https://www.elixir-denmark.org

Odense, August 27-28, 2015

Annual Danish Bioinformatics Conference

ELIXIR Denmark has organized the first Annual Danish Bioinformatics Conference. Via this event, the scientific community had the opportunity to form an overview of and get updated about what is happening within the fields of Bioinformatics and Systems Biology in Denmark.

ELIXIR Denmark is one of the nodes of ELIXIR, the European infrastructure for biological information that supports life science research and its translation to medicine, agriculture, bioindustries and society.

PROGRAMME

POSTER ABSTRACTS

LIST OF PARTICIPANTS

SCIENTIFIC COMMITTEE

MANAGEMENT COMMITTEE

#elixirdk

Supported by:

elixir DENMARK

Odense, August 27-28, 2015

Annual Danish Bioinformatics Conference

PROGRAMME

Thursday	Session	Session chair	Speakers
10:00 - 12:00	*Registration & brunch buffet*		
12:30 - 13:00	**Welcome**		• **View on Danish Bioinformatics**. Søren Brunak, Univ. of Copenhagen & Tech. Univ. of Denmark, Anders Krogh, University of Copenhagen
13:00 - 14:30	**Systems Biology and Medical Informatics**	Søren Brunak	• **Genomic epidemiology.** Ole Lund, Technical University of Denmark • **Adding context to genetics: ongoing studies in childhood ALL leukaemia and in diabetes resolution after bariatric surgery**. Ramneek Gupta, Technical University of Denmark • **Computational Breath Analysis.** Jan Baumbach, University of Southern Denmark
14:30 - 15:00	*Coffee break & poster walk*		
15:00 - 16:30	**Proteomics Informatics**	Ole Nørregaard	• **A global platform for community-wide interpretation of all mass spectrometry data.** Nuno Bandeira, University of California, San Diego • **Analytic framework for peptidomics applied to large-scale neuropeptide identification .** Christian Kelstrup, University of Copenhagen • **The pig protein atlas and an optimized porcine proteome database for clinical studies.** Tue Bennike, Aalborg University • **Systems level analysis of histone H3 posttranslational modifications reveals features of PTM crosstalk in chromatin regulation.** Veit Schwämmle, University of Southern Denmark
16:30 - 17:00	*Refreshments & poster walk*		
17:00 - 18:30	**RNA Bioinformatics**	Jan Gorodkin	• **A model for scoring damaging mutations in noncoding RNA.** Daniel Gautheret, Université Paris-Sud • **Prediction of conserved RNA structures and their functional implications**. Stefan Seemann, University of Copenhagen • **RNA expression and RNA structure**. Jakob Skou Pedersen, Aarhus University
20:00	*Dinner at Danmarks Jernbanemuseum, Dannebrorgsgade 24, 5000 Odense.* *The Museum opens at 19:00 for those of you interested in visiting the exhibition prior to the dinner.*		

Supported by:

elixir DENMARK

Odense, August 27-28, 2015

Annual Danish Bioinformatics Conference

PROGRAMME

Friday	Session	Session chair	Speakers
9:00 - 10:30	**Population Genetics**	Mikkel Schierup	• **The Danish pan-genome project.** Jacob Malte Jensen, Aarhus University • **Human adaptation to life in the high arctic.** Anders Albrechtsen, University of Copenhagen • **Incomplete lineage sorting reveals the action of selection and recombination in ancestral species.** Kasper Munch, Aarhus University
10:30 - 11:00	*Coffee break & poster walk*		
11:00 - 12:30	**Medical Genomics**	Albin Sandelin	• **Predictive quantitave modelling for stem cell engineering**. Franz-Josef Müller, University Medical Center Schleswig-Holstein • **The promoter and enhancer landscape of inflammatory bowel disease.** Morana Vitezic, University of Copenhagen • **Whole genome sequencing of antibiotic-resistant bacteria at Danish hospitals**. Peder Worning, Hvidovre Hospital
12:30 - 12:35	**CBIO vikings**		• **Connecting early career bioinformaticians in Denmark**. Alexander Junge, University of Copenhagen
12:35 - 13:30	*Lunch*		
13:30 - 15:00	**Industrial View on Danish Bioinformatics: Challenges and Opportunities**	Peter Løngreen	• **Bioinformatics – the Novo Nordisk way**. Dennis Madsen, Novo Nordisk • **Differential expression analysis and experimental validation with XploreRNA™** Niels Tolstrup, Exiqon • **Enzyme and microbe discovery in Novozymes**. Henrik M Geertz-Hansen, Novozymes.
15:00 - 15:30	**ELIXIR Denmark** **Closing of Meeting**	Peter Løngreen	• **ELIXIR Denmark, Computerome and Cloud computing**. Peter Løngreen, Technical University of Denmark
15:30	*Coffee & goodbye*		

Supported by:

POSTER ABSTRACTS 1 – 35

1

Error correction and Assembly of PacBio Reads on a Laptop in the new CLC Genome Finishing Module.

Martin Simonsen, Andreas Sand, Aske Simon Christensen, Martin Bundgaard, Leif Schauser, Arne Materna.

QIAGEN

Introduction: SMRT sequencing technologies, as implemented by Pacific Biosciences™, have the potential to vastly improve the completeness of genome sequence assemblies, as read lengths often exceed the length of most repeats in the genome. A major obstacle is the high (10-15%) error rate of SMRT reads. A second obstacle is the presence of chimeric reads and sequences derived from untrimmed adapters, which can be hard to recognize given the rate of errors and fractionations. However, as sequencing errors are mostly random and reads are randomly sampled across the genome, it is possible to correct SMRT sequencing reads if coverage is sufficiently high and ultimately assemble them into high-quality contigs.

Contribution and Availability: We have designed and implemented two novel software tools for i) error-correction of SMRT sequencing reads, and ii) de novo assembly of long high-quality reads. The two new tools will be available in the new CLC Genome Finishing Module for CLC Genomics Workbench from late June 2015: www.clcbio.com/clc-plugin/

Results: The two new tools in CLC Genome Finishing Module make it feasible to assemble microbial genomes on a laptop in less than half an hour. We demonstrate that we can correct and assemble a typical SMRT sequencing read data set from E. coli into one gold-standard quality contig in less than 30 minutes, with a peak memory usage of 2.2GB RAM. On a larger system with 32 cores, the same result is obtained in just 12 minutes 45 seconds. Furthermore, the assembly pipeline scales well to larger genomes such as S. cerevisae (12.15Mb) and C. elegans (100.3Mb).

2

IsoformSwitchAnalyzeR: Enabling Identification and Analysis of Isoform Switches with Functional Consequences from RNA-sequencing data.

Kristoffer Vitting-Seerup, Albin Sandelin.

The Bioinformatics Centre, University of Copenhagen, Denmark.

RNA-sequencing data is currently under utilized, in part because it is difficult to predict the functional consequences of changes in alternative transcription. The recent software improvements in full-length transcript deconvolution came with the promise of better understanding of these events, but unfortunately such analyses are still hard to obtain and only rarely done. To solve this problem we developed IsoformSwitchAnalyzeR. IsoformSwitchAnalyzeR is an easy to use R package that enables annotation of full-length RNA-seq derived transcripts with protein domains, signal peptides, coding potential as well as NMD sensitivity. Furthermore IsoformSwitchAnalyzeR supports identification (though a statistical model) and visualization of isoform switches, which together with the obtained annotation, allows for prediction of functional consequences due to alternative transcription events.

3

The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides.

Konstantinos Tsirigos , Christoph Peters , Nanjiang Shu , Lukas Käll, Arne Elofsson.

Stockholm University, Department of Biochemistry and Biophysics, Sweden.

The TOPCONS web server (http://topcons.net) was launched in 2009 and it is one of the most commonly used web servers regarding membrane protein topology prediction. In this work, we present a major update to the method, which, apart from a new combination of topology predictors used, can now additionally predict the presence of a signal peptide in the query sequence. Moreover, TOPCONS can differentiate between membrane and non-membrane proteins more efficiently, although a much larger database is used. The implementation of a batch submission and a queuing system additionally facilitates whole-proteome analyses.

4

Integrative visualisation and analysis of genomic track-data.

Dan Amlund Thomsen, Francesco Strino, Michael Lappe.

QIAGEN.

5

Identifying peaks in -seq data using shape information.

Francesco Strino, Michael Lappe.

QIAGEN.

6

Tools for gaining novel epigenetic insights from NGS data.

Igor Kardailsky, Jens Nielsen.

QIAGEN.

The Bisulfite Sequencing tools consists of a set of new software tools for analysing NGS data from bisulfite treated DNA samples to assist in calling methylated cytosines. Workflows for analysing bisulfite converted reads, including mapping reads to a reference using the state of the art CLC read mapper, calling differential levels of methylation from multiple samples can be set up. Analysis of Whole Genome Bisulfite Shotgun Sequencing (MethylC-seq) are supported as well as the commonly used Reduced Representation Bisulfite Sequencing (RRBS) high-throughput technique.

We present benchmarking results comparing the quality of the read mapping produce by our read mapper to those produced by state of the art open source projects (Bismark and BSSeeker). A preliminary benchmark on our differential methylation caller is also presented.

