Direct acting antiviral treatment of chronic hepatitis C in Denmark: factors associated with and barriers to treatment initiation

Objectives: We describe factors associated with and barriers to initiation of Direct Acting Antiviral (DAA) treatment in patients with chronic hepatitis C, who fulfill national fibrosis treatment guidelines in Denmark. Materials and Methods: In this nationwide cohort study, we included patients with chronic hepatitis C from The Danish Database for Hepatitis B and C (DANHEP) who fulfilled fibrosis treatment criteria. Factors associated with treatment initiation and treatment failure were determined by logistic regression analyses. Medical records were reviewed from patients who fulfilled fibrosis treatment criteria, but did not initiate DAA treatment to determine the cause. Results: In 344 (49%) of 700 patients, who fulfilled treatment criteria, factors associated with DAA treatment initiation were transmission by other routes than injecting drug use odds ratio (OR) 2.13 (CI: 1.38–3.28), previous treatment failure OR 2.58 (CI: 1.84–3.61) and ALT above upper limit of normal OR 1.60 (CI: 1.18–2.17). The most frequent reasons for not starting treatment among 356 (51%) patients were non-adherence to medical appointments (n=107/30%) and ongoing substance use (n=61/17%). Treatment failure with viral relapse occurred in 19 (5.5%) patients, who were more likely to have failed previous treatment OR 4.53 (CI: 1.59–12.91). Conclusions: In this nationwide cohort study, we found non-adherence to medical appointments and active substance use to be major obstacles for DAA treatment initiation. Our findings highlight the need for interventions that can overcome these barriers and increase the number of patients who can initiate and benefit from curative DAA treatment.
Accurate genotyping across variant classes and lengths using variant graphs

Genotype estimates from short-read sequencing data are typically based on the alignment of reads to a linear reference, but reads originating from more complex variants (for example, structural variants) often align poorly, resulting in biased genotype estimates. This bias can be mitigated by first collecting a set of candidate variants across discovery methods, individuals and databases, and then realigning the reads to the variants and reference simultaneously. However, this realignment problem has proved computationally difficult. Here, we present a new method (BayesTyper) that uses exact
alignment of read k-mers to a graph representation of the reference and variants to efficiently perform unbiased, probabilistic genotyping across the variation spectrum. We demonstrate that BayesTyper generally provides superior variant sensitivity and genotyping accuracy relative to existing methods when used to integrate variants across discovery approaches and individuals. Finally, we demonstrate that including a ‘variation-prior’ database containing already known variants significantly improves sensitivity.

**General information**

State: Published

Organisations: Department of Bio and Health Informatics, Metagenomics, Integrative Systems Biology, Genomic Epidemiology, Disease Intelligence and Molecular Evolution, University of Copenhagen, South China University of Technology, BGI-Europe, BGI-Shenzhen, Technical University of Denmark, University of Oslo, University of Bergen, University of North Carolina, Karolinska Institutet, Aarhus Universitet, Aarhus University


Number of pages: 11

Publication date: 2018

Main Research Area: Technical/natural sciences

**Publication information**

Journal: Nature Genetics

ISSN (Print): 1061-4036

Ratings:

- BFI (2018): BFI-level 3
- Web of Science (2018): Indexed yes
- BFI (2017): BFI-level 2
- Scopus rating (2017): CiteScore 21.12 SJR 22.243 SNIP 5.867
- Web of Science (2017): Indexed yes
- BFI (2016): BFI-level 2
- Scopus rating (2016): CiteScore 20.83 SJR 21.979 SNIP 6.709
- Web of Science (2016): Indexed yes
- BFI (2015): BFI-level 2
- Web of Science (2015): Indexed yes
- BFI (2014): BFI-level 2
- Scopus rating (2014): SJR 23.98 SNIP 6.332 CiteScore 22.76
- Web of Science (2014): Indexed yes
- BFI (2013): BFI-level 2
- Scopus rating (2013): SJR 24.193 SNIP 6.287 CiteScore 24.17
- ISI indexed (2013): ISI indexed yes
- Web of Science (2013): Indexed yes
- BFI (2012): BFI-level 2
- Scopus rating (2012): SJR 25.621 SNIP 7.137 CiteScore 27.17
- ISI indexed (2012): ISI indexed yes
- Web of Science (2012): Indexed yes
- BFI (2011): BFI-level 2
- Scopus rating (2011): SJR 25.298 SNIP 7.206 CiteScore 25.75
- ISI indexed (2011): ISI indexed yes
- Web of Science (2011): Indexed yes
- BFI (2010): BFI-level 2
A near full-length open reading frame next generation sequencing assay for genotyping and identification of resistance-associated variants in hepatitis C virus

BACKGROUND: The current treatment options for hepatitis C virus (HCV), based on direct acting antivirals (DAA), are dependent on virus genotype and previous treatment experience. Treatment failures have been associated with detection of resistance-associated substitutions (RASs) in the DAA targets of HCV, the NS3, NS5A and NS5B proteins. OBJECTIVE: To develop a next generation sequencing based method that provides genotype and detection of HCV NS3, NS5A, and NS5B RASs without prior knowledge of sample genotype. STUDY DESIGN: In total, 101 residual plasma samples from patients with HCV covering 10 different viral subtypes across 4 genotypes with viral loads of 3.84-7.61 Log IU/mL were included. All samples were de-identified and consequently prior treatment status for patients was unknown. Almost full open reading frame amplicons (~9kb) were generated using RT-PCR with a single primer set. The resulting amplicons were sequenced with high throughput sequencing and analysed using an in-house developed script for detecting RASs. METHODS: The method successfully amplified and sequenced 94% (95/101) of samples with an average coverage of 14,035; four of six failed samples were genotype 4a. Samples analysed twice yielded reproducible nucleotide frequencies across all sites. RASs were detected in 21/95 (22%) samples at a 15% threshold. The method identified one patient infected with two genotype 2b variants, and the presence of subgenomic deletion variants in 8 (8.4%) of 95 successfully sequenced samples. CONCLUSIONS: The presented method may provide identification of HCV genotype, RASs detection, and detect multiple HCV infection without prior knowledge of sample genotype.
Benchmarking the HLA typing performance of Polysolver and Optitype in 50 Danish parental trios

Background: The adaptive immune response intrinsically depends on hypervariable human leukocyte antigen (HLA) genes. Concomitantly, correct HLA phenotyping is crucial for successful donor-patient matching in organ transplantation. The cost and technical limitations of current laboratory techniques, together with advances in next-generation sequencing (NGS) methodologies, have increased the need for precise computational typing methods. Results: We tested two widespread HLA typing methods using high quality full genome sequencing data from 150 individuals in 50 family trios from the Genome Denmark project. First, we computed descendant accuracies assessing the agreement in the inheritance of alleles from parents to offspring. Second, we compared the locus-specific homozygosity rates as well as the...
allele frequencies; and we compared those to the observed values in related populations. We provide guidelines for testing the accuracy of HLA typing methods by comparing family information, which is independent of the availability of curated alleles. Conclusions: Although current computational methods for HLA typing generally provide satisfactory results, our benchmark – using data with ultra-high sequencing depth – demonstrates the incompleteness of current reference databases, and highlights the importance of providing genomic databases addressing current sequencing standards, a problem yet to be resolved before benefiting fully from personalised medicine approaches HLA phenotyping is essential.

