, promoting two works HIIT is involved: the Digivaalit 2015 and the Digital Humanities of Public Policy Making. Both of these research projects have applied machine learning to solve questions critical in social science, such as: Who influences the media agenda during elections? What kind of interaction social media affords between the citizen and politicians? How have the party alignments changed during the last 100 years? How can we automate detection of deliberation?

HIIT is also strongly involved in teaching these skills to the new generations of social scientists at University of Helsinki through the computational social science study program, where we teach social scientists to program and apply data-driven tools.

---

## *Big Data can help leukemia patients and prematurely born babies*

D4Health-project tries to figure out how already existing data masses could be utilized efficiently in the treatment of different diseases. The D4Health – Data-Driven Decision Support for Digital Health Care project is funded by the Academy of Finland. It includes the Department of Computer Science and the Department of Electrical Engineering and Automation from the Aalto University, as well as Helsinki Institute for Information Technology HIIT – a joint research institute of the University of Helsinki and the Aalto University - and FIMM – Institute for Molecular Medicine Finland. The responsible researchers include Professor Juho Rousu (Department of Computer Science, Aalto University), Professor Samuel Kaski (Helsinki Institute for Information Technology HIIT, Aalto University), Professor Giulio Jacucci (Helsinki Institute for Information Technology, University of Helsinki), Tero Aittokallio (Institute for Molecular Medicine Finland FIMM, University of Helsinki), and Professor Raimo Sepponen (Department of Electrical Engineering and Automation, Aalto University).




The funding of the project is 1,2 million euros. It aims at developing methods for combining large medical datasets and prediction models, as well as creating flexible user interfaces that support physicians in diagnosing diseases and planning the treatment. Medical staff has an important role in this project.

---

## *HIIT presence in ACM conference on Intelligent User Interfaces 2016*

Results from the HIIT wide focus area research on user modeling in exploratory search will be published in the IUI'16 conference this year. Our papers are titled: "Interactive intent modeling from multiple feedback domains", "Beyond Relevance: Adapting Exploration/Exploitation in Information Retrieval", "RelaWorld: Neuroadaptive and Immersive Virtual Reality Meditation System", and "Dealing with Concept Drift in Exploratory Search: An Interactive Bayesian Approach".

---

## *Professor Jukka Corander becomes Visiting Fellow at University of Cambridge*

Professor Jukka Corander has been elected as a Visiting Fellow at the Churchill College of University of Cambridge. He spent 4 months at the College during 2016 to do research in infectious disease epidemiology and genomics. Churchill College is renowned for its excellence in Science and Technology; in total 29 Nobel prize winners have been Fellows of the College.

## *Spaceify: A new HIIT start up to webify spaces*


The aim of Spaceify is to commercialise solutions that provide immediate interaction with smart spaces through web technologies. The core of the solution is a client-edge-server ecosystem that seamlessly integrates physical spaces with the web. Spaceify can give web apps controlled access to resources such as big screens, sensors and lighting in the space.

These solutions can provide benefit in a variety of settings in retail, exhibitions, meeting rooms, museums, and education. For example, Spaceify to turn passive advertising screens into a social games platform, where the audience can interact with the advertisements using their smart phones.

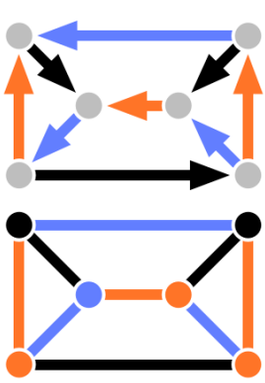
The startup originates from the Network Society Program at HIIT, from research carried out starting in 2012 in the TEKES FiDiPro project with Professor Sumi Helal,University of Florida, “Smart Spaces: Critical Software Enablers for Real-World Deployments”.

The founding of Spaceify Oy is an example of long-term strategic research resulting in an impact beyond academic results. The open-sourced Spaceify smart space ecosystem the company aims to commercialise, is still under active development in the EIT Digital Street Smart Retail High Impact Initiative. (http://www.spaceify.org/)

From the NS program, key persons include: Petri Savolainen, Kai Kuikkaniemi, Max Vilkki, Jouni Ojala, Sasu Tarkoma, Jouni Vepsäläinen, Giulio Jacucci, Sumi Helal, and Marko Turpeinen.

---

## *Breakthroughs in Foundations of Distributed Computing*



In distributed computing, many computational problems are inherently global: to solve such a problem, we need to get full information on the structure of the computer network. There are also some problems that can be solved in a strictly local manner: it is enough that each computer is aware of its local neighbourhood (e.g, of other computers that are within some small constant number of hops away).

While the two extremes — global problems and local problems — are nowadays fairly well understood, much less is known about problems that fall between the two extremes. Our recent work “A lower bound for the distributed Lovász local lemma” now gives the first genuine example of a problem of an intermediate time complexity, falling strictly between local

problems and global problems, even in the case of networks in which all nodes have only a small number of neighbours. This work was accepted for presentation at STOC 2016, which is one of the most prestigious and selective conferences in theoretical computer science.

---

## *IWOCA 2016 - 27th International Workshop on Combinatorial Algorithms*


IWOCA 2016 will be held in Helsinki, Finland, 17-19 August, 2016.

IWOCA is a forum for researchers who design algorithms for the myriad combinatorial problems that underlie computer applications in science, engineering and business.

- A special issue of selected papers will be published in the journal Theory of Computing Systems.
- There will be a Best Student Paper award, sponsored by the EATCS. Please see the rules for eligibility.
- As in past editions, the proceedings of IWOCA 2016 will be published by Springer in the Lecture Notes in Computer Science (LNCS) series.
- The main conference will be 17 August to 19 August (Wednesday to Friday).
- There will be an edition of the StringMasters workshop on 16 August (Tuesday), free to all IWOCA attendees.

The workshop is partially funded by HIIT ADA.

## *Film screening: Code - Debugging Gender Gap*

The Silicon Valley wisdom says that software is eating the world. However, in major US companies women are not working on software. Facebook reports 15% of its technical workforce being female, Twitter 10% and the US national avarage is around 23%. This is known as the gender gap in technology fields. It is widely discussed both in US media and pops up every now and then in Finnish media as well.

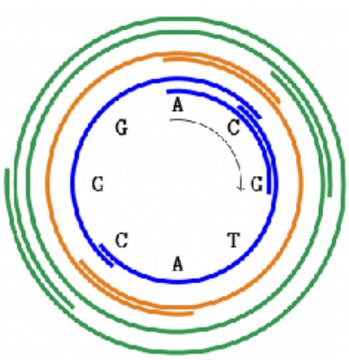

To start to consider this problem as a crown and highlighting it's importance, we organize a film screening of "Debugging the Gender Gap" in Helsinki.

Tech jobs are growing three times faster than our colleges are producing computer science graduates. By 2020, there will be one million unfilled software engineering jobs in the USA. Through compelling interviews, artistic animation and clever flashpoints in popular culture, CODE documentary examines the reasons why more girls and people of color are not seeking opportunities in computer science and explores how cultural mindsets, stereotypes, educational hurdles and sexism all play roles in this national crisis. Expert voices from the worlds of tech, psychology, science, and education are intercut with inspiring stories of women who are engaged in the fight to challenge complacency in the tech industry and have their voices heard. CODE aims to inspire change in mindsets, in the educational system, in startup culture and in the way women see themselves in the field of coding

The film has also been reviewed e.g. in IEEE Spectrum. The screen is followed with a discussion around this theme, iwith members from the industry, the academia and students themselves reflecting their experiences.

---

## *Theoretical Breakthrough in Genome Assembly*




One of the oldest bioinformatics problems is to reconstruct the genome of a species from short fragments, such as those produced by high-throughput sequencing. Due to various technical limitations, it is currently impossible to fully reconstruct an entire genome. State-of-the-art genome assemblers in fact produce long genomic fragments that are "guaranteed" to occur in the genome that generated the data. A major question, originating more than 20 years ago, is to characterize all the information that can be safely assembled in this way.

Research conducted by Alexandru Tomescu (Genome-Scale Algorithmics group, HIIT sub-programme Algorithmic Data Analysis) and Paul Medvedev (Penn State University, USA) solved this problem, by obtaining a mathematical characterization of all these long fragments. As a

consequence, this result also provides the first tight upper-bound on what can be safely assembled from input data. We expect these theoretical results to gradually make their way into practical genome assemblers.

This research was published in a paper "Safe and complete contig assembly via omnitigs" and presented at the conference RECOMB 2016 - The 20th Annual International Conference on Research in Computational Molecular Biology. This is one of the most prestigious and selective conferences in Bioinformatics.

At the same conference, Ahmed Sobih, Alexandru Tomescu, and Veli Mäkinen from the Genome-Scale Algorithmics group also presented another paper on finding the bacterial composition of a high-throughput sequencing sample taken from an environment, such as the human gut. Their method "MetaFlow: Metagenomic profiling based on whole-genome coverage analysis with min-cost flows" was shown to be more precise and sensitive than current popular tools.

---

## *Social media image analysis for public health*

Researchers in HIIT, CMU, and QCRI use deep-learning techniques and images from social-media platforms to monitor public health. More details are given in the research paper that will appear in CHI 2016:

Several projects have shown the feasibility to use textual social media data to track public health concerns, such as temporal influenza patterns or geographical obesity patterns. In this paper, we look at whether geo-tagged images from Instagram also provide a viable data source. Especially for "lifestyle" diseases, such as obesity, drinking or smoking, images of social gatherings could provide information that is not necessarily shared in, say, tweets. In this study, we explore whether (i) tags provided by the users and (ii) annotations obtained via automatic image tagging are indeed valuable for studying public health. We find that both user-provided and machine-generated tags provide information that can be used to infer a county's health statistics. Whereas for most statistics user-provided tags are better features, for predicting excessive drinking machine-generated tags such as "liquid" and "glass" yield better models. This hints at the potential of using machine-generated tags to study substance abuse.

---

## *Popularize computational social science*

Helsingin Sanomat, the leading newspaper in Finland, run a full page story of Digivaalit 2015 and computational tools in social science. The story is based on interview with Matti Nelimarkka (HIIT) and Salla-Maaria Laaksonen (University of Helsinki, Faculty of Social Science). They describe details of the methods used to examine agenda influencers during the 2015 elections - results which were published e.g. in Helsingin Sanomat.

The aim is to demystify data science and present to the public how results are computed. As Salla said, the press and the public are interested on the results, but not on the methods how the results was achieved. This is unfortunate, as data science can deliver almost any result one might imagine. According to Matti, the methods nor the code in this study are complicated nor fancy. Instead, the value emerges from integrating with social science context and theories and this way, critically evaluating what is measured.

---

## *#DHH16 Helsinki Digital Humanities Hackathon*

#DHH16 — Helsinki Digital Humanities Hackathon 2016 — was jointly organised by HIIT, University of Helsinki, and Aalto University on 16–20 May 2016. The hackathon brought together students and researchers of humanities, social sciences and computer science for one week of intensive collaboration.

The groups studied emigration discourse in Finnish newspapers in 1870–1910, Finnish public tv broadcasting in 1957–1990, the development of the Pasila neighbourhood in Helsinki in 1972–1989, and text recycling in English books in 1473–1700.

DIGITAL HUMANITIES HACKATHON
MAY 2016 HELSINKI

## *HIIT Open Programming Contest*

13 teams took part in HIIT Open 2016 programming contest on 28 May 2016 in Otaniemi. The contest was open to everyone interested in programming and algorithmic challenges — we had 34 participants in total, among them students or faculty from at least 6 different universities or universities of applied sciences and 3 different high schools, as well as some participants from the industry.

The winning team was “Game of Nolife”, with Tuukka Korhonen, Otte Heinävaara, and Olli Hirviniemi, all of them students at the University of Helsinki. The 2nd place went to “Ace of Spades”, with Janne Junnila (University of Helsinki), Tommi Pesu (Imperial College), and Kalle Luopajärvi (Seinäjoki upper secondary school), and the 3rd place went to “Noname 01”, with Ilya Nikolaevskiy (Aalto University) and Khaled Gamal Abdelnaser (German University in Cairo).

The teams had 12 tasks to solve, and 5 hours of time. The winning team solved 9/12 tasks correctly, and the two runner-ups solved 8/12 tasks each, with “Ace of Spades” using less time in total than “Noname 01”.

---

## *Renewed Carat app gives a new smart boost to battery*

The Carat Project Team at the University of Helsinki, Department of Computer Science, has published a new version of the popular mobile energy-awareness application.

After launch in June 2012, Carat has helped over 850,000 users, of which 41 per cent have been Android and 59 per cent iOS users, respectively. The new user interface follows modern application design guidelines and presents battery information in a more intuitive and easy to use manner.

- In addition to the new user interface, we have increased the accuracy of the energy saving recommendations of Carat, says Professor Sasu Tarkoma, the leader of this research done at the university.

The user interface features the number of energy intensive applications (Hogs), energy anomalies (Bugs) and user recommendations (Actions) at a glance on the main screen as well as global energy statistics for the device community.

Carat is a free application that tells you what is using up the battery of your mobile device, whether that’s normal, and what you can do about it. Carat helps you avoid certain energy-hungry applications and compare your battery lifetime to other users. Based on the energy usage data, the Carat Project research team continuously develops new techniques to solve mobile device related energy consumption and battery problems.

The new application version has better support for new versions of the Android operating system, including Android 5.0 and 6.0. The new version of Carat is currently available for download in Google Play for most Android devices, and the iOS version will be soon published for iOS devices in the App Store.

## *Radiation influences Aalto-1 nanosatellite software design*

In his Master's Thesis, Joonas Javanainen examined the level of preparedness the additional challenges of space impose on software design. The Master's Thesis was supervised by HIIT's programme director Keijo Heljanko.

The Aalto-1 nanosatellite uses fairly ordinary electronic components which are not specially designed to withstand radiation in space. However, the degree of reliability can be increased through good software architecture which takes into account possible failure scenarios caused by the environment.

