Publications

List of Publications

Click on title for complete information.

The use of wearable sensor technology to enhance supportive care in hospitalized palliative patients (Support trial): a prospective preliminary pilot study

CONCLUSION: This pilot study demonstrates the potential of continuous monitoring technologies in palliative care, but inconsistent data availability limits the ability to recommend their routine use at this stage. Despite these challenges, the promising results highlight the need for further studies to improve device reliability and explore the broader applicability of this technology in palliative care settings.

Postoperative use of fitness trackers for continuous monitoring of vital signs: a survey of hospitalized patients

Wearables and Internet of Things (IoT) technologies are increasingly incorporated into healthcare, including perioperative settings. These devices offer continuous non-invasive monitoring of vital signs, patient position, and mobilization. Nonetheless, there is currently little information about tolerance and acceptance of wearables in postoperative patients. We therefore assessed opinions and user experience in postoperative patients who used three popular fitness trackers during their entire...

Evaluation of Non-Invasive Hemoglobin Monitoring in Perioperative Patients: A Retrospective Study of the Rad-67TM (Masimo)

Background: Hemoglobin (Hb) is a crucial parameter in perioperative care due to its essential role for oxygen transport and tissue oxygenation. Accurate Hb monitoring allows for timely interventions to address perioperative anemia and, thus, prevent morbidity and mortality. Traditional Hb measurements rely on invasive blood sampling, which significantly contributes to iatrogenic anemia and poses discomfort and increased infection risks. The advent of non-invasive devices like Masimo's Rad-67TM,...

TAM receptors mediate the Fpr2-driven pain resolution and fibrinolysis after nerve injury

Nerve injury causes neuropathic pain and multilevel nerve barrier disruption. Nerve barriers consist of perineurial, endothelial and myelin barriers. So far, it is unclear whether resealing nerve barriers fosters pain resolution and recovery. To this end, we analysed the nerve barrier property portfolio, pain behaviour battery and lipidomics for precursors of specialized pro-resolving meditators (SPMs) and their receptors in chronic constriction injury of the rat sciatic nerve to identify...

Comparative Methods for Demystifying Spatial Transcriptomics

Spatial Transcriptomics (ST), coined as the term for parallel RNA-Seq on cell populations ordered spatially on a histological tissue section, has recently become increasingly popular, especially in experiments where microfluidics-based single-cell sequencing fails, such as assays on neurons. ST platforms, like the 10x Visium technology investigated herein, therefore produce in a single experiment simultaneously thousands of RNA readouts, captured by an array of micrometer scale spots under the...

Reliability of continuous vital sign monitoring in post-operative patients employing consumer-grade fitness trackers: A randomised pilot trial CONCLUSION: According to the results of the present study, tracker devices are generally reliable and accurate for HR monitoring, whereas SpO(2) and RR measurements should be interpreted carefully, considering the clinical context of the respective patients.

Evaluating blood oxygen saturation measurements by popular fitness trackers in postoperative patients: A prospective clinical trial

Blood oxygen saturation is an important clinical parameter, especially in postoperative hospitalized patients, monitored in clinical practice by arterial blood gas (ABG) and/or pulse oximetry that both are not suitable for a long-term continuous monitoring of patients during the entire hospital stay, or beyond. Technological advances developed recently for consumer-grade fitness trackers could-at least in theory-help to fill in this gap, but benchmarks on the applicability and accuracy of these...

Accuracy and Systematic Biases of Heart Rate Measurements by Consumer-Grade Fitness Trackers in Postoperative Patients: Prospective Clinical Trial

CONCLUSIONS: Consumer-grade fitness trackers appear promising in hospitalized patients for monitoring heart rate.