**General information**

State: Published

Organisations: Department of Bio and Health Informatics, Metagenomics, Integrative Systems Biology, Genomic Epidemiology, Disease Intelligence and Molecular Evolution, Technical University of Denmark, Aarhus University, BGI-Shenzhen, University of Copenhagen, Københavns Universitet, Aarhus Universitet, South China University of Technology, BGI-Europe


Number of pages: 12
Publication date: 2018

**Publication information**

Journal: B M C Bioinformatics
Volume: 19
Article number: 239
ISSN (Print): 1471-2105

Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): CiteScore 2.49 SJR 1.479 SNIP 0.878
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.54 SJR 1.581 SNIP 0.974
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.737 SNIP 1.079 CiteScore 2.77
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.916 SNIP 1.185 CiteScore 2.91
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.999 SNIP 1.323 CiteScore 3.38
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.9 SNIP 1.145 CiteScore 3.24
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Characterization of the enhancer and promoter landscape of inflammatory bowel disease from human colon biopsies

Inflammatory bowel disease (IBD) is a chronic intestinal disorder, with two main types: Crohn’s disease (CD) and ulcerative colitis (UC), whose molecular pathology is not well understood. The majority of IBD-associated SNPs are located in non-coding regions and are hard to characterize since regulatory regions in IBD are not known. Here we profile transcription start sites (TSSs) and enhancers in the descending colon of 94 IBD patients and controls. IBD-upregulated promoters and enhancers are highly enriched for IBD-associated SNPs and are bound by the same transcription factors. IBD-specific TSSs are associated to genes with roles in both inflammatory cascades and gut epithelia while TSSs distinguishing UC and CD are associated to gut epithelia functions. We find that as few as 35 TSSs can distinguish active CD, UC, and controls with 85% accuracy in an independent cohort. Our data constitute a foundation for understanding the molecular pathology, gene regulation, and genetics of IBD.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution, Department of Biotechnology and Biomedicine, University of Copenhagen, Roskilde University, Zealand University Hospital
Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox Nitrospira

The description of comammox Nitrospira spp., performing complete ammonia-to-nitrate oxidation, and their co-occurrence with canonical β-proteobacterial ammonia oxidizing bacteria (β-AOB) in the environment, calls into question the metabolic potential of comammox Nitrospira and the evolutionary history of their ammonia oxidation pathway. We report four new comammox Nitrospira genomes, constituting two novel species, and the first comparative genomic analysis on comammox Nitrospira. Unlike canonical Nitrospira, comammox Nitrospira genomes lack genes for assimilatory nitrite reduction, suggesting that they have lost the potential to use external nitrite nitrogen sources. By contrast, compared to canonical Nitrospira, comammox Nitrospira harbor a higher diversity of urea transporters and copper homeostasis genes and lack cyanate hydratase genes. Additionally, the two comammox clades differ in their ammonium uptake systems. Contrary to β-AOB, comammox Nitrospira genomes have single copies of the two central ammonia oxidation pathway
operons. Similar to ammonia oxidizing archaea and some oligotrophic AOB strains, they lack genes involved in nitric oxide reduction. Furthermore, comammox Nitrospira genomes encode genes that might allow efficient growth at low oxygen concentrations. Regarding the evolutionary history of comammox Nitrospira, our analyses indicate that several genes belonging to the ammonia oxidation pathway could have been laterally transferred from β-AOB to comammox Nitrospira. We postulate that the absence of comammox genes in other sublineage II Nitrospira genomes is the result of subsequent loss.

**General information**

State: Accepted/In press
Organisations: Department of Environmental Engineering, Water Technologies, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution
Authors: Palomo, A. (Intern), Pedersen, A. G. (Intern), Fowler, J. (Intern), Dechesne, A. (Intern), Sicheritz-Pontén, T. (Intern), Smets, B. F. (Intern)
Number of pages: 15
Publication date: 2018
Main Research Area: Technical/natural sciences

**Publication information**

Journal: *ISME Journal*
ISSN (Print): 1751-7362
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 2.284 SJR 4.813 CiteScore 9.5
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 8.91 SJR 4.938 SNIP 2.248
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 6.385 SNIP 2.473 CiteScore 9.64
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 5.369 SNIP 2.288 CiteScore 8.42
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 5.012 SNIP 2.271 CiteScore 8.62
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 4.941 SNIP 2.161 CiteScore 8.02
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 3.732 SNIP 1.826 CiteScore 6.5
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 3.361 SNIP 1.652
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 2.658 SNIP 1.47
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 2.047 SNIP 0.788
Web of Science (2008): Indexed yes
Original language: English
DOIs:
Integrative network analysis highlights biological processes underlying GLP-1 stimulated insulin secretion: A DIRECT study

Glucagon-like peptide 1 (GLP-1) stimulated insulin secretion has a considerable heritable component as estimated from twin studies, yet few genetic variants influencing this phenotype have been identified. We performed the first genome-wide association study (GWAS) of GLP-1 stimulated insulin secretion in non-diabetic individuals from the Netherlands Twin register (n = 126). This GWAS was enhanced using a tissue-specific protein-protein interaction network approach. We identified a beta-cell protein-protein interaction module that was significantly enriched for low gene scores based on the GWAS P-values and found support at the network level in an independent cohort from Tübingen, Germany (n = 100). Additionally, a polygenic risk score based on SNPs prioritized from the network was associated (P <0.05) with glucose-stimulated insulin secretion phenotypes in up to 5,318 individuals in MAGIC cohorts. The network contains both known and novel genes in the context of insulin secretion and is enriched for members of the focal adhesion, extracellular-matrix receptor interaction, actin cytoskeleton regulation, Rap1 and PI3K-Akt signaling pathways. Adipose tissue is, like the beta-cell, one of the target tissues of GLP-1 and we thus hypothesized that similar networks might be functional in both tissues. In order to verify peripheral effects of GLP-1 stimulation, we compared the transcriptome profiling of ob/ob mice treated with liraglutide, a clinically used GLP-1 receptor agonist, versus baseline controls. Some of the upstream regulators of differentially expressed genes in the white adipose tissue of ob/ob mice were also detected in the human beta-cell network of genes associated with GLP-1 stimulated insulin secretion. The findings provide biological insight into the mechanisms through which the effects of GLP-1 may be modulated and highlight a potential role of the beta-cell expressed genes RYR2, GD12, KIAA0232, COL4A1 and COL4A2 in GLP-1 stimulated insulin secretion.

General information

State: Published
Organisations: Department of Bio and Health Informatics, Integrative Systems Biology, Disease Intelligence and Molecular Evolution, Vrije Universiteit Amsterdam, Sanofi Aventis Deutschland GmbH, University of Copenhagen, NIHR Oxford Biomedical Research Centre, University of Oxford, University of Dundee, Eberhard-Karls-Universität Tübingen, VU University Medical Centre, Leiden University Medical Center, Vrije Universiteit Amsterdam, Imperial College London
Number of pages: 12
Publication date: 2018
Main Research Area: Technical/natural sciences

Publication information

Journal: P L o S One
Volume: 13
Issue number: 1
Article number: e0189886
ISSN (Print): 1932-6203
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): CiteScore 3.01 SJR 1.164 SNIP 1.111
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.11 SJR 1.236 SNIP 1.101
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.427 SNIP 1.136 CiteScore 3.32
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.559 SNIP 1.148 CiteScore 3.54
Web of Science (2014): Indexed yes
Metabolite ratios as potential biomarkers for type 2 diabetes: a DIRECT study