The Bisulfite Sequencing tools will be freely available as a plugin for the CLC Genomics - and Biomedical Workbench, and the CLC Genomics Server.

7 Insights from sequencing of Bifidobacterium from lab grown cultures and commercial fermentations.

Mette Jørgensen, Anders Krogh, Adam Baker.

The Bioinformatics Centre, University of Copenhagen, Denmark; Chr.Hansen A/S, Hørsholm, Denmark

Introduction: At Chr Hansen we have the largest bacterial fermentation production units in the world and a commitment to implement cutting edge technologies in our entire production process. One focus is that we are starting to sequence some of our strains through production. Bacteria evolve fast and mutations happen all the time. We are therefore interested in how many mutations that reside in the population during production and if the extreme environment in the fermenters influences the mutation rate? To study these questions we have sequenced a bacteria strain at three time points during the production process. In addition we are exploring the available software and the influence of the used tools.

Materials and methods: DNA was extracted from Bifidobacterium both grown in culture and as isolates on a plate. The DNA is harvested from frozen samples, the inoculation material and from freeze-dried product. The DNA has been sequenced on both a miseq(300bp paired end) and hiseq(100bp paired end). Various software packages were used to analyse the data.

Results: We managed to sequence the whole genome under all conditions. The assemblies from both the lab grown and commercial bacteria cover more than 99% of the reference genome. We also show that the software used to analyse the data has a huge impact on the results. An interesting result is that for miseq data error correction is important to reach a reasonable number of contigs while is it not as important for hiseq data. The contigs breaks were often found at the borders of rRNAs. Another result is that this specific strain seems to be homogenous.

Conclusion: It is possible to recover the whole genome from freeze-dried bacteria, but the software used has a high impact on the resulting assembly. The results also indicate that this strain is homogenous, but more experiments need to be done to validate this.

8 Development of Methods for Bulk Segregate Analysis in Polyploids to facilitate Marker Assisted Selection in Tetraploid Potato.

Mads Sønderkær, Kacper Kaminski, Mette S. Andersen, Hanne Grethe Kirk, Kåre L. Nielsen.

Aalborg University, Department of Chemistry and Bioscience, Denmark; Danish Potato Breeding Foundation, Vandel, Denmark.

Breeding for increased space and resource efficient crops is important to feed the world's increasing population and support a societal shift from a fossil fuel-based to a bio-based society. Crops with storage organs in the soil, like potato (S. tuberosum), produce twice the amount of calories per hectare with the same or less input compared to cereals, but breeding tools have been limited and thus a high unexploited potential is likely to present in the elite germplasm. The completion of the genome sequence of potato has enabled applications such as bulk segregate analysis (BSA) to rapidly identify genomic loci of agronomical interest and associated molecular markers to use in e.g. marker assisted selection (MAS) and Genomics Selection (GS). We have developed a next generation sequencing-assisted ultra-rapid BSA method, which can be applied on any species regardless of ploidity. There are three overall steps in the method: 1) Calling and filtering of marker positions. 2) Statistical testing of each marker position. 3) Identification of genomic regions with non-random nt distributions. We have applied the BSA method on a population of a ~ 5,000 individuals derived from a poly parental cross consisting of 19 tetraploid breeding clones. The population is derived from elite germplasm and is being phenotyped for six major traits in potato.

Here we present the BSA method developed, details regarding each step of the method; marker selection, choice of statistical test, and parameters for the sliding window analysis. Moreover, initial results from different data sets will be presented. The identified markers will be used to perform BSAs on the off spring enabling the development of a MAS and GS platform for potato.

9 Analysis of gene essentiality in Escherichia coli across strains and growth conditions.

Ida Bonde, Rebecca Lennen, João Cardoso, Anna Koza, Markus Herrgård.

DTU Biosustain, Technical University of Denmark, Hørsholm, Denmark.

Different types of knock-out studies have for years been applied in addressing the question of gene essentiality in various organisms. The development within the field of next generation sequencing has paved the way for more extensive studies due to the high throughput. One of these fairly resent methods is transposon insertion sequencing (Tn-Seq), in which a mutant library is constructed by randomly inserting transposons into the genome, the position of which is determined by sequencing. By knowing the number of inserts in each gene in the initial library it is possible to determine if genes are either essential or detrimental for growth in the test condition in question. In this study the TN-Seq method was used to investigate the differences in gene essentiality between four laboratory strains of E.coli subjected to four different growth conditions to investigate the reason for the differences in osmotic and chemical stress tolerance that exists between the strains as well as to assess the commonalities. Based on the sequencing data we identified genes that were essential for growth under the different conditions, some of which are essential in all conditions across strains and others that are specifically essential under certain growth conditions and/or in certain strains. This knowledge is important in the effort to engineer more stress tolerant strains, which are highly relevant for industrial purposes. Here is presented the bioinformatics analysis of the data, which includes one to one comparisons for each strain in each condition to the control condition and a multivariate analysis including all strains across conditions.

10 Maternal airway exposure to TiO2-nanoparticles: gene expression and promoter usage in the liver and placenta.

Mette Boyd, Morana Vitezic, Jette Bornholdt, Anne Thoustrup Saber, Ulla Vogel, Karin Sørig Hougaard, Albin Sandelin.

The Bioinformatics Centre and Biotech Research and Innovation Centre (BRIC), University of Copenhagen, and National Research Centre for the Working Environment, Denmark.

Maternal airway exposure to nanosized particles may interfere negatively with fetal development and affect organ function after birth. The placenta is the interface between mother and fetus, and has been suggested to be an important player in the transfer of effects of maternal exposure to particles to the fetus. In other words, the placenta has been hypothesized to adapt to the maternal environment in order to influence fetal development in a way that prepares the offspring for the environment later in life. To investigate the underlying molecular processes, we have generated genome-wide maps of active transcription start sites (TSSs) and enhancers in placenta, maternal and fetal liver tissue from pregnant mice with and without exposing the mice to nanoparticles (TiO2, a common component of paints and sunscreen lotions). We identified a large number of active promoters and enhancers, many of which change expression in response to the TiO2 exposure. Surprisingly, we find that maternal lung exposure leads to a rapid and dramatic down-regulation of gene expression in the placenta, while inducing pronounced up-regulation of inflammatory genes in maternal and embryonic liver.

11 Lipidome-Based Early Predictors of Type 2 Diabetes

Tommi Suvitaival, Isabel Bondia-Pons, Henna Cederberg, Alena Stančáková, Johanna Kuusisto, John Nolan, Tuulia Hyötyläinen, Markku Laakso, Matej Oresic.

Steno Diabetes Center, Gentofte, Denmark; Institute of Clinical Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

Background: Type 2 diabetes (T2D) is a multifactorial metabolic disorder, which is preceded by a long period of pre-symptomatic development. There is a need for early identification of progression from a state of normal glucose tolerance through pre-diabetes to T2D. Changes in serum lipid levels have previously been shown to precede the progression to T2D. In this work we studied, which lipids are the most promising candidates as early predictors of T2D, and what is their predictive power compared to established clinical risk factors.

Methods: In the METSIM (Metabolic Syndrome in Men) study, 954 participants were followed up for five years and their lipidomic serum profiles were measured with ultra-high performance liquid chromatography (UHPLC) coupled to quadrupole time-of-flight (QTOF) mass spectrometry (MS). In a data-driven way, we selected marker lipids for early prediction using nested cross-validated stepwise forward selection for the logistic regression model. We compared the selected new lipidomic models to established clinical models both within the cross-validation and with an independent validation data set.

Results: The most promising candidate markers for the early prediction of T2D were from the phosphatidylcholine and triglyceride lipid groups. The lipidomic markers consistently improved the prediction of T2D compared to established clinical risk-models throughout the validation. The marker lipids were found to be especially useful for identifying progressors among the high-risk subjects.

12 FoldAlign: multithreaded implementation for pairwise structural RNA alignment.

Daniel Sundfeld, Jakob H. Havgaard.

Center for non-coding RNA in Technology and Health, and IKVH, University of Copenhagen, Denmark; Department of Computer Science, University of Brasilia, Brasília, Brazil.

Foldalign is a Sankoff implementation for simultaneously aligning and folding two RNA sequences using several constrains to improve on resource consumption. Here, we a present a new multithreaded version, which substantially improve execution time while maintaining all previous functionalities and localization performance. The improvements allow for comparing longer RNAs and increase with the sequence length. For example, for sequences in the range 500-7870 nucleotides the parallel version improve execution of up to 4.91× while only increasing the memory consumption slightly (1.6×). This new version: (i) improves the quality of the predicted alignment by relaxing the constraints, such as increasing the size of the final alignment, (ii) produces the same results in a fraction of time and (iii) reduces the required memory while executing multiple jobs in parallel.

13 A finite mixture model under BLUP approach to detect differentially expressed genes among tissues in morbidly obese individuals.

Lisette J.A. Kogelman, Daniah Trabzuni, Marc Jan Bonder, Lude Franke, Peter C. Thomson, Haja N. Kadarmideen.

Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, UK; Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia; University of Groningen, University Medical Center Groningen, Department of Genetics, NL; ReproGen – Animal Bioscience Group, Faculty of Veterinary Science, The University of Sydney, Australia.