'Joonas Javanainen, an information technology student, carried out some excellent work and in his thesis he analyzed effectively the Aalto-1 satellite's software and its failure recovery mechanisms. The satellite's software uses the Linux system and the software has been designed and written by students from Aalto University', explains Professor Jaan Praks who is responsible for the Aalto-1 nanosatellite project.

'As the result of radiation in space, a single bit can be changed somewhere in the computer. The errors are caused by the environment and they would happen anyway, even if the programming was perfect', points out Joonas Javanainen.

Even if a bit is changed because of radiation, it does not necessarily affect the functioning of the satellite, but in the worst case it could cause calculation errors or crash the software. On the other hand, radiation can also cause permanent damage to the functioning of the equipment.

## *Järvisalo to deliver IJCAI-16 Early Career Spotlight as one of the most active early career AI researchers in the world*


Dr. Matti Järvisalo has been invited to deliver one of the Early Career Spotlight Talks at IJCAI-16, 22nd International Joint Conference on Artificial Intelligence taking place In New York City, USA, in July 2016. IJCAI is the prestigious number-one conference on artificial intelligence world-wide. The Early Career Spotlight Track at IJCAI-16 brings in---as some of the most active early career researchers in AI in the world---22 top talents in AI, most already holding professorships at world-leading research universities, as representatives of different modern branches of AI research to present on their individual research directions.

Quoting the official IJCAI-16 website:

Since its inception in 1969, IJCAI took great pride in representing all aspects of AI research. Even now, with a myriad specialized conferences, IJCAI remains uniquely suited to give the attendees an integrated snapshot of the whole of AI enterprise. In keeping with this goal, we are excited to announce a special track of “Early Career Spotlight” talks. The talks are aimed at providing an accessible introduction to the research directions of some of the most active early career researchers in all representative areas of AI.

The talks are by invitation, based on nominations from the IJCAI program committee, and will run in parallel to the technical sessions. It is our great pleasure to introduce the stellar speakers giving the Early Career Spotlight tracks at this leap year IJCAI.

---

## *BREW is a learning workshop organized by PhD students in bioinformatics*

BREW 2016 Bioinformatics Research and Education Workshop was organized in Helsinki in May. BREW is a series of annual workshops for PhD students in bioinformatics and computational biology, which aims to give an introduction to scientific conferences, including submission, peer review and presentation of scientific papers. All participants both present a paper and participate in the programme committee, reviewing papers from other participants.

‘

The first BREW was organized in 2002, 14 years ago. The basic format is the same, it is just carried out in different places each year. There are a limited number of universities participating: Aalto University, University of Helsinki through the HICT Doctoral Network, as well as the universities of Bielefeld, Berlin, Bergen, Novosibirsk and Tartu. This year there were altogether 28 participants', tells professor Juho Rousu from Department of Computer Science at Aalto University, who oversaw the organization with professor Veli Mäkinen from University of Helsinki.

The organizing committee in Finland included three PhD students, Anna Kuosmanen from University of Helsinki, Viivi Uurtio and Anna Cichonska from Aalto University, as well as post-doctoral researcher Travis Gagie from University of Helsinki. Invited talks include Matti Pirinen and Simon Anders from the Institute for Molecular Medicine Finland, FIMM, and Esko Ukkonen from the University of Helsinki. Ukkonen is one of the founding members of BREW together with professor Robert Giegerich from University of Bielefeld.

'I participated in BREW the first time as a PhD student in Bergen, Norway, in 2007. That was one of my first presentations, a very good opportunity for practicing. The feedback was honest and direct, I personally consider BREW a very good learning experience. In 2009 I organized BREW myself in Helsinki. BREW allows you to connect with researchers in the field', explains Academy of Finland Research Fellow Matti Pirinen.

The distinctive factor about BREW is that it is almost completely organized by the PhD students themselves. The role of the steering group of BREW is to involve the professors, provide support for the organising team, make sure that researchers from the departments will participate in BREW, and also put positive energy into the workshop. The aim of BREW is to mimic workshops as realistically as possible.

'In the beginning of my academic career, I was asked to chair a real workshop session myself without any advice and without ever being to a workshop before. I had no questions prepared and I did not receive any feedback afterwards', explains Rousu.

'In BREW everyone learns how to be a chair person in a workshop, and it has been surprising how good all the participants are at chairing. There is also a feedback form for the presentation skills', adds Viivi Uurtio, Doctoral Candidate at Aalto University.

BREW is open for students and supervisors from the PhD programs in Bioinformatics and Computational Biology in the participating universities. The workshop in May was hosted by the University of Helsinki at the Kumpula Campus and sponsored by Aalto University and the University of Helsinki through the Helsinki Doctoral Education Network in Information and Communications Technology HICT and by the University of Helsinki through the Doctoral School in Computer Science DoCS.

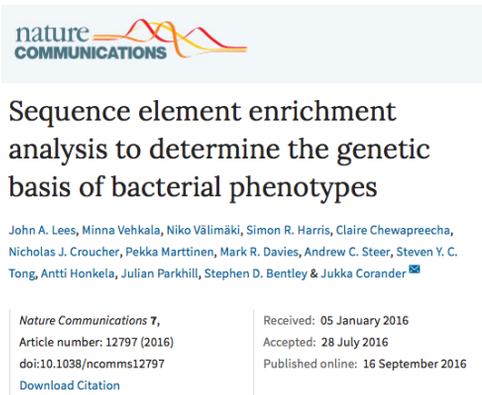
## *HIIT participated to the 10th International Workshop on Machine Learning in Systems Biology*


The tenth edition of MLSB was organized as a two-day satellite meeting before ECCB2016, the European Conference on Computational Biology (September 5-7, 2016) by Dick de Ridder and Aalt-Jan van Dijk (Wageningen University, The Netherlands) and Juho Rousu and Harri Lähdesmäki (Aalto University, Finland) at the World Forum in The Huage. The workshop was generously sponsored by contributions from NWO Exact Sciences, the Helsinki Institute of Information Technology (HIIT) and five companies: BaseClear, Bayer, Enza Zaden, Philips and RijkZwaan.

Biology is rapidly turning into an information science, thanks to enormous advances in the ability to observe the molecular properties of cells, organs and individuals. This wealth of data allows us to model molecular systems at an unprecedented level of detail and to start to understand the underlying biological mechanisms. This field of systems biology creates a huge need for methods from machine learning, which find statistical dependencies and patterns in these large-scale datasets and that use them to establish models of complex molecular systems. MLSB, a series of workshops, is a scientific forum for the exchange between researchers from Systems Biology and Machine Learning, to promote the exchange of ideas, interactions and collaborations between these communities.

The programme contained four keynote talks by internationally recognized top scientists in the field: (1) Prof. Lodewyk Wessels, The Delft Bioinformatics Lab, Delft University of Technology, The Netherland (2) Prof. **Jukka Corander,** Department of Mathematics and Statistics, University of Helsinki, Finland (3) Prof. Yvan Saeys, VIB Inflammation Research Center, Ghent University, Belgium (4) Prof. Ziv Bar-Joseph, School of Computer Science Carnegie--Mellon University, USA.

## *HIIT scientists present the most advanced GWAS method for bacteria to date*



nature COMMUNICATIONS

Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes

John A. Lees, Minna Vehkala, Niko Välimäki, Simon R. Harris, Claire Chewapreecha, Nicholas J. Croucher, Pekka Marttinen, Mark R. Davies, Andrew C. Steer, Steven Y. C. Tong, Antti Honkela, Julian Parkhill, Stephen D. Bentley & Jukka Corander

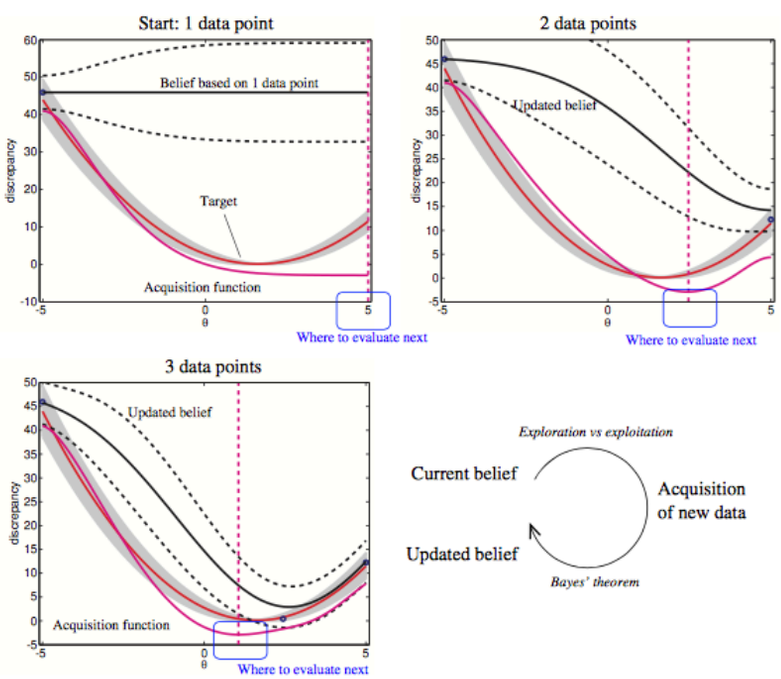
*Nature Communications* **7**, Article number: 12797 (2016)
doi:10.1038/ncomms12797
Download Citation

Received: 05 January 2016
Accepted: 28 July 2016
Published online: 16 September 2016

A HIIT-wide team led by professor Jukka Corander included several scientists from both UH and Aalto with a joint mission to create the most advanced and computationally best scalable method for genome-wide association (GWAS) studies in bacteria. The team had a close collaboration with the Pathogen Genomics Group at the Wellcome Trust Sanger Institute where GWAS is an important step towards unraveling the secrets behind evolution and success of numerous major human pathogens from large-scale population genomic data. The new method, SEER, was published in Nature Communications on September 16 and it is already rapidly gaining worldwide popularity among microbiologists. Since SEER is scalable to even tens of thousands of bacterial genome sequences, it will open a totally new era in bacterial GWAS.

---

## *Giant leap in ABC inference scalability*



HIIT scientists Michael Gutmann and Corander published a machine learning based ABC inference approach in the Journal of Machine Learning Research. Their method (BOLFI) is based on Bayesian optimization with Gaussian processes and is generally applicable to simulator models with intractable likelihoods. Without sacrificing accuracy, BOLFI speeds up posterior computation by 3-4 orders of magnitude compared with the state-of-the-art sequential Monte Carlo algorithms. It is expected to become a new standard in ABC inference, paving way for a multitude of new applications where the earlier methods have been too expensive computationally. Details of the method can be found in the article: Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models. Michael U. Gutmann, Jukka Corander; 17(125):1−47, 2016.

---

## *Regression modelling reconstructs weather forecasts for the past from animal teeth*

In the new study, the annual rainfall and average temperatures in the national parks are inferred from the teeth of herbivorous mammals. Such reverse engineering opens up new opportunities for interpreting fossil records. The results were recently published in the journal PNAS.

Exact data on the number and geographical spread of the animals in Kenya's national parks have been collected over the course of the past 60 years. The data used for the modelling in this study include the mammals in Kenya's 13 national parks, dental traits, and rainfall, temperature and the amount of vegetation.

"We computed the average dental traits for each area and related them in relation to environmental factors using regression models. The model's generalizability was tested, and the predictability of individual environmental factors was compared," explains researcher Indrė Žliobaitė, who was responsible for the modelling.

The research shows that features in animal teeth are particularly good at detecting where the weather has been unfavorable for the species in question. Such weather conditions include long dry periods, heavy rains or exceptionally low temperatures – anything that could result in the animal's primary food source becoming unavailable, forcing the animals to turn to less preferred plants to survive.

The researchers were particularly interested in why animals were absent from a particular geographical area.

"African national parks frequently endure poor years, which seem to prevent the establishment of permanent populations of certain animals. Animals live where the conditions allow them to live and reproduce over the span of decades or centuries," says Mikael Fortelius, professor of evolutionary paleontology at the University of Helsinki.

"We would be able to go millions of years past and see what the weather conditions were like back then, and that we have actually done in another research article published earlier this year in The Royal Society Publishing including the temperature curves of the past millions of years", adds Žliobaitė.

The role of a computer scientist in studies like this is to deal with high-dimensional datasets and make the estimates, as evolutionary paleontologists make the analysis and interpretations based on the estimates. This work follows the collaboration between professor of evolutionary paleontology Mikael Fortelius in the University of Helsinki and professor Heikki Mannila from the Aalto University Department of Computer Science, now the President of the Academy of Finland, started about 10 years ago.

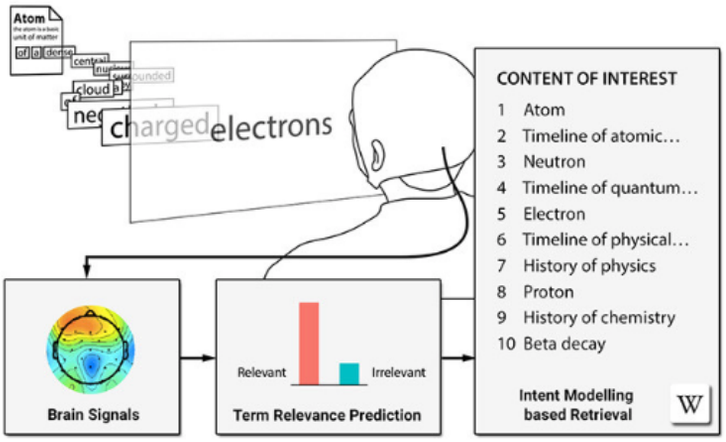
---

## *EEG reveals information essential to users*

In a study conducted by the Helsinki Institute for Information Technology (HIIT) and the Centre of Excellence in Computational Inference (COIN), laboratory test subjects read the introductions of Wikipedia articles of their own choice. During the reading session, the test subjects' EEG was recorded, and the readings were then used to model which key words the subjects found interesting.

'The aim was to study if EEG can be used to identify the words relevant to a test subject, to predict a subject's search intentions and to use this information to recommend new relevant and interesting documents to the subject. There are millions of documents in the English Wikipedia, so the recommendation accuracy was studied against this vast but controllable corpus', says HIIT researcher Tuukka Ruotsalo.

