<table>
<thead>
<tr>
<th>Title</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERV evolution - A bioinformatics pipeline to investigate retroviral integration in museum koala samples</td>
<td>The pipeline comprises three main stages: 1) sequencing errors using de Bruijn graphs, 2) defining virus and bacteria concentrations used to evaluate the method, and 3) combined RNA and DNA sequencing of native and host depleted surrogate and patients' CSF samples was run using IonTorrent™ and Illumina® technology.</td>
</tr>
</tbody>
</table>
A Novel Method to identify Significant DNA motifs in the human genome associated with virus sequences.

Background: The discovery of virus sequences is essential to support virological surveillance and patient care. The “Wing-flying framework,” an automated classification method that exploits the patterns associated with virus sequences and discriminates them from other motifs and sites, has been identified and investigated in previous studies. The method has been applied to identify significant virus sequences in the human genome.

Methods: We have developed a new approach for identifying significant virus sequences. This method uses a combination of statistical analysis and machine learning techniques. The approach involves mapping the sequences to the human genome, estimating the background frequency of the sequences, and then identifying statistically significant sequences.

Results: We applied our method to the human genome and identified a large number of significant virus sequences. The identified sequences included known virus sequences as well as novel sequences not previously identified.

Conclusion: Our method provides a new and effective approach for identifying significant virus sequences in the human genome. This approach can be used to support virological surveillance and patient care.

Methods:

1. Mapping: The sequences were mapped to the human genome using a string alignment approach.
2. Background frequency estimation: The background frequency of the sequences was estimated using a statistical model.
3. Identification: Significant sequences were identified using a statistical test.

Results:

1. Known sequences: Known virus sequences were included in the identified sequences.
2. Novel sequences: Novel sequences not previously identified were also included.

Conclusion: The method provides a new and effective approach for identifying significant virus sequences in the human genome.
Comparing alignment and assembly

Although the advancements in high-throughput sequencing technologies are increasing the sequencing depth and speed, this enables concomitant full-length transcriptome and genome recovery of diverse microbial communities, the comparison of methods used to align and assemble the reads is still relevant. Several methods have been proposed to align and assemble the reads, but the choice of method depends on the data type and the characteristics of the organism. The study aimed to compare the performance of different alignment and assembly strategies with respect to runtime and quality scores. The analysis was performed on data which were sequenced on the Illumina platform.

Different approaches were used to align the reads, including mapping to reference genomes, de novo assembly, and hybrid approaches. The study compared several alignment methods, such as BWA, Bowtie, and Bowtie2, and several assembly methods, such as SPades, IDBA-UD, and Canu.

The results showed that Bowtie2 and SPades performed better than the other methods in terms of runtime and quality scores. Bowtie2 was the fastest method, while SPades produced the highest quality assemblies. The study also showed that the choice of alignment method had a significant impact on the quality of the assemblies.

In conclusion, the study highlighted the importance of choosing the appropriate alignment and assembly method for the specific data type and organism. The results also underscored the importance of evaluating the performance of different methods in terms of both runtime and quality scores.
Fatemeh Behjati Ardakani, Marita A. Isokallio and Jerzy Tiuryn

Epigenetic marks of the chromatin 3D evolutionarily conserved regulatory elements

We present a comparative evaluation of metagenomic analysis methods in which we use sequence simulators to generate gold-standard data against which to benchmark the efficacy of routine use. We describe a Bayesian network classifier to discover causative link between chromatin marks and loop distinct. In this work, we formalize the problem of enzyme selection for misassembly detection, suggest suffix array algorithmic solutions, and analyze their computational complexity.

WES as well as RNA-seq data. We propose to use this opportunity for an efficient and reliable consistency check for human HTS data, as it allows to match samples and detect sample swaps. Compared to conventional approaches for swap detection based on matching SNVs our approach has several major advantages: first it is easily applicable and reliable even for whole-mt-genome level, and by revealing purifying selection also in the soma. With the improved protocol, we will clarify the developmental timing of purifying selection in the mouse.

We present a new comprehensive tool including steps from read mapping to accurate differential expression of both entire cellular populations and single cells. The majority of RNA-Seq analyses begin by mapping each experimentally produced sequence (i.e., read) to a set of annotated coding genome. The occurrence of motifs within a window length. While this tool gave us interesting predictions of known and novel regulatory elements it was very slow in operation as the scoring function tool, called Billboard, for detection of such elements using a sliding window approach and a scoring function penalizing non-matching motif occurrences between species and rewarding co-

The increasing quantity of ATAC-seq gives the information about chromatin marks and DNA accessibility. We propose a Bayesian network classifier to discover causative link between chromatin marks and loop distinct. In this work, we formalize the problem of enzyme selection for misassembly detection, suggest suffix array algorithmic solutions, and analyze their computational complexity.