Aims/hypothesis: Circulating metabolites have been shown to reflect metabolic changes during the development of type 2 diabetes. In this study we examined the association of metabolite levels and pairwise metabolite ratios with insulin responses after glucose, glucagon-like peptide-1 (GLP-1) and arginine stimulation. We then investigated if the identified metabolite ratios were associated with measures of OGTT-derived beta cell function and with prevalent and incident type 2 diabetes. Methods: We measured the levels of 188 metabolites in plasma samples from 130 healthy members of twin families (from the Netherlands Twin Register) at five time points during a modified 3 h hyperglycaemic clamp with glucose, GLP-1 and arginine stimulation. We validated our results in cohorts with OGTT data (n = 340) and epidemiological case-control studies of prevalent (n = 4925) and incident (n = 4277) diabetes. The data were analysed using regression models with adjustment for potential confounders. Results: There were dynamic changes in metabolite levels in response to the different secretagogues. Furthermore, several fasting pairwise metabolite ratios were associated with one or multiple clamp-derived measures of insulin secretion (all p < 9.2 × 10^{-7}). These associations were significantly stronger compared with the individual metabolite components. One of the ratios, valine to phosphatidylcholine acyl-alkyl C32:2 (PC ae C32:2), in addition showed a directionally consistent positive association with OGTT-derived measures of insulin secretion and resistance (p = 9.2 × 10^{-7}) and prevalent type 2 diabetes (ORVal_PC ae C32:2 2.64 [β 0.97 ± 0.09], p = 1.0 × 10^{-27}). Furthermore, Val_PC ae C32:2 predicted incident diabetes independent of established risk factors in two epidemiological cohort studies (HRVal_PC ae C32:2 1.57 [β 0.45 ± 0.06], p = 1.3 × 10^{-15}), leading to modest improvements in the receiver operating characteristics when added to a model containing a set of established risk factors in both cohorts (increases from 0.780 to 0.801 and from 0.862 to 0.865 respectively, when added to the model containing traditional risk factors + glucose). Conclusions/interpretation: In this study we have shown that the Val_PC ae C32:2 metabolite ratio is associated with an increased risk of type 2 diabetes and measures of insulin secretion and resistance. The observed effects were stronger than that of the individual metabolites and independent of known risk factors.
Transcriptome analysis of root-knot nematode (Meloidogyne incognita)-infected tomato (Solanum lycopersicum) roots reveals complex gene expression profiles and metabolic networks of both host and nematode during susceptible and resistance responses

Root knot nematodes (RKNs, Meloidogyne incognita) are economically important endoparasites having a wide-host range. We have taken a comprehensive transcriptomic approach to investigate the expression of both tomato and RKN genes in tomato roots at five infection time intervals from susceptible plants and two infection time intervals from resistant plants, grown under soil conditions. Differentially expressed genes during susceptible (1827-tomato, 462-RKN) and resistance (25-tomato, 160-RKN) interactions were identified. In susceptible responses, tomato genes involved in cell wall structure, development, primary and secondary metabolites and defense signalling pathways along with RKN genes involved in host parasitism, development and defense are discussed. In resistance responses, tomato genes involved in secondary metabolite and hormone-mediated defense responses along with RKN genes involved in starvation stress-induced apoptosis are discussed. Also, forty novel differentially expressed RKN genes encoding secretory proteins were identified. Our findings, for the first time, provide novel insights into temporal regulation of genes involved in various biological processes from tomato and RKN simultaneously during susceptible and resistance responses, and reveals involvement of a complex network of biosynthetic pathways during disease development.
Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion

The human Y-chromosome does not recombine across its male-specific part and is therefore an excellent marker of human migrations. It also plays an important role in male fertility. However, its evolution is difficult to fully understand because of repetitive sequences, inverted repeats and the potentially large role of gene conversion. Here we perform an evolutionary analysis of 62 Y-chromosomes of Danish descent sequenced using a wide range of library insert sizes and high coverage, thus allowing large regions of these chromosomes to be well assembled. These include 17 father-son pairs, which we use to validate variation calling. Using a recent method that can integrate variants based on both mapping and de novo assembly, we genotype 10898 SNVs and 2903 indels (max length of 27241 bp) in our sample and show by father-son concordance and experimental validation that the non-recurrent SNP and indel variation on the Y chromosome tree is called very accurately. This includes variation called in a 0.9 Mb centromeric heterochromatic region, which is by far the most variable in the Y chromosome. Among the variation is also longer sequence-stretches not present in the reference genome but shared with the chimpanzee Y chromosome. We analyzed 2.7 Mb of large inverted repeats (palindromes) for variation patterns among the two palindrome arms and identified 603 mutation and 416 gene conversions events. We find clear evidence for GC-biased gene conversion in the palindromes (and a balancing AT mutation bias), but irrespective of this, also a strong bias towards gene conversion towards the ancestral state, suggesting that palindromic gene conversion may alleviate Muller’s ratchet. Finally, we also find a large number of large-scale gene duplications and deletions in the palindromic regions (at least 24) and find that such events can consist of complex combinations of simultaneous insertions and deletions of long stretches of the Y chromosome.

General information

State: Published
Organisations: Department of Bio and Health Informatics, Integrative Systems Biology, Metagenomics, Department of Systems Biology, Center for Biological Sequence Analysis, Integrative Systems Biology, Disease Intelligence and Molecular Evolution, Genomic Epidemiology, Functional Human Variation, Aarhus University, Technical University of Denmark, University of Bergen, Karolinska Institutet, BGI-Europe, University of Bristol, University of Copenhagen, Københavns Universitet, BGI-Shenzhen
An introduction to Deep learning on biological sequence data - Examples and solutions

Deep neural network architectures such as convolutional and long short-term memory networks have become increasingly popular as machine learning tools during the recent years. The availability of greater computational resources, more data, new algorithms for training deep models and easy to use libraries for implementation and training of neural networks are the drivers of this development. The use of deep learning has been especially successful in image recognition; and the development of tools, applications and code examples are in most cases centered within this field rather than within biology. Here, we aim to further the development of deep learning methods within biology by providing application examples and ready to apply and adapt code templates. Given such examples, we illustrate how architectures consisting of convolutional and long short-term memory neural networks can relatively easily be designed and trained to state-of-the-art performance on three biological sequence problems: prediction of subcellular localization, protein secondary structure and the binding of peptides to MHC Class II molecules. All implementations and datasets are available online to the scientific community at https://github.com/vanessajurtz/lasagne4bio. Supplementary data are available at Bioinformatics online.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Immunoinformatics and Machine Learning, Department of Applied Mathematics and Computer Science, Department of Electrical Engineering, Disease Intelligence and Molecular Evolution, Copenhagen Center for Health Technology, Cognitive Systems, University of Copenhagen
Pages: 3685-3690
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Bioinformatics
Volume: 33
Issue number: 22
ISSN (Print): 1367-4803
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 7.84
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
Bacterial whole genome-based phylogeny: construction of a new benchmarking dataset and assessment of some existing methods

Background
Whole genome sequencing (WGS) is increasingly used in diagnostics and surveillance of infectious diseases. A major application for WGS is to use the data for identifying outbreak clusters, and there is therefore a need for methods that can accurately and efficiently infer phylogenies from sequencing reads. In the present study we describe a new dataset that we have created for the purpose of benchmarking such WGS-based methods for epidemiological data, and also present an analysis where we use the data to compare the performance of some current methods.

