ArrayPitope: Automated Analysis of Amino Acid Substitutions for Peptide Microarray-Based Antibody Epitope Mapping

Identification of epitopes targeted by antibodies (B cell epitopes) is of critical importance for the development of many diagnostic and therapeutic tools. For clinical usage, such epitopes must be extensively characterized in order to validate specificity and to document potential cross-reactivity. B cell epitopes are typically classified as either linear epitopes, i.e. short consecutive segments from the protein sequence or conformational epitopes adapted through native protein folding. Recent advances in high-density peptide microarrays enable high-throughput, high-resolution identification and characterization of linear B cell epitopes. Using exhaustive amino acid substitution analysis of peptides originating from target antigens, these microarrays can be used to address the specificity of polyclonal antibodies raised against such antigens containing hundreds of epitopes. However, the interpretation of the data provided in such large-scale screenings is far from trivial and in most cases it requires advanced computational and statistical skills. Here, we present an online application for automated identification of linear B cell epitopes, allowing the non-expert user to analyse peptide microarray data. The application takes as input quantitative peptide data of fully or partially substituted overlapping peptides from a given antigen sequence and identifies epitope residues (residues that are significantly affected by substitutions) and visualize the selectivity towards each residue by sequence logo plots. Demonstrating utility, the application was used to identify and address the antibody specificity of 18 linear epitope regions in Human Serum Albumin (HSA), using peptide microarray data consisting of fully substituted peptides spanning the entire sequence of HSA and incubated with polyclonal rabbit anti-HSA (and mouse anti-rabbit-Cy3). The application is made available at: www.cbs.dtu.dk/services/ArrayPitope.
Chromosome copy number variation in telomerized human bone marrow stromal cells; insights for monitoring safe ex-vivo expansion of adult stem cells

Adult human bone marrow stromal cells (hBMSC) cultured for cell therapy require evaluation of potency and stability for safe use. Chromosomal aberrations upsetting genomic integrity in such cells have been contrastingly described as "Limited" or "Significant". Previously reported stepwise acquisition of a spontaneous neoplastic phenotype during three-year continuous culture of telomerized cells (hBMSC-TERT20) didn't alter a diploid karyotype measured by spectral karyotype analysis (SKY). Such screening may not adequately monitor abnormal and potentially tumorigenic hBMSC in clinical scenarios. We here used array comparative genomic hybridization (aCGH) to more stringently compare non-tumorigenic parental hBMSC-TERT strains with their tumorigenic subcloned populations. Confirmation of a known chromosome 9p21 microdeletion at locus CDKN2A/B, showed it also impinged upon the adjacent MTAP gene. Compared to reference diploid human fibroblast genomic DNA, the non-tumorigenic hBMSC-TERT4 cells had a copy number variation (CNV) in at least 14 independent loci. The pre-tumorigenic hBMSC-TERT20 cell strain had further CNV including 1q44 gain enhancing SMYD3 expression and 11q13.1 loss downregulating MUS81 expression. Bioinformatic analysis of gene products reflecting 11p15.5 CNV gain in tumorigenic hBMSC-TERT20 cells highlighted networks implicated in tumorigenic progression involving cell cycle control and mis-match repair. We provide novel biomarkers for prospective risk assessment of expanded stem cell cultures.

General information
State: Published
Organisations: Center for Biological sequence analysis, Department of Systems Biology, Center for Biological Sequence Analysis, DTU Multi Assay Core, Odense University Hospital
Authors: Burns, J. S. (Ekstern), Harkness, L. (Ekstern), Aldahmash, A. (Ekstern), Gautier, L. (Intern), Kassem, M. (Ekstern)
Pages: 6-17
Chromosome-wise Protein Interaction Patterns and Their Impact on Functional Implications of Large-Scale Genomic Aberrations

Gene copy-number changes influence phenotypes through gene-dosage alteration and subsequent changes of protein complex stoichiometry. Human trisomies where gene copy numbers are increased uniformly over entire chromosomes provide generic cases for studying these relationships. In most trisomies, gene and protein level alterations have fatal consequences. We used genome-wide protein-protein interaction data to identify chromosome-specific patterns of protein interactions. We found that some chromosomes encode proteins that interact infrequently with each other, chromosome 21 in particular. We combined the protein interaction data with transcriptome data from human brain tissue to investigate how this pattern of global interactions may affect cellular function. We identified highly connected proteins that also had coordinated gene expression. These proteins were associated with important neurological functions affecting the characteristic phenotypes for Down syndrome and have previously been validated in mouse knockout experiments. Our approach is general and applicable to other gene-dosage changes, such as arm-level amplifications in cancer.
MIToS.jl: mutual information tools for protein sequence analysis in the Julia language

Motivation: MIToS is an environment for mutual information analysis and a framework for protein multiple sequence alignments (MSAs) and protein structures (PDB) management in Julia language. It integrates sequence and structural information through SIFTS, making Pfam MSAs analysis straightforward. MIToS streamlines the implementation of any measure calculated from residue contingency tables and its optimization and testing in terms of protein contact prediction. As an example, we implemented and tested a BLOSUM62-based pseudo-count strategy in mutual information analysis.

Availability and Implementation: The software is totally implemented in Julia and supported for Linux, OS X and Windows. It's freely available on GitHub under MIT license: http://mitos.leloir.org.ar.

Contacts: diegozea@gmail.com or cmb@leloir.org.ar

Supplementary information:

Supplementary data are available at Bioinformatics online.