Due to the noise in brain signals, machine learning was used for modelling, so that relevance and interest could be identified by learning the EEG responses. With the help of machine learning methods, it was possible to identify informative words, so they were also useful in the information retrieval application.




‘Information overload is a part of everyday life, and it is impossible to react to all the information we see. And according to this study, we don’t need to; EEG responses measured from brain signals can be used to predict a user’s reactions and intent’, tells HIIT researcher Manuel Eugster.

Based on the study, brain signals could be used to successfully predict other Wikipedia content that would interest the user.

‘Applying the method in real information retrieval situations seems promising based on the research findings. Nowadays, we use a lot of our working time searching for information, and there is much room in making knowledge work more effective, but practical applications still need more work. The main goal of this study was to show that this kind of new thing was possible in the first place’, tells Professor at the Department of Computer Science and Director of COIN Samuel Kaski.

‘It is possible that, in the future, EEG sensors can be worn comfortably. This way, machines could assist humans by automatically observing, marking and gathering relevant information by monitoring EEG responses’, adds Ruotsalo.

The study was carried out in cooperation by the Helsinki Institute for Information Technology (HIIT), which is jointly run by Aalto University and the University of Helsinki, and the Centre of Excellence in Computational Inference (COIN). The study has been funded by the EU, the Academy of Finland as a part of the COIN study on machine learning and advanced interfaces, and the Revolution of Knowledge Work project by Tekes.

## *Supporting the Helsinki ICT community*

HIIT collaborates actively both within Helsinki area and internationally, and acts as a link between researchers, companies and public administration. In Helsinki area, HIIT coordinates common activities of computer science research in several ways: education, recruitments, and joint initiatives.

In doctoral education, HIIT coordinates the Helsinki Doctoral Education Network in Information and Communications Technology (HICT). HICT is hosted jointly by Aalto University and the University of Helsinki, the two leading universities within this area in Finland. The network serves as a collaboration platform for doctoral education combining all the relevant subfields of computer science and information technology at Aalto University and the University of Helsinki. It involves at present 65 professors and almost 300 doctoral students, and the participating units produce altogether more than 40 new doctors each year.

Recruitment is an important means of collaboration. HIIT arranges joint Helsinki IT postdoc and senior researcher calls together with several departments in the field of computer science at Aalto University and University of Helsinki. These joint calls leverage on and further enhance the brand of Helsinki as a hub of computer science, and attract a wide range of high-level applicants.

Digi Platform is Aalto University's collaborative initiative in the field of ICT and digitalization. It is chartered to innovate, initiate and increase digitalization related cooperation broadly inside Aalto and with its industrial and academic partners. The Digi Platform facilitates and brings together Aalto's competences in this multidisciplinary area, in a bottom up manner, and increases Aalto's visibility. Expertise across Aalto is brought together to maximize the potential of IT to boost other sciences, technologies and society. The Digi Platform provides seed funding, arranges matchmaking events, acts as a collaboration platform within Aalto, and provides a "front desk" for external contacts. HIIT coordinates the Digi Platform, and University of Helsinki participates in the platform via HIIT.

All HIIT's activities are inherently international, mostly in a bottom- up manner. Cooperation with key international research institutes and universities is active, and the collaborators include Berkeley (ICSI and UC Berkeley), MIT, Centre for Computational Statistics and Machine Learning (CSML) at UCL, European Bioinformatics Institute EBI, Human Technology Lab (HTLab) at University of Padova, and Waseda University, to name a few. HIIT also coordinates Aalto University's and University of Helsinki's strategic partnership with NSF Science and Technology Center for Science of Information (SoI). The member universities of the SoI center are Purdue (leader), Bryn Mawr, Howard, MIT, Princeton, Stanford, Texas A&M, UC Berkeley, UCSD and Urbana-Champaign. Other strategic partners in Europe are ETH (Zürich, Switzerland) and the LINCS network (Paris, France).

EIT Digital is an European initiative intended to turn Europe into a global leader in ICT innovation. It aims to fulfill this mission by establishing a new type of partnership between leading companies, research centres, and universities in Europe. One of the 5 nodes is Helsinki. Several HIIT researchers have research projects with EIT Digital.

HICT

Aalto Digi Platform

Center for Science of Information
NSF Science and Technology Center

National Science Foundation

eit Digital

For increasing the international visibility of the Helsinki ICT community, HIIT is coordinating a high-profile lecture series on Future Information Technologies (see http://www.hiit.fi/HelsinkiITLectures). The idea is not to run yet another series of scientific guest lectures, but to attract a more versatile audience and focus on highlighting the research challenges and solutions faced by current and future information technologies, as seen by the internationally leading experts of the field. The speakers are typically esteemed visionaries with an academic background (e.g. the CTO of an IT company), or university professors with an exceptionally high societal or industrial impact.

## *HIIT evaluations in 2016*

The Scientific Advisory Board (SAB) of HIIT visited Helsinki on September 22-23 for a thorough evaluation of HIIT's progress. In addition, the Scientific Advisory Board of the School of Science of Aalto visited Otaniemi on October 25-26 to evaluate the whole School, and HIIT was again evaluated as part of this process.

**The HIIT SAB consisted of the following persons**:

- Prof. John Shawe-Taylor, UCL (Chair)
- Prof. Jodi Forlizzi, Carnegie Mellon University
- Prof. Kari-Jouko Räihä, University of Tampere
- Prof. Michael Mahoney, UC Berkeley
- Prof. Wojciech Szpankowski, Purdue University
- Prof. Seif Haridi, KTH SICS
- Prof. Elisabeth André, University of Augsberg
- Prof. Richard Durbin, Sanger Institute

**The Aalto School of Science SAB consisted of the following persons:**

- Prof. Arto Nurmikko, Brown University (Chair)
- Dr. George Drettakis, REVES/INRIA Sophia-Antipolis
- Prof. Paula Eerola, University of Helsinki
- Prof. Martin Grötschel, The Berlin-Brandenburg Academy of Sciences and Humanities
- Prof. Martin Ingvar, Karolinska Institutet
- Prof. Carlo Ghezzi, Politecnico di Milano
- Prof. Sir Peter Knight, Imperial College
- Prof. Jan van Leeuwen, Utrecht University
- Prof. Georges Romme, Eindhoven University of Technology
- Prof. John Shawe-Taylor, University College London

Overall, the feedback from both SABs was very positive. The HIIT SAB noted that “HIIT has a track record of outstanding research at very high international standards, reflected by an excellent publication record at top tier international conferences and highly reputed journals, a striking number of awards (ERC grant, IJCAI Early Career Spotlight, Best Paper Awards etc.), and successful acquisition of project funding.” Also the role of HIIT and the new organizational model of HIIT was commended: the Aalto SCI SAB noted that “HIIT is well-positioned for playing a crucial role as a catalyst for increased collaboration and an attractor of further funded activities.” and the HIIT SAB noted that “The new challenge-driven objectives provide an excellent framework within which to foster collaboration among researchers and to show the added value of HIIT.”

The main concerns raised by the SABs were related to gender imbalance, transparency and openness of HIIT’s operations, the relationship with respect to hosting organizations, and how to monitor and reward success while maintaining continuous renewal. These are all areas where we can improve a lot, and we have already initiated processes aiming for this.

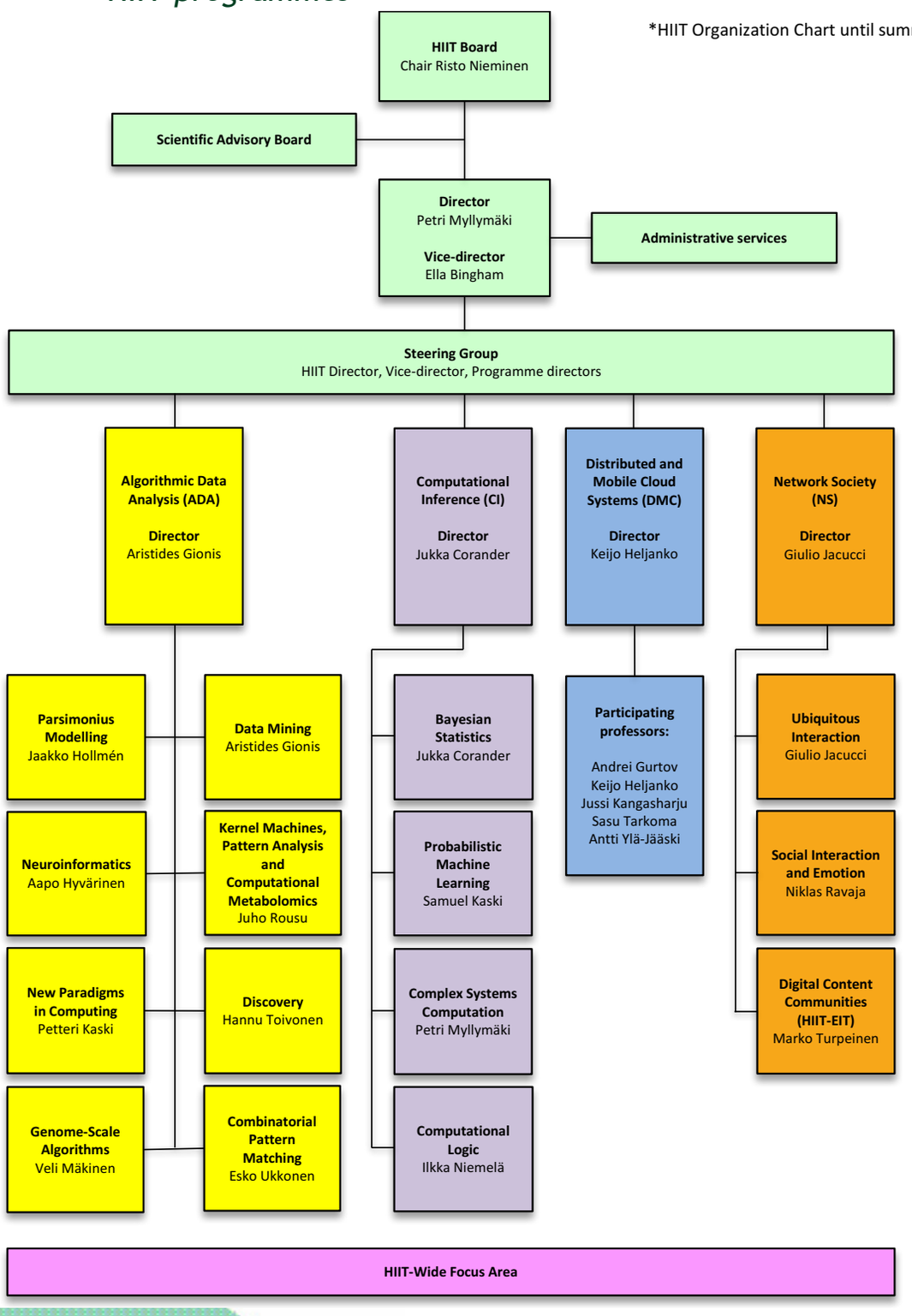
## *Looking forward: outlook for year 2017*

We are quite happy with our new research programmes addressing high-risk grand challenges that represent high-quality high-impact collaborative research, demonstrating clear added value and supporting strongly the “Helsinki ICT” mission of HIIT, and the strategies of the hosting universities. We would like to continue supporting these programmes and other similar initiatives with the hope that they will participate in the future CoE calls. At the same time, it should be emphasized that the current HIIT research programmes represent only a tip of an ice berg: they are top-quality research, but there is also a lot of good research done outside the research programmes. The mission for 2017 is the recognize the areas and forms of collaboration where HIIT can make a big impact, and create suitable instruments for supporting these activities.

# Research Programmes

## *HIIT programmes*

*HIIT Organization Chart until summer 2016



## *Algorithmic Data Analysis (ADA) Programme*

The mission of the Algorithmic Data Analysis (ADA) research programme at HIIT is to develop useful algorithmic data analysis methods for other sciences and for industry. The work involves both basic research in computer science and applied work on problems arising from applications. The research focuses on the algorithmic and modeling problems of combinatorial pattern matching, data mining, and machine learning. The work is strongly interdisciplinary. Developing new concepts and algorithms is an iterative process consisting of interacting extensively with the application experts, formulating computational concepts, analyzing the properties of the concepts, designing algorithms and analyzing their performance, implementing and experimenting with the algorithms, and applying the results in practice. The main application areas of ADA are in biology, medicine, environmental studies, creativity, social networks, and neuroscience and natural language processing.

## Combinatorial Pattern Matching, Professor Esko Ukkonen

The combinatorial pattern matching group develops combinatorial algorithms for pattern search and synthesis problems for sequential and higher dimensional data. Probabilistic modeling and machine learning methods for pattern analysis are also considered. The group is interested in the basic research of the theoretical aspects of the area as well as in various applications, mostly in bioinformatics and information retrieval. Recent results include error correction methods for the so-called long reads in DNA sequencing, fast scanner for first-order Markov models of DNA binding as well as a probabilistic model and learning algorithm for dimeric binding motifs of transcription factors in DNA.