The Use of Non-Invasive Continuous Blood Pressure Measuring (ClearSight[®]) during Central Neuraxial Anaesthesia for Caesarean Section-A Retrospective Validation Study

The close monitoring of blood pressure during a caesarean section performed under central neuraxial anaesthesia should be the standard of safe anaesthesia. As classical oscillometric and invasive blood pressure measuring have intrinsic disadvantages, we investigated a novel, non-invasive technique for continuous blood pressure measuring. Methods: In this monocentric, retrospective data analysis, the reliability of continuous non-invasive blood pressure measuring using ClearSight^(®) (Edwards...

Analysis of ovarian transcriptomes reveals thousands of novel genes in the insect vector Rhodnius prolixus

Rhodnius prolixus is a Triatominae insect species and a primary vector of Chagas disease. The genome of R. prolixus has been recently sequenced and partially assembled, but few transcriptome analyses have been performed to date. In this study, we describe the stage-specific transcriptomes obtained from previtellogenic stages of oogenesis and from mature eggs. By analyzing ~ 228 million paired-end RNA-Seq reads, we significantly improved the current genome annotations for 9206 genes. We provide...

Echocardiographic Measurements in a Preclinical Model of Chronic Chagasic Cardiomyopathy in Dogs: Validation and Reproducibility

Background: The failure to translate preclinical results to the clinical setting is the rule, not the exception. One reason that is frequently overlooked is whether the animal model reproduces distinctive features of human disease. Another is the reproducibility of the method used to measure treatment effects in preclinical studies. Left ventricular (LV) function improvement is the most common endpoint in preclinical cardiovascular disease studies, while echocardiography is the most frequently...

Characterization of HPV integration, viral gene expression and E6E7 alternative transcripts by RNA-Seq: A descriptive study in invasive cervical cancer

Scarce data are available on the expression of papillomavirus genome and the frequency of alternatively spliced E6E7 mRNAs in invasive cervical cancer. We carried out a comprehensive characterization of HPV expression by RNA-Seq analysis in 22 invasive cervical cancer with HPV16 or HPV18, characterizing the presence of integrated/episomal viral DNA, the integration sites in human genome and the proportion of alternative splicing products of E6 and E7 genes. The expression patterns suggested the...

Transcriptomic and functional analyses of the piRNA pathway in the Chagas disease vector Rhodnius prolixus

The piRNA pathway is a surveillance system that guarantees oogenesis and adult fertility in a range of animal species. The pathway is centered on PIWI clade Argonaute proteins and the associated small non-coding RNAs termed piRNAs. In this study, we set to investigate the evolutionary conservation of the piRNA pathway in the hemimetabolous insect Rhodnius prolixus. Our transcriptome profiling reveals that core components of the pathway are expressed during previtellogenic stages of oogenesis....

An automated method for detecting alternatively spliced protein domains

MOTIVATION: Alternative splicing (AS) has been demonstrated to play a role in shaping eukaryotic gene diversity at the transcriptional level. However, the impact of AS on the proteome is still controversial. Studies that seek to explore the effect of AS at the proteomic level are hampered by technical difficulties in the cumbersome process of casting forth and back between genome, transcriptome and proteome space coordinates, and the naïve prediction of protein domains in the presence of AS...

Landscape of the spliced leader trans-splicing mechanism in Schistosoma mansoni

Spliced leader dependent trans-splicing (SLTS) has been described as an important RNA regulatory process that occurs in different organisms, including the trematode Schistosoma mansoni. We identified more than seven thousand putative SLTS sites in the parasite, comprising genes with a wide spectrum of functional classes, which underlines the SLTS as a ubiquitous mechanism in the parasite. Also, SLTS gene expression levels span several orders of magnitude, showing that SLTS frequency is not...

The effects of death and post-mortem cold ischemia on human tissue transcriptomes

Post-mortem tissues samples are a key resource for investigating patterns of gene expression. However, the processes triggered by death and the postmortem interval (PMI) can significantly alter physiologically normal RNA levels. We investigate the impact of PMI on gene expression using data from multiple tissues of post-mortem donors obtained from the GTEx project. We find that many genes change expression over relatively short PMIs in a tissuespecific manner, but this potentially confounding...