Background: Morbid obesity, the excessive accumulation of body fat, is a complex disease with major consequences for human health e.g. due to the various effects of excreted proteins by adipose tissues (e.g. adipokines). Morbid obesity is associated with several severe diseases, like Type 2 Diabetes, cardiovascular problems and several types of cancer. The biological mechanisms behind this association are mostly unclear, however, the biological complexity of morbid obesity indicates an important role for pathogenesis arising from different tissues/organs and interactions among them. Studying the different involved tissues and their interactions might provide a better insight in the genetic architecture of morbid obesity and its associated diseases. We therefore hypothesized that detecting differentially expressed (DE) genes between different organs in morbidly obese individuals provide important knowledge about morbid obesity.

Material and methods: We used publically available whole-transcriptome expression levels of subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), liver tissue, and muscle tissue of 93 morbidly obese individuals (26 males, 67 females) who were all phenotyped for 13 different metabolic parameters (e.g. glucose, cholesterol, and CRP). We estimated genetic random effects of the interactions between tissues and probes using BLUP (Best Linear Unbiased Prediction) linear models, which were subsequently used in a finite mixture model to detect DE genes in each tissue. This approach evades the multiple-testing problem and is able to detect variation among different biological states.

Results: Preliminary results show evidence of DE genes operating within tissues, but variation in these effect sizes was similar between the four tissues. We detected 17% of the transcripts to be DE in liver, 16% in muscle, 13% in SAT and 23% in VAT (Probability (DE) > 0.9).

Conclusion: Preliminary results show a decent number of differentially expressed genes, giving the potential of studying tissue-specific genes and interactions among tissues. Results will be used to gain insight into the underlying genetic and biological mechanisms by functional annotation of DE genes, detection of pathways. Furthermore, it can potentially be used as input for a network construction, leading to more insight in gene-gene interactions across tissues.

14 Genome-wide detection of regulatory active elements and their directionality using the histone read profile pattern.

Sachin Pundhir, Felicia Lauridsen, Frederik Otzen Bagger, Nicolas Rapin, Bo Porse.

The Bioinformatics Centre, Finsen Laboratory, Biotech Research and Innovation Centre (BRIC) at University of Copenhagen, Denmark.

H3K4me1 and H3K4me3 histone modifications are widely acknowledged as important distinguishing features between enhancer and promoter elements, respectively. However, the classical approach based on the enrichment of histone peaks cannot distinguish between regulatory active and inactive elements. We propose a novel approach utilizing peak-valley-peak read density profile of the two histone marks, reflecting the presence of nucleosome free regions (NFRs), to detect regulatory active elements. We show that our predictions are enriched for active roles in transcriptional regulation, and belong to a subset of previously reported enhancers and promoters. On extending the analysis across four human cell lines, predicted enhancers showed excellent correlation with H3K27ac enrichment, Pol-II binding and up-regulation in the expression of proximally located genes in cell line specific manner. We also show the distinct ratio of H3K4me3 and H3K4me1 up- and down-stream to promoters that is useful to predict directionality of stable transcription. Being solely based on single histone mark, we anticipate that our approach will be of interest to detect regulatory active elements for tissue samples where limited ChIP-seq assays are available. Also, due to its simple formulation, the proposed method is applicable to virtually any ChIP-based assay.

15 Designing Antisense LNA™ GapmeRs for efficient silencing of mRNA and lncRNA with high specificity

Asli Ozen, Jesper Culmsee Tholstrup, Francesco Favero, Niels M Frandsen, Johnathan Lai, Peter Mouritzen, Niels Tolstrup.

Exiqon, Vedbæk, Denmark.

Locked Nucleic Acid (LNA™) GapmeRs are powerful antisense tools for loss of function studies of proteins, mRNAs and lncRNAs. They are short single strand oligonucleotides and catalyze RNase H-dependent degradation of the complementary long RNA targets. GapmeRs contain a central region of DNA flanked by LNAs which increases target binding affinity regardless of the GC content. When bound to their fully complementary target LNA™) GapmeRs activates RNase H cleavage of the target RNA upon binding. Exiqon provides antisense LNA gapmers for any human, mouse and rat RNA targets that are longer than 80 bp. The design of the gapmers are performed by an empirically derived algorithm that is developed by Exiqon's Bioinformatics team. In this poster we present our improved design tool (available at http://www.exiqon.com/gapmer) that optimizes LNA gapmers against more than 30 parameters to identify the most effective and specific antisense oligonucleotides. Some of the important parameters we optimize against are melting temperature, target accessibility, GC content self-complementarity for achieving high potency and off-target analysis is performed in both spliced and unspliced transcriptomes for achieving specificity. Here we present the increased specificity of our gapmer designs by showing results from our improved off-target analysis.

16 NGS data analysis with XploreRNA.

Francesco Favero, Jesper Culmsee Tholstrup, Asli Ozen, Yuan Mang, Jacob Bock Axelsen, Jacopo Fabiani, Niels H Sørensen, Christopher Shaun F. Ramos, Tomasz Maciazek, Tomasz Kamiński, Troels Thomsen, Mihai-Marius Vlasceanu, Iben Plate, Johan Wahlin, Charlotte Klempel, Michael Thorsen, Peter Mouritzen, Niels Tolstrup.

Exiqon, Vedbæk, Denmark.

Transcriptomics is revolutionized by the next-generation sequencing technologies as they provide high-throughput expression data that gives the opportunity to perform multidimensional analyses of transcriptomes. Although it is becoming a very common approach in many research areas, an NGS analysis is heavy on computational resources and requires bioinformatics expertise which means the whole analysis part becomes extremely costly. Here we introduce XploreRNA – a high quality, yet cost-effective Next Generation Sequencing data analysis service for mRNA sequencing projects. It is designed as an extremely user-friendly service where the users simply upload the sequencing samples in FASTQ format, define their comparative analysis steps and the computational analysis is taken care of the Exiqon's data analysis pipeline. The pipeline provides a sequence quality control, performs mapping, differential expression analysis (including statistical analysis), and Gene Ontology enrichment analysis. In 5-7 working days, the whole analysis is finalized and the results are presented in a comprehensive report so that the researchers can have an overview of the important results. In the results section, we provide all the result files through a neatly designed file explorer. In the results exploration, we have also implemented an intuitive Transcript Sorting Wizard to select candidate transcripts and order custom reagents for qPCR validation or functional analysis that are Exiqon's expertise. XploreRNA can be accessed through our web interface at xplorerna.exiqon.com

17 **Fully automated RNA-seq Analysis Pipeline.**

Jesper Culmsee Tholstrup, Francesco Favero, Asli Ozen, Yuan Mang, Jacob Bock Axelsen, Michael Thorsen, Maria Wrang Teilum, Peter Mouritzen, Niels Tolstrup.

Exiqon, Vedbæk, Denmark.

Modern sequencing technologies generate large amounts of raw data and even comparably small experiments generate data in quantities, which pushes or exceeds the computational capabilities of personal computer sized machines. Consequently, to deliver overnight analysis results larger computers are required. However, with larger computers comes more complexity and the analysis scheduling mechanism can quickly turn into a bottleneck.

Exiqon has developed a RNA-Seq analysis pipeline, which features completely automated analysis all the way from raw fastq input to analysis results, such as differential expression analysis and unsupervised analysis. The pipeline features an internal scheduling system, which support asynchronous analysis of samples in parallel as well as analysis merge and branch points. The pipeline is able to handle projects containing hundreds of individual samples and is able to utilize large computers efficiently. The modular design of the pipeline allows Exiqon to offer many different types of analyses and we currently support analysis of miRNA, smallRNA, mRNA, and totalRNA samples. In addition, the modular design gives us the flexibility to offer new types of analysis while staying in a robust and tested core framework. In the poster, we will present an overview of the pipeline, resource usage graphs, and actual analysis results

18 **Integration and visualization of non-coding RNA and protein interaction networks**.

Alexander Junge, Jan C. Refsgaard, Christian Garde, Xiaoyong Pan, Alberto Santos, Christian Anthon, Ferhat Alkan, Christian von Mering, Christopher T. Workman, Lars Juhl Jensen, Jan Gorodkin.

Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark; Center for Biological Sequence Analysis, Technical University of Denmark; Center for non-coding RNA in Technology and Health, University of Copenhagen, Denmark; Institute of Molecular Life Sciences, Swiss Institute of Bioinformatics, University of Zurich, Switzerland.

Non-coding RNAs (ncRNAs) fulfill a diverse set of biological functions relying on interactions with other molecular entities. The advent of new experimental and computational approaches makes it possible to study ncRNAs and their associations on an unprecedented scale. We present RAIN (RNA Association and Interaction Networks) - a database that combines ncRNA-ncRNA, ncRNA-mRNA and ncRNA-protein interactions with large-scale protein association networks available in the STRING database. By integrating ncRNA and protein networks, RAIN provides a more complete picture of the cell's complex interaction network. RAIN aggregates associations and (predicted) interactions of a vast collection of ncRNA classes, including microRNAs and long ncRNAs, collected from a wide range of resources: a) curated knowledge, b) experimentally supported interactions, c) predicted microRNA-target interactions, and d) co-occurrences found by text mining Medline abstracts. Each resource was assigned a reliability score by assessing its agreement with a gold standard set of microRNA-target interactions. RAIN is available at: http://rth.dk/resources/rain

19 Efficient Methods for reference-free identification of SNVs

Lars Andersen, Fabio Vandin.