### *Key publications:*

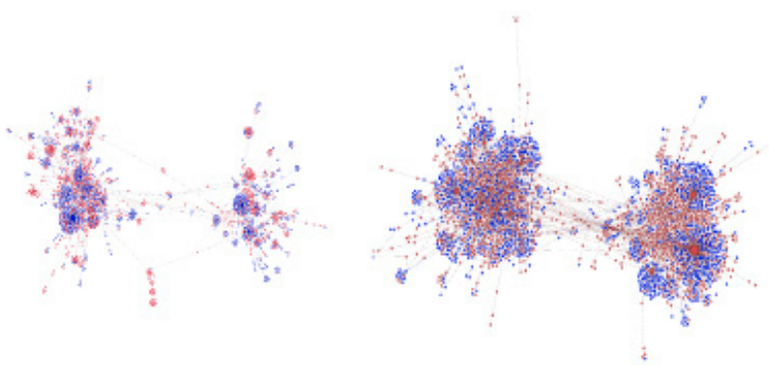
- Leena Salmela, Riku Walve, Eric Rivals, and Esko Ukkonen. Accurate self-correction of errors in long reads using de Bruijn graphs. Bioinformatics 33, 799-806, 2017.
- Tomas Flouri, Emanuele Giaquinta, Kassian Kolbert, and Esko Ukkonen. Longest common substrings with k mismatches. Information Processing Letters. 115, 6-8, 643–647, 2015.
- Virpi Ahola, Rainer Lehtonen, Panu Somervuo, Leena Salmela et al. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nature Communications 5, 4737 (9 pages), 2014.
- Arttu Jolma, Jian Yan, Thomas Whitington, Jarkko Toivonen, Kazuhiro R. Nitta, Pasi Rastas, Ekaterina Morgunova, Martin Enge, Mikko Taipale, Gonghong Wei, Kimmo Palin, Juan M. Vaquerizas, Renaud Vincentelli, Nicholas M. Luscombe, Timothy R. Hughes, Patrick Lemaire, Esko Ukkonen, Teemu Kivioja, and Jussi Taipale. Dna-binding specifities of human transcription factors. Cell 152, 1-2, 327–339, 2013.

## Data Mining,
## Professor Aristides Gionis

The data-mining group focuses on developing novel methods to extract knowledge from data, designing algorithms to summarize large volumes of data efficiently and effectively, and exploring new ways of using the extracted information. Specific areas of interest include: pattern discovery, clustering and outlier detection, graph mining, social-network analysis, analysis of information networks and social-network dynamics, and analysis of smart-city sensor data.

### *Key publications:*


- Albert Arockiasamy, Aristides Gionis, and Nikolaj Tatti. A Combinatorial Approach to Role Discovery. ICDM: 787-792, 2016.
- Gianmarco De Francisci Morales, and Aristides Gionis. Streaming Similarity Self-Join. PVLDB 9(10): 792-803, 2016.
- Esther Galbrun, Aristides Gionis, and Nikolaj Tatti. Top-k overlapping densest subgraphs. Data Min. Knowl. Discov. 30(5): 1134-1165, 2016.
- Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying Controversy in Social Media. WSDM: 33-42, 2016.
- Polina Rozenshtein, and Aristides Gionis. Temporal PageRank. ECML/PKDD (2): 674-689, 2016.
- Polina Rozenshtein, Aristides Gionis, B. Aditya Prakash, and Jilles Vreeken. Reconstructing an Epidemic Over Time. KDD: 1835-1844, 2016.

## Discovery Group: Data Mining and Computational Creativity, Professor Hannu Toivonen

The Discovery research group develops novel methods and tools for data science and for computational creativity. Our focus is on algorithmic methods for discovering links and patterns in data, and on automated creativity in different areas. In the intersection of these fields, we develop creative systems and applications that learn to create.

### *Key publications:*

- Oskar Gross, Antoine Doucet, and Hannu Toivonen. Language-Independent Multi-Document Text Summarization with Document-Specific Word Associations. The 31st Annual ACM Symposium on Applied Computing (ACM SAC), 853-860. Pisa, Italy, April 2016.
- Anna Kantosalo and Hannu Toivonen. Modes for Creative Human-Computer Collaboration: Alternating and Task-Divided Co-Creativity. The Seventh International Conference on Computational Creativity (ICCC), 77-84. Paris, France, June-July 2016.
- Simo Linkola, Tapio Takala, Hannu Toivonen. Novelty-Seeking Multi-Agent Systems. The Seventh International Conference on Computational Creativity (ICCC), 1-8. Paris, France, June-July 2016.
- Eric Malmi, Pyry Takala, Hannu Toivonen, Tapani Raiko, Aristides Gionis. DopeLearning: A Computational Approach to Rap Lyrics Generation. 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 195-204. San Francisco, CA, 2016.
- Alessandro Valitutti, Antoine Doucet, Jukka M. Toivanen, Hannu Toivonen. Computational Generation and Dissection of Lexical Replacement Humor. Natural Language Engineering. 22 (5): 727–749. 2016.
- Tony Veale, Khalid Al-Najjar. Grounded for life: creative symbol-grounding for lexical invention. Connection Science 28 (2), 139-154. 2016.
- Ping Xiao, Khalid Alnajjar, Mark Granroth-Wilding, Kathleen Agres, Hannu Toivonen. Meta4meaning: Automatic Metaphor Interpretation Using Corpus-Derived Word Associations. The Seventh International Conference on Computational Creativity (ICCC), 230-237. Paris, France, June-July 2016.
- Ping Xiao and Hannu Toivonen. LayerFolding: Discovering Creative Links in Word Association Networks. The 31st Annual ACM Symposium on Applied Computing (ACM SAC), 894-897. Pisa, Italy, April 2016.

## Neuroinformatics, Professor Aapo Hyvärinen

Neuroinformatics is widely defined as the cross-fertilization of information-processing and mathematical sciences on the one hand, and neural and cognitive sciences on the other.

Our group works on different aspects of neuroinformatics related to machine learning. We model the visual system in the brain by analyzing the statistical structure of the natural input images. We apply machine learning models on neuroimaging data, in particular MEG. We also develop the relevant theory of statistical machine learning, typically unsupervised.

### *Key publications:*

- Haruo Hosoya and Aapo Hyvärinen. Learning Visual Spatial Pooling by Strong PCA Dimension Reduction. Neural Computation, 28:1249-1264, 2016.
- Aapo Hyvärinen, Jun-ichiro Hirayama , Vesa Kiviniemi, and Motoaki Kawanabe. Orthogonal Connectivity Factorization: Interpretable decomposition of Variability in Correlation Matrices. Neural Computation, 28:445-484, 2016.

## Parsimonious Modeling, Dr. Jaakko Hollmén

The research group Parsimonious Modelling develops novel computational data analysis methods and applies these methods on two application fields: cancer genomics and environmental informatics. Parsimonious modeling aims at simple, compact, or sparse models as a result of learning from data in the presence of very little or no a priori information about the modeled problem. Simplicity of the models facilitates understanding of the problem domain by humans.

Both application fields present similar challenges to the data analysis problems: the high dimensionality of observed data and the presence of moderate or large noise levels are both factors that bear fundamental problems for any data analysis. Seeking new areas of application and interfacing the newest application domains with lots of novel types of generated data helps in finding new, unsolved settings of problems.

### *Key publications:*


- Prem Raj Adhikari, Anze Vavpetic, Jan Kralj, Nada Lavrač, and Jaakko Hollmén. Explaining Mixture Models through Semantic Pattern Mining and Banded Matrix Visualization. Machine Learning, 105(1), 3-39, 2016.
- Liisa Kulmala, Jesse Read, Pekka Nöjd, Cyrille B.K. Rathgeber, Henri E. Cuny, Jaakko Hollmén, and Harri Mäkinen. Identifying the main drivers for the production and maturation of Scots pine tracheids along a temperature gradient, Agricultural and Forest Meteorology, 232:210-224, In press.
- Liisa Kulmala, Indre Žliobaitė, Eero Nikinmaa, Pekka Nöjd, Pasi Kolari, Kourosh Kabiri Koupaei, Jaakko Hollmén, and Harri Mäkinen. Environmental control of growth variation in a boreal Scots pine stand - a data-driven approach. Silva Fennica, In Press.
- Jesse Read, Luca Martino, and Jaakko Hollmén. Multi-label Methods for Prediction with Sequential Data, Pattern Recognition, In press.
- Jesse Read, Peter Reutemann, Bernhard Pfahringer, and Geoff Holmes. MEKA: A Multi-label/Multi-target Extension to Weka. Journal of Machine Learning Research. Vol. 17(21). pp 1-5. 2016.
- Jesse Read, Indre Zliobaite, and Jaakko Hollmén. Labeling sensing data for mobility modeling. Information Systems, 57, 207-222, 2016.
- Indre Žliobaitė, Janne Rinne, Anikó Toth, Michael Mechenich, Liping Liu, Anna K. Behrensmeyer, and Mikael Fortelius. Herbivore teeth predict climatic limits in Kenyan ecosystems. Proceedings of the National Academy of Sciences. 113(45), p. 12751-12756, 2016.

## Kernel Methods, Pattern Analysis and Computational Metabolomics (KEPACO), Professor Juho Rousu

The KEPACO group develops machine learning methods, models and tools for data science, in particular computational metabolomics. The methodological backbone of the group is formed by kernel methods and regularized learning. The group focusses in learning with multiple and structured targets, multiple views and ensembles. Machine learning applications of interest include metabolite identification, metabolic network reconstruction and pathway analysis, chemogenomics as well as biomarker discovery.

In year 2016, the group had success by winning an automatic small molecule identification category in the CASMI 2016 contest (www.casmi-contest.org) using the IOKR technology (Brouard et al. 2016a, 2016b). Other significant results were made in multivariate meta-analysis methods fro genome-wide association studies (Cichonska et al, 2016) as well as in machine learning to complete partially observed kernel matrices (Bhadra et al. 2016), a problem frequently encountered in bioinformatics. The group co-organized the 10th edition of the Machine Learning in Systems Biology with more than 80 international attendees (van Dijk et al. 2016). The group continues its work in the new HIIT programme Foundations of Computational Health.

### *Key publications:*


- Sahely Bhadra, Samuel Kaski, and Juho Rousu. Multi-view kernel completion. Machine Learning, 1-27, 2016.
- Celine Brouard, Huibin Shen, Kai Dührkop, Florence D'Alche-buc, Sebastian Böcker and Juho Rousu. Fast metabolite identification with Input Output Kernel Regression. Proc. Intelligent Systems for Molecular Biology, ISMB 2016, Bioinformatics 32, 12, i28–i36, 2016a.
- Céline Brouard,, Marie Szafranski, and F. d'Alché Buc. Input output kernel regression: Supervised and semi-supervised structured output prediction with operator-valued kernels." Journal of Machine Learning Research 17.176: 1-48, 2016b.
- Anna Cichonska, Juho Rousu, Pekka Marttinen et al. metaCCA: Summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. Bioinformatics, 32 (13): 1981-1989, 2016.
- Aalt-Jan van Dijk,, Harri Lähdesmäki, Dick de Ridder, and Juho Rousu. Selected proceedings of Machine Learning in Systems Biology: MLSB 2016. BMC Bioinformatics 17, no. 16: 51, 2016.

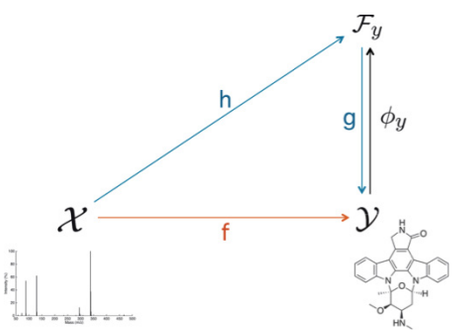



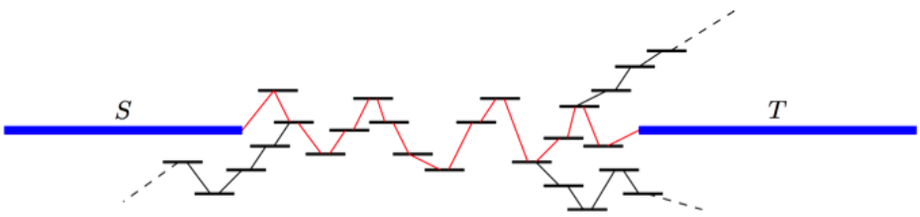

*Figure: Overview of the Input-Output Kernel regression approach for metabolite identification, predicting molecular structures form tandem mass spectra (Brouard et al. 2016a, 2016b)*

## Genome-Scale Algorithmics, Professor Veli Mäkinen

We develop algorithms and data structures for the analysis of genome-scale data. Such data is abundant due to modern molecular biology measurement techniques like high-throughput sequencing. We are especially interested in applications of compressed data structures, that make it possible to analyse the often highly redundant data within the space of their information content. Our latest developments focus on pan-genome indexing, space-efficient sequence analysis, alignments on sequence-like structures, and transcript and genome assembly.

### *Key publications:*

- Djamal Belazzougui and Simon J. Puglisi. Range Predecessor and Lempel-Ziv Parsing, Proc. ACM-SIAM Symposium on Discrete Algorithms (SODA 2016), pp. 2053-2071, 2016.
- Leena Salmela, Kristoffer Sahlin, Veli Mäkinen, and Alexandru I. Tomescu. Gap Filling as Exact Path Length Problem. Journal of Computational Biology 23(5): 347-361, 2016.
- Ahmed Sobih, Veli Mäkinen, and Alexandru I. Tomescu. MetaFlow: Metagenomic profiling based on whole-genome coverage analysis with min-cost flows. In Proc. RECOMB 2016, Springer, LNCS 9649, pp. 111-121, 2016.
- Yasuo Tabei, Hiroto Saigo, Yoshihiro Yamanishi, and Simon J. Puglisi. Scalable Partial Least Squares Regression on Grammar-Compressed Data Matrices, Proc. 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2016), pp. 1875-1884, 2016.
- Alexandru I. Tomescu and Paul Medvedev. Safe and complete contig assembly via omnitigs. In Proc. RECOMB 2016, Springer, LNCS 9649, pp. 152-163, 2016.



## New Paradigms in Computing, Professor Petteri Kaski

The group performs basic research at the intersection of core computer science (algorithm design and analysis) and discrete mathematics, with an emphasis towards novel techniques and less studied models of computation. We invest substantial effort to high-risk, high-yield research problems of relatively broad theoretical interest, selected on both problem and method driven basis. The current research themes of the group are threefold. (1) Exploring the interplay between algebraic, combinatorial, and geometric techniques in the design of exact deterministic algorithms. For example, many combinatorial problems can be cast in algebraic form, whereby a nontrivial algebraic algorithm yields a more efficient solution compared with direct combinatorial tools. (2) Restricted models of computation and tradeoffs in resources and/or objectives. (3) While we are a theory group, we occasionally engage in practical algorithm implementation. Examples include attacks on combinatorial classification problems and applications in computational geometry.