Comparative Genomics in Homo sapiens

Genomes can be compared at different levels of divergence, either between species or within species. Within species genomes can be compared between different subpopulations, such as human subpopulations from different continents. Investigating the genomic differences between different human subpopulations is important when studying complex diseases that are affected by many genetic variants, as the variants involved can differ between populations. The 1000 Genomes Project collected genome-scale...

Comparative Genomics in Drosophila

Since the pioneering studies of Thomas Hunt Morgan and coworkers at the dawn of the twentieth century, Drosophila melanogaster and its sister species have tremendously contributed to unveil the rules underlying animal genetics, development, behavior, evolution, and human disease. Recent advances in DNA sequencing technologies launched Drosophila into the post-genomic era and paved the way for unprecedented comparative genomics investigations. The complete sequencing and systematic comparison of...

Landscape of X chromosome inactivation across human tissues

X chromosome inactivation (XCI) silences transcription from one of the two X chromosomes in female mammalian cells to balance expression dosage between XX females and XY males. XCI is, however, incomplete in humans: up to one-third of X-chromosomal genes are expressed from both the active and inactive X chromosomes (Xa and Xi, respectively) in female cells, with the degree of 'escape' from inactivation varying between genes and individuals. The extent to which XCI is shared between cells and...

Genetic effects on gene expression across human tissues

Characterization of the molecular function of the human genome and its variation across individuals is essential for identifying the cellular mechanisms that underlie human genetic traits and diseases. The Genotype-Tissue Expression (GTEx) project aims to characterize variation in gene expression levels across individuals and diverse tissues of the human body, many of which are not easily accessible. Here we describe genetic effects on gene expression levels across 44 human tissues. We find that...

Dynamic landscape and regulation of RNA editing in mammals

Adenosine-to-inosine (A-to-I) RNA editing is a conserved post-transcriptional mechanism mediated by ADAR enzymes that diversifies the transcriptome by altering selected nucleotides in RNA molecules. Although many editing sites have recently been discovered, the extent to which most sites are edited and how the editing is regulated in different biological contexts are not fully understood. Here we report dynamic spatiotemporal patterns and new regulators of RNA editing, discovered through an...

The impact of rare variation on gene expression across tissues

Rare genetic variants are abundant in humans and are expected to contribute to individual disease risk. While genetic association studies have successfully identified common genetic variants associated with susceptibility, these studies are not practical for identifying rare variants. Efforts to distinguish pathogenic variants from benign rare variants have leveraged the genetic code to identify deleterious protein-coding alleles, but no analogous code exists for non-coding variants. Therefore,...

Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis

The impact of inherited genetic variation on gene expression in humans is well-established. The majority of known expression quantitative trait loci (eQTLs) impact expression of local genes (cis-eQTLs). More research is needed to identify effects of genetic variation on distant genes (trans-eQTLs) and understand their biological mechanisms. One common trans-eQTLs mechanism is "mediation" by a local (cis) transcript. Thus, mediation analysis can be applied to genome-wide SNP and expression data...

Co-expression networks reveal the tissue-specific regulation of transcription and splicing

Gene co-expression networks capture biologically important patterns in gene expression data, enabling functional analyses of genes, discovery of biomarkers, and interpretation of genetic variants. Most network analyses to date have been limited to assessing correlation between total gene expression levels in a single tissue or small sets of tissues. Here, we built networks that additionally capture the regulation of relative isoform abundance and splicing, along with tissue-specific connections...

Correction: Gene Expansion Shapes Genome Architecture in the Human Pathogen Lichtheimia corymbifera: An Evolutionary Genomics Analysis in the Ancient Terrestrial Mucorales (Mucoromycotina)

[This corrects the article DOI: 10.1371/journal.pgen.1004496.].