The increasing quantity of ATAC-seq gives the information about chromatin marks and DNA accessibility. We propose a Bayesian network classifier to discover causative link between chromatin marks and loop distinct. In this work, we formalize the problem of enzyme selection for misassembly detection, suggest suffix array algorithmic solutions, and analyze their computational complexity.

Genomes

P5061

P5063

P5060

P5058

P5057

P5056

P5055

P5054

P5053

P5052

P5051

P5049

P5048

P5047

P5046

P5045

P5044

P5043

P5042

P5041

P5040

P5039

P5038

Claudia Calabrese, Nuno Exploiting Next Generation Sequencing to
Genome sequencing and assembly of the
Leon Kuchenbecker, Knut Shay Ben-Elazar, Benny de Castro, Delphine
Fanin, Alessandro den Broek, Melanie Jankowski, Jan Korbel,
Finkers and Dick de Ridder

rearrangements on chromatin organization
medium-sized deletions in clinical application
resolution contact maps that we use to position topological domains in relation to each other. Then we take the advantage of the ChIA-PET specificity that allows to target a particular protein
based on ChIA-PET data. We base our modeling on underlying biological structures, i.e. chromatin loops and topological domains. First, we employ the weak interactions to create the low-
Spatial organization of the genome plays an important role in its functioning and is closely related to gene expression level, DNA replication and repair and others. The basic units of this

routinely provides annotation of Swiss-Prot quality for millions of unreviewed protein sequences in UniProtKB/TrEMBL. In addition, HAMAP can be used directly for the annotation of

very low frequency and escaped identification due to the sensitivity limitations of WES, we used Duplex Sequencing to identify mutations at very low variant allele frequency (<1/10000).

36-fold reduction in the number of contigs and a 18-fold increase in the N50 in comparison to a previous short-read assembly of CENPK113-7D. Interestingly, we show

CEN.PK113-7D is a haploid strain of Saccharomyces cerevisiae that is used widely in biotechnology because of its robust growth characteristics in industrial settings. Although previous

reported for S. halodurans followed by 71 for Salinicoccus sp BAB_3246 and 46 for S. carnicancri Crm strain. Total 27 subsystem annotations were resulted from the RAST based annotation

by 1691,863, 668, 449 and 334 correspondingly for Salinicoccus sp BAB_3246, S. carnicancri Crm, S. luteus DSM 17002, S. roseus and S.albus DSM 19776 strain. Maximum 73 RNAs was

S.carnicancri Crm, Salinicoccus sp BAB_3246, S. luteus DSM 17002, S. roseus and S.albus DSM 19776 strain. Maximum 2839 coding sequences were reported for S. halodurans followed

efforts to exploit NGS-based TCR profiling for the characterization of antigen specificity and clinical classification applications.

Adaptive immunity is driven by a highly diverse population of T and B cells expressing unique antigen receptor proteins. The genetic mechanism allowing for this diversity is the somatic

regulation in cis.

expression in adult flies as well as embryos and compared the rearranged chromosomes to their normal state in a heterozygous cross, which intrinsically normalizes for trans regulatory

etc.) at a genome-wide scale. Yet our understanding of how these interactions form and under which circumstances they regulate gene expression is only rudimentary. Recent studies

involved in circRNA biogenesis. FUCHS is an easy to use python-based pipeline that contributes to new aspects of the circRNA research.

Circular RNAs (circRNAs) belong to a recently re-discovered class of RNA species that emerge during RNA maturation by a process called back-splicing. Circular transcripts, as opposed to

point at transcription factor targets nuclear co-localization.

Circular RNAs (circRNAs) are non-coding RNAs that are generated by a process called back-splicing. Accumulating evidence in the recent years shows that circRNAs take part in various

motifs in the circRNA sequences. This is the first study that uses a similarity-based approach to comprehensively analyze the non-coding RNA transcripts of D. melanogaster.

The extra features provided by FUCHS enable the user to perform differential motif enrichment and miRNA seed analysis to determine potential regulators

regulated and not a mere by-product of splicing. Though functional studies have shown that some circRNAs could act as miRNA-sponges, the function of most circRNAs remains unknown. To