Results
Our aim was to create a benchmark data set that mimics sequencing data of the sort that might be collected during an outbreak of an infectious disease. This was achieved by letting an E. coli hypermutator strain grow in the lab for 8 consecutive days, each day splitting the culture in two while also collecting samples for sequencing. The result is a data set consisting of 101 whole genome sequences with known phylogenetic relationship. Among the sequenced samples 51 correspond to internal nodes in the phylogeny because they are ancestral, while the remaining 50 correspond to leaves. We also used the newly created data set to compare three different online available methods that infer phylogenies from whole-genome sequencing reads: NDtree, CSI Phylogeny and REALPHY. One complication when comparing the output of these methods with the known phylogeny is that phylogenetic methods typically build trees where all observed sequences are placed as leafs, even though some of them are in fact ancestral. We therefore devised a method for post processing the inferred trees by collapsing short branches (thus relocating some leafs to internal nodes), and also present two new measures of tree similarity that takes into account the identity of both internal and leaf nodes.

Conclusions
Based on this analysis we find that, among the investigated methods, CSI Phylogeny had the best performance, correctly identifying 73% of all branches in the tree and 71% of all clades. We have made all data from this experiment (raw sequencing reads, consensus whole-genome sequences, as well as descriptions of the known phylogeny in a variety of formats) publicly available, with the hope that other groups may find this data useful for benchmarking and exploring the performance of epidemiological methods. All data is freely available at: https://cge.cbs.dtu.dk/services/evolution_data.php.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Genomic Epidemiology, Disease Intelligence and Molecular Evolution, National Food Institute, Research Group for Genomic Epidemiology, University of Copenhagen Authors: Ahrenfeldt, J. (Intern), Skaarup, C. (Intern), Hasman, H. (Ekstern), Pedersen, A. G. (Intern), Aarestrup, F. M. (Intern), Lund, O. (Intern)
Breadth of T cell responses after immunization with adenovirus vectors encoding ancestral antigens or polyvalent papillomavirus antigens

Oncogenic human papillomaviruses (HPVs) are in most cases eliminated by intervention of T cells. As many other pathogens, these oncogenic HPVs belong to an ancient and diverse virus family. Therefore, we found it relevant to investigate the potential and limitations of inducing a broad response - either by inducing cross-reactive T cells or by administering a polyvalent vaccine. To test these strategies, we designed 3 ancestral and 2 circulating sequences based on the two domains of the E1 and E2 proteins of papillomaviruses (PVs) that exhibit the highest degree of conservation in comparison to the other PV proteins. The PV sequences were fused to a T cell adjuvant, the murine invariant chain and encoded in a recombinant adenoviral vector which was administered to naïve outbred mice. By measuring T cell responses induced by these different vaccines and towards peptide pools representing 3 circulating strains and a putative ancestor of oncogenic HPVs, we showed that the ancestral vaccine antigen has to be approximately 90% identical to the circulating PVs before a marked drop of ~90% mean CD8+ T cell responses ensues. Interestingly, the combination of two or three type-specific PV vaccines did not induce a significant decrease of the CD8+ T cell response to the individual targeted PV types. Polyvalent HPV vaccine based on the E1 and E2 proteins seem to be capable of triggering responses towards more than one type of PV while the cross-reactivity of ancestral vaccine seems insufficient in consideration of the sequence diversity between HPV types.
Cerebellar mutism syndrome in children with brain tumours of the posterior fossa

Background: Central nervous system tumours constitute 25% of all childhood cancers; more than half are located in the posterior fossa and surgery is usually part of therapy. One of the most disabling late effects of posterior fossa tumour surgery is the cerebellar mutism syndrome (CMS) which has been reported in up to 39% of the patients but the exact incidence is uncertain since milder cases may be unrecognized. Recovery is usually incomplete. Reported risk factors are tumour type, midline location and brainstem involvement, but the exact aetiology, surgical and other risk factors, the clinical course and strategies for prevention and treatment are yet to be determined.

Methods: This observational, prospective, multicentre study will include 500 children with posterior fossa tumours. It opened late 2014 with participation from 20 Nordic and Baltic centres. From 2016, five British centres and four Dutch centres will join with a total annual accrual of 130 patients. Three other major European centres are invited to join from 2016/17. Follow-up will run for 12 months after inclusion of the last patient. All patients are treated according to local practice. Clinical data are collected through standardized online registration at pre-determined time points pre- and postoperatively. Neurological status and
speech functions are examined pre-operatively and postoperatively at 1-4 weeks, 2 and 12 months. Pre- and postoperative speech samples are recorded and analysed. Imaging will be reviewed centrally. Pathology is classified according to the 2007 WHO system. Germline DNA will be collected from all patients for associations between CMS characteristics and host genome variants including pathway profiles. Discussion: Through prospective and detailed collection of information on 1) differences in incidence and clinical course of CMS for different patient and tumour characteristics, 2) standardized surgical data and their association with CMS, 3) diversities and results of other therapeutic interventions, and 4) the role of host genome variants, we aim to achieve a better understanding of risk factors for and the clinical course of CMS - with the ultimate goal of defining strategies for prevention and treatment of this severely disabling condition.

General information
State: Published
Organisations: Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution, Rigshospitalet, St Olav University Hospital, Skåne University Hospital, Aarhus University Hospital, Karolinska University Hospital, Nordsjællands University Hospital, Helsinki University Central Hospital, Akademiska sjukhuset, Habilitation and Technical Aid, Karolinska Institutet
Number of pages: 7
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: B M C Cancer
Volume: 17
Article number: 439
ISSN (Print): 1471-2407
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.066 SJR 1.464 CiteScore 3.49
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 3.56 SJR 1.488 SNIP 1.071
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.652 SNIP 1.14 CiteScore 3.72
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.719 SNIP 1.27 CiteScore 3.73
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.694 SNIP 1.282 CiteScore 3.84
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.654 SNIP 1.203 CiteScore 3.79
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 1
Scopus rating (2011): SJR 1.541 SNIP 1.074 CiteScore 3.55
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 1
Scopus rating (2010): SJR 1.508 SNIP 1.123
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 1
Scopus rating (2009): SJR 1.437 SNIP 1.069
Clustering on baseline clinical variables identifies subgroups of type 2 diabetes patients with different rate of progression over 18 months: a DIRECT study

General information
State: Published
Organisations: Department of Bio and Health Informatics, Integrative Systems Biology, Disease Intelligence and Molecular Evolution, National Research Council of Italy, Lund University, University of Geneva, University of Manchester, University of Dundee
Pages: S117
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Diabetologia
Volume: 60
Issue number: Suppl. 1
Article number: 250
ISSN (Print): 0012-186X
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SJR 3.228 SNIP 1.619 CiteScore 5.09
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 5.23 SJR 3.25 SNIP 1.721
DeepLoc: prediction of protein subcellular localization using deep learning

The prediction of eukaryotic protein subcellular localization is a well-studied topic in bioinformatics due to its relevance in proteomics research. Many machine learning methods have been successfully applied in this task, but in most of them, predictions rely on annotation of homologues from knowledge databases. For novel proteins where no annotated homologues exist, and for predicting the effects of sequence variants, it is desirable to have methods for predicting protein properties from sequence information only. Here, we present a prediction algorithm using deep neural networks to predict protein subcellular localization relying only on sequence information. At its core, the prediction model uses a recurrent neural network that processes the entire protein sequence and an attention mechanism identifying protein regions important for the subcellular localization. The model was trained and tested on a protein dataset extracted from one of the latest UniProt releases, in which experimentally annotated proteins follow more stringent criteria than previously. We demonstrate that our model achieves a good accuracy (78% for 10 categories; 92% for membrane-bound or soluble), outperforming current state-of-the-art algorithms, including those relying on homology information. The method is available as a web server at http://www.cbs.dtu.dk/services/DeepLoc. Example code is available at https://github.com/JJAlmagro/subcellular_localization. The dataset is available at http://www.cbs.dtu.dk/services/DeepLoc/data.php. jjalma@dtu.dk.
Evolutionary analysis of whole-genome sequences confirms inter-farm transmission of Aleutian mink disease virus