General information

State: Published
Organisations: Center for Biological sequence analysis, Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Fundación Instituto Leloir
Authors: Zea, D. J. (Ekstern), Anfossi, D. (Ekstern), Nielsen, M. (Intern), Marino-Buslje, C. (Ekstern)
Number of pages: 2
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information

Journal: Bioinformatics
Volume: 33
Issue number: 4
Article number: btw646
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 6.42
Web of Science (2016): Indexed yes
Neuroplasticity pathways and protein-interaction networks are modulated by vortioxetine in rodents

Background: The identification of biomarkers that predict susceptibility to major depressive disorder and treatment response to antidepressants is a major challenge. Vortioxetine is a novel multimodal antidepressant that possesses pro-cognitive properties and differentiates from other conventional antidepressants on various cognitive and plasticity measures. The aim of the present study was to identify biological systems rather than single biomarkers that may underlie vortioxetine's treatment effects. Results: We show that the biological systems regulated by vortioxetine are overlapping between mouse and rat in response to distinct treatment regimens and in different brain regions. Furthermore, analysis of complexes of physically-interacting proteins reveal that biomarkers involved in transcriptional regulation, neurodevelopment, neuroplasticity, and endocytosis are modulated by vortioxetine. A subsequent qPCR study examining the expression of targets in the protein-protein interactome space in response to chronic vortioxetine treatment over a range of doses provides further biological validation that vortioxetine engages neuroplasticity networks. Thus, the same biology is regulated in different species and sexes, different brain regions, and in response to distinct routes of administration and regimens. Conclusions: A recurring theme, based on the present study as well as previous findings, is that networks related to synaptic plasticity, synaptic transmission, signal transduction, and neurodevelopment are modulated in response to vortioxetine treatment. Regulation of these signaling pathways by vortioxetine may underlie vortioxetine's cognitive-enhancing properties.
Bioinformatics prediction of swine MHC class I epitopes from Porcine Reproductive and Respiratory Syndrome Virus

Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) causes one of the most important diseases in all swine producing countries. The infection has a high impact on animal welfare, food safety and production economics. PRRSV possesses multiple immunoevasive strategies, from suppression of the host cell antiviral machinery, to the deceptive induction of a non-neutralizing antibody response through decoy antigen presentation. This, combined with a very high mutation rate, has hampered the development of safe and effective vaccines. With the overall aim to design a vaccine that induces an effective CTL response against PRRSV, we have taken a bioinformatics approach to identify common PRRSV epitopes predicted to react broadly with predominant swine MHC (SLA) alleles. First, the genomic integrity and sequencing method was examined for 334 available complete PRRSV type 2 genomes leaving 104 strains of high quality. For each strain, a library of all possible 9- and 10-mer peptides was generated considering the known ribosomal frame shift sites and sites for post translational cleavage. All peptides were in silico analyzed for binding affinity to either of five common SLA class I alleles. A quantitative rank score was generated for each peptide by combining two algorithms based on the NetMHCpan neural network and lab determined SLA binding affinity of each amino acid at any position in the peptide, respectively. Peptides with a rank score above a predefined threshold were further analyzed by the PopCover algorithm, providing a final list of 54 epitopes prioritized according to maximum coverage of PRRSV strains and SLA alleles. This bioinformatics approach provides a rational strategy for selecting peptides for a CTL-activating vaccine with broad coverage of both virus and swine diversity. The immunogenicity of the selected peptides is in the process of being verified in vivo.
species (MGS). We use these to assemble 238 high-quality microbial genomes and identify affiliations between MGS and hundreds of viruses or genetic entities. Our method provides the means for comprehensive profiling of the diversity within complex metagenomic samples.

**General information**

**State:** Published

**Organisations:** Department of Systems Biology, Center for Biological Sequence Analysis, Novo Nordisk Foundation Center for Biosustainability, Center for Biological sequence analysis, Comparative Microbial Genomics, National Food Institute, Division of Epidemiology and Microbial Genomics, INRA Institut National de La Recherche Agronomique, South China University of Technology, European Molecular Biology Laboratory, Centre National de la Recherche Scientifique, University of Southern Denmark, University Hospital Vall d’Hebron, University of Copenhagen, Vrije Universiteit Brussel, Beijing Genomics Institute Hong Kong, Wageningen IMARES, Tokyo Institute of Technology


**Pages:** 822-828

**Publication date:** 2014

**Main Research Area:** Technical/natural sciences

**Publication information**

**Journal:** Nature Biotechnology

**Volume:** 32

**Issue number:** 8

**ISSN (Print):** 1087-0156

**Ratings:**

- BFI (2018): BFI-level 3
- Web of Science (2018): Indexed yes
- BFI (2017): BFI-level 2
- Web of Science (2017): Indexed Yes
- BFI (2016): BFI-level 2
- Scopus rating (2016): CiteScore 13.16 SJR 20.253 SNIP 6.303
- Web of Science (2016): Indexed yes
- BFI (2015): BFI-level 2
- Scopus rating (2015): SJR 17.892 SNIP 5.505 CiteScore 11.88
- Web of Science (2015): Indexed yes
- BFI (2014): BFI-level 2
- Scopus rating (2014): SJR 16.443 SNIP 5.433 CiteScore 11.4
- Web of Science (2014): Indexed yes
- BFI (2013): BFI-level 2
- Scopus rating (2013): SJR 13.849 SNIP 5.416 CiteScore 10.45
- ISI indexed (2013): ISI indexed yes
- Web of Science (2013): Indexed yes
- BFI (2012): BFI-level 2
- Scopus rating (2012): SJR 10.76 SNIP 4.96 CiteScore 8.44
- ISI indexed (2012): ISI indexed yes
- Web of Science (2012): Indexed yes
- BFI (2011): BFI-level 2
- Scopus rating (2011): SJR 11.627 SNIP 6.248 CiteScore 8.21
- ISI indexed (2011): ISI indexed yes
- Web of Science (2011): Indexed yes
- BFI (2010): BFI-level 2
LocTree3 prediction of localization

The prediction of protein sub-cellular localization is an important step toward elucidating protein function. For each query protein sequence, LocTree2 applies machine learning (profile kernel SVM) to predict the native sub-cellular localization in 18 classes for eukaryotes, in six for bacteria and in three for archaea. The method outputs a score that reflects the reliability of each prediction. LocTree2 has performed on par with or better than any other state-of-the-art method. Here, we report the availability of LocTree3 as a public web server. The server includes the machine learning-based LocTree2 and improves over it through the addition of homology-based inference. Assessed on sequence-unique data, LocTree3 reached an 18-state accuracy $Q_{18} = 80 \pm 3\%$ for eukaryotes and a six-state accuracy $Q_{6} = 89 \pm 4\%$ for bacteria. The server accepts submissions ranging from single protein sequences to entire proteomes. Response time of the unloaded server is about 90 s for a 300-residue eukaryotic protein and a few hours for an entire eukaryotic proteome not considering the generation of the alignments. For over 1000 entirely sequenced organisms, the predictions are directly available as downloads. The web server is available at http://www.rostlab.org/services/loctree3.