University of Southern Denmark; Brown University, USA.

Recent advances in DNA sequencing technologies have enabled the sequencing of whole genomes for a large number of individuals, including a large number of diseased individuals (e.g., cancer patients). Current DNA technologies produce millions or billions of short sequencing reads, and our ability to efficiently and rigorously analyze these large datasets plays a crucial part in advancing our understanding of the genetic basis of diseases like cancer. A first step in studying the genetic basis of diseases is the measurement of the variation, such as single nucleotide variants (SNVs), of a diseased genome compared to a normal genome. Current methods for SNVs identification rely on the accurate mapping of reads to a reference genome. Mapping of reads is a computationally expensive process, and a conspicuous fraction of the reads cannot be mapped (uniquely) to a reference genome and are therefore discarded from downstream analyses.

We present a new efficient reference-free approach for identifying SNVs by directly comparing two sequencing reads datasets. Our approach does use neither a reference genome nor complex data structures such as a de Bruin Graph. Our approach relies on an efficient locality-sensitive hashing scheme tailored for finding pairs of similar reads, and an algorithm used for detecting SNVs starting from groups of similar reads. We provide experimental evidence that the approach identifies SNVs with high sensitivity and high specificity.

20 How we aim to combat the most important disease in the mink industry.

Emma Elisabeth Hagberg.

Kopenhagen Diagnostics, Kopenhagen Fur / DTU Systems Biology, Denmark

Aleutian Mink Disease Virus (AMDV) is a frequently encountered pathogen associated with commercial mink fur breeding. AMVD infection leads to increased mortality and compromised animal health and welfare. Currently little is known about the molecular evolution of the virus, and the few existing studies have focused on limited regions of the viral genome. We have developed a robust, reliable, and fast protocol for amplification of AMDV genome using long-range PCR. We have amplified a number of AMDV isolates using this protocol, and sequenced their genomes using next generation sequencing (NGS). Evolutionary origin of the viral strains will be analysed using a range of model-based methods, mostly within a probabilistic (Bayesian) paradigm, including phylogenetic and geophylogenetic approaches. To our knowledge, this is the first study to apply next generation sequencing on the entire AMDV genome and thoroughly investigate its phylogeny. The results from the study will facilitate the development of new diagnostic tools and can form the basis for more detailed molecular epidemiological analysis of the virus.

21 Optimization of flanking region lengths for folding into constrained RNA structures: RNAcop.

Nikolai Hecker, Mikkel Christensen-Dalsgaard, Stefan E. Seemann, Jakob H. Havgaard, Peter F. Stadler, Ivo L. Hofacker, Henrik Nielsen, Jan Gorodkin.

Center for non-coding RNA in Technology and Health and Department of Veterinary Clinical and Animal Science and Department of Cellular and Molecular, Medicine, University of Copenhagen, Denmark; Bioinformatics Group, Department of Computer Science & IZBI-Interdisciplinary Center for Bioinformatics & LIFE-Leipzig Research Center for Civilization Diseases, University of Leipzig, Germany; Department of Theoretical Chemistry, University of Vienna, Austria.

RNA secondary structure prediction is often performed with the aim to identify functional elements. Especially for screens for structured non-coding RNAs or predicted structures in untranslated regions of a mRNAs, the length of such an element is not well defined. For in vitro experiments, however, a smaller part of an entire transcript comprising the predicted structure is usually extracted. Here, the composition of flanking nucleotides can disturb folding into the structure of interest.

Therefore, we developed a computational tool, RNAcop (RNA context optimization by probability), that optimizes folding into the structure of interest. Using constrained folding, our approach computes probabilities for folding into the structure of interest for all pair-wise combinations of flanking region lengths. Our analysis suggests that proper choices of flanking regions are crucial for a number of structures. The results are supported by in vitro experiments. RNAcop is available as web server and command-line tool at http://rth.dk/resources/rnacop.

22 Identification of promoters and enhancers induced by carbon nanotube exposure .

Jette Bornholdt, Berit Lilje, Anne Thoustrup Saber, Mette Boyd, Mette Jørgensen, Yun Chen, Morana Vitezic, Nicklas Raun Jacobsen, Sarah Søs Poulsen, Robin Andersson, Karin Sørig Hougaard, Carole L Yauk, Sabina Halappanavar, Håkan Wallin, Ulla Vogel, Albin Sandelin.

The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre, University of Copenhagen, Denmark; National Research Centre for the Working Environment, Copenhagen, Denmark; Environmental and Radiation Health Science Directorate, Ontario, Canada; Institute of Public Health, University of Copenhagen, Denmark; Department of Micro and Nanotechnology, Technical University of Denmark.

Usage of carbon nanotubes (CNTs) is increasing in industry due to their mechanical and electrical properties. However, pulmonary exposure to CNTs induces, an asbestos-like toxicological response characterized by persistent inflammation, granuloma formation and fibrosis with low no-effect levels. Little is known about the regulation of the response to CNTs. To this end, we have profiled transcription start sites and enhancers in mouse lung tissues following CNT exposure using Cap Analysis Gene Expression Assay (CAGE). This revealed a massive transcriptome response, with over 100-fold expression increases for key promoters, and a large change in transcription of enhancer regions linked to similarly responding genes. The response included key genes involved in inflammation, phagocytosis, cell and proliferation. We found a clear correlation between the overall CNT response strength and the number of alternative promoters in a given gene, but not the number of proximal enhancers. Upregulated genes after CNT exposure, where only the most annotated upstream promoter was upregulated, were associated to inflammation. Also NFkB binding sites were over-represented among these promoters. Conversely, upregulated genes where the upregulation could be attributed to promoters within the gene were not in particular linked to inflammation, and these promoters had distinct DNA motif enrichment patterns, not including the NFkB binding sites. Interestingly, NFkB binding sites were not over-represented in upregulated enhancer regions.

23 Convolutional LSTM Networks for Subcellular Localization of Proteins.

Henrik Nielsen, Søren Kaae Sønderby, Casper Kaae Sønderby, Ole Winther.

Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark; The Bioinformatics Centre, University of Copenhagen, Denmark.

Machine learning is widely used to analyze biological sequence data. Non-sequential models such as SVMs or feed-forward neural networks are often used although they have no natural way of handling sequences of varying length. Recurrent neural networks such as the long short term memory (LSTM) model on the other hand are designed to handle sequences. In this study we demonstrate that LSTM networks predict the subcellular location of proteins given only the protein sequence with high accuracy (0.902) outperforming current state of the art algorithms. We further improve the performance by introducing convolutional filters and experiment with an attention mechanism which lets the LSTM focus on specific parts of the protein. Lastly we introduce new visualizations of both the convolutional filters and the attention mechanisms and show how they can be used to extract biologically relevant knowledge from the LSTM networks.

24 The Bioinformatics Centre: Ongoing research and hiring opportunities.

Henriette Husum Bak-Jensen.

The Bioinformatics Centre, Section for Computational and RNA Biology, University of Copenhagen, Denmark.

25 Rhea, a manually curated resource of biochemical reactions.

Kristian B. Axelsen, Anne Morgat, Thierry Lombardot, Lucila Aimo, Anne Niknejad, Nevila Hyka-Nouspikel, Elisabeth Coudert, Nicole Redaschi, Lydie Bougeleret, Christoph Steinbeck, Ioannis Xenarios, Alan Bridge.

SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; Institute for Plant and Environmental Science, University of Copenhagen, Denmark.

Rhea (http://www.ebi.ac.uk/rhea) is a comprehensive and non-redundant resource of expert-curated biochemical reactions described using species from the ChEBI (Chemical Entities of Biological Interest) ontology of small molecules. Rhea is designed for the functional annotation of enzymes and the description of genome-scale metabolic networks, providing stoichiometrically balanced enzyme-catalyzed reactions, transport reactions and spontaneously occurring reactions. Rhea reactions are used as a reference for the reconciliation of genome-scale metabolic networks in the MetaNetX resource (www.metanetx.org) and serve as the basis for the computational generation of a library of theoretically feasible lipid structures in SwissLipids (www.swisslipids.org). Recent developments include the provision of reactions involving complex macromolecules such as proteins, nucleic acids and other polymers and substantial growth of Rhea through sustained literature curation efforts.

26 Transcription start site positioning is not nucleosome-dependent.

Axel Thieffry, Mette Boyd, Jette Bornholdt, Babett Steglich, Olaf Nielsen, Karl Ekwall, Albin Sandelin.

The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre, University of Copenhagen; Department of Biosciences and Nutrition, Karolinska Institutet, Sweden. Cell Cycle and Genome Stability Group, Department of Biology, University of Copenhagen.