Aleutian mink disease virus (AMDV) is a frequently encountered pathogen associated with mink farming. Previous phylogenetic analyses of AMDV have been based on shorter and more conserved parts of the genome, e.g. the partial NS1 gene. Such fragments are suitable for detection but are less useful for elucidating transmission pathways while sequencing entire viral genomes provides additional informative sites and often results in better-resolved phylogenies. We explore how whole-genome sequencing can benefit investigations of AMDV transmission by reconstructing the relationships between AMDV field samples from a Danish outbreak. We show that whole-genome phylogenies are much better resolved than those based on the partial NS1 gene sequences extracted from the same alignment. Well-resolved phylogenies contain more information about the underlying transmission trees and are useful for understanding the spread of a pathogen. In the main case investigated here, the transmission path suggested by the tree structure was supported by epidemiological data. The use of molecular clock models further improved tree resolution and provided time estimates for the viral ancestors consistent with the proposed direction of spread. It was however impossible to infer transmission pathways from the partial NS1 gene tree, since all samples from the case farms branched out from a single internal node. A sliding window analysis showed that there were no shorter genomic regions providing the same phylogenetic resolution as the entire genome. Altogether, these results suggest that phylogenetic analyses based on whole-genome sequencing taking into account sampling dates and epidemiological data is a promising set of tools for clarifying AMDV transmission.

General information
State: Published
Organisations: Molecular Evolution, Department of Biotechnology and Biomedicine, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution, National Veterinary Institute, Virology, Kopenhagen Fur
Authors: Hagberg, E. E. (Intern), Pedersen, A. G. (Intern), Larsen, L. E. (Intern), Krarup, A. (Ekstern)
Pages: 1360-1371
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Journal of General Virology
Volume: 98
Issue number: 6
ISSN (Print): 0022-1317
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 0.877 SJR 1.325 CiteScore 2.68
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 2.93 SJR 1.544 SNIP 0.891
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 1.738 SNIP 0.998 CiteScore 3.26
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 1.69 SNIP 1.057 CiteScore 3.25
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 1.764 SNIP 1.154 CiteScore 3.64
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 1.525 SNIP 1.034 CiteScore 3.28
ISI indexed (2012): ISI indexed yes
Forskellige virusstammer var årsag til udbrud af plasmacytose i danske mink (Neovison vison) i 2015

General information
State: Published
Organisations: National Veterinary Institute, Virology, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution, Diagnostic & Development, Department of Biotechnology and Biomedicine, Kopenhagen Fur
Pages: 163-167
Publication date: 2017

Host publication information
Title of host publication: Faglig årsberetning 2016 : Kopenhagen Fur
Place of publication: Aarhus N
Publisher: Kopenhagen Fur
Main Research Area: Technical/natural sciences
Electronic versions:
How Much of the Human Genome is Functional?

General information
State: Published
Organisations: Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution
Authors: Nielsen, H. (Intern)
Number of pages: 1
Publication date: 2017

Host publication information
Title of host publication: Abstract from Seventeenth Annual Gatherings in Biosemiotics, Lausanne, Switzerland.
Main Research Area: Technical/natural sciences
Electronic versions:
NIELSEN.pdf

Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor

The international Testicular Cancer Consortium (TECAC) combined five published genome-wide association studies of testicular germ cell tumor (TGCT; 3,558 cases and 13,970 controls) to identify new susceptibility loci. We conducted a fixed-effects meta-analysis, including, to our knowledge, the first analysis of the X chromosome. Eight new loci mapping to 2q14.2, 3q26.2, 4q35.2, 7q36.3, 10q26.13, 15q21.3, 15q22.31, and Xq28 achieved genome-wide significance (P < 5 × 10⁻⁸). Most loci harbor biologically plausible candidate genes. We refined previously reported associations at 9p24.3 and 19p12 by identifying one and three additional independent SNPs, respectively. In aggregate, the 39 independent markers identified to date explain 37% of father-to-son familial risk, 8% of which can be attributed to the 12 new signals reported here. Our findings substantially increase the number of known TGCT susceptibility alleles, move the field closer to a comprehensive understanding of the underlying genetic architecture of TGCT, and provide further clues to the etiology of TGCT.

General information
State: Published
Organisations: Department of Biotechnology and Biomedicine, DTU Multi Assay Core, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution, Department of Systems Biology, National Institutes of Health, University of Leeds, Cancer Registry of Norway, Oslo and Akershus University College of Applied Sciences, Karolinska Institutet, The Institute of Cancer Research, University of Pennsylvania, Fred Hutchinson Cancer Research Center, Institute for Systems Biology, Harvard University, H. Lee Moffitt Cancer Center and Research Institute, Copenhagen University Hospital
Molecular diagnostics of aleutian mink disease virus: applied use of next generation sequencing and phylogenetics

Aleutian Mink Disease virus (AMDV) is a parvovirus causing Aleutian Mink Disease (AMD), often referred to as plasmacytosis. It is a systemic infection affecting mink of all ages, and is globally the most important pathogen impacting mink farming. In Denmark AMDV has since 1999 been monitored by a national control program, which is based on serological screening of all animals and encourages infected farms to stamp out. Historically there has been no consensus about which genomic region of the virus to analyse e.g. in relation to surveillance, and most previous studies in this regard, have been based either on partial or entire genes, or on pure epidemiological data. Thus, when initiating this project, little was known about AMDV’s total genomic diversity and how the virus was spread between farms.

Recent advances in the field of molecular diagnostics have made high throughput tools such as next generation sequencing cheaper and more easily available. Whole genome sequencing and advanced phylogenetic analyses have successfully been applied to describe the molecular evolution and transmission patterns for viruses such as Foot and Mouth Disease Virus (FMDV), Ebola, and avian influenza virus, however not previously for AMDV. The overall aim with this thesis was to investigate if next generation sequencing and phylogenetic analyses of full length isolates could improve our understanding of the total genomic diversity and evolution of AMDV. Additionally, we wanted to evaluate if this knowledge could contribute to the elucidation of AMDV transmission between farms and improve molecular diagnostics. During the first phase of this project a method for performing whole genome sequencing of AMDV was developed. This protocol enabled the sequencing of a large number of in vivo infectious AMDV isolates and provided the necessary dataset to act as foundation for the remaining analyses in the thesis. The first original paper (Manuscript 1) describes this protocol.

Manuscript 2 is a proof-of-concept study which demonstrated the advantage of using the whole genome sequence approach, compared to the in Denmark traditionally used partial NS1 gene sequencing, for the elucidation of transmission pathways between farms. The study was performed on samples from a small local AMDV outbreak, and clearly illustrated that the phylogenies based on partial NS1 gene sequencing were uninformative and could not be used for determining transmission pathways, even in the light of supporting epidemiological data. The whole-genome approach on the other hand, confirmed the epidemiological hypothesis about the direction of spread.