General information

State: Published
Organisations: Center for Biological sequence analysis, Department of Systems Biology, Center for Biological Sequence Analysis
Pages: W350-W355
Publication date: 2014
Main Research Area: Technical/natural sciences

Publication information
Journal: Nucleic Acids Research
Volume: 42
Issue number: W1
ISSN (Print): 0305-1048
Ratings:
The microbiome of New World vultures

Vultures are scavengers that fill a key ecosystem niche, in which they have evolved a remarkable tolerance to bacterial toxins in decaying meat. Here we report the first deep metagenomic analysis of the vulture microbiome. Through face and gut comparisons of 50 vultures representing two species, we demonstrate a remarkably conserved low diversity of gut microbial flora. The gut samples contained an average of 76 operational taxonomic units (OTUs) per specimen, compared with 528 OTUs on the facial skin. Clostridia and Fusobacteria, widely pathogenic to other vertebrates, dominate the vulture's gut microbiota. We reveal a likely faecal-oral-gut route for their origin. DNA of prey species detectable on facial swabs was completely degraded in the gut samples from most vultures, suggesting that the gastrointestinal tracts of vultures are extremely selective. Our findings show a strong adaption of vultures and their bacteria to their food source, exemplifying a specialized host-microbial alliance.
Whole Genome Epidemiological Typing of Salmonella

Salmonella is one of the most common foodborne pathogens worldwide. In the US alone, salmonellosis was estimated to cause 1.4 million cases effecting 17,000 hospitalization and almost 600 deaths each year. Particularly, Salmonella enterica is a common cause of minor and large food borne outbreaks. Technological advances and effective price in high throughput genome sequencing are making whole genome sequencing (WGS) available as a routine tool for bacterial typing.

Typing of Salmonella, especially sub-typing within the same serotype or even the same clone, the genetic variation of the target genes being used for typing is crucial for successful discrimination. The core genes or the genes that are conserved in all members of a genus or species are potentially good candidates for investigating genomic variation in phylogeny and epidemiology. A total of 2,882 core genes have been observed among 73 available Salmonella enterica genomes (accessed in April 2011). A consensus tree based on variation of the core genes gives better resolution than 16S rRNA and MLST that rarely provide separation between closely related strains. The performance of the pan-genome tree which is based on the presence/absence of all genes across genomes, is similar to the consensus tree but with higher branching confidence value. The core genes can be divided into two categories: a few highly variable genes and a larger set of conserved core genes, with low variance. These core genes are useful for investigating molecular evolution and remain useful as candidate genes for bacterial genome typing-even if they cannot be expected to differentiate highly clonal isolates e.g. outbreak cases of Salmonella [I].

To achieve successful ‘real-time’ monitoring and identification of outbreaks, rapid and reliable sub-typing is essential. A collection of thirty-four human S. Typhimurium strains from six different outbreaks together with background strains plus eight S. Enteritidis isolates from two outbreaks and five S. Derby isolates from a single outbreak were used to evaluate the strengths and drawbacks of different WGS approaches compared to the traditional typing, PFGE, for retrospectively outbreak typing of Salmonella. The resulting outcome showed that SNP analysis and nucleotide difference approach seem to be the superior methods for outbreak detection compared to other phylogenetic analytic approaches of WGS.

Furthermore, WGS approaches were also superior to the more classical typing method, PFGE. Meanwhile, k-mer method constructs a tree in high speed and giving high accuracy in clade level [II].

SNP analysis has successfully applied in recent epidemiological studies of Salmonella. Currently, there are different tools and methods to identify SNPs including various cut-off values. In addition, all the tools require bioinformatics skill. In order to apply WGS in routine typing, an automatic and user-friendly tool is needed. Therefore, snpTree has been developed as a server for online-automatic SNP analysis. snpTree can identify SNPs and construct phylogenetic tree from WGS raw reads as well as from assembled genomes or contigs. The tool is freely accessible at http://cge.cbs.dtu.dk/services/snpTree/ [III].

Globally, Salmonella enterica serovar Typhimurium is the most commonly isolated serovar. S. Typhimurium consists of a number of subtypes that conventionally have been divided by phagotyping. During the last three decades, S. Typhimurium phage type DT104 emerged as the most prevalent phage type and one of the best-studied because of its rapid global dissemination. Nonetheless, the origin and transmission route of this particular phage type have not been revealed. To bridge the gaps in epidemiology of DT104, WGS and temporally structured sequence analysis within Bayesian framework have been incorporated for reconstructing temporal and spatial phylogenies, estimating rate of mutation and divergence time of global and local S. Typhimurium DT104 isolates sampled from 1969 to 2012 from twenty-one countries in six continents. The DT104 was estimated to initially emerge as antimicrobial-susceptible strains in ~1946 (1931-1959) and further became multidrug-resistant (MDR) DT104 in ~1974 (1966-1981) through horizontal transfer of 13-kb SG1 MDR region into SG1-contained susceptible strains. Changes in population size over time supported global occurrences of MDR DT104. Besides, using WGS is capable to confirm local epidemiology especially the transmission between animal herds of DT104 isolates from Denmark. Interestingly, the demographic history of Danish MDR DT104 provided an evidence for the accomplishment of an eradicating program across pig herds in Denmark during 1996 to 2000 [IV].

Overall, this Ph.D. thesis has assessed the usefulness of WGS epidemiological typing in Salmonella as well as evaluated the different WGS approaches for outbreak investigation compared to the traditional typing, PFGE. An online tool to construct phylogenetic tree based on SNPs has also been developed. Furthermore, it has revealed the application of WGS in epidemiological study of global and local occurrences of S. Typhimurium DT104.
Hyperactivity of the Ero1α Oxidase Elicits Endoplasmic Reticulum Stress but No Broad Antioxidant Response