The relation between the positioning of nucleosomes and the pre-initiation complex, and thereby active transcription start sites (TSSs) within core promoters, is poorly understood. Models where the TSS dictates the placement of the +1 nucleosome, and vice versa, have been proposed. Here, we show that depletion of members of the Fun30 class of chromatin remodelers in Schizosaccharomyces pombe leads to global shifts in chromatin, where the +1 nucleosome is often shifted one nucleosomal unit downstream. We combine this with genome-wide profiling of capped 5'-ends of RNAs to identify transcription and investigate the relation between the nucleosomal edge and TSS positioning. Strikingly, in the large majority of cases, the positioning of the nucleosomal edge has no impact on the placement of the TSS. Thus, our data strongly favor a model where TSS positioning is primarily sequence-dependent.

27 A Promoter Atlas of Schizosaccharomyces pombe Reveals Stress-induced Usage of Alternative Promoters.

Malte Thodberg, Ajuna Azad, Axel Thieffry, Jette Lange, Mette Boyd, Olaf Nielsen, Albin Sandelin.

The Bioinformatics Centre and Cell Cycle and Genome Stability Group, University of Copenhagen.

A large number of studies have shown that alternative promoter usage facilitates diversity and flexibility in gene expression and thus complexity in the final transcript and proteome. Still, the fine details of this regulatory mechanism remain unknown.

We performed Cap Analysis of Gene Expression (CAGE) on 15 samples of S. pombe, a common unicellular model organism, to investigate promoter usage under various stress conditions: nitrogen starvation, heat stress, oxidative stress, and change of growth medium. The resulting atlas of almost 14.000 promoters shows a consistent bias in the position of transcription start sites compared to reference annotation from PomBase as well as extensive usage of multiple promoters by single genes. This includes cases where alternative promoters are preferred under specific types of stress, so-called alternative promoter shifts.

This promoter atlas, along with differential expression information and alternative promoters, the first of its kind, will be made publically available through the PomBase online database to facilitate future research into S. pombe transcriptional regulation and promoter architecture.

28 An interactive environment for agile analysis and visualization of ChIP-sequencing data.

Mads Lerdrup, Jens Vilstrup Johansen, Shuchi Agrawal-Singh, and Klaus Hansen.

Biotech Research and Innovation Centre, Centre for Epigenetics, and The Bioinformatics Centre at University of Copenhagen, Denmark.

To empower experimentalists with means for fast and comprehensive ChIP-sequencing data analyses, we here introduce an integrated computational environment, EaSeq, which provides previously unseen levels of interactivity. The software combines the exploratory power of genome browsers with an extensive set of user-friendly cutting-edge tools for genome-wide abstraction and visualization. It thereby enables experimentalists to easily extract knowledge and generate hypotheses from their own data and hundreds of public genome-wide datasets. To demonstrate this, we performed meta-analyses of public Polycomb ChIP-sequencing data from mouse ES cells and established a novel screening approach, which identified both predicted and novel factors linked to Polycomb recruitment from all public mouse ES cell ChIP-seq. EaSeq, which is freely available and works on a standard PC, will radically increase the throughput of many analysis workflows, enable a broader group of scientists to gain insights, and increase transparency and reproducibility by automatically documenting and organizing analyses.

29 Bayesian transcriptome assembly.

Lasse Maretty, Jonas Andreas Sibbesen, Anders Krogh.

The Bioinformatics Centre, Department of Biology, University of Copenhagen, Denmark.

The massive throughput of second-generation RNAsequencing methods allows for simultaneous discovery and quantification of transcript variants and has thus dramatically increased our ability to explore complex transcriptomic landscapes. However, the short sequencing fragments and noise typical of second generation sequencing protocols complicate the assembly of transcripts from these data. The problem of transcriptome assembly is generally divided into the subproblems of first constructing a splicegraph and subsequently estimating which combination of transcripts or paths in the graph and associated abundance levels best explain the data. However, finding efficient solutions to the latter problem remains a major challenge. Indeed, current algorithms depend at least partly on heuristics to maintain accuracy and more robust, probabilistic approaches are thus needed. We introduce a new, Bayesian approach to splicegraph inference. Our main contribution is the derivation of a Bayesian model of the RNAsequencing process, which uses a novel prior distribution over transcript abundances to model the number of expressed transcripts. Importantly, this prior does not penalise lowly abundant transcripts in contrast to existing methods. Inference of the posterior distribution over possible transcripts and abundance values is conducted using an efficient Gibbs sampling method. Samples from the posterior distribution are then used to estimate both a confidence and an abundance estimate for each possible transcript. The confidence estimates in turn determine which transcripts are included in the final assembly and thus provide a rigorous method for controlling the tradeoff between recall and precision. Using this method, we demonstrate significant improvements in both recall and precision over state-of-the-art assemblers on simulated RNAsequencing data. More importantly, we also show marked improvements in assembly accuracy on multiple real RNAseq datasets as determined using annotations and thirdgeneration (PacBio) RNA sequencing data. Bayesembler is available at http://bioinformatics-centre.github.io/bayesembler/

30 Probabilistic genotyping without alignment.

Lasse Maretty, Jonas Andreas Sibbesen, Anders Krogh.

The Bioinformatics Centre, Department of Biology, University of Copenhagen, Denmark.

Current approaches to discovery and genotyping of complex variants from high-throughput genome sequencing data generally employ a combination of orthogonal variant callers to ensure variant sensitivity and thus genotyping accuracy. However, approaches capable of integrating such sets of variant calls with the raw data (and possibly previously annotated variants) to produce a consistent set of genotypes with confidence estimates are missing. We here present BayesTyper, a fully probabilistic approach to genotyping a population of individuals on a fixed set of arbitrarily complex variants (SNVs, insertions, deletions etc.). The method is based on exact alignment of read k-mers to a variant graph and hence has no intrinsic bias towards the reference sequence. The algorithm first counts all nonsingleton k-mers in the sequencing reads for each individual. Next, loci less than k-nucleotides apart are joined to form a variant graph in which all possible haplotypes are enumerated except for larger graphs, where a heuristic is used to limit the number of haplotypes. Finally, the multiset containing all k-mers found in haplotypes are enumerated and combined with the corresponding sample table to provide a vector containing the occurrences of haplotype k-mers for each sample. We model this observed count vector as generated by combining counts obtained from an individual's diplotype with counts originating from a carefully designed noise process. An individual's diplotype is in turn modelled as drawn from a shared population of haplotypes whose frequencies are modelled using a novel sparse prior. The posterior distribution over genotypes is inferred using collapsed Gibbs sampling of diplotypes, haplotype frequencies and noise states. Using simulations we demonstrate that our method can accurately and rapidly estimate genotypes composed of arbitrarily complex alleles including nested arrangements.

31 microRNA target sites act as regulatory hotspots in 3'UTRs.

Simon H. Ramussen, Mireya Plass, Anders Krogh.

The Bioinformatics Centre, Department of Biology, University of Copenhagen, Denmark.

microRNAs (miRNAs) are endogenous short non-coding RNAs that target mRNAs leading to degradation of the mRNA. miRNAs associate with Argonaute (AGO) proteins to perform their regulatory function, and other RNA binding proteins (RBPs) modulate it. Recently, it has been proposed that mRNAs compete for miRNA regulators, this is known as the competing endogenous RNA (ceRNA) hypothesis. If RBPs can modulate target site accessibility and miRNA association, they can regulate the ceRNA networks that miRNAs form. CLIP-seq is a high-throughput assay, which is used to map binding sites of a particular RBPs transcriptome wide. In this study, we have analyzed 119 previously published CLIP-seq datasets of 49 RBPs in HEK293 cells with the aim of understanding the role of RBPs in miRNA-mediated regulation. We mapped all RBP binding sites to 3'UTRs of protein coding mRNAs and found that similarly to miRNAs, most RBPs analyzed preferentially bind to the edges of 3'UTRs. An analysis of CLIP-seq read distribution around miRNA target sites showed that most RBPs are highly enriched right on the miRNA target sites. Interestingly, the binding sites of many RBPs show a strong positional correlation not only with AGO2 binding sites but also with other RBPs. This result points towards the existence of regulatory hotspots where many RBPs, including AGO proteins, bind. In summary, our results suggest the presence of regulatory hotspots enriched on miRNA target sites. In an AGO2 knock down analysis of a ceRNA network with 961 genes we show that a subset of genes that has no hotspots overlapping miRNA target sites generally are more down-regulated by miRNAs. Our interpretation is that most RBPs act as miRNA competitors at most target sites adding an extra layer of regulation of ceRNA networks.

32 Enriching siRNA Off-target Volume Discovery with Transcript Abundance and Accessibility .

Ferhat Alkan, Oana Palasca, Anne Wenzel, Peter F. Stadler, Ivo L. Hofacker, Jan Gorodkin.

Center for non-coding RNA in Technology and Health, University of Copenhagen, Denmark; Department of Veterinary Clinical and Animal Science, University of Copenhagen, Denmark; Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark; Bioinformatics Group, Department of Computer Science & IZBI-Interdisciplinary Center for Bioinformatics & LIFE-Leipzig Research Center for Civilization Diseases, University Leipzig, Leipzig, Germany; Institute for Theoretical Chemistry, University of Vienna, Wien, Austria; Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany; Santa Fe Institute, Santa Fe, USA.