### *Key publications:*


- Per Austrin, Petteri Kaski, Mikko Koivisto, Jesper Nederlof. Sharper upper bounds for unbalanced Uniquely Decodable Code Pairs. The 2016 IEEE International Symposium on Information Theory (ISIT 2016), 335-339, 2016.
- Andreas Björklund and Petteri Kaski. How Proofs are Prepared at Camelot: Extended Abstract. ACM Symposium on Principles of Distributed Computing (PODC 2016), 391-400, 2016.
- Sebastian Brandt, Orr Fischer, Juho Hirvonen, Barbara Keller, Tuomo Lempiäinen, Joel Rybicki, Jukka Suomela, and Jara Uitto. A lower bound for the distributed Lovász local lemma. In proceedings of the 48th annual ACM symposium on Theory of Computing (STOC' 2016), 479-488, 2016.
- Kustaa Kangas, Teemu Hankala, Teppo Mikael Niinimäki, and Mikko Koivisto. Counting Linear Extensions of Sparse Posets. The 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), 603-609, 2016.
- Matti Karppa, Petteri Kaski, and Jukka Kohonen. A Faster Subquadratic Algorithm for Finding Outlier Correlations. ACM-SIAM Symposium on Discrete Algorithms (SODA 2016), 1288-1305, 2016.

## *Computational Inference (CI) Programme*

The four groups of the CI programme are all members of the Finnish Centre of Excellence in Computational Inference Research (COIN), and the objectives of the programme are closely intertwined with those of COIN.

The main objective of CI is to develop methods for transforming the data produced by the current data revolution into useful information. The key methodology for achieving this goal is statistical and computational inference based on the data. The emphasis is on large data collections and computationally demanding modelling and inference algorithms. Our mission is to push the boundary towards both more complex problems, requiring more structured data models, and towards extremely rapid inference. We address a set of carefully chosen interdisciplinary “grand challenge” -level problems with high societal impact where solving the data intensive problems requires novel methodologies that can only result from combining the expertise of separate subfields

Our mission brings up four partially overlapping methodological focus areas: Learning of massive data-driven models; Learning from multiple sources; Statistical inference in highly structured stochastic models; and Extreme inference engine.

In applied research our work is also motivated by the big data and ubiquitous computing vision, where adaptivity, context-awareness and personalisation are key enablers. We see that our four methodological research areas support strongly each other, and they all address from a different perspective the key technological problems we face in our future “big data” information society. In our applied research we link our strong basic research work in machine learning and constraint reasoning to well-motivated applied research activities involving prototype applications and real-world deployments.

## Probabilistic Machine Learning, Professor Samuel Kaski

We develop new methods for probabilistic modeling, Bayesian inference and machine learning. Our current focuses are in particular learning from multiple data sources, Bayesian model assessment and selection, approximate inference and information visualization. Our primary application areas are digital health and biology, neuroscience and user interaction.

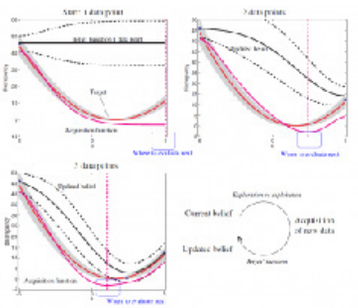
### *Key publications:*


- Paul Blomstedt, Ritabrata Dutta, Sohan Seth, Alvis Brazma, and Samuel Kaski. Modelling-based experiment retrieval: A case study with gene expression clustering. Bioinformatics, 32(9):1388-1394, 2016.
- Kerstin Bunte, Eemeli Leppäaho, Inka Saarinen, and Samuel Kaski. Sparse group factor analysis for biclustering of multiple data sources. Bioinformatics. 32, 16, p. 2457-2463, 2016.
- Manuel J. A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. Scientific Reports, 6:38580, 2016.
- Jussi Gillberg, Pekka Marttinen, Matti Pirinen, Antti J. Kangas, Pasi Soininen, Mehreen Ali, Aki S. Havulinna, Marjo-Riitta Järvelin, Mika Ala-Korpela, and Samuel Kaski. Multiple Output Regression with Latent Noise. Journal of Machine Learning Research, 17:1-35, 2016.
- Suleiman A. Khan, Eemeli Leppäaho and Samuel Kaski. Bayesian multi-tensor factorization. Machine Learning, 105(2):233-253, 2016.
- Juho Piironen and Aki Vehtari. Comparison of Bayesian predictive methods for model selection. Statistics and Computing, 27(3):711-735, 2016.
- Aravind Sankar, Brandon Malone, Sion C. Bayliss, Ben Pascoe, Guillaume Méric, Matthew D. Hitchings, Samuel K. Sheppard, Edward J. Feil, Jukka Corander, and Antti Honkela. Bayesian identification of bacterial strains from sequencing data. Microbial genomics 2(8):e000075, 2016.
- Hande Topa and Antti Honkela. Analysis of differential splicing suggests different modes of short-term splicing regulation. Bioinformatics 32(12):i147–i155, 2016.
- Aki Vehtari, Andrew Gelman and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. In Statistics and Computing, 27(5):1413–1432, 2016.
- Aki Vehtari, Tommi Mononen, Ville Tolvanen, Tuomas Sivula, and Ole Winther. Bayesian leave-one-out cross-validation approximations for Gaussian latent variable models. Journal of Machine Learning Research, 17(103):1−38, 2016.

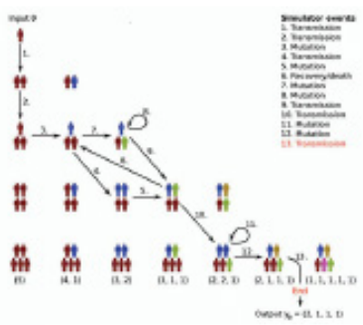
## Bayesian Statistics, Professor Jukka Corander

BSG works on theoretical and applied machine learning, simulation based inference and computational biology. A core focus area is approximate inference for intractable statistical models, including both ABC for simulator models and structural learning of graphical and context-specific Markov models. Our main application area is computational biology, in particular models for evolution and transmission of bacteria and viruses, where BSG methods have enabled numerous highlight discoveries.

### *Key publications:*



- Michael Gutmann, Jukka Corander. Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models. Journal of Machine Learning Research, 17(121), 1-46, 2016.
- John A. Lees, Minna Vehkala, Niko Välimäki, Simon R. Harris, Claire Chewapreecha, Nicholas J. Croucher, Pekka Marttinen, Mark R. Davies, Andrew C. Steer, Steven Y. C. Tong, Antti Honkela, Julian Parkhill, Stephen D. Bentley, and Jukka Corander. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. Nature Communications, doi:10.1038/ncomms12797, 2016.
- Jarno Lintusaari, Michael U. Gutmann, Ritabrata Dutta, Samuel Kaski, and Jukka Corander. Fundamentals and Recent Developments in Approximate Bayesian Computation. Systematic Biology, doi: 10.1093/sysbio/syw077, 2016.
- Alan McNally et al. Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. PloS Genetics, 12(9): e1006280, 2016.
- Renaud-Alexandre Pitaval et al. Volume of Metric Balls in High-Dimensional Complex Grassmann Manifolds. IEEE Transactions on Information Theory, 62(9):5105 - 5116, 2016.
- Zhirong Yang, Jukka Corander, and Erkki Oja. Low-rank doubly stochastic matrix decomposition for cluster analysis. Journal of Machine Learning Research, 17(187):1−25, 2016.
- Zhong Zheng, Lu Wei, Roland Speicher, Ralf Muller, Jyri Hämäläinen, Jukka Corander. Asymptotic Analysis of Rayleigh Product Channels: A Free Probability Approach. IEEE Transactions on Information Theory, doi: 10.1109/TIT.2016.2646699, 2016.

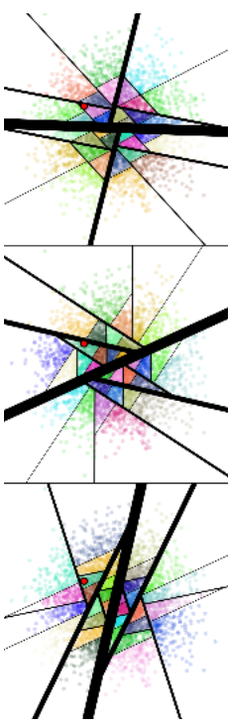


## Complex Systems Computation (CoSCo), Professor Petri Myllymäki

The CoSCo research group investigates computational problems related to complex systems, focusing on prediction and modeling tasks. The basic research areas are closely related to artificial intelligence, machine learning, probabilistic modeling and data analysis, information theoretical approaches to inference and constraint reasoning and optimization. Central themes in the applied research include methods for analyzing and visualizing multidimensional and multimodal Big Data, intelligent information retrieval and context- awareness for ubiquitous computing. The research work is implemented through the following four sub groups “Constraint Reasoning and Optimization” (led by Matti Järvisalo), “Information, Complexity and Learning”, (led by Teemu Roos), “Intelligent Interactive Information Access” (led by Patrik Floréen) and “Multi-Source Probabilistic Inference” (led by Arto Klami).

### *Key publications:*


- Kumaripaba Athukorala, Alan Medlar, Antti Oulasvirta, Giulio Jacucci, and Dorota Glowacka. Beyond Relevance: Adapting Exploration/Exploitation in Information Retrieval. In Proc. of the 21st International Conference on Intelligent User Interfaces (IUI’16), 359-369, 2016.
- Pedram Daee, Joel Pyykkö, Dorota Glowacka, Samuel Kaski. Interactive Intent Modeling from Multiple Feedback Domains. In Proc. of the 21st International Conference on Intelligent User Interfaces (IUI’16), 71-75, 2016.
- Jussi Korpela, Andreas Henelius, Lauri Ahonen, Arto Klami, and Kai Puolamäki. Using regression makes extraction of shared variation in multiple datasets easy. Data Mining and Knowledge Discovery, 30(5):1112-1133, 2016.
- Arto Klami and Aditya Jitta. Probabilistic size-constrained microclustering. In Proc. of Uncertainty in Artificial Intelligence (UAI), 329-338, 2016.
- Jussi Määttä, Daniel F. Schmidt, and Teemu Roos. Subset selection in linear regression using sequentially normalized least squares. Asymptotic theory, Scandinavian Journal of Statistics 43(2):382–395, 2016.
- Jamshid Tehrani, Quan Nguyen, and Teemu Roos. Oral fairy tale or literary fake? Investigating the origins of Little Red Riding Hood using phylogenetic network analysis. Digital Scholarship in the Humanities 31(3):611–636, 2016.
- Johannes P. Wallner, Andreas Niskanen, and Matti Järvisalo. Complexity Results and Algorithms for Extension Enforcement in Abstract Argumentation. In proc. of the 30th AAAI Conference on Artificial Intelligence (AAAI 2016), 1088-1094, 2016. Extended version accepted for publication in Journal of Artificial Intelligence Research.
- Andreas Niskanen, Johannes P. Wallner, and Matti Järvisalo. Synthesizing Argumentation Frameworks from Examples. In proc. of the 22nd European Conference on Artificial Intelligence (ECAI 2016), 551-559, 2016. [Runner-up for the ECAI 2016 Best Student Paper Award]
- Matti Järvisalo. Boolean Satisfiability and Beyond: Algorithms, Analysis, and AI Applications. In proc. of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), 4066-4069, 2016. [IJCAI 2016 Early Career Spotlight]

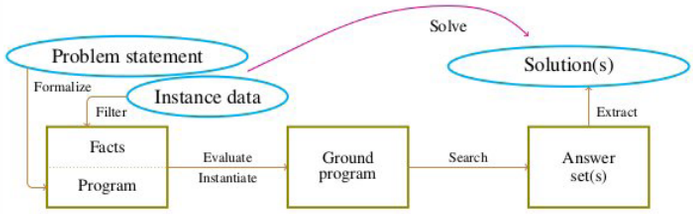
## Computational Logic, Professor Ilkka Niemelä

The group develops automated reasoning techniques for solving challenging computational problems in engineering and science. The current focus is on efficient computational methods for solving large constraint satisfaction problems formally represented as Boolean satisfiability (SAT) problems or their generalizations, or expressed in terms of rule-based constraints used in answer-set programming (ASP). The group has a strong track record in research on verification and testing of automation systems and software, as well as applying formal methods in the analysis of distributed systems.

In 2016, the group continued its fundamental work on ASP by developing new extensions for the paradigm. The novel stable-unstable semantics provides a systematic way to address reasoning problems that go beyond the complexity class NP. The approach was implemented using SAT solvers recursively as oracles, which turned out to have further applications in the evaluation of quantified Boolean formulas (QBFs) and argumentation frameworks (AFs). New static symmetry breaking method was developed for SAT solvers. Moreover, the group devised alternative techniques for the instantiation of propositional schemata based on second-order logic and state-of-the-art ASP grounders.

### *Key publications:*


- Bart Bogaerts, Tomi Janhunen, and Shahab Tasharrofi. Declarative Solver Development: Case Studies. In Proceedings of KR 2016, 74-83.
- Bart Bogaerts, Tomi Janhunen, and Shahab Tasharrofi. Stable-Unstable Semantics: Beyond NP with Normal Logic Programs. Theory and Practice of Logic Programming, 16(5-6), 570-586, 2016.
- Jori Bomanson, Martin Gebser, Tomi Janhunen, Benjamin Kaufmann, and Torsten Schaub. Answer Set Programming modulo Acyclicity. Fundamenta Informaticae, 147(1), 63-91, 2016.
- Jo Devriendt, Bart Bogaerts, Maurice Bruynooghe, and Marc Denecker: Improved Static Symmetry Breaking for SAT. In Proceedings of SAT 2016, 104-122.
- Martin Gebser, Tomi Janhunen, Roland Kaminski, Torsten Schaub, and Shahab Tasharrofi: Writing Declarative Specifications for Clauses. In Proceedings of JELIA 2016, 256-271.
- Tomi Janhunen and Ilkka Niemelä. The Answer Set Programming Paradigm. AI Magazine, 37(3), 13-24, 2016.
- Tomi Janhunen, Shahab Tasharrofi, and Evgenia Ternovska. SAT-to-SAT: Declarative Extension of SAT Solvers with New Propagators. In Proceedings AAAI 2016, 978-984.