Sequence variation between 462 human individuals fine-tunes functional sites of RNA processing

Recent advances in the cost-efficiency of sequencing technologies enabled the combined DNA- and RNA-sequencing of human individuals at the population-scale, making genome-wide investigations of the inter-individual genetic impact on gene expression viable. Employing mRNA-sequencing data from the Geuvadis Project and genome sequencing data from the 1000 Genomes Project we show that the computational analysis of DNA sequences around splice sites and poly-A signals is able to explain several...

Corrigendum: Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases No abstract

A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project

The Genotype-Tissue Expression (GTEx) project, sponsored by the NIH Common Fund, was established to study the correlation between human genetic variation and tissue-specific gene expression in non-diseased individuals. A significant challenge was the collection of high-quality biospecimens for extensive genomic analyses. Here we describe how a successful infrastructure for biospecimen procurement was developed and implemented by multiple research partners to support the prospective collection,...

Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases

Aging is one of the most important biological processes and is a known risk factor for many age-related diseases in human. Studying age-related transcriptomic changes in tissues across the whole body can provide valuable information for a holistic understanding of this fundamental process. In this work, we catalogue age-related gene expression changes in nine tissues from nearly two hundred individuals collected by the Genotype-Tissue Expression (GTEx) project. In general, we find the aging gene...

Sharing and Specificity of Co-expression Networks across 35 Human Tissues

To understand the regulation of tissue-specific gene expression, the GTEx Consortium generated RNA-seq expression data for more than thirty distinct human tissues. This data provides an opportunity for deriving shared and tissue specific gene regulatory networks on the basis of co-expression between genes. However, a small number of samples are available for a majority of the tissues, and therefore statistical inference of networks in this setting is highly underpowered. To address this problem,...

Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome

Accurate prediction of the functional effect of genetic variation is critical for clinical genome interpretation. We systematically characterized the transcriptome effects of protein-truncating variants, a class of variants expected to have profound effects on gene function, using data from the Genotype-Tissue Expression (GTEx) and Geuvadis projects. We quantitated tissue-specific and positional effects on nonsense-mediated transcript decay and present an improved predictive model for this...

Human genomics. The human transcriptome across tissues and individuals

Transcriptional regulation and posttranscriptional processing underlie many cellular and organismal phenotypes. We used RNA sequence data generated by Genotype-Tissue Expression (GTEx) project to investigate the patterns of transcriptome variation across individuals and tissues. Tissues exhibit characteristic transcriptional signatures that show stability in postmortem samples. These signatures are dominated by a relatively small number of genes—which is most clearly seen in blood—though few are...

Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans

Understanding the functional consequences of genetic variation, and how it affects complex human disease and quantitative traits, remains a critical challenge for biomedicine. We present an analysis of RNA sequencing data from 1641 samples across 43 tissues from 175 individuals, generated as part of the pilot phase of the Genotype-Tissue Expression (GTEx) project. We describe the landscape of gene expression across tissues, catalog thousands of tissue-specific and shared regulatory expression...

Analysis of alternative splicing events in custom gene datasets by AStalavista

Alternative splicing (AS) is a eukaryotic principle to derive more than one RNA product from transcribed genes by removing distinct subsets of introns from a premature polymer. We know today that this process is highly regulated and makes up a large part of the differences between species, cell types, and states. The key to compare AS across different genes or organisms is to tokenize the AS phenomenon into atomary units, so-called AS events. These events then usually are grouped by common...

Nova1 is a master regulator of alternative splicing in pancreatic beta cells

Alternative splicing (AS) is a fundamental mechanism for the regulation of gene expression. It affects more than 90% of human genes but its role in the regulation of pancreatic beta cells, the producers of insulin, remains unknown. Our recently published data indicated that the 'neuron-specific' Nova1 splicing factor is expressed in pancreatic beta cells. We have presently coupled specific knockdown (KD) of Nova1 with RNA-sequencing to determine all splice variants and downstream pathways...