Oxidizing equivalents for the process of oxidative protein folding in the endoplasmic reticulum (ER) of mammalian cells are mainly provided by the Ero1α oxidase. The molecular mechanisms that regulate Ero1α activity in order to harness its oxidative power are quite well understood. However, the overall cellular response to oxidative stress generated by Ero1α in the lumen of the mammalian ER is poorly characterized. Here we investigate the effects of overexpressing a hyperactive mutant (C104A/C131A) of Ero1α. We show that Ero1α hyperactivity leads to hyperoxidation of the ER oxidoreductase ERp57 and induces expression of two established unfolded protein response (UPR) targets, Bip (immunoglobulin-binding protein) and HERP (homocysteine-induced ER protein). These effects could be reverted or aggravated by N-acetylcysteine and buthionine sulfoximine, respectively. Because both agents manipulate the cellular glutathione redox buffer, we conclude that the observed effects of Ero1α-C104A/C131A overexpression are likely caused by an oxidative perturbation of the ER glutathione redox buffer. In accordance, we show that Ero1α hyperactivity affects cell viability when cellular glutathione levels are compromised. Using microarray analysis, we demonstrate that the cell reacts to the oxidative challenge caused by Ero1α hyperactivity by turning on the UPR. Moreover, this analysis allowed the identification of two new targets of the mammalian UPR, CRELD1 and c18orf45. Interestingly, a broad antioxidant response was not induced. Our findings suggest that the hyperoxidation generated by Ero1α-C104A/C131A is addressed in the ER lumen and is unlikely to exert oxidative injury throughout the cell.
Metatranscriptomics of the human gut microbiome

Our ‘other’ genome is the collective genetic information in all of the microorganisms that are living on and within us. Collectively known as the microbiome, these microbial cells outnumber human cells in the body by more than 10 to 1, and the genes carried by these organisms outnumber the genes in the human genome by more than 100 to 1.

How these organisms contribute to and affect human health is poorly understood, but the emerging field of metagenomics promises a more comprehensive and complete understanding of the human microbiome. In the European-funded Metagenomics of the Human Intestinal Tract (MetaHIT) project [1], we combined next-generation sequencing with high-density microarrays, generating metagenomic and metatranscriptomic data for more than 400 individuals.

The combined data reveal clusters of coexisting species with differences in pathway and gene function activity, suggesting that there is a division of labor between the bacterial species in the human gut microbiome.
Bifidobacterium bifidum Actively Changes the Gene Expression Profile Induced by Lactobacillus acidophilus in Murine Dendritic Cells

Dendritic cells (DC) play a pivotal regulatory role in activation of both the innate as well as the adaptive immune system by responding to environmental microorganisms. We have previously shown that Lactobacillus acidophilus induces a strong production of the pro-inflammatory and Th1 polarizing cytokine IL-12 in DC, whereas bifidobacteria do not induce IL-12 but inhibit the IL-12 production induced by lactobacilli. In the present study, genome-wide microarrays were used to investigate the gene expression pattern of murine DC stimulated with Lactobacillus acidophilus NCFM and Bifidobacterium bifidum Z9. L. acidophilus NCFM strongly induced expression of interferon (IFN)-beta, other virus defence genes, and cytokine and chemokine genes related to the innate and the adaptive immune response. By contrast, B. bifidum Z9 up-regulated genes encoding cytokines and chemokines related to the innate immune response. Moreover, B. bifidum Z9 inhibited the expression of the Th1-promoting genes induced by L. acidophilus NCFM and had an additive effect on genes of the innate immune response and Th2 skewing genes. The gene encoding Jun dimerization protein 2 (JDP2), a transcription factor regulating the activation of JNK, was one of the few genes only induced by B. bifidum Z9. Neutralization of IFN-beta abrogated L. acidophilus NCFM-induced expression of Th1-skewing genes, and blocking of the JNK pathway completely inhibited the expression of IFN-beta. Our results indicate that B. bifidum Z9 actively inhibits the expression of genes related to the adaptive immune system in murine dendritic cells and that JPD2 via blocking of IFN-beta plays a central role in this regulatory mechanism.
Lactobacillus acidophilus induces virus immune defence genes in murine dendritic cells by a Toll-like receptor-2-dependent mechanism

Lactobacilli are probiotics that, among other health-promoting effects, have been ascribed immunostimulating and virus-preventive properties. Certain Lactobacillus spp. have been shown to possess strong interleukin-12 (IL-12) -inducing properties. As IL-12 production depends on the up-regulation of type I interferons (IFNs), we hypothesized that the strong IL-12-inducing capacity of Lactobacillus acidophilus NCFM in murine bone-marrow-derived dendritic cells (DCs) is caused by an up-regulation of IFN-beta, which subsequently induces IL-12 and the double-stranded RNA binding Toll-like receptor-3 (TLR-3). The expression of the genes encoding IFN-beta, TLR-3, IL-12 and IL-10 in DCs upon stimulation with L. acidophilus NCFM was determined. Lactobacillus acidophilus NCFM induced a much stronger expression of Ifn-beta, Il-12 and Il-10 compared with the synthetic double-stranded RNA ligand Poly I:C, whereas the levels of expressed Tlr-3 were similar. Whole genome microarray gene expression analysis revealed that other genes related to viral defence were significantly up-regulated and among the strongest induced genes in DCs stimulated with L. acidophilus NCFM. The ability to induce IFN-beta was also detected in another L. acidophilus strain (X37), but was not a property of other probiotic strains tested, i.e. Bifidobacterium bifidum Z9 and Escherichia coli Nissle 1917. The IFN-beta expression was markedly reduced in TLR-2(-/-) DCs, dependent on endocytosis, and the major cause of the induction of Il-12 and Tlr-3 in DCs stimulated with L. acidophilus NCFM. Collectively, our results reveal that certain lactobacilli trigger the expression of viral defence genes in DCs in a TLR-2 manner dependent on IFN-beta.
InterMap3D: predicting and visualizing co-evolving protein residues

InterMap3D predicts co-evolving protein residues and plots them on the 3D protein structure. Starting with a single protein sequence, InterMap3D automatically finds a set of homologous sequences, generates an alignment and fetches the most similar 3D structure from the Protein Data Bank (PDB). It can also accept a user-generated alignment. Based on the alignment, co-evolving residues are then predicted using three different methods: Row and Column Weighing of Mutual Information, Mutual Information/Entropy and Dependency. Finally, InterMap3D generates high-quality images of the protein with the predicted co-evolving residues highlighted.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological sequence analysis, Center for Biological Sequence Analysis
Pages: 1963-1965
Publication date: 2009
Main Research Area: Technical/natural sciences

Publication information
Journal: Bioinformatics
Volume: 25
Issue number: 15
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 6.42
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 6.06
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 5.5
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 5.78
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): CiteScore 6.73
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): CiteScore 5.61
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 2
Web of Science (2008): Indexed yes
Web of Science (2007): Indexed yes
Web of Science (2006): Indexed yes
Web of Science (2005): Indexed yes
Prediction of Glycosylation Sites in Proteins