## *Distributed and Mobile Cloud Systems (DMC)*

The focus of the programme is to bridge the gap between mobile devices and the cloud based server backend systems into a single seamless distributed and mobile computing platform. The main motivation behind this is that mobile devices are by their very nature very resource constrained in available battery power, CPU, memory, network, as well as storage capacity compared to the server hardware available in the cloud backend systems. Thus mobile devices need to be tightly integrated to the cloud backend systems in order to do computational tasks that are too heavy for them. However, this basic setup is not yet sufficient for highly interactive applications. The wide area network (WAN) communication latencies between the mobile device and the possibly quite physically remote cloud backend can often be too large for interactive mobile applications, e.g., for interactive augmented reality applications such as Google Glass, as well as computationally intensive mobile intelligent information access applications. Therefore, sometimes an additional layer of computing called cloudlets is called for, that is a collection of local cloud servers that acts as local computing service for mobile devices in its own network neighborhood area. This idea is also known as cyber foraging in the literature.

The research programme operates via close collaboration of the groups of Professors Keijo Heljanko, Jussi Kangasharju, Sasu Tarkoma and Antti Ylä-Jääski.

*Key publications of the DMC programme:*


- Danny Dolev, Keijo Heljanko, Matti Järvisalo, Janne H. Korhonen, Christoph Lenzen, Joel Rybicki, Jukka Suomela, and Siert Wieringa. Synchronous counting and computational algorithm design. J. Comput. Syst. Sci. 82(2): 310-332, 2016.
- Markku Hinkka, Teemu Lehto, and Keijo Heljanko. Assessing Big Data SQL Frameworks for Analyzing Event Logs. PDP 2016: 101-108, 2016.
- Mohammad Hoque, Matti Siekkinen, Yu Xiao, Kashif N. Khan, and Sasu Tarkoma. Modeling, Profiling, and Debugging the Energy Consumption of Mobile Devices. ACM Comput. Surv. 48(3): 39, 2016.
- Jarno Leppänen, Mikko Pelkonen, Haipeng Guo, Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. Collaborative and Energy-Efficient Speech Monitoring on Smart Devices. IEEE Computer 49(12): 22-30, 2016.
- Julien Mineraud, Oleksiy Mazhelis, Xiang Su, and Sasu Tarkoma. A gap analysis of Internet-of- Things platforms. Computer Communications 89: 5-16, 2016.
- Nitinder Mohan and Jussi Kangasharju. Edge-Fog Cloud: A Distributed Cloud for Internet of Things Computations. In Proceedings of Cloudification of the Internet of Things Confer- ence, Paris, France, November 2016.
- Ella Peltonen, Eemil Lagerspetz, Petteri Nurmi, and Sasu Tarkoma. Constella: Crowdsourced system setting recommendations for mobile devices. Pervasive and Mobile Computing 26: 71-90, 2016.
- Olli Saarikivi and Keijo Heljanko. LCTD: Test-guided proofs for C programs on LLVM. J. Log. Algebr. Meth. Program. 85(6): 1292-1317, 2016.
- Jouni Vepsäläinen, Petri Savolainen, Jouni Ojala, Antonella Di Rienzo, Matti Nelimarkka, Kai Kuikkaniemi, Sasu Tarkoma, and Giulio Jacucci. Web-Based Public-Screen Gaming: Insights from Deployments. IEEE Pervasive Computing 15(3): 40-46, 2016.
- Otto Waltari and Jussi Kangasharju. The Wireless Shark: Identifying WiFi Devices Based on Probe Fingerprints. In Proceedings of ACM MobiSys Workshop on Mobile Data, Singapore, June 2016.

## *Network Society (NS) Programme*

The mission of the Network Society research programme is to empower ubiquitous users with transparent and resourceful ICT in particular:

**Multimodal interaction and adaptive information.** The research explore new interaction paradigms between computers and users towards synergic and symbiotic relations. Exploiting recent advances in interaction technologies such as gestures, eye tracking, physiological, tangible computing, and haptics novel adaptive applications can be explored in urban exploration, immersive meditation environments, and search user interfaces.

**Experience of mediated Interaction**. In particular the focus is on emotional and cognitive processes during mediated social interaction, including mediated touch, interaction with virtual humans, and knowledge work. The research uses a wide methodological approach, from surveys and qualitative methods to neurophysiological recordings (e.g., EEG), to study users in the laboratory as well as in real-life contexts. By providing both fundamental and applied insight into social interaction, we help to create ICT and services that take the human social and emotional processes into account.

**Emergent social media platforms and practices**. The creation of in particular digital content and groups and communities using this data. New platforms and associated research include support for live participation in events, collocated and collaborative interaction with public screens. A particular focus is on MyData ecosystem through research and development efforts in empowering individuals to control their personal data and thus to manage their own lives and privacy both online and offline.

## Digital Content Communities, Professor Marko Turpeinen

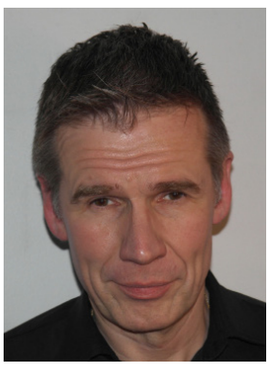
The Digital Content Communities (DCC) is a multidisciplinary group focused to study the intersection of information technology and society. Our mission is to ensure that the technical solutions benefit also its users and the wider society. This means that we must consider the design decisions, legal and social aspects of technology and requires end-to-end research: from technology and product development to in-the-wild social studies of technologies implications.

During 2016, the focus of DCC was on large techno-social transformation driven through platformization of technology. We focused on the platform economics and data governance, especially about new human centric management models about personal data (MyData). We also address the changes of media caused through social media platform (hybrid media environment) and aim to understand how these platforms could be also used for social good.

### *Key publications:*

- Anette Alen-Savikko and Olli Pitkänen. Rights and Entitlements in Information: Proprietary Perspectives and Beyond. Data Protection, Privacy and European Regulation in the Digital Age, Forum Iuris, 2016
- Luca Chittaro and Andrea Vianello. Evaluation of a mobile mindfulness app distributed through on-line stores: a 4-week study. International Journal of Human-Computer Studies, 2016.
- Luca Chittaro and Andrea Vianello. Mobile Mindfulness and User's Worry: A Qualitative Study of Using a Smartphone App for Distancing from Negative Thoughts. Interacting with Computers, 2016.
- Mari Marttila, Salla-Maaria Laaksonen, Arto Kekkonen, Mari Tuokko, and Matti Nelimarkka. Digitaalinen vaaliteltta: Twitter politiikan areenana eduskuntavaaleissa 2015. Eduskuntavaalitutkimus 2015: Poliittisen osallistumisen eriytyminen (pp. 117–137). Helsinki: Oikeusministeriö, 2016.
- Matti Nelimarkka et al. Live Participation: Augmenting Events with Audience-Performer Interaction Systems. In Proc. of the 2016 ACM Conference on Designing Interactive Systems, pp. 509-520, ACM 2016.
- Jouni Vepsäläinen at al. Web-Based Public-Screen Gaming: Insights from Deployments. IEEE Pervasive Computing, 15(3), 40-46, 2016.
- Andrea Vianello, Yves Florack, Andrea Bellucci, and Giulio Jacucci. T4Tags 2.0: A Tangible System for Supporting Users' Needs in the Domestic Environment. In Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '16). ACM, New York, NY, USA, 38-43, 2016.

## Social Interaction and Emotion (SIE), Professor Niklas Ravaja



The mission of SIE group is to increase our understanding of ICT-mediated social interaction. Our studies focus on: (a) emotional and cognitive processes during mediated social interaction, including mediated touch, interaction with virtual humans, and knowledge work, (b) the neuroscience of social ICT, (c) socially-mediated behavioral phenomena in SNSs, (d) the use of gamification to support the adoption of healthier lifestyles, and (e) user experience of ICT.

We use a wide methodological approach, from surveys and qualitative methods to neurophysiological recordings (e.g., EEG), to study users in the laboratory as well as in real-life contexts. By providing both fundamental and applied insight into social interaction, we help to create ICT and services that take the human social and emotional processes into account.

### *Key publications:*


- Niklas Ravaja, Gary Bente, Jari Kätsyri, Mikko Salminen, and Tapio Takala. Virtual character facial expressions influence human brain and facial EMG activity in a decision-making game. IEEE Transactions on Affective Computing, 2016.
- Niklas Ravaja, Pekka Korhonen, Murat Köksalan, Jari Lipsanen, Mikko Salminen, Outi Somervuori, and Jyrki Wallenius. Emotional-motivational responses predicting choices: The role of asymmetrical frontal cortical activity. Journal of Economic Psychology. Vol. 52. pp. 56-70, 2016.
- Mikko Salminen, Pentti Henttonen, and Niklas Ravaja. The role of personality in dyadic interaction: A psychophysiological study. International Journal of Psychophysiology. Vol. 109. pp. 45-50, 2016.
- Manuel J. Eugster, Tuukka Ruotsalo, Michiel M. Spape, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. Scientific Reports. Vol. 6, 2016.
- Jari Kätsyri, Teemu Kinnunen, Kenta Kusumoto,Pirkko Oittinen, and Niklas Ravaja. Negativity Bias in Media Multitasking: The Effects of Negative Social Media Messages on Attention to Television News Broadcasts. PLoS One. Vol. 11, No. 5, 2016.
- Simo Jarvela, Jari Kätsyri, Niklas Ravaja, Guillaume Chanel, and Pentti Henttonen. Intragroup Emotions : Physiological Linkage and Social Presence. Frontiers in Psychology. Vol. 7, 2016.
- Oswald Barral, Ilkka Kosunen, Tuukka Ruotsalo, Michiel M. Spapé, Manuel J. Eugster, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. Extracting relevance and affect information from physiological text annotation. User Modeling and User-Adapted Interaction. Vol. 26, No. 5. pp. 493-520, 2016.
- Imtiaj Ahmed, Ville Harjunen, Giulio Jacucci, Eve Hoggan, Niklas Ravaja, and Michiel M. Spapé. Reach out and touch me: effects of four distinct haptic technologies on affective touch in virtual reality. Proc of 18th ACM International Conference on Multimodal Interaction. pp. 341-348, 2016.
- Michiel M. Spape and Niklas Ravaja. Not My Problem: Vicarious Conflict Adaptation with Human and Virtual Co-actors. Frontiers in Psychology. Vol. 7, 2016.

## Ubiquitous Interaction (UiX), Professor Giulio Jacucci

Ubiquitous Interaction investigates two orthogonal approaches in HCI and their combination: 1) symbiotic relations between computers and humans leveraging implicit signals and user modeling 2) resourceful interaction promoting the active user through skilled and ingenious use of available resources. These approaches capture important qualities of our vision of the role of computers in the future accounting for their co-adaptivity in symbiosis and for the centrality of the users' skills and agency through resourcefulness. Guided by such objectives we develop further multimodal interaction techniques for example combining touch and eye tracking, the use of physiological computing for information seeking, UI for interactive intent modeling, adaptive systems for behavior change, novel forms of social computing merging physical and digital interaction.

### *Key publications:*


- Imtiaj Ahmed, Ville Harjunen, Giulio Jacucci, Eve Hoggan, Niklas Ravaja, and Michiel M. Spapé. Reach out and touch me: effects of four distinct haptic technologies on affective touch in virtual reality. In Proceedings of the 18th ACM International Conference on Multimodal Interaction (pp. 341-348), 2016.
- Kumaripaba Athukorala, Dorota Glowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. Is Exploratory Search Different? American Society for Information Science and Technology. Journal, volume 67 issue 11, 2635-2651, 2016.
- Oswald Barral, Ilkka Kosunen, Tuukka Ruotsalo, Michiel M. Spapé, Manuel J. Eugster, Niklas Ravaja, ... and Giulio Jacucci. Extracting relevance and affect information from physiological text annotation. User Modeling and User-Adapted Interaction, 26(5), 493-520, 2016.
- Andrea Bellucci, Andrea Vianello, Yves Florack, and Giulio Jacucci Supporting the Serendipitous Use of Domestic Technologies. IEEE Pervasive Computing, 15(2), 16-25, 2016.
- Yi-Ta Hsieh et al. Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4203-4215, ACM 2016.
- Ilkka Kosunen, Mikko Salminen, Simo Järvelä, Antti Ruonala, Niklas Ravaja, and Giulio Jacucci. RelaWorld: Neuroadaptive and Immersive Virtual Reality Meditation System. In Proceedings of the 21st International Conference on Intelligent User Interfaces (pp. 208-217). ACM 2016.
- Baris Serim, and Giulio Jacucci. Pointing while Looking Elsewhere: Designing for Varying Degrees of Visual Guidance during Manual Input In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). "Best of CHI" awards honor, ACM 2016
- Salu Ylirisku, Giulio Jacucci, Abigail Sellen, and Richard Harper. Design Research as Conceptual Designing: The Manhattan Design Concept. Interacting with Computers, 28(5), 648-663, 2016.

## *HIIT-Wide Focus Area: Augmented Science (AR)*

Augmented search, research, and knowledge work are the main themes of the multidisciplinary HIIT-wide research initiative that is a strategic research focus of HIIT, which is since August 2016 also a Research Program for two years. Several research groups ranging from Human-Computer Interaction to Machine Learning and Complex Systems Computation collaborate to produce cutting edge research and demonstrations that go beyond individual research papers. The project investigates how Human- Computer Interaction and Probabilistic Machine Learning can be combined increase, by order of magnitude, the effectiveness of search and knowledge work. We are applying our research to scientific data and work environments. The HWFA Augmented Research has been successful in attracting external funding in particular in two large projects coordinated by HIIT: MindSee: Symbiotic Mind Computer Interaction for Information Seeking a EU FET project with 7 partners from 4 countries (mindsee.eu), Re:Know a TEKES strategic opening project on the Revolution of Knowledge Work (www.reknow.fi). The research investigates methods and tools to better utilize the massive explosion of raw data, documents, distributed information and link structures between these, and sensory information recorded from the users. The methods and pilot applications are expected to revolutionize our work practices in data- driven fields such as modern biology, business intelligence, and others. In particular improving the general problem solving method of science, research, and development, in collaboration with the other fields, is the best way for our research community to contribute to solving the grand challenges of humanity.