Sequence-specific off-target effects of siRNAs are caused by unintended interactions between the transfected siRNA and functional transcripts other than the intended target. These off-target interactions can cause cleavage of off-targeted transcripts with near-perfect complementarity or can trigger miRNA-like silencing. Given the multiple possibilities of forming stable interactions, off-target effects cannot be avoided, however, these effects could be minimized by comparing various siRNAs by their potential off-target volumes. Prediction of these unintended interactions on large scale recently became feasible with suffix-array based RIsearch 2.0 algorithm, and here we present a comprehensive method that makes use of these predictions to measure off-target volume of given siRNAs by also taking transcript expression levels and binding site accessibilities into account.

33 GPCRdb Structural Data and Tools for G Protein-Coupled Receptors .

Vignir Isberg, Stefan Mordalski, Christian Munk, Alexander Hauser, Kasper Harpsøe, Krzysztof Rataj, Andrzej Bojarski, Gerrit Vriend, David E. Gloriam.

Dept of Drug Design and Pharmacology, University of Copenhagen, Denmark; Department of Medicinal Chemistry, Polish Academy of Sciences, Krakow, Poland; CMBI, Radboud University Medical Centre, Nijmegen, The Netherlands.

GPCRdb (gpcrdb.org) has been a popular resource for the G protein-coupled receptor community for the past 20 years and today has around 1500 users every month (1). GPCRdb contains experimental data on crystal structures, mutants, as well as computationally derived sequence alignments and homology models. The latest release has added user-friendly web browser structural tools (2,3).

References

1 Isberg, V. et al. GPCRDB: an information system for G protein-coupled receptors. Nucleic Acids Res 42, D422-425, doi:10.1093/nar/gkt1255 (2014).

2 Isberg, V. et al. Generic GPCR residue numbers - aligning topology maps while minding the gaps. Trends Pharmacol. Sci. 36, 22-31, doi:10.1016/j.tips.2014.11.001 (2015).

3 Fidom, K. et al. A new crystal structure fragment-based pharmacophore method for G protein-coupled receptors. Methods 71, 104-112, doi:10.1016/j.ymeth.2014.09.009 (2015).

34 Tryggve project - IT Services for sensitive biomedical data.

Antti Pursula.

CSC - IT Center for Science Ltd., Espoo, Finland; Nordic e-Infrastructure Collaboration (NeIC), Oslo, Norway.

Tryggve is a Nordic project established to support cross-border collaboration on sensitive biomedical data. The aim is to develop and connect existing capacities and services to enable researchers better to transfer, store and process the data in high-security environment across Nordic countries. The project is supported by the ELIXIR Nodes in Denmark, Finland, Norway and Sweden, and the Nordic e-Infrastructure Collaboration NeIC.

Outcomes of the Tryggve project will improve the research environment for biomedical science and in this way facilitate improvement of human well-being. The experiences of the Nordic project are extendable for wider international application. The project works on several areas to support research on sensitive data, including development of secure processing environments in different countries, improving their interoperable use and analysing legal requirements for cross-Nordic studies.

Tryggve supports also concrete research use cases, which can be proposed by research teams spanning more than one Nordic country. More information on project and use case proposals are found on the project website wiki.neic.no/Tryggve.

35 Protein raftophilicity. How bioinformatics can help membranologists.

Henrik Nielsen, Maria Maddalena Sperotto.

Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark.

Protein raftophilicity is the affinity of proteins for lipid ‘rafts’. Rafts denote nano- and submicro-sized biomembrane domains that are enriched in cholesterol and sphingolipids. These domains are considered relevant for maintaining specialized structures that constitute suitable sites for bioprocesses (1).

Protein raftophilicity depends on features such as lipidation and GPI-anchoring. Can this affinity be inferred solely by knowing such features, without knowing the physical and physico-chemical properties of biomembranes?

We tried to answer the question by an artificial neural network (ANN)-based bioinformatics approach. The ANN was trained to recognize feature-based patterns in proteins that are considered to be associated with lipid rafts. The trained ANN was then used to predict protein raftophilicity.

We found that, in the case of α-helical membrane proteins, their hydrophobic length does not affect their raftophilicity. This is in agreement with confocal microscopy experiments on DOPC/SM/cholesterol bilayers with reconstituted model peptides, P-23 and P-29 (2).

(1) Nicolson, G.L.: The Fluid-Mosaic Model of Membrane Structure: still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years. Biochim. Biophys. Acta 1838, 1451-1466 (2014)

(2) Vidal A. and T.J. McIntosh (2005). Transbilayer Peptide Sorting between Raft and Nonraft Bilayers: Comparisons of Detergent Extraction and Confocal Microscopy. Biophys J. 89, 1102-1108. 2005.

Annual Danish Bioinformatics Conference

LIST OF PARTICIPANTS

No.	Last name	Name	Company/ Institution	Email	Country
1	Ahrenfeldt	Johanne	Technical University of Denmark	johah@cbs.dtu.dk	Denmark
2	Albrechtsen	Anders	University of Copenhagen	albrecht@binf.ku.dk	Denmark
3	Alcaraz	Nicolas	University of Southern Denmark	nalcaraz@imada.sdu.dk	Denmark
4	Alkan	Ferhat	University of Copenhagen	ferro@rth.dk	Denmark
5	Almagro Armenteros	Jose Juan	University of Copenhagen	jjalmagro@gmail.com	Denmark
6	Al-Nakeeb	Kosai	Technical University of Denmark	kosai@cbs.dtu.dk	Denmark
7	Andersen	Sara Eiriksdottir	Technical University of Denmark	saraeandersen@gmail.com	Denmark
8	Andersson	Robin	University of Copenhagen	robin@binf.ku.dk	Denmark
9	Anthon	Christian	University of Copenhagen	anthon@rth.dk	Denmark
10	Aparicio Puerta	Ernesto Luis	University of Copenhagen	erapariciopuerta@hotmail.com	Denmark
11	Argelaguet	Ricard	University of Copenhagen	crt474@alumni.ku.dk	Denmark
12	Axelsen	Kristian	University of Copenhagen	kristian.axelsen@isb-sib.ch	Denmark
13	Azad	Ajuna	University of Copenhagen	azadajuna@gmail.com	Denmark
14	Bandeira	Nuno	University of California	bandeira@ucsd.edu	The United States
15	Baumbach	Jan	University of Southern Denmark	jan.baumbach@imada.sdu.dk	Denmark
16	Bellod Cisneros	Jose Luis	Technical University of Denmark	Cisneros@cbs.dtu.dk	Denmark
17	Bennike	Tue Bjerg	Aalborg University	tbe@hst.aau.dk	Denmark
18	Bergsten	Mads	University of Copenhagen	obi.sbc@gmail.com	Denmark
19	Besenbacher	Søren	Aarhus University	besenbacher@clin.au.dk	Denmark
20	Bjerregaard	Anne-Mette	Technical University of Denmark	ambj@cbs.dtu.dk	Denmark
21	Blair	Joseph	University of Copenhagen	joeblair2000@hotmail.com	Denmark
22	Bonde	Ida	Technical University of Denmark	idab@biosustain.dtu.dk	Denmark
23	Boomsma	Wouter	University of Copenhagen	wb@bio.ku.dk	Denmark
24	Boyd	Mette	University of Copenhagen	metteboyd@binf.ku.dk	Denmark
25	Brunak	Søren	Tech. Univ. of Denmark, Univ. of Copenhagen	brunak@cbs.dtu.dk	Denmark
26	Bøgsted	Martin	Aalborg University	martin.boegsted@rn.dk	Denmark
27	Calle Sánchez	Xabier	University of Copenhagen	xcalle91@gmail.com	Denmark
28	Canbäck	Björn	Lund University	bjorn.canback@biol.lu.se	Sweden
29	Cantwell	Lisa	University of Copenhagen	lissacantwell@gmail.com	Denmark
30	Castro	Josue	University of Copenhagen	jcame@food.ku.dk	Denmark
31	Chalabi Hajkarim	Morteza	University of Southern Denmark	MCHALABI@IMADA.SDU.DK	Denmark
32	Chandrasai	Potla Durthi		chandrasaip@gmail.com	India
33	Chen	Yun	University of Copenhagen	cylsae@gmail.com	Denmark
34	Chmura	Piotr	Technical University of Denmark	pjchmura@cbs.dtu.dk	Denmark
35	Dalby	Maria	University of Copenhagen	maria.dalby@bio.ku.dk	Denmark
36	Doktor	Thomas Koed	University of Southern Denmark	thomaskd@bmb.sdu.dk	Denmark
37	Duan	Jinjie	Aarhus University	jjduan@birc.au.dk	Denmark
38	Dworzynski	Piotr	Statens Serums Institut	p.dworzynski@gmail.com	Denmark
39	Dybdahl	Marianne	Technical University of Denmark	mdyb@food.dtu.dk	Denmark
40	Eklund	Aron	Technical University of Denmark	eklund@cbs.dtu.dk	Denmark
41	Ellabaan	Mostafa	Technical University of Denmark	mohel@bio.dtu.dk	Denmark
42	Etches	Zsuzsanna Sükösd	Qiagen Aarhus A/S	zsuzsanna.etches@qiagen.com	Denmark
43	Falkenby	Lasse	University of Southern Denmark	lassef@bmb.sdu.dk	Denmark
44	Favero	Francesco	Exiqon	fra@exiqon.com	Denmark
45	Ferreira	Lucas	University of Southern Denmark	lmferreira@imada.sdu.dk	Denmark
46	Fløyel	Tina	Herlev and Gentofte Hospital	tina.floeyel.01@regionh.dk	Denmark
47	Folkersen	Lasse	Technical University of Denmark	lasfol@cbs.dtu.dk	Denmark
48	Fonseca	Rute	University of Copenhagen	rute.r.da.fonseca@gmail.com	Denmark
49	Freitas	Flavia	University of Copenhagen	flaviacpfreitas@gmail.com	Denmark
50	Friborg	Rune	Aarhus University	runef@birc.au.dk	Denmark
51	Friis Jørgensen	Isabella	Technical University of Denmark	isfj@bio.dtu.dk	Denmark
52	Friis-Nielsen	Jens	Technical University of Denmark	jensfriisnielsen@gmail.com	Denmark
53	Gautheret	Daniel	Université Paris-Sud	daniel.gautheret@u-psud.fr	France