General information
State: Published
Organisations: Center for Biological sequence analysis, Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Julenius, K. (Intern), Johansen, M. B. (Intern), Zhang, Y. (Intern), Brunak, S. (Intern), Gupta, R. (Intern)
Pages: 163-192
Publication date: 2009

Host publication information
Title of host publication: Bioinformatics for Glycobiology and Glycomics: An Introduction
Publisher: Wiley
Editors: von der Lieth, C., Lütteke, T., Frank, M.
ISBN (Print): 9780470016671
ISBN (Electronic): 9780470029619
Main Research Area: Technical/natural sciences
DOIs: 10.1002/9780470029619.ch9
Source: orbit
Source-ID: 265742
Publication: Research - peer-review › Book chapter – Annual report year: 2009

CycleBase.org - a comprehensive multi-organism online database of cell-cycle experiments
The past decade has seen the publication of a large number of cell-cycle microarray studies and many more are in the pipeline. However, data from these experiments are not easy to access, combine and evaluate. We have developed a centralized database with an easy-to-use interface, Cyclebase.org, for viewing and downloading these data. The user interface facilitates searches for genes of interest as well as downloads of genome-wide results. Individual genes are displayed with graphs of expression profiles throughout the cell cycle from all available experiments. These expression profiles are normalized to a common timescale to enable inspection of the combined experimental evidence. Furthermore, state-of-the-art computational analyses provide key information on both individual experiments and combined datasets such as whether or not a gene is periodically expressed and, if so, the time of peak expression. Cyclebase is available at http://www.cyclebase.org.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, Center for Biological sequence analysis
Pages: D854-D859
Publication date: 2007
Main Research Area: Technical/natural sciences

Publication information
Journal: Nucleic Acids Res
Volume: 36
ISSN (Print): 0305-1048
Ratings:
BFI (2018): BFI-level 2
Integrative analysis for finding genes and networks involved in diabetes and other complex diseases
We have developed an integrative analysis method combining genetic interactions, identified using type 1 diabetes genome scan data, and a high-confidence human protein interaction network. Resulting networks were ranked by the significance of the enrichment of proteins from interacting regions. We identified a number of new protein network modules and novel candidate genes/proteins for type 1 diabetes. We propose this type of integrative analysis as a general method for the elucidation of genes and networks involved in diabetes and other complex diseases.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Center for Biological sequence analysis, Department of Systems Biology
Authors: Bergholdt, R. (Ekstern), Størling, Z. M. (Intern), Hansen, K. L. (Intern), Karlberg, E. O. L. (Intern), Ólason, P. Í. (Intern), Aalund, M. (Ekstern), Nerup, J. (Ekstern), Brunak, S. (Intern), Workman, C. (Intern), Pociot, F. (Ekstern)
Pages: R253
Publication date: 2007
Main Research Area: Technical/natural sciences

Publication information
Journal: Genome Biology
Volume: 8
Issue number: 11
ISSN (Print): 1465-6906
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 11.12 SJR 10.484 SNIP 2.846
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 8.65 SNIP 2.608 CiteScore 9.08
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 8.188 SNIP 2.224 CiteScore 8.34
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 8.557 SNIP 2.418 CiteScore 8.55
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 7.9 SNIP 2.372 CiteScore 8.14
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 5.989 SNIP 1.973 CiteScore 6.69
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 5.012 SNIP 1.819
BFI (2009): BFI-level 1
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 3.713 SNIP 1.154
Scopus rating (2007): SJR 3.218 SNIP 0.721
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 2.523 SNIP 0.331
An environmental signature for 323 microbial genomes based on codon adaptation indices

Background: Codon adaptation indices (CAIs) represent an evolutionary strategy to modulate gene expression and have widely been used to predict potentially highly expressed genes within microbial genomes. Here, we evaluate and compare two very different methods for estimating CAI values, one corresponding to translational codon usage bias and the second obtained mathematically by searching for the most dominant codon bias. Results: The level of correlation between these two CAI methods is a simple and intuitive measure of the degree of translational bias in an organism, and from this we confirm that fast replicating bacteria are more likely to have a dominant translational codon usage bias than are slow replicating bacteria, and that this translational codon usage bias may be used for prediction of highly expressed genes. By analyzing more than 300 bacterial genomes, as well as five fungal genomes, we show that codon usage preference provides an environmental signature by which it is possible to group bacteria according to their lifestyle, for instance soil bacteria and soil symbionts, spore formers, enteric bacteria, aquatic bacteria, and intercellular and extracellular pathogens. Conclusion: The results and the approach described here may be used to acquire new knowledge regarding species lifestyle and to elucidate relationships between organisms that are far apart evolutionarily.
Protein evolution is faster outside the cell

General information
State: Published
Organisations: Center for Biological sequence analysis, Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Julenius, K. (Intern), Pedersen, A. G. (Intern)
Pages: 2039-2048
Publication date: 2006
Main Research Area: Technical/natural sciences

Publication information
Journal: Molecular Biology and Evolution
Volume: 23
ISSN (Print): 0737-4038
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): SJR 8.724 SNIP 7.289 CiteScore 13.93
Transcriptome analysis of FSH and FSH variant stimulation in granulosa cells from IVM patients reveals novel regulated genes

General information
State: Published
Organisations: Center for Biological sequence analysis
Authors: Perlman, S. (Ekstern), Bouquin, T. (Ekstern), van den Hazel, B. (Ekstern), Jensen, T. (Ekstern), Schambye, H. (Ekstern), Knudsen, S. (Intern), Okkels, J. (Ekstern)
Pages: 135-144
Publication date: 2006
Main Research Area: Technical/natural sciences

Publication information
Journal: MOLECULAR HUMAN REPRODUCTION
Volume: 12
Issue number: 3
ISSN (Print): 1360-9947
Ratings:
Original language: English
Source: orbit
Source-ID: 189980
Publication: Research - peer-review › Journal article – Annual report year: 2006

**Genome update: 2D clustering of bacterial genomes**

**General information**

State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, Center for Biological sequence analysis
Authors: Willenbrock, H. (Intern), Binnewies, T. (Intern), Hallin, P. F. (Intern), Ussery, D. (Intern)
Pages: 333-336
Publication date: 2005
Main Research Area: Technical/natural sciences

**Publication Information**
Journal: MICROBIOLOGY-SGM
Genome update: distribution of two-component transduction systems in 250 bacterial genomes