In Manuscript 3, the methodologies from Manuscript 1 and 2 were applied to generate the to-date most comprehensive phylogenetic and genetic analysis of full-length AMDV isolates, composed of more than 200 field strains. The study shed light on the diversity and evolutionary behaviour of two distinct AMDV strains, in addition to providing the first robust evolutionary rate-estimates. Altogether, the work presented in this thesis provides a contribution to the molecular diagnostics of AMDV, enables us better to understand the virus’ evolutionary behaviour in the context of mink farming, and is anticipated to be of value for more accurately tracing back in time the emergence of future outbreaks.
Niche differentiation and evolution of comammox Nitrospira through a comparative genomics analysis

Nitrification, the biological oxidation of ammonium to nitrate, is a fundamental process in the nitrogen cycle and plays an important role in natural and engineered systems. Throughout the last century, nitrification was assumed to be a two-step process executed by two different functional groups, ammonia oxidizing prokaryotes (AOP) and nitrite oxidizing bacteria (NOB). Recently, several articles have shown the capability of a single microorganism, belonging to the genus Nitrospira, to carry out the complete oxidation of ammonia to nitrate (comammox). Nitrospira spp. are widespread in both natural and engineered ecosystems associated with nitrogen cycling and different species are frequently observed to coexist in the same environment. Besides recent discoveries pointing towards versatile metabolism in some Nitrospira species, little is known about the functional potential of the two comammox Nitrospira clades, and the factors involved in niche-partitioning between comammox and canonical Nitrospira.

A comparative genomics analysis was conducted with five genomes recovered from a groundwater-fed rapid sand filter (including both comammox clades and a nitrite-oxidizing Nitrospira population genome) and high quality published Nitrospira genomes, to reveal distinct genomic features within Nitrospira. In addition, we investigated the evolution of the ammonia oxidation pathway in comammox Nitrospira. This analysis revealed distinct genetic capabilities of the different comammox clades and canonical Nitrospira which can help to explain the coexistence and niche partitioning of Nitrospira spp. These divergences range from the nitrogen source utilization capacity to the ability for electron donor versatility, and other characteristics such as stress response. With respect to the evolutionary history of comammox Nitrospira, our analysis indicates transfer events with betaproteobacterial ammonia oxidizers. In addition, transfer events between comammox clade A and clade B were also detected for genes belonging to the ammonium oxidation pathway. Together, these results expand the actual knowledge of the ecology and evolution of the recently discovered comammox Nitrospira.

General information
State: Published
Organisations: Department of Environmental Engineering, Water Technologies, Department of Biotechnology and Biomedicine, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution, Metagenomics
Authors: Palomo, A. (Intern), Fowler, J. (Intern), Pedersen, A. G. (Intern), Sicheritz-Pontén, T. (Intern), Smets, B. F. (Intern)
Number of pages: 1
Publication date: 2017
Event: Abstract from ICoN5: 5th International Conference on Nitrification, Vienna, Austria.
Main Research Area: Technical/natural sciences
Electronic versions:
Alex_ICON2017_JF.pdf
Source: PublicationPreSubmission
Source-ID: 136926853
Publication: Research - peer-review › Conference abstract for conference – Annual report year: 2017

Obesity is associated with depot-specific alterations in adipocyte DNA methylation and gene expression

The present study aimed to identify genes exhibiting concomitant obesity-dependent changes in DNA methylation and gene expression in adipose tissues in the mouse using diet-induced obese (DIO) C57BL/6J and genetically obese ob/ob mice as models. Mature adipocytes were isolated from epididymal and inguinal adipose tissues of ob/ob and DIO C57BL/6J mice. DNA methylation was analyzed by MeDIP-sequencing and gene expression by microarray analysis. The majority of differentially methylated regions (DMRs) were hypomethylated in obese mice. Global methylation of long interspersed elements indicated that hypomethylation did not reflect methyl donor deficiency. In both DIO and ob/ob mice, we observed more obesity-associated methylation changes in epididymal than in inguinal adipocytes. Assignment of DMRs to promoter, exon, intron and intergenic regions demonstrated that DIO-induced changes in DNA methylation in C57BL/6J mice occurred primarily in exons, whereas inguinal adipocytes of ob/ob mice exhibited a higher enrichment of DMRs in promoter regions than in other regions of the genome, suggesting an influence of leptin on DNA methylation in inguinal adipocytes. We observed altered methylation and expression of 9 genes in epididymal adipocytes, including the known obesity-associated genes, Ehd2 and Kcnd15, and a novel candidate gene, Irf8, possibly involved in immune type 1/type2 balance. The use of 2 obesity models enabled us to dissociate changes associated with high fat feeding from those associated with obesity per se. This information will be of value in future studies on the mechanisms governing the development of obesity and changes in adipocyte function associated with obesity.

General information
State: Published
Organisations: Department of Systems Biology, DTU Multi Assay Core, Department of Biotechnology and Biomedicine, DTU Multi Assay Core, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution, University of Copenhagen, BGI-Shenzhen, National Institute for Nutrition and Seafood Research, University of California, San Francisco
Authors: Sonne, S. B. (Ekstern), Yadav, R. (Intern), Yin, G. (Ekstern), Dalgaard, M. D. (Intern), Myrmel, L. S. (Ekstern), Gupta, R. (Intern), Wang, J. (Ekstern), Madsen, L. (Ekstern), Kajimura, S. (Ekstern), Kristiansen, K. (Ekstern)
Predicting Secretory Proteins with SignalP

SignalP is the currently most widely used program for prediction of signal peptides from amino acid sequences. Proteins with signal peptides are targeted to the secretory pathway, but are not necessarily secreted. After a brief introduction to the biology of signal peptides and the history of signal peptide prediction, this chapter will describe all the options of the current version of SignalP and the details of the output from the program. The chapter includes a case study where the scores of SignalP were used in a novel way to predict the functional effects of amino acid substitutions in signal peptides.

General information

State: Published
Organisations: Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution
Authors: Nielsen, H. (Intern)
Pages: 59-73
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information

Journal: Methods in Molecular Biology
Volume: 1611
ISSN (Print): 1064-3745
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SJR 0.616 SNIP 0.318 CiteScore 0.96
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 0.76 SJR 0.585 SNIP 0.278
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 0.627 SNIP 0.319 CiteScore 0.82
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 0.735 SNIP 0.374 CiteScore 1.02
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 0.751 SNIP 0.351 CiteScore 1.17
ISI indexed (2013): ISI indexed no
BFI (2012): BFI-level 1
Scopus rating (2012): SJR 0.753 SNIP 0.427 CiteScore 1.26
ISI indexed (2012): ISI indexed no
Protein Sorting Prediction

Many computational methods are available for predicting protein sorting in bacteria. When comparing them, it is important to know that they can be grouped into three fundamentally different approaches: signal-based, global-property-based and homology-based prediction. In this chapter, the strengths and drawbacks of each of these approaches is described through many examples of methods that predict secretion, integration into membranes, or subcellular locations in general. The aim of this chapter is to provide a user-level introduction to the field with a minimum of computational theory.
Sequencing and de novo assembly of 150 genomes from Denmark as a population reference

Hundreds of thousands of human genomes are now being sequenced to characterize genetic variation and use this information to augment association mapping studies of complex disorders and other phenotypic traits. Genetic variation is identified mainly by mapping short reads to the reference genome or by performing local assembly. However, these approaches are biased against discovery of structural variants and variation in the more complex parts of the genome. Hence, large-scale de novo assembly is needed. Here we show that it is possible to construct excellent de novo assemblies from high-coverage sequencing with mate-pair libraries extending up to 20 kilobases. We report de novo assemblies of 150 individuals (50 trios) from the GenomeDenmark project. The quality of these assemblies is similar to those obtained using the more expensive long-read technology. We use the assemblies to identify a rich set of structural variants including many novel insertions and demonstrate how this variant catalogue enables further deciphering of known association mapping signals. We leverage the assemblies to provide 100 completely resolved major histocompatibility complex haplotypes and to resolve major parts of the Y chromosome. Our study provides a regional reference genome that we expect will improve the power of future association mapping studies and hence pave the way for precision medicine initiatives, which now are being launched in many countries including Denmark.