Tandem RNA chimeras contribute to transcriptome diversity in human population and are associated with intronic genetic variants

Chimeric RNAs originating from two or more different genes are known to exist not only in cancer, but also in normal tissues, where they can play a role in human evolution. However, the exact mechanism of their formation is unknown. Here, we use RNA sequencing data from 462 healthy individuals representing 5 human populations to systematically identify and in depth characterize 81 RNA tandem chimeric transcripts, 13 of which are novel. We observe that 6 out of these 81 chimeras have been...

Gene expansion shapes genome architecture in the human pathogen Lichtheimia corymbifera: an evolutionary genomics analysis in the ancient terrestrial mucorales (Mucoromycotina)

Lichtheimia species are the second most important cause of mucormycosis in Europe. To provide broader insights into the molecular basis of the pathogenicity-associated traits of the basal Mucorales, we report the full genome sequence of L. corymbifera and compared it to the genome of Rhizopus oryzae, the most common cause of mucormycosis worldwide. The genome assembly encompasses 33.6 MB and 12,379 protein-coding genes. This study reveals four major differences of the L. corymbifera genome to R....

RNA sequencing identifies dysregulation of the human pancreatic islet transcriptome by the saturated fatty acid palmitate

Pancreatic -cell dysfunction and death are central in the pathogenesis of type 2 diabetes (T2D). Saturated fatty acids cause -cell failure and contribute to diabetes development in genetically predisposed individuals. Here we used RNA sequencing to map transcripts expressed in five palmitate-treated human islet preparations, observing 1,325 modified genes. Palmitate induced fatty acid metabolism and endoplasmic reticulum (ER) stress. Functional studies identified novel mediators of adaptive ER...

Assessment of transcript reconstruction methods for RNA-seq

We evaluated 25 protocol variants of 14 independent computational methods for exon identification, transcript reconstruction and expression-level quantification from RNA-seq data. Our results show that most algorithms are able to identify discrete transcript components with high success rates but that assembly of complete isoform structures poses a major challenge even when all constituent elements are identified. Expression-level estimates also varied widely across methods, even when based on...

Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories

RNA sequencing is an increasingly popular technology for genome-wide analysis of transcript sequence and abundance. However, understanding of the sources of technical and interlaboratory variation is still limited. To address this, the GEUVADIS consortium sequenced mRNAs and small RNAs of lymphoblastoid cell lines of 465 individuals in seven sequencing centers, with a large number of replicates. The variation between laboratories appeared to be considerably smaller than the already limited...

Transcriptome and genome sequencing uncovers functional variation in humans

Genome sequencing projects are discovering millions of genetic variants in humans, and interpretation of their functional effects is essential for understanding the genetic basis of variation in human traits. Here we report sequencing and deep analysis of messenger RNA and microRNA from lymphoblastoid cell lines of 462 individuals from the 1000 Genomes Project--the first uniformly processed high-throughput RNA-sequencing data from multiple human populations with high-quality genome sequences. We...

The Genotype-Tissue Expression (GTEx) project

Genome-wide association studies have identified thousands of loci for common diseases, but, for the majority of these, the mechanisms underlying disease susceptibility remain unknown. Most associated variants are not correlated with protein-coding changes, suggesting that polymorphisms in regulatory regions probably contribute to many disease phenotypes. Here we describe the Genotype-Tissue Expression (GTEx) project, which will establish a resource database and associated tissue bank for the...

The GEM mapper: fast, accurate and versatile alignment by filtration

Because of ever-increasing throughput requirements of sequencing data, most existing short-read aligners have been designed to focus on speed at the expense of accuracy. The Genome Multitool (GEM) mapper can leverage string matching by filtration to search the alignment space more efficiently, simultaneously delivering precision (performing fully tunable exhaustive searches that return all existing matches, including gapped ones) and speed (being several times faster than comparable...