Annual Danish Bioinformatics Conference

No.	Last name	Name	Company/ Institution	Email	Country
54	Geertz-Hansen	Henrik M.	Novozymes	hmgh@novozymes.com	Denmark
55	Gorodkin	Jan	University of Copenhagen	gorodkin@rth.dk	Denmark
56	Gupta	Ramneek	Technical University of Denmark	ramneek@cbs.dtu.dk	Denmark
57	Hadsbjerg	Johanne	Technical University of Denmark	johad@vet.dtu.dk	Denmark
58	Hagberg	Emma	Kopenhagen Fur	p.hertz@outlook.com	Denmark
59	Hallin	Peter Fischer	Novozymes	pfha@novozymes.com	Denmark
60	Hannani	Monica	Technical University of Denmark	monicahannani@gmail.com	Denmark
61	Harder	Lea Mørch	University of Southern Denmark	lmh@bmb.sdu.dk	Denmark
62	Hauschild	Anne-Christin	University of Southern Denmark	hauschild@imada.sdu.dk	Denmark
63	Hauser	Alexander	University of Münster	alexshauser@gmail.com	Denmark
64	Havgaard	Jakob Hull	University of Copenhagen	hull@rth.dk	Denmark
65	Hecker	Nikolai	University of Copenhagen	hecker@rth.dk	Denmark
66	Henius	Anna	Technical University of Denmark	A.henius@gmail.com	Denmark
67	Hoffmeyer	Alexandra	Technical University of Denmark	s123645@student.dtu.dk	Denmark
68	Holtsmark Nielsen	Sofie	Technical University of Denmark	Sofie.holtsmark@gmail.com	Denmark
69	Hother	Christoffer Egeberg	University of Copenhagen	rzs216@alumni.ku.dk	Denmark
70	Hu	Jessica	University of Copenhagen	jessicaxhu@gmail.com	Denmark
71	Husum Bak-Jensen	Henriette	University of Copenhagen	hhu@bio.ku.dk	Denmark
72	Højholt Kristensen	Katrine	Slagelse Sygehus	katrine.hojholt@gmail.com	Denmark
73	Ibrisimovic	Dino	Technical University of Denmark	s103500@student.dtu.dk	Denmark
74	Isberg	Vignir	University of Copenhagen	vis@sund.ku.dk	Denmark
75	Ison	Jon	Technical University of Denmark	jison@cbs.dtu.dk	Denmark
76	Izarzugaza	Jose MG	Technical University of Denmark	txema@cbs.dtu.dk	Denmark
77	Jakobsen	Lasse Hjort	Aalborg University	lasse.j@rn.dk	Denmark
78	Jansen	Peter Gildsig	Novo Nordisk	pcgj@novonordisk.com	Denmark
79	Jensen	Jacob Malte	Aarhus University	jmj@birc.au.dk	Denmark
80	Jensen	Ole Nørregaard	University of Southern Denmark	jenseno@bmb.sdu.dk	Denmark
81	Jespersen	Marie Louise	Technical University of Denmark	Marie.l.jespersen@gmail.com	Denmark
82	Jessen	Leon Eyrich	Steno Diabetes Center A/S	leyj@steno.dk	Denmark
83	Jiang	Li	Technical University of Denmark	jiang@cbs.dtu.dk	Denmark
84	Johnsen	Hans E	Aalborg University	haej@rn.dk	Denmark
85	Joshi	Hiren	University of Copenhagen	joshi@sund.ku.dk	Denmark
86	Juncker	Agnieszka	Technical University of Denmark	ag@cbs.dtu.dk	Denmark
87	Junge	Alexander	University of Copenhagen	ajunge@rth.dk	Denmark
88	Jørgensen	Mette	KU and Chr Hansen A/S	mette@binf.ku.dk	Denmark
89	Jørsboe	Emil	University of Copenhagen	jorsboe@hotmail.com	Denmark
90	Kang	Taewook	University of Southern Denmark	systemkang@gmail.com	Denmark
91	Karadoulama	Evdoxia	Aarhus University	evka@mbg.au.dk	Denmark
92	Kelstrup	Christian	University of Copenhagen	christian.kelstrup@cpr.ku.dk	Denmark
93	Kirsch	Rebecca	University of Copenhagen	rebecca@rth.dk	Denmark
94	Kjær	Jonas	Technical University of Denmark	jonak@vet.dtu.dk	Denmark
95	Kogelman	Lisette	University of Copenhagen	lkog@sund.ku.dk	Denmark
96	Kovacevic	Milos	University of Copenhagen	milos.cz@gmail.com	Denmark
97	Krogh	Anders	University of Copenhagen	krogh@binf.ku.dk	Denmark
98	Krych	Lukasz	University of Copenhagen	krych@life.ku.dk	Denmark
99	Lange	Jette Bornholdt	University of Copenhagen	bornholdt@binf.ku.dk	Denmark
100	Lappe	Michael	Qiagen Aarhus A/S	michael.lappe@qiagen.com	Denmark
101	Lilje	Berit	Statens Serum Institut	beritlilje@hotmail.com	Denmark
102	List	Markus	University of Southern Denmark	mlist@health.sdu.dk	Denmark
103	Liu	Siyang	University of Copenhagen	siyang.liu@bio.ku.dk	Denmark
104	Lund	Ole	Technical University of Denmark	lund@cbs.dtu.dk	Denmark
105	Løngreen	Peter	Technical University of Denmark	peterl@cbs.dtu.dk	Denmark
106	Madsen	Dennis	Novo Nordisk	dnnm@novonordisk.com	Denmark