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological sequence analysis
Pages: 3447-3452
Publication date: 2005
Main Research Area: Technical/natural sciences

Publication information
Journal: MICROBIOLOGY-SGM
Volume: 151
ISSN (Print): 1350-0872
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 1.56 SJR 0.805 SNIP 0.648
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.136 SNIP 0.834 CiteScore 2.05
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.448 SNIP 0.978 CiteScore 2.69
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 1.652 SNIP 1.031 CiteScore 3.34
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 1.596 SNIP 0.974 CiteScore 3.12
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 1.636 SNIP 1.036 CiteScore 3.18
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 1.774 SNIP 0.988
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 1.69 SNIP 0.994
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 1.709 SNIP 1.009
Genome Update: 2D clustering of bacterial genomes

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, Center for Biological sequence analysis
Authors: Willenbrock, H. (Intern), Binnewies, T. T. (Intern), Hallin, P. F. (Intern)
Pages: 33-336
Publication date: 2005
Main Research Area: Technical/natural sciences

Publication information
Journal: Microbiology-sgm
Volume: 151
ISSN (Print): 1350-0872
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 1.56 SJR 0.805 SNIP 0.648
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.136 SNIP 0.834 CiteScore 2.05
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.448 SNIP 0.978 CiteScore 2.69
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 1.652 SNIP 1.031 CiteScore 3.34
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
Implementation of a gene expression index calculation method based on the PDNN model

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, Center for Biological sequence analysis
Authors: Nielsen, H. B. (Intern), Gautier, L. (Intern), Knudsen, S. (Intern)
Pages: 687-688
Publication date: 2005
Main Research Area: Technical/natural sciences

Publication information
Journal: BIOINFORMATICS
Volume: 21
Issue number: 5
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
Integrating protein feature based predictions into Systems Biology

General information
State: Published
Organisations: Center for Biological sequence analysis
Authors: Fausbøll, A. (Intern)
Publication date: 2005

Publication information
Place of publication: Kgs. Lyngby, Denmark
Publisher: Technical University of Denmark (DTU)
New weakly expressed cell cycle-regulated genes in yeast

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, Center for Biological sequence analysis
Pages: 1191-1201
Publication date: 2005

Publication information
Journal: YEAST
Volume: 22
Issue number: 15
ISSN (Print): 0749-503X
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): SJR 0.816 SNIP 0.811 CiteScore 1.87
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 0.962 SNIP 0.745 CiteScore 2.01
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 0.875 SNIP 0.792 CiteScore 1.67
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.232 SNIP 0.72 CiteScore 2.09
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.197 SNIP 0.762 CiteScore 2.05
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.063 SNIP 0.701 CiteScore 1.77
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.049 SNIP 0.835
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 1.584 SNIP 0.81
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 1.8 SNIP 0.87
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 1.635 SNIP 0.945
Non-classical protein secretion in bacteria

**General information**
State: Published
Organisations: Department of Systems Biology, Center for Biological sequence analysis, Center for Biological Sequence Analysis
Authors: Bendtsen, J. D. (Intern), Kiemer, L. (Intern), Fausbøll, A. (Intern), Brunak, S. (Intern)
Pages: 58
Publication date: 2005
Main Research Area: Technical/natural sciences

**Publication information**
Journal: BMC Microbiology
Volume: 5
ISSN (Print): 1471-2180
Ratings:
- BFI (2018): BFI-level 1
- Web of Science (2018): Indexed yes
- BFI (2017): BFI-level 1
- Web of Science (2017): Indexed yes
- BFI (2016): BFI-level 1
- Scopus rating (2016): CiteScore 2.82 SJR 1.23 SNIP 0.992
- Web of Science (2016): Indexed yes
- BFI (2015): BFI-level 1
- Scopus rating (2015): SJR 1.401 SNIP 0.998 CiteScore 2.93
- Web of Science (2015): Indexed yes
- BFI (2014): BFI-level 1
- Scopus rating (2014): SJR 1.472 SNIP 1.039 CiteScore 2.95
- Web of Science (2014): Indexed yes
- BFI (2013): BFI-level 1
- Scopus rating (2013): SJR 1.527 SNIP 1.143 CiteScore 3.32
- ISI indexed (2013): ISI indexed yes
- Web of Science (2013): Indexed yes
- BFI (2012): BFI-level 1
- Scopus rating (2012): SJR 1.454 SNIP 1.12 CiteScore 3.38
- ISI indexed (2012): ISI indexed yes
Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites

O-GalNAc-glycosylation is one of the main types of glycosylation in mammalian cells. No consensus recognition sequence for the O-glycosyltransferases is known, making prediction methods necessary to bridge the gap between the large number of known protein sequences and the small number of proteins experimentally investigated with regard to glycosylation status. From O-GLYCBASE a total of 86 mammalian proteins experimentally investigated for in vivo O-GalNAc sites were extracted. Mammalian protein homolog comparisons showed that a glycosylated serine or threonine is less likely to be precisely conserved than a nonglycosylated one. The Protein Data Bank was analyzed for structural information, and 12 glycosylated structures were obtained. All positive sites were found in coil or turn regions. A method for predicting the location for mucin-type glycosylation sites was trained using a neural network approach. The best overall network used as input amino acid composition, averaged surface accessibility predictions together with substitution matrix profile encoding of the sequence. To improve prediction on isolated (single) sites, networks were trained on isolated sites only. The final method combines predictions from the best overall network and the best isolated site network; this prediction method correctly predicted 76% of the glycosylated residues and 93% of the nonglycosylated residues. NetOGlyc 3.1 can predict sites for completely new proteins without losing its performance. The fact that the sites could be predicted from averaged properties together with the fact that glycosylation sites are not precisely conserved indicates that mucin-type glycosylation in most cases is a bulk property and not a very site-specific one. NetOGlyc 3.1 is made available at www.cbs.dtu.dk/services/netoglyc.
Prediction of twin-arginine signal peptides

Background: Proteins carrying twin-arginine (Tat) signal peptides are exported into the periplasmic compartment or extracellular environment independently of the classical Sec-dependent translocation pathway. To complement other methods for classical signal peptide prediction we here present a publicly available method, TatP, for prediction of bacterial Tat signal peptides. Results: We have retrieved sequence data for Tat substrates in order to train a computational method for discrimination of Sec and Tat signal peptides. The TatP method is able to positively classify
91% of 35 known Tat signal peptides and 84% of the annotated cleavage sites of these Tat signal peptides were correctly predicted. This method generates far less false positive predictions on various datasets than using simple pattern matching. Moreover, on the same datasets TatP generates less false positive predictions than a complementary rule based prediction method. Conclusion: The method developed here is able to discriminate Tat signal peptides from cytoplasmic proteins carrying a similar motif, as well as from Sec signal peptides, with high accuracy. The method allows filtering of input sequences based on Perl syntax regular expressions, whereas hydrophobicity discrimination of Tat- and Sec- signal peptides is carried out by an artificial neural network. A potential cleavage site of the predicted Tat signal peptide is also reported. The TatP prediction server is available as a public web server at http://www.cbs.dtu.dk/services/TatP/.