Modelling and simulating generic RNA-Seq experiments with the flux simulator

High-throughput sequencing of cDNA libraries constructed from cellular RNA complements (RNA-Seq) naturally provides a digital quantitative measurement for every expressed RNA molecule. Nature, impact and mutual interference of biases in different experimental setups are, however, still poorly understood-mostly due to the lack of data from intermediate protocol steps. We analysed multiple RNA-Seq experiments, involving different sample preparation protocols and sequencing platforms: we broke them...

Landscape of transcription in human cells

Eukaryotic cells make many types of primary and processed RNAs that are found either in specific subcellular compartments or throughout the cells. A complete catalogue of these RNAs is not yet available and their characteristic subcellular localizations are also poorly understood. Because RNA represents the direct output of the genetic information encoded by genomes and a significant proportion of a cell's regulatory capabilities are focused on its synthesis, processing, transport, modification...

An integrated encyclopedia of DNA elements in the human genome

The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions. Many discovered candidate regulatory...

The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines

Type 1 diabetes (T1D) is an autoimmune disease in which pancreatic beta cells are killed by infiltrating immune cells and by cytokines released by these cells. Signaling events occurring in the pancreatic beta cells are decisive for their survival or death in diabetes. We have used RNA sequencing (RNA-seq) to identify transcripts, including splice variants, expressed in human islets of Langerhans under control conditions or following exposure to the pro-inflammatory cytokines interleukin-1...

Evaluating characteristics of de novo assembly software on 454 transcriptome data: a simulation approach

CONCLUSION: Our evaluation of four assemblers suggested that MIRA and Newbler slightly outperformed the other programs, while showing contrasting characteristics. Oases did not perform very well on the 454 reads. Our evaluation indicated that the software was either conservative (MIRA) or liberal (Newbler) about merging reads into contigs. This suggested that in choosing an assembly program researchers should carefully consider their follow up analysis and consequences of the chosen approach to...

Estimation of alternative splicing variability in human populations

DNA arrays have been widely used to perform transcriptome-wide analysis of gene expression, and many methods have been developed to measure gene expression variability and to compare gene expression between conditions. Because RNA-seq is also becoming increasingly popular for transcriptome characterization, the possibility exists for further quantification of individual alternative transcript isoforms, and therefore for estimating the relative ratios of alternative splice forms within a given...

A user's guide to the encyclopedia of DNA elements (ENCODE)

The mission of the Encyclopedia of DNA Elements (ENCODE) Project is to enable the scientific and medical communities to interpret the human genome sequence and apply it to understand human biology and improve health. The ENCODE Consortium is integrating multiple technologies and approaches in a collective effort to discover and define the functional elements encoded in the human genome, including genes, transcripts, and transcriptional regulatory regions, together with their attendant chromatin...

Transcriptome genetics using second generation sequencing in a Caucasian population

Gene expression is an important phenotype that informs about genetic and environmental effects on cellular state. Many studies have previously identified genetic variants for gene expression phenotypes using custom and commercially available microarrays. Second generation sequencing technologies are now providing unprecedented access to the fine structure of the transcriptome. We have sequenced the mRNA fraction of the transcriptome in 60 extended HapMap individuals of European descent and have...

Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing

Technical advances such as the development of molecular cloning, Sanger sequencing, PCR and oligonucleotide microarrays are key to our current capacity to sequence, annotate and study complete organismal genomes. Recent years have seen the development of a variety of so-called 'next-generation' sequencing platforms, with several others anticipated to become available shortly. The previously unimaginable scale and economy of these methods, coupled with their enthusiastic uptake by the scientific...

Complete alternative splicing events are bubbles in splicing graphs

Eukaryotic splicing structures are known to involve a high degree of alternative forms derived from a premature transcript by alternative splicing (AS). With the advent of new sequencing technologies, evidence for new splice forms becomes increasingly available-bit by bit revealing that the true splicing diversity of "AS events" often comprises more than two alternatives and therefore cannot be sufficiently described by pairwise comparisons as conducted in analyzes hitherto. Especially, I...