Annual Danish Bioinformatics Conference

No.	Last name	Name	Company/ Institution	Email	Country
107	Madsen	Kim	Qiagen Aarhus A/S	kim.madsen@qiagen.com	Denmark
108	Marcussen	Mette Haagen	Technical University of Denmark	mehma@bio.dtu.dk	Denmark
109	Maretty	Lasse	University of Copenhagen	lassemaretty@binf.ku.dk	Denmark
110	Marinho Almeida	Diogo	University of Southern Denmark	diogoma@bmb.sdu.dk	Denmark
111	Martinez	Sara Christina	Technical University of Denmark	saracm@dtu.dk	Denmark
112	Menzel	Peter	University of Copenhagen	ptr@binf.ku.dk	Denmark
113	Moltke	Ida	University of Copenhagen	ida@binf.ku.dk	Denmark
114	Munch	Kasper	Aarhus University	kaspermunch@me.com	Denmark
115	Munk	Christian	University of Copenhagen	christian@munk.be	Denmark
116	Ng	Kim	Statens Serum Institut	kimn@ssi.dk	Denmark
117	Nguyen	Myhanh	Technical University of Denmark	myng@bio.dtu.dk	Denmark
118	Nielsen	Henrik	Technical University of Denmark	hnielsen@cbs.dtu.dk	Denmark
119	Nielsen	Jens	Qiagen Aarhus A/S	jens.nielsen@qiagen.com	Denmark
120	Nielsen	Lars Bo	University of Copenhagen	lars.bo.nielsen@regionh.dk	Denmark
121	Novotny	Peter	University of Copenhagen	peternovot@gmail.com	Denmark
122	Nygaard	Sanne	University of Copenhagen	sanne@bio.ku.dk	Denmark
123	Nørgaard	Lasse	CIMT	lasse.noergaard@regionh.dk	Denmark
124	Nørregaard-Madsen	Mads	Nuevolution A/S	mn@nuevolution.com	Denmark
125	Ohlsen	Signe Krogh	University of Southern Denmark	siohl12@student.sdu.dk	Denmark
126	Okstoft Carmelo	Victor Adriano	University of Copenhagen	victor.carmelo@sund.ku.dk	Denmark
127	Ozen	Asli Ismihan	Exiqon	aoz@exiqon.com	Denmark
128	Palasca	Oana	University of Copenhagen	oana.palasca@gmail.com	Denmark
129	Pan	Xiaoyong	University of Copenhagen	panxy@rth.dk	Denmark
130	Pan	Xiaoyu	University of Copenhagen	xiaoyu.pan@sund.ku.dk	Denmark
131	Papaleo	Elena	Danish Cancer Society	elena.papaleo.78@gmail.com	Denmark
132	Pedersen	Casper-Emil Tingskov	University of Copenhagen	capedersen@bio.ku.dk	Denmark
133	Pedersen	Jakob Skou	Aarhus University	jmj@birc.au.dk	Denmark
134	Petersen	Bent	Technical University of Denmark	bent@cbs.dtu.dk	Denmark
135	Pickles	Rob	Roche Diagnostics	robert.pickles@roche.com	Denmark
136	Piedrahita Abella	Sonia Maria	University of Copenhagen	spiedrahita@yahoo.com	Denmark
137	Pociot	Flemming	Herlev and Gentofte Hospital	flemming.pociot.01@regionh.dk	Denmark
138	Poulsen	Nina	Aarhus University	nina.poulsen@food.au.dk	Denmark
139	Pultz	Dennis	University of Southern Denmark	dennispultz@gmail.com	Denmark
140	Pursula	Antti	NeIC	antti.pursula@csc.fi	Finland
141	Rapacki	Kristoffer	Technical University of Denmark	rapacki@cbs.dtu.dk	Denmark
142	Rasmussen	Jane	Technical University of Denmark	jlnr@bio.dtu.dk	Denmark
143	Rasmussen	Simon	Technical University of Denmark	simon@cbs.dtu.dk	Denmark
144	Rasmussen	Simon H.	University of Copenhagen	simras@binf.ku.dk	Denmark
145	Ravn	Christian		christian.ravnen@gmail.com	Denmark
146	Rennie	Sarah	University of Copenhagen	sarah.rennie@igmm.ed.ac.uk	Denmark
147	Rigbolt	Kristoffer	Gubra	kri@gubra.dk	Denmark
148	Robles Tascón	Lucía	University of Copenhagen	luciaroblestascon@gmail.com	Denmark
149	Rosenberg	Sine	Technical University of Denmark	siro@food.dtu.dk	Denmark
150	Rud Nielsen	Minna	Technical University of Denmark	minna.rud.nielsen@gmail.com	Denmark
151	Rundsten	Carsten	University of Copenhagen	carsten.rundsten@gmail.com	Denmark
152	Rydza	Emil	Technical University of Denmark	ekr@cbs.dtu.dk	Denmark
153	Ryssel	Mia	University of Copenhagen	mry@food.ku.dk	Denmark
154	Rønn Olsen	Lars	Technical University of Denmark	lro@cbs.dtu.dk	Denmark
155	Röttger	Richard	University of Southern Denmark	roettger@imada.sdu.dk	Denmark
156	Salamaca	Juan	University of Copenhagen	juan.salamanca.viloria@gmail.com	Denmark
157	Sandelin	Albin	University of Copenhagen	albin@binf.ku.dk	Denmark
158	Sanusi	Samsudeen	University of Tampere	samsudeensanusi2012@yahoo.com	Finland
159	Schauser	Leif	Qiagen Aarhus A/S	leif.schauser@qiagen.com	Denmark
160	Schwammle	Veit	University of Southern Denmark	veits@bmb.sdu.dk	Denmark

Annual Danish Bioinformatics Conference

No.	Last name	Name	Company/ Institution	Email	Country
161	Seemann	Stefan	University of Copenhagen	seemann@rth.dk	Denmark
162	Sibbesen	Jonas Andreas	University of Copenhagen	jasi@binf.ku.dk	Denmark
163	Sicheritz-Ponten	Thomas	Technical University of Denmark	thomas@cbs.dtu.dk	Denmark
164	Siegismund	Hans R.	University of Copenhagen	HSiegismund@bio.ku.dk	Denmark
165	Simonsen	Martin	Qiagen Aarhus A/S	martin.simonsen@qiagen.com	Denmark
166	Skov	Laurits	BIRC (bioinformatics research center) Aarhus	lauritsskov2@gmail.com	Denmark
167	Sperotto	Maria Maddalena	Technical University of Denmark	maria@cbs.dtu.dk	Denmark
168	Stanchev	Kiril	Technical University of Denmark	stanchev.kiril@yahoo.com	Denmark
169	Staunstrup	Line Mærsk	University of Copenhagen	linestaunstrup@gmail.com	Denmark
170	Stensballe	Allan	Aalborg University	as@hst.aau.dk	Denmark
171	Stæhr	Katrine		kagetrine90@hotmail.com	Denmark
172	Sundfeld	Daniel	University of Copenhagen	danielsundfeld@gmail.com	Denmark
173	Suvitaival	Tommi	Steno Diabetes Center A/S	tsvv@steno.dk	Denmark
174	Syed	Ali	Technical University of Denmark	Alisyed@cbs.dtu.dk	Denmark
175	Særmark	Pia	University of Copenhagen	psaermark@gmail.com	Denmark
176	Søndergaard	Dan	Aarhus University	das@birc.au.dk	Denmark
177	Sønderkær	Mads	Aalborg University	mson@bio.aau.dk	Denmark
178	Saaby Schmidt	Anne	University of Copenhagen	annesaabyschmidt@live.dk	Denmark
179	Teku	Gabriel	Lund University	gabriel.teku@med.lu.se	Sweden
180	Tetzschner	Anna Maria Malberg	Technical University of Denmark	ammte@cbs.dtu.dk	Denmark
181	Theobald	Sebastian	Technical University of Denmark	setd@bio.dtu.dk	Denmark
182	Thieffry	Axel	University of Copenhagen	axel.thieffry@bio.ku.dk	Denmark
183	Thodberg	Malte	University of Copenhagen	maltethodberg@gmail.com	Denmark
184	Tholstrup	Jesper Culmsee	Exiqon	jet@exiqon.com	Denmark
185	Thompson	Wesley	Region Hovedstadens Psykiatri	wkthompson@ucsd.edu	Denmark
186	Thorn Ekstrom	Claus	University of Copenhagen	ekstrom@sund.ku.dk	Denmark
187	Timshel	Pascal	SSI/KU/DTU	pascal.timshel@gmail.com	Denmark
188	Tolstrup	Niels	Exiqon	nt@exiqon.com	Denmark
189	Tsirigos	Konstantinos	Stockholm University	kostas.tsirigos@scilifelab.se	Sweden
190	Tsitlakidis	Evangelos	University of Copenhagen	vagosmae@gmail.com	Denmark
191	Tsolakou	Anna	University of Copenhagen	anna.tsolakou@gmail.com	Denmark
192	Van Hauwaert	Elvira Laila	University of Southern Denmark	elvan12@student.sdu.dk	Denmark
193	Vieira Barbosa	Eudes Guilherme	University of Southern Denmark	eudes@imada.sdu.dk	Denmark
194	Vilstrup Johansen	Jens	University of Copenhagen	jens@binf.ku.dk	Denmark
195	Vinther	Jeppe	University of Copenhagen	jvinther@bio.ku.dk	Denmark
196	Vitezic	Morana	University of Copenhagen	mvitezic@binf.ku.dk	Denmark
197	Vitting-Seerup	Kristoffer	University of Copenhagen	kristoffer.vittingseerup@bio.ku.dk	Denmark
198	Wenzel	Anne	University of Copenhagen	wenzel@rth.dk	Denmark
199	Werge	Thomas Mears	Region Hovedstadens Psykiatri	thomas.werge@regionh.dk	Denmark
200	Westergaard	David	Novo Nordisk	david.westergaard@cpr.ku.dk	Denmark
201	Wingaard Thrane	Sandra	Technical University of Denmark	swth@bio.dtu.dk	Denmark
202	Wiwie	Christian	University of Southern Denmark	wiwiec@sdu.dk	Denmark
203	Workman	Christopher	Technical University of Denmark	workman@cbs.dtu.dk	Denmark
204	Worning	Peder	Hvidovre Hospital	peder.worning@hvh.regionh.dk	Denmark
205	Xu	Ruiqi	BGI Europe	ruiqi.xu@genomics.cn	Denmark
206	Yadav	Rachita	Technical University of Denmark	rachita@cbs.dtu.dk	Denmark
207	Zachariasen	Martin	University of Southern Denmark	dekan-science@sdu.dk	Denmark
208	Zhang	Shiqi	University of Copenhagen	zhang@rth.dk	Denmark
209	Øster	Bodil	Qiagen Aarhus A/S	bodil.oester@qiagen.com	Denmark

Annual Danish Bioinformatics Conference

SCIENTIFIC COMMITTEE

- **Albin Sandelin**, UCPH, DK
- **Anders Krogh**, UCPH, DK
- **Jan Gorodkin**, UCPH, DK
- **Mikkel Heide Schierup**, AU, DK
- **Ole Nørregaard Jensen**, SDU, DK
- **Søren Brunak**, DTU and UCPH, DK

MANAGEMENT COMMITTEE

- **Henriette Husum Bak-Jensen,** UCPH, DK
- **Maria Maddalena Sperotto**, DTU, DK
- **Myhanh Nguyen**, DTU, DK
- **Peter Løngreen**, DTU, DK

Supported by:

UNIVERSITY OF COPENHAGEN

DTU Technical University of Denmark

UNIVERSITY OF SOUTHERN DENMARK

AARHUS UNIVERSITY