**General information**
State: Published
Organisations: Center for Biological sequence analysis, Center for Biological Sequence Analysis, Department of Systems Biology
Authors: Bendtsen, J. D. (Intern), Nielsen, H. (Intern), Widdick, D. (Ekstern), Palmer, T. (Ekstern), Brunak, S. (Intern)
Pages: 167
Publication date: 2005
Main Research Area: Technical/natural sciences

**Publication information**
Journal: BMC Bioinformatics
Volume: 6
ISSN (Print): 1471-2105
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.54 SJR 1.467 SNIP 0.946
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.656 SNIP 1.077 CiteScore 2.77
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.836 SNIP 1.202 CiteScore 2.91
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.932 SNIP 1.335 CiteScore 3.38
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.857 SNIP 1.155 CiteScore 3.24
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.655 SNIP 1.215 CiteScore 3.34
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.756 SNIP 1.15
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 1.89 SNIP 1.32
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 1.945 SNIP 1.146
Web of Science (2008): Indexed yes
Presence of a functional but dispensable Nuclear Export Signal in the HTLV-2 Tax protein

General information
State: Published
Organisations: Center for Biological sequence analysis, Institut Pasteur
Authors: Chevalier, S. (Ekstern), Meertens, L. (Ekstern), Calattini, S. (Ekstern), Gessain, A. (Ekstern), Kiemer, L. (Intern), Mahieux, R. (Ekstern)
Pages: 70
Publication date: 2005
Main Research Area: Technical/natural sciences

Publication information
Journal: Retrovirology
Volume: 2
ISSN (Print): 1742-4690
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): SJR 2.417 SNIP 0.875 CiteScore 3.67
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 2.527 SNIP 0.875 CiteScore 3.67
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 2.412 SNIP 0.878 CiteScore 3.49
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 2.111 SNIP 0.944 CiteScore 4.05
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 2.313 SNIP 1.211 CiteScore 1.9
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 2.804 SNIP 1.228 CiteScore 2.03
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 2.049 SNIP 1.087
Systems biology: in the broadest sense of the word

General information
State: Published
Organisations: Center for Biological sequence analysis, Center for Biological Sequence Analysis, Department of Systems Biology, European Molecular Biology Laboratory
Authors: de Lichtenberg, U. N. (Intern), Ussery, D. (Intern), Jensen, L. (Ekstern)
Pages: 482-483
Publication date: 2005
Main Research Area: Technical/natural sciences

Publication information
Journal: Environmental Microbiology
ISSN (Print): 1462-2912
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 5.02 SJR 2.221 SNIP 1.406
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 2.999 SNIP 1.584 CiteScore 5.61
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 2.85 SNIP 1.616 CiteScore 5.6
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 3.246 SNIP 1.843 CiteScore 6.37
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 3.128 SNIP 1.646 CiteScore 5.94
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 3.33 SNIP 1.708 CiteScore 6.1
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
Alternative mapping of probes to genes for Affymetrix chips

Background: Short oligonucleotide arrays have several probes measuring the expression level of each target transcript. Therefore the selection of probes is a key component for the quality of measurements. However, once probes have been selected and synthesized on an array, it is still possible to re-evaluate the results using an updated mapping of probes to genes, taking into account the latest biological knowledge available. Methods: We investigated how probes found on recent commercial microarrays for human genes (Affymetrix HG-U133A) were matching a recent curated collection of human transcripts: the NCBI RefSeq database. We also built mappings and used them in place of the original probe to genes associations provided by the manufacturer of the arrays. Results: In a large number of cases, 36%, the probes matching a reference sequence were consistent with the grouping of probes by the manufacturer of the chips. For the remaining cases there were discrepancies and we show how that can affect the analysis of data. Conclusions: While the probes on Affymetrix arrays remain the same for several years, the biological knowledge concerning the genomic sequences evolves rapidly. Using up-to-date knowledge can apparently change the outcome of an analysis.
Background: Despite the passing of more than a year since the first outbreak of Severe Acute Respiratory Syndrome (SARS), efficient counter-measures are still few and many believe that reappearance of SARS, or a similar disease caused by a coronavirus, is not unlikely. For other virus families like the picornaviruses it is known that pathology is related to proteolytic cleavage of host proteins by viral proteinases. Furthermore, several studies indicate that virus proliferation can be arrested using specific proteinase inhibitors supporting the belief that proteinases are indeed important during
infection. Prompted by this, we set out to analyse and predict cleavage by the coronavirus main proteinase using computational methods. Results: We retrieved sequence data on seven fully sequenced coronaviruses and identified the main 3C proteinase cleavage sites in polyproteins using alignments. A neural network was trained to recognise the cleavage sites in the genomes obtaining a sensitivity of 87.0% and a specificity of 99.0%. Several proteins known to be cleaved by other viruses were submitted to prediction as well as proteins suspected relevant in coronavirus pathology. Cleavage sites were predicted in proteins such as the cystic fibrosis transmembrane conductance regulator (CFTR), transcription factors CREB-RP and OCT-I, and components of the ubiquitin pathway. Conclusions: Our prediction method NetCorona predicts coronavirus cleavage sites with high specificity and several potential cleavage candidates were identified which might be important to elucidate coronavirus pathology. Furthermore, the method might assist in design of proteinase inhibitors for treatment of SARS and possible future diseases caused by coronaviruses.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Center for Biological sequence analysis, Department of Systems Biology
Authors: Kiemer, L. (Intern), Lund, O. (Intern), Brunak, S. (Intern), Blom, N. (Intern)
Pages: 72
Publication date: 2004
Main Research Area: Technical/natural sciences