Nucleosome positioning as a determinant of exon recognition

Chromatin structure influences transcription, but its role in subsequent RNA processing is unclear. Here we present analyses of high-throughput data that imply a relationship between nucleosome positioning and exon definition. First, we have found stable nucleosome occupancy within human and Caenorhabditis elegans exons that is stronger in exons with weak splice sites. Conversely, we have found that pseudoexons--intronic sequences that are not included in mRNAs but are flanked by strong splice...

A general definition and nomenclature for alternative splicing events

Understanding the molecular mechanisms responsible for the regulation of the transcriptome present in eukaryotic cells is one of the most challenging tasks in the postgenomic era. In this regard, alternative splicing (AS) is a key phenomenon contributing to the production of different mature transcripts from the same primary RNA sequence. As a plethora of different transcript forms is available in databases, a first step to uncover the biology that drives AS is to identify the different types of...

Based Upon Repeat Pattern (BURP): an algorithm to characterize the long-term evolution of Staphylococcus aureus populations based on spa polymorphisms

CONCLUSION: BURP is the first automated and objective tool to infer clonal relatedness from spa repeat regions. It is able to extract an evolutionary signal rather congruent to MLST and micro-array data.

ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets

In the process of establishing more and more complete annotations of eukaryotic genomes, a constantly growing number of alternative splicing (AS) events has been reported over the last decade. Consequently, the increasing transcript coverage also revealed the real complexity of some variations in the exon-intron structure between transcript variants and the need for computational tools to address 'complex' AS events. ASTALAVISTA (alternative splicing transcriptional landscape visualization tool)...

Comparing tandem repeats with duplications and excisions of variable degree

Traditional sequence comparison by alignment employs a mutation model comprised of two events, substitutions and indels (insertions or deletions) of single positions. However, modern genetic analysis knows a variety of more complex mutation events (e.g., duplications, excisions, and rearrangements), especially regarding DNA. With ever more DNA sequence data becoming available, the need to accurately compare sequences which have clearly undergone more complicated types of mutational processes is...

Global multiple-sequence alignment with repeats

Repeating fragments in biological sequences are often essential for structure and function. Over the years, many methods have been developed to recognize repeats or to multiply align protein sequences. However, the integration of these two methodologies has been largely unexplored to date. Here, we present a new method capable of globally aligning multiple input sequences under the constraints of a given repeat analysis. The method supports different stringency modes to adapt to various levels...

Panta rhei (QAlign2): an open graphical environment for sequence analysis

MOTIVATION: The first version of the graphical multiple sequence alignment environment QAlign was published in 2003. Heavy response from the molecular-biological user community clearly demonstrated the need for such a platform.

RIDOM: comprehensive and public sequence database for identification of Mycobacterium species

CONCLUSIONS: The data from this analysis show that it is possible to differentiate most mycobacterial species by sequence analysis of partial 16S rDNA. The high-quality sequences reported here, together with ancillary information (e.g., taxonomic, medical), are available in a public database, which is currently being expanded in the RIDOM project http://www.ridom-rdna.de), for similarity searches.

Divide-and-conquer multiple alignment with segment-based constraints

A large number of methods for multiple sequence alignment are currently available. Recent benchmarking tests demonstrated that strengths and drawbacks of these methods differ substantially. Global strategies can be outperformed by approaches based on local similarities and vice versa, depending on the characteristics of the input sequences. In recent years, mixed approaches that include both global and local features have shown promising results. Herein, we introduce a new algorithm for multiple...

QAlign: quality-based multiple alignments with dynamic phylogenetic analysis Integrating different alignment strategies, a layout editor and tools deriving phylogenetic trees in a 'multiple alignment environment' helps to investigate and enhance results of multiple sequence alignment by hand. QAlign combines algorithms for fast progressive and accurate simultaneous multiple alignment with a versatile editor and a dynamic phylogenetic analysis in a convenient graphical user interface.