### *Key publications:*


- Kumaripaba Athukorala, Alan Medlar, Antti Oulasvirta, Giulio Jacucci, and Dorota Glowacka. Beyond relevance: Adapting exploration/exploitation in information retrieval. In Proceedings of the 21st International Conference on Intelligent User Interfaces (pp. 359-369), ACM 2016.
- Oswald Barral, Ilkka Kosunen, Tuukka Ruotsalo, Michiel M. Spapé, Manuel J. Eugster, Niklas Ravaja, ... and Giulio Jacucci. Extracting relevance and affect information from physiological text annotation. User Modeling and User-Adapted Interaction, 26(5), 493-520, 2016.
- Manuel J. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. Scientific Reports, 6, 2016.
- Ilkka Kosunen, Mikko Salminen, Simo Järvelä, Antti Ruonala, Niklas Ravaja, and Giulio Jacucci. RelaWorld: Neuroadaptive and Immersive Virtual Reality Meditation System. In Proceedings of the 21st International Conference on Intelligent User Interfaces, pp. 208-217, ACM 2016.
- Matti Nelimarkka, Kai Kuikkaniemi, Antti Salovaara, and Giulio Jacucci. Live Participation: Augmenting Events with Audience-Performer Interaction Systems. In Proceedings of the ACM Conference on Designing Interactive Systems, pp. 509-520, 2016.

# New programmes 2016-2021

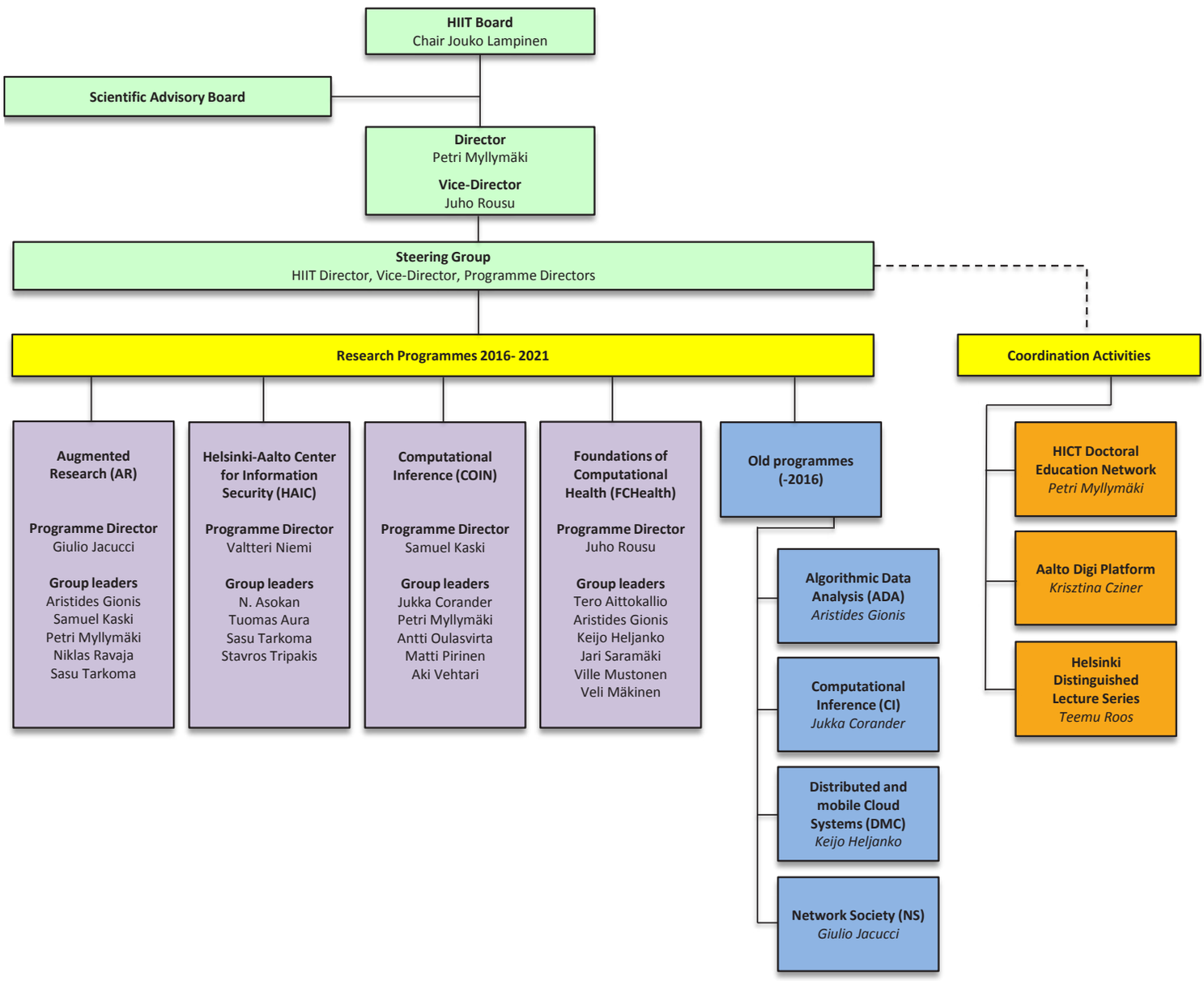
When a call for new Centers of Excellence was opened by AoF in Spring 2016, the Board of HIIT decided to align the process of forming new HIIT programmes with this AoF call: the idea was that HIIT will make the decisions about its programmes independently, but for setting the scientific ambition level on the right level, initially only candidates participating in the AoF call would be considered. Naturally, as the AoF CoE call is open for all sciences, not all candidates were eligible as HIIT programmes, but they were expected to directly support the “Helsinki ICT” mission of HIIT and the strategies of the two hosting universities. In addition, they were expected to show high potential impact, providing clear added value to the normal activities of the participating units via inter-university and inter-disciplinary collaboration.

As a result of this process, the HIIT Board decided in their meeting on June 1st, 2016, to launch the following four new HIIT research programs in August 2016:

- Foundations of Computational Health, program director Juho Rousu
- Computational Inference, program director Samuel Kaski
- Building Trust in Secure Computing Systems, program director Valtteri Niemi
- Augmented Research, program director Giulio Jacucci

Following the AoF CoE model, the new programmes are fixed-term: the first 3 programs are initially operational for 5 years, 1.8.2016-31.7.2021, and the fourth program, which continues the activities of the ongoing HIIT-Wide Focus Area, will run for 2 years. Each of the four programs has participating professors from both Aalto University and University of Helsinki, and is strongly committed to support the joint “Helsinki ICT” mission of HIIT.

Unfortunately the outcome of the AoF CoE call was that none of the 3 new HIIT programmes participating in the call were accepted in the second round of the call. This was a major disappointment, and as a matter of fact the outcome was disastrous for the whole discipline: some computer scientists were involved with applications continuing in the second round, but no ICT-led consortium made it through. We believe that the quality of the Finnish ICT research is already on a high international level, and that HIIT in general and the new research programmes in particular, can play a major role in boosting the quality, visibility and impact of Finnish ICT on an even higher level.


HIIT Board
Chair Jouko Lampinen
Scientific Advisory Board
Director
Petri Myllymäki
Vice-Director
Juho Rousu
Steering Group
HIIT Director, Vice-Director, Programme Directors
Research Programmes 2016- 2021
Coordination Activities
Augmented Research (AR)
Programme Director
Giulio Jacucci
Group leaders
Aristides Gionis
Samuel Kaski
Petri Myllymäki
Niklas Ravaja
Sasu Tarkoma
Helsinki-Aalto Center for Information Security (HAIC)
Programme Director
Valtteri Niemi
Group leaders
N. Asokan
Tuomas Aura
Sasu Tarkoma
Stavros Tripakis
Computational Inference (COIN)
Programme Director
Samuel Kaski
Group leaders
Jukka Corander
Petri Myllymäki
Antti Oulasvirta
Matti Pirinen
Aki Vehtari
Foundations of Computational Health (FCHealth)
Programme Director
Juho Rousu
Group leaders
Tero Aittokallio
Aristides Gionis
Keijo Heljanko
Jari Saramäki
Ville Mustonen
Veli Mäkinen
Old programmes (-2016)
Algorithmic Data Analysis (ADA)
Aristides Gionis
Computational Inference (CI)
Jukka Corander
Distributed and mobile Cloud Systems (DMC)
Keijo Heljanko
Network Society (NS)
Giulio Jacucci
HICT Doctoral Education Network
Petri Myllymäki
Aalto Digi Platform
Krisztina Cziner
Helsinki Distinguished Lecture Series
Teemu Roos

# Helsinki-Aalto Center for Information on Security (HAIC)

Programme Director: Valtteri Niemi (UH)

Team leaders: N. Asokan (Aalto), Tuomas Aura (Aalto), Sasu Tarkoma (UH), Stavros Tripakis (Aalto)

## Research Objectives

Digitalization of all walks of life is proceeding at a relentless and ever-increasing pace. Computing systems form the backbone of everything, including business, entertainment, national defence and critical infrastructure. Systems must meet various security, privacy and safety requirements. Ensuring trust in them in an open and interconnected world is a fundamental concern. Computing systems by their very nature are prone to design and implementation flaws that may be exploited by adversaries. Recent revelations about widespread surveillance and deliberate inclusion of backdoors in widely used technology have further eroded confidence in computing systems' ability to provide adequate security.

We posit that the way to facilitate trust in computing systems is by building in the means and mechanisms for verifiability at all stages in their lifecycles: design, implementation, configuration and run time. We use the term "verifiability" broadly, to describe different types of tools, techniques and mechanisms that can be used to increase trust in a system. This includes traditional cryptographic proof systems and formal verification paradigms but also covers assurance mechanisms like authentication protocols and remote attestation schemes. The objective of this CoE is to make fundamental advances in a broad range of platform security technologies that make it possible to verify the correctness of computing systems built on platforms secured by such technologies.

As an example, consider the emerging scenario of autonomsystems. During the next few decades, different types of autonomous systems like self-driving vehicles, collaborating drones and robots will become part of our daily lives. By definition, these systems will need to autonomously interact and collaborate with one another. Establishment of trust is a prerequisite for collaboration. This can be done in a variety of ways. Each entity can verify the identity of the peers with which it collaborates (authentication). It can also verify the software and hardware configuration of its peers (remote attestation). The design and implementation of the systems can be subject to rigorous verification (cryptographic proofs, formal verification, side-channel analysis).

Our long-term mission, aiming towards the year 2025, is to enable the design, building and deployment of distributed large-scale systems where each node (component, device, network element etc.) contributes in verifying trustworthiness of the entire system. Each node would also be able to verify whether any other node or even the entire system is trustworthy.

The division of responsibilities and complementary expertise among team leaders is roughly as follows (although most will contribute in several areas): authentication (Aura), attestation (Asokan), cryptography (Niemi), formal verification (Tripakis), networks (Tarkoma).

## *Scientific Merits*

**Valtteri Niemi** is a Professor at UH. He has more than 25 years of experience in information security, including 15 years in industry. He was a leading contributor for 3G and 4G mobile system security, serving also as the chairman of 3GPP security standards group during 2003-2009. Niemi has co-authored four books and around ten essential patents (that have been used in litigation, e.g., in Nokia vs Qualcomm and Nokia vs Apple). Already in 1990s he made the cryptographic design for Nokia's first hardware-based security solution for mobile phones, later known as M-shield by Texas Instruments.

**N. Asokan** is a Professor at Aalto. He has been a leading system security researcher with two decades of experience in industrial and academic research. While at Nokia Research Center, he pioneered the first publicly reported research on smartphone trusted execution environments. He also led the design of the first practical property-based attestation scheme which is included in the specifications of the Car Connectivity Consortium. His work has appeared in leading system security venues like ACM CCS, IEEE S&P and ISOC NDSS. Asokan is an Associate Editor of ACM TISSEC, IEEE S&P and PoPETS. He is the lead academic PI of Intel Collaborative Research Institute for Secure Computing (ICRI-SC) in Finland. He is an ACM Distinguished Scientist.

**Tuomas Aura** is a Professor at Aalto. His research area is secure protocol design and analysis. He is interested in finding creative and low overhead solutions to security issues of new networking and communication technologies. Some of his research results including infrastructure-less authentication with return routability (e.g., Mobile IPv6), cryptographically generated addresses (CGA), stateless connections, and client puzzles (e.g., HIP) have become common design patterns for secure network protocols. His recent focus areas include future Internet protocols and the IoT. Before joining Aalto, he worked as a researcher at Microsoft Research in 2001-2009. He continues to collaborate with the ICT industry and standardization to ensure the continuing practical impact of his research.

**Stavros Tripakis** is an Associate Professor at Aalto, and an Adjunct Associate Professor at the University of California, Berkeley. He received a PhD and held positions at UC Berkeley (1998-2001 and 2009-), worked at CNRS (2001-06), and Cadence Design Systems (2006-08). He has made fundamental contributions to verification and model-based design for safety-critical, real-time, embedded, and cyber-physical systems. He was co-Chair of the 10th ACM & IEEE EMSOFT 2010, and Secretary/Treasurer (2009-11) and Vice-Chair (2011-13) of ACM SIGBED. His h-index is 40.

**Sasu Tarkoma** is a Professor at UH, and Head of the Department of CS. He is a group leader at HIIT (see later). He has authored 4 textbooks and published over 160 scientific articles. He has managed several national and international research projects at UH, Aalto and HIIT. He was the Academic Coordinator of the large-scale IoT research program (2012-2015). He has seven granted US Patents. He has received several Best Paper awards, such as IEEE PerCom, ACM CCR, and ACM OSR. He is an editor of the Computer Networks Journal and member of organizing and scientific committees of many international conferences.