General information

State: Published
Organisations: Department of Bio and Health Informatics, Metagenomics, Department of Systems Biology, Center for Biological Sequence Analysis, Integrative Systems Biology, Integrative Systems Biology, Disease Intelligence and Molecular Evolution, Genomic Epidemiology, High Performance Computing, Functional Human Variation, University of Copenhagen, Aarhus University, BGI-Shenzhen, BGI-Europe, Technical University of Denmark, University of Oslo, University of Bergen, Karolinska Institutet


Pages: 87-91
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information

Journal: Nature
Volume: 548
ISSN (Print): 0028-0836
Ratings:
BFI (2018): BFI-level 3
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 14.59
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 13.33
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 14.38
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 14.22
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 14.96
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Currently available prophylactic vaccines have no therapeutic efficacy for preexisting human papillomavirus (HPVs) infections, do not target all oncogenic HPVs and are insufficient to eliminate the burden of HPV induced cancer. We aim to develop an alternative HPV vaccine which is broadly effective and capable of clearing preexisting infection. In an initial attempt to develop a broadly reactive therapeutic vaccine, we designed a putative papillomavirus (PV) ancestor antigen (circulating sequence derived antigenic sequences E1E2-CDSE1E2) based on the conserved E1 and E2 protein sequences from existing oncogenic HPV strains. This antigen was found to be as related to circulating oncogenic Macaca fascicularis papillomaviruses (MfPVs) as to oncogenic HPVs. The CDSE1E2 antigen was fused to a T-cell adjuvant and encoded in chimpanzee 3 and 63 adenoviral vectors. We first showed that the combination of these 2 vaccines induced long-lasting potent CDSE1E2 specific T cell responses in outbred mice. This prime-boost regimen was then tested in female macaques naturally infected with MfPVs. All immunized animals (16/16) responded to the vaccine antigen but with reduced cross-reactivity against existing PVs. Preexisting MfPV infections did not prime vaccine inducible immune responses. Importantly, immunized oncogenic MfPV type 3 (MfPV3) infected animals that responded toward MfPV3 were able to diminish cervical MfPV3 DNA content. Although insufficient breadth was achieved, our results suggest that a relevant level of E1E2 specific T cell immunity is achievable and might be sufficient for the elimination of PV infection. Importantly, naturally infected macaques, offer a relevant model for testing vaccines aimed at eliminating mucosal PV infections.
Whole grain-rich diet reduces body weight and systemic low-grade inflammation without inducing major changes of the gut microbiome: a randomised cross-over trial

**Objective** To investigate whether a whole grain diet alters the gut microbiome and insulin sensitivity, as well as biomarkers of metabolic health and gut functionality. Design 60 Danish adults at risk of developing metabolic syndrome were included in a randomised cross-over trial with two 8-week dietary intervention periods comprising whole grain diet and refined grain...
diet, separated by a washout period of ≥6 weeks. The response to the interventions on the gut microbiome composition and insulin sensitivity as well on measures of glucose and lipid metabolism, gut functionality, inflammatory markers, anthropometry and urine metabolomics were assessed. Results 50 participants completed both periods with a whole grain intake of 179±50 g/day and 13±10 g/day in the whole grain and refined grain period, respectively. Compliance was confirmed by a difference in plasma alkylresorcinols (p<0.0001). Compared with refined grain, whole grain did not significantly alter glucose homeostasis and did not induce major changes in the faecal microbiome. Also, breath hydrogen levels, plasma short-chain fatty acids, intestinal integrity and intestinal transit time were not affected. The whole grain diet did, however, compared with the refined grain diet, decrease body weight (p=0.0001), serum inflammatory markers, interleukin (IL)-6 (p=0.009) and C-reactive protein (p=0.003). The reduction in body weight was consistent with a reduction in energy intake, and IL-6 reduction was associated with the amount of whole grain consumed, in particular with intake of rye. Conclusion Compared with refined grain diet, whole grain diet did not alter insulin sensitivity and gut microbiome but reduced body weight and systemic low-grade inflammation.

General information
State: Accepted/In press
Organisations: National Food Institute, Research Group for Gut Microbiology and Immunology, Department of Bio and Health Informatics, Metagenomics, Disease Intelligence and Molecular Evolution, Department of Biotechnology and Biomedicine, Disease Systems Immunology, Department of Chemical and Biochemical Engineering, Organic Chemistry, Center for BioProcess Engineering, DTU Multi Assay Core, Research Group for Analytical Food Chemistry, Copenhagen Center for Health Technology, University of Copenhagen, Chalmers University of Technology, Chalmers University of Technology, Bispebjerg University Hospital, Herlev and Gentofte Hospital, University of Auckland
Number of pages: 12
Publication date: 2017
Main Research Area: Technical/natural sciences

Publication information
Journal: Gut
ISSN (Print): 0017-5749
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 3.832 SJR 7.44 CiteScore 9.82
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 9.29 SJR 7.074 SNIP 3.946
BFI (2015): BFI-level 2
Scopus rating (2015): SJR 6.809 SNIP 3.968 CiteScore 9.1
BFI (2014): BFI-level 2
Scopus rating (2014): SJR 6.104 SNIP 3.865 CiteScore 8.76
BFI (2013): BFI-level 2
Scopus rating (2013): SJR 5.58 SNIP 3.459 CiteScore 7.6
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): SJR 4.066 SNIP 2.737 CiteScore 6.36
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): SJR 3.626 SNIP 2.612 CiteScore 5.74
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 3.527 SNIP 2.719
Web of Science (2010): Indexed yes
Asparaginase-associated pancreatitis: a study on phenotype and genotype in the NOPHO ALL2008 protocol

Asparaginase (ASP)-associated pancreatitis (AAP) occurs during acute lymphoblastic leukemia treatment. Among 1285 children (1.0-17.9 years) diagnosed during July 2008-December 2014 and treated according to the Nordic/Baltic ALL2008 protocol, 86 (cumulative incidence = 6.8%) developed AAP. Seventy-three cases were severe (diagnostic AAP criteria persisting 472 h) and 13 mild. Cases were older than controls (median: 6.5 vs 4.5 years; P = 0.001). Pseudocysts developed in 28%. Of the 20 re-exposed to ASP, 9 (45%) developed a second AAP. After a median follow-up of 2.3 years, 8% needed permanent insulin therapy, and 7% had recurrent abdominal pain. Germline DNA on 62 cases and 638 controls was genotyped on Omni2.5exome-8-v1.2 BeadChip arrays. Overall, the ULK2 variant rs281366 showed the strongest association with AAP (P = 5.8x10(-7); odds ratio (OR) = 6.7). Cases with the rs281366 variant were younger (4.3 vs 8 years; P = 0.015) and had lower risk of AAP-related complications (15% vs 43%; P = 0.13) compared with cases without this variant. Among 45 cases and 517 controls.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Functional Human Variation, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution, Copenhagen University Hospital
Pages: 325-332
Publication date: 2016
Main Research Area: Technical/natural sciences

Publication Information
Journal: Leukemia
Volume: 31
Issue number: 2
ISSN (Print): 0887-6924
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): SNIP 2.085 SJR 5.131 CiteScore 6.45
Web of Science (2017): Indexed Yes
BFI (2016): BFI-level 2
Scopus rating (2016): SJR 5.041 SNIP 2.226 CiteScore 6.47
Quality of life (QoL) and neurotoxicity in germ-cell cancer survivors (GCCS)

Background: The majority of patients with testicular cancer become long-term survivors. However, treatment is associated with late effects which may hamper QoL. The aims of the present study were to assess the impact of treatment on long-term QoL and evaluate the influence of neurotoxicity on QoL.