Publication information
Journal: BMC Bioinformatics
Volume: 5
ISSN (Print): 1471-2105
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.54 SJR 1.467 SNIP 0.946
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.656 SNIP 1.077 CiteScore 2.77
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.836 SNIP 1.202 CiteScore 2.91
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.932 SNIP 1.335 CiteScore 3.38
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.857 SNIP 1.155 CiteScore 3.24
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.655 SNIP 1.215 CiteScore 3.34
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.756 SNIP 1.15
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 1.89 SNIP 1.32
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 1
Scopus rating (2008): SJR 1.945 SNIP 1.146
Analysis of two large functionally uncharacterized regions in the Methanopyrus kandleri AV19 genome

Background: For most sequenced prokaryotic genomes, about a third of the protein coding genes annotated are "orphan proteins", that is, they lack homology to known proteins. These hypothetical genes are typically short and randomly scattered throughout the genome. This trend is seen for most of the bacterial and archaeal genomes published to date.

Results: In contrast we have found that a large fraction of the genes coding for such orphan proteins in the Methanopyrus kandleri AV19 genome occur within two large regions. These genes have no known homologs except from other M. kandleri genes. However, analysis of their lengths, codon usage, and Ribosomal Binding Site (RBS) sequences shows that they are most likely true protein coding genes and not random open reading frames.

Conclusions: Although these regions can be considered as candidates for massive lateral gene transfer, our bioinformatics analysis suggests that this is not the case. We predict many of the organism specific proteins to be transmembrane and belong to protein families that are non-randomly distributed between the regions. Consistent with this, we suggest that the two regions are most likely unrelated, and that they may be integrated plasmids.
A new non-linear normalization method for reducing variability in DNA microarray experiments

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Department of Systems Biology, Bacterial Physiology and Genetics Group, Department of Microbiology, Center for Biological sequence analysis, Statens Serum Institut
Publication date: 2002
Main Research Area: Technical/natural sciences
Identification of Wheat Varieties Using Matrix-assisted Laser Desorption/Ionisation Time-of-flight Mass Spectrometry and an Artificial Neural network

A novel tool for variety identification of wheat (Triticum aestivum L.) has been developed: an artificial neural network (ANN) is used to classify the gliadin fraction analysed by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOFMS). The robustness of this novel method with respect to various experimental parameters has been tested. The results can be summarised: (i) With this approach 97% of the wheat varieties can be classified correctly with a corresponding correlation coefficient of 1.0, (ii) The method is fast since the time of extracting gliadins from flour can be reduced to 20 min without significant decrease in overall performance, (iii) The storage of flour or extracts under standard conditions does not influence the classification ability (i.e. the generalisation ability) of the method, and (iv) The classification obtained is not influenced by the identity of the operator making the analysis. This study demonstrates that a combination of an ANN and MALDI-TOFMS analysis of the gliadin fraction provides a fast and reliable tool for the variety identification of wheat. Copyright (C) 1999 John Whey & Sons, Ltd.
Prediction of signal peptides and signal anchors by a hidden Markov model.

A hidden Markov model of signal peptides has been developed. It contains submodels for the N-terminal part, the hydrophobic region, and the region around the cleavage site. For known signal peptides, the model can be used to assign objective boundaries between these three regions. Applied to our data, the length distributions for the three regions are significantly different from expectations. For instance, the assigned hydrophobic region is between 8 and 12 residues long in almost all eukaryotic signal peptides. This analysis also makes obvious the difference between eukaryotes, Gram-positive bacteria, and Gram-negative bacteria. The model can be used to predict the location of the cleavage site, which it finds correctly in nearly 70% of signal peptides in a cross-validated test—almost the same accuracy as the best previous method. One of the problems for existing prediction methods is the poor discrimination between signal peptides and uncleaved signal anchors, but this is substantially improved by the hidden Markov model when expanding it with a very simple signal anchor model.
A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites

We have developed a new method for the identification of signal peptides and their cleavage sites based on neural networks trained on separate sets of prokaryotic and eukaryotic sequences. The method performs significantly better than previous prediction schemes, and can easily be applied to genome-wide data sets. Discrimination between cleaved signal peptides and uncleaved N-terminal signal-anchor sequences is also possible, though with lower precision.
Neural Network Prediction of Translation Initiation Sites in Eukaryotes: Perspectives for EST and Genome analysis

Translation in eukaryotes does not always start at the first AUG in an mRNA, implying that context information also plays a role. This makes prediction of translation initiation sites a non-trivial task, especially when analysing EST and genome data where the entire mature mRNA sequence is not known. In this paper, we employ artificial neural networks to predict which AUG triplet in an mRNA sequence is the start codon. The trained networks correctly classified 88% of Arabidopsis and 85% of vertebrate AUG triplets. We find that our trained neural networks use a combination of local start codon context and global sequence information. Furthermore, analysis of false predictions shows that AUGs in frame with the actual start codon are more frequently selected than out-of-frame AUGs, suggesting that our networks use reading frame detection. A number of conflicts between neural network predictions and database annotations are analysed in detail, leading to identification of possible database errors.
The gntP Gene of Escherichia coli Involved in Gluconate Uptake. 

GntP, a high-affinity gluconate permease.

The gntP gene, located between the fim and uxu loci in Escherichia coli K-12, has been cloned and characterized. Nucleotide sequencing of a region encompassing the gntP gene revealed an open reading frame of 447 codons with significant homology to the Bacillus subtilis gluconate permease. Northern (RNA) blotting indicated that the gntP gene was monocistronic and was transcribed as an mRNA with an apparent molecular size of 1.54 kb. The transcriptional start point was determined by primer extension analysis. The gntP gene was found to be under catabolite repression and was not induced by gluconate. Also, expression seemed to be stringently controlled. Several observations indicated that the GntP protein is an inner membrane protein; it contains characteristic membrane-spanning regions and was isolated redominantly from the inner-membrane fraction of fractionated host cells. A topology analysis predicted a protein with 14 membrane-spanning segments. The inability of a mutant strain to grow on gluconate minimal medium could be relieved by introduction of a plasmid encoding the gntP gene. Finally, the kinetics of GntP-mediated gluconate uptake were investigated, indicating an apparent Km for gluconate of 25 mM.