# *Computational Inference (COIN)*

Programme Director: Samuel Kaski (Aalto)

Team leaders: Jukka Corander (vice director, UH), Petri Myllymäki (UH), Antti Oulasvirta (Aalto), Matti Pirinen (FIMM), Aki Vehtari (Aalto)

## *Research Objectives*

We develop the core technologies of data science, that is, probabilistic modeling and machine learning, which are needed to harness the global societal changes caused by digitalization. Machine learning is already revolutionizing the world through solutions to problems where access to a wealth of data is commonplace and reasonably straightforward models can be employed. The existing solutions are not sufficient for the key problems in which a human is needed in the loop, which we focus on by combining existing excellence in the leading groups of the current Centre of Excellence in Computational Inference Research COIN (Kaski, Corander, Myllymäki), renewed with new rising talent in HCI (Oulasvirta), machine learning (strengthened by Vehtari) and precision medicine (through collaboration with a proposed CoE in the Institute for Molecular Medicine Finland FIMM; Pirinen). This combination of top-level expertise is rare and hence has breakthrough potential when matched to our four interrelated flagship challenges aimed at high impact by augmenting researchers and services with new computational-inference-based capabilities: (F1) Revolutionize human-computer interaction by making interface design modeling-based; (F2) Help revolutionize healthcare by developing new statistical methods needed for precision medicine; (F3) Develop open advanced inference engines for probabilistic programming, to make the techniques available and used across other fields of science and to industry and the rest of society, with (F4) information exploration as a prime case. To solve these problems, we will address a set of core methodological challenges. Inference needs to be made on (C1) complex and intractable models, for which our new scalable Approximate Bayesian Computation (ABC) is the key. Data-efficient learning from (C2) multiple data sources will be needed to solve the ubiquitous problem of “large p, small n but globally larger data.” Problem solving in practically important settings requires (C3) interaction between users and models, in which advanced modeling is currently under-utilized. We will scale up solutions to large data and more complex problems while still keeping the models as interpretable “white-box models.” In each challenge, our plan (below) goes clearly beyond the state of the art, and since the necessary expertise comes as a novel interdisciplinary combination of research groups in the strong already collaborating community in the Helsinki region, the planned significant new outcomes are also feasible.

## *Scientific merits*

**Samuel Kaski** (director; statistical machine learning) is a professor of computer science at Aalto SCI and Academy Professor (2016-2020). He has ample research leadership experience as the director of the centre of excellence this proposal renews (2015-2017) and of Helsinki Institute for Information Technology HIIT (2010-2015). He is well-known for his work on statistical machine learning for multiple data sources, for biology, medicine3 and information exploration19. He has published 224 peer-reviewed papers (GS h-index 45), incl. 2 in Nature Biotechnology, and supervised 19 PhDs. He has given 32 invited talks including 7 plenaries in conferences.

**Jukka Corander** (vice director; biostatistics) is a professor of statistics at UH and of biostatistics at UiO. He has supervised 14 PhDs and published over 160 articles (GS h-index 34), including 10 papers in the highest ranking scientific forums. He has received the ERC StG (SmartBayes, 2009-2014) and the Cozzarelli Prize from PNAS in 2015. The methods introduced by the group of Corander have led to numerous important discoveries on the evolution, resistance, virulence and transmission of pathogenic bacteria and viruses.

**Antti Oulasvirta** (PI; HCI) led the first HCI group at Max Planck Institute for Informatics before starting as an Assoc. Prof at Aalto ELEC. Oulasvirta was awarded an ERC StG for computational design of user interfaces (2015-2020). At 36 years of age, Oulasvirta's work has been cited 5,481 times and his GS h-index is 39. He has published 30 papers at ACM CHI, the flagship of HCI, and received eight Best Paper Awards and Honorable Mentions. He serves in the committees of CHI, IEEE Computer and International Journal of Human-Computer Studies. He has given 3 invited keynote talks in CS conferences.

**Petri Myllymäki** (PI; intelligent systems) is a professor of computer science at UH/CS, and also the current Director of the HIIT research institute. He has published over 150 peer-reviewed scientific articles in the area of intelligent systems and machine learning, with around 3500 citations, and his GS h-index is 27. His and his team member Prof. Giulio Jacucci's (GS h-index 28) extensive co-operation with the industry has led to a number of fielded applications, patents and spin-offs (Ekahau, cloudnsci.fi, BayesIT, screen.io, Multitaction, Etsimo, Spaceify).

**Matti Pirinen** (PI; computational genomics) is an Academy Research Fellow at the Institute for Molecular Medicine Finland (FIMM), UH. Pirinen's method development has been crucial in several international consortia including landmark studies on multiple sclerosis (Nature 2011) and gene expression (Science 2015). Pirinen has published 40 peer-reviewed articles, including 15 in Nature, Science or Nature Genetics and 6 in Bioinformatics, has h-index of 19 and over 3300 citations (GS). Outside academia, Pirinen provides consultancy services for GENOMICS plc, a company developing genome analytics.

**Aki Vehtari** (PI; Bayesian data analysis) is an associate professor at Aalto SCI and visiting professor at Technical University of Denmark. He is well known for his work on Bayesian model assessment and selection, Gaussian processes, approximative inference methods and survival analysis. He is a co-author of the best-selling book Bayesian Data Analysis, 3rd ed, and 69 articles (GS h-index 24, citations >19,400). He is a member of the core development team for Stan and the corresponding developer of free software for Gaussian processes (GPstuff21; in top 4% most downloaded machine learning open source software in mloss.org).

# *Foundations of Computational Health (FCHealth)*

Programme Director: Juho Rousu (Aalto)

Team leaders: Veli Mäkinen (vice director, UH), Tero Aittokallio (FIMM), Aristides Gionis (Aalto), Keijo Heljanko (Aalto), Ville Mustonen (UH), Jari Saramäki (Aalto)

## *Research Objectives*

We propose a Centre of Excellence in Foundations of Computational Health, with the aim of overcoming hard computational problems that need to be solved before data-driven healthcare can become reality. To this end, we have formed a consortium that comprises leading experts in computational genomics, predictive analytics, network science, and Big Data. This consortium is well-embedded in the Helsinki area healthcare and biomedical research ecosystems, and has a wide international network of collaborators. The key strength of our CoE lies in its unique, interdisciplinary combination of expertise, as well as its goal-oriented approach: the core computational challenges are chosen because of their relevance to improved analytics and, ultimately, to clinical practice.

Digitalization and the adoption of data-driven approaches is widely seen as one of the key remedies to the cost crisis that healthcare in the developed world is currently facing. This is a disruptive change that will require a new generation of computational models, methods and clinical tools. Great opportunities lie in the use genomic and other information in predictive analytics tools that allow personalized treatments and facilitate better and faster clinical decision-making. We identify three major challenges that stand in the way:

- **Big genomic data needs to be stored efficiently and made easily accessible.** This is required for new healthcare infrastructure that must be capable of handling personalized genomic information for large masses of patients.
- **Predictive analytics need to be improved**, in particular for complex diseases, which requires new ways of learning from structured heterogeneous data, and modelling complex relationships between individual genomes and medical phenotypes (e.g. disease outcomes and drug responses).
- **The gap to clinical practice needs to be bridged** and the developed solutions need to find their way to the hospital floor.

Our consortium is exceptionally well-positioned to overcome these challenges, with its proven track record in high-quality research, and collaboration with medical experts for rooting computational methods into practice. We propose a research programme organized in three major thrusts: (A) Core methods and theory, where we develop a well-founded, scientific basis for analytics tools and applications, (B) Next-generation

## *Scientific merits*

**Juho Rousu**, CoE Director, is an experienced leader with 10 externally funded projects with total budget of 3.6 Meur. Rousu acts as the Vice Chair of the Aalto Health Platform. He focuses on machine learning for structured data, especially small molecules (drugs, metabolites), with several path-breaking conceptual and methodological contributions in metabolite identification such as the state-of-the-art method CSI:FingerID, and metabolic network reconstruction and analysis.

**Veli Mäkinen**, Vice-director, is a co-author of a textbook on genome-scale algorithm design and mentor of several prominent postdocs on the topic. He has contributed to some early results on compressed text indexing and has more recently focused on tailoring those techniques to genomic data.

**Tero Aittokallio** is an EMBL-FIMM group leader at Institute for Molecular Medicine Finland (FIMM), where his group has a track record of collaborative projects that combine mathematical modelling with genome-wide profiling and high-throughput drug screening to provide personalized treatment predictions for patients with complex diseases. He is a PI in FIMM/HUS individualized systems medicine project, and PI in 5 AoF and 2 EU/H2020 projects.

**Aristides Gionis** has co-authored numerous highly-cited papers in data mining. He has contributed in areas such as graph mining, social-media analysis, web mining, data clustering, and privacy-preserving data mining. His work combines basic research with strong focus on applications, supported by six-year experience in industrial research.

**Keijo Heljanko** is an Associate Professor in distributed systems with focus on Big Data processing. His group has pioneered the use of Big Data platforms (Hadoop, Spark) for processing genomics data. He has extensive experience in SAT-based parallel NP-complete problem solving. He is the chairman of the Aalto Science-IT HPC infrastructure and a member of the advisory board for the Finnish national HPC infrastructure.

**Ville Mustonen** is a computational genomics faculty member at the Sanger Institute, UK. His group develops evolutionary theory and computational methods to analyse large scale genomic data. The work is done on systems of direct relevance to human health, e.g., cancer, infectious disease and evolution of drug resistance.

**Jari Saramäki** is a leading expert in network science, currently PI of 2 AoF projects. He has co-authored highly cited papers in weighted network methods, network-based computational social science, and recently, in temporal networks, where he is one of the best-known pioneers of the field.

# Administration

As noted earlier, administratively the organization of HIIT was completely renewed in Aalto in 2016. After the re-organization HIIT no longer maintains its own administrative staff, but relies of services of the hosting universities and their departments, but in some cases (IT services in particular), the migration of services is a relatively complicated process that could not be completed in full by the end of 2016. HIIT also restructured its research programmes: the new programmes started in August 2016 while the old programmes were allowed to maintain their operations until the end of 2016. To ensure compatibility with the earlier annual reports, the numbers presented in this report and its appendix are based on the old research programmes only. From 2017 onwards, the annual report will be based on the new research programmes.

## *Personnel and funding*

At Aalto University, HIIT researchers currently work in the Department of Computer Science, School of Science. An administrative merge was made in the beginning of 2016 to provide HIIT's services via the Computer Science department. This does not restrict researchers from other departments in joining HIIT's activities. At University of Helsinki, our operations are at the Department of Computer Science, Department of Mathematics and Statistics, and Department of Social Sciences. The personnel of HIIT are employed by the two parent universities. Many of HIIT's personnel have double affiliations. The most common is an affiliation with other Departments of one or both of the parent universities, but there are also some who share their time between HIIT and some other organisation. Thus the total number of HIIT affiliated personnel is much higher than the number of person-years, 126.

The total funding of HIIT in 2016 was 10 MEur, of which 80 per cent was competitive. The main funding sources were Academy of Finland, Tekes, EU and the mother universities. More details will be given in the Facts and Figures document found via http://www.hiit.fi/abouthiit

Aalto University

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

## *Board*

The highest decision-making body of HIIT is the Board. It decides on HIIT's overall research strategy and research programmes. The statutory tasks of the Board are to approve the annual budget and activity plans, and follow up and comment on the work of HIIT through regular activity updates given by the Director of HIIT. In 2016 the Board convened four times.











The HIIT Board consists of nine members who are appointed for four years at a time. Each university appoints four board members, two of which are university staff and two are not employed by the university, and personal deputies to each of them. The staff of HIIT selects one board member and his or her deputy from among their colleagues. In 2016, the Board consisted of the following members (personal deputies in parentheses):

- Chairman: Dean, Professor Risto Nieminen, Aalto until 31 May 2016 and starting from 1 June 2016 Dean, Professor Jouko Lampinen, Aalto (Vice Dean Esko Kauppinen, Aalto)
- Vice chairman: Dean, Professor Jouko Väänänen, UH (Vice Dean, Professor Esko Ukkonen, UH)
- Professor Pekka Orponen, Aalto until 31 July 2016 and starting from 1 August 2016 Professor Lauri Savioja (Professor Antti Ylä-Jääski, Aalto)
- Professor Sasu Tarkoma, UH (Professor Valtteri Niemi, UH)
- Vice President Hannu Kauppinen, Nokia Oyj (Director Jyri Huopaniemi, Nokia Oyj)
- Docent Kari-Pekka Estola (Director Ville Peltola, The Federation of Finnish Technology Industries)
- Director Mervi Karikorpi, The Federation of Finnish Technology Industries (Director Petri Vasara, Pöyry Management Consulting Oy)
- CEO Harri Valpola, Curious AI (CEO Kimmo Kiviluoto, Enreach Solutions)
- Personnel representative: Docent Antti Honkela, HIIT (Kai Kuikkaniemi, HIIT)

The Director of HIIT, Professor Petri Myllymäki, is responsible for preparing and submitting propositions to the Board. The Vice Director of HIIT starting from 14 September 2016 is Professor Juho Rousu. Board Secretary is Vice Director, Docent Ella Bingham. From 5 December 2016 the Board Secretary is DSc. Krisztina Cziner.




## *Scientific Advisory Board (SAB)*

The SAB consists of internationally prominent scholars who are invited by the HIIT Board once in approximately 4 years. The SAB convened in 22-23 September 2016 in Helsinki and its recommendations have been actively taken into account afterwards. The members of the SAB in 2016 were:

- Professor John Shawe-Taylor, University College London, UK; Chairman of the SAB
- Professor Elisabeth André, University of Augsburg, Germany
- Professor Richard Durbin, Wellcome Trust Sanger Institute, UK
- Professor Jodi Forlizzi, Carnegie Mellon University, USA
- Professor Seif Haridi, Swedish Institute of Computer Science, Sweden
- Professor Michael Mahoney, University of California Berkeley, USA
- Professor Kari-Jouko Räihä, University of Tampere, Finland
- Professor Wojciech Szpankowski, Purdue University, USA