Methods: All GCCS identified in the Danish DaTeCa database were asked to fill in a questionnaire concerning late-effects Nov 2014 – Jan 2016. QoL was assessed with EORTC-QLQ C30 including 30 items divided into 15 subscales. Neurotoxicity was assessed with the FACT/GOG NTX12-scale including 12 items, divided into 4 subscales (neuropathy, ototoxicity, motor impairment, and dysfunction). Patients were divided into treatment groups; surveillance only (reference), n = 1092, radiotherapy (RT), n = 299, BEP chemotherapy (CT), n = 790, and more than one line of treatment (MTOL), n = 82. Outcomes were compared with ordinal logistic regression using treatment and attained age as covariates.

Results: In total, 2308 patients answered the questionnaire. Median attained age was 53.5 years (range: 24.9 - 94.5), and median time from treatment was 18.8 years (range: 7.0 - 32.2). Overall, Global health status was good, mean: 75.4, SD: 20.0. Treatments were significantly negatively associated with QoL in many subscales; CT: dyspnea, financial difficulties, impaired cognitive function, impaired social function, MTOL: impaired global health status, fatigue, dyspnea, financial
difficulties, impaired physical function, impaired cognitive function, and impaired social function. Neurotoxicity was closely correlated to treatment. RT was associated with three of four subscales; CT and MTOL were associated with all subscales. When adjusting QoL outcomes for neurotoxicity, all negative associations between QoL and treatment disappeared except dyspnea and impaired social function in the MTOL-group. Neurotoxicity was associated with all EORTC-subscales (p < .001).

Conclusions: Treatment with BEP and MTOL were associated with several QoL subscales in GCCS. However, when adjusting for neurotoxicity the associations generally disappeared. Neurotoxicity correlated strongly with QoL.
The Nopho-European Study on Cerebellar Mutism Syndrome (CMS)

**General information**

State: Published
Organisations: Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution, Rigshospitalet, Alder Hey Children's Hospital, University Hospital Linköping, Karolinska University Hospital, Helsinki University Central Hospital, Lithuanian University of Health Sciences, Turku University Hospital, University Hospital of Umeå, Uppsala University Hospital, Tampere University Hospital, BarnReHab Skåne, Kuopio University Hospital, Sahlgrenska University Hospital, Haukeland University Hospital, Radboud University Medical Centre, Children Brain Tumour Research Centre
Number of pages: 1
Pages: 17
Publication date: 2016
Conference: 17th International Symposium on Pediatric Neuro-Oncology (ISPNO), Liverpool, United Kingdom, 12/06/2016 - 12/06/2016
Main Research Area: Technical/natural sciences

**Publication information**

Journal: Neuro-Oncology
Volume: 18
ISSN (Print): 1522-8517
Ratings:
BFI (2018): BFI-level 1
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 1
Scopus rating (2017): SNIP 1.962 SJR 4.064 CiteScore 6.76
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 1
Scopus rating (2016): CiteScore 5.64 SJR 3.048 SNIP 1.877
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 1
Scopus rating (2015): SJR 3.246 SNIP 2.004 CiteScore 6
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 1
Scopus rating (2014): SJR 3.079 SNIP 1.905 CiteScore 5.66
BFI (2013): BFI-level 1
Scopus rating (2013): SJR 2.98 SNIP 1.788 CiteScore 5.91
ISI indexed (2013): ISI indexed yes
Whole genome sequencing for childhood cancer in Denmark

The talk will describe our involvement in the Danish project STAGING, “Sequencing Three Actionable Genomes – Implications & National Guidelines”, an interdisciplinary, multi-tiered 3-year study of 600 consecutive childhood cancer patients and their families, with extensive genomic sequencing of host, tumour and gut microbiome’s genomes. In Europe, cancer accounts for approximately 25% of all deaths in children >1 year. Most cured patients are burdened by late effects, including risk of second cancer and debilitating toxicities. Recent advancements in genetic sequencing technology and reduction in costs have led to new strategies for identification of cancer predisposition and targeted treatment. STAGING is a nation-wide programme offering full, up-front genetic testing for childhood cancer patients and implements the findings into health care. Paediatric oncology provides a unique proof-of-principle framework for such a program, since it is one of the best organized medical specialties with nation-wide strategies for diagnostics, therapy, deep response phenotyping, and follow-up.

General information
State: Published
Organisations: Center for Biological Sequence Analysis, Functional Human Variation, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution
Authors: Gupta, R. (Intern)
Number of pages: 1
Publication date: 2016
Main Research Area: Technical/natural sciences
Links:
http://www.sustain.dtu.dk/

Bibliographical note
Sustain Abstract H-1
Publication: Research - peer-review › Conference abstract for conference – Annual report year: 2016

Activities:
DeepLoc: Prediction of protein subcellular localization using deep learning
Period: 3 Nov 2017
Henrik Nielsen (Guest lecturer)
Jose Juan Almagro Armenteros (Guest lecturer)
Department of Bio and Health Informatics
Disease Intelligence and Molecular Evolution
Department of Applied Mathematics and Computer Science

Related external organisation
Intomics A/S
Denmark
Activity: Talks and presentations › Conference presentations

DeepLoc: Prediction of protein subcellular localization using deep learning
Period: 29 Aug 2017
Henrik Nielsen (Guest lecturer)
Department of Bio and Health Informatics
Disease Intelligence and Molecular Evolution

Related external organisation
Stockholm University
Sweden
Activity: Talks and presentations › Conference presentations

DeepLoc: Prediction of protein subcellular localization using deep learning
Period: 25 Aug 2017
Henrik Nielsen (Guest lecturer)
Department of Bio and Health Informatics
Disease Intelligence and Molecular Evolution

Related event
Annual Danish Bioinformatics Conference 2017: Elixir
23/08/2017 → 25/08/2017
Odense, Denmark
Activity: Talks and presentations › Conference presentations

How Much of the Human Genome is Functional?
Period: 8 Jun 2017
Henrik Nielsen (Guest lecturer)
Department of Bio and Health Informatics
Disease Intelligence and Molecular Evolution
Documents:
Abstract

Related event
Seventeenth Annual Gatherings in Biosemiotics
06/06/2017 → 10/06/2017
Lausanne, Switzerland
Activity: Talks and presentations › Conference presentations

Press clippings:

Bedre smittesporing med supercomputer
Emma Elisabeth Hagberg
Subject
genar og genomer; husdyrsygdomme; produktionsdyr; dataanalyse
Molecular Evolution, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution

Media contribution (1)

Bedre smittesporing med supercomputer
17/09/2016
Dynamo, Print
Julie Iben Schmidt
http://www.dtu.dk/Om-DTU/Nyheder-og-presse/Dynamo
Emma Elisabeth Hagberg
Molecular Evolution, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution

Computerome - Kopenhagen Fur
Emma Elisabeth Hagberg
10/12/2015
Molecular Evolution, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution

Media contribution (1)

Computerome - Kopenhagen Fur
10/12/2015
Youtube, Web
Julie Iben Schmidt
https://www.youtube.com/watch?v=HPsWZzi5Gkg
Emma Elisabeth Hagberg
Molecular Evolution, Department of Bio and Health Informatics, Disease Intelligence and Molecular Evolution

Press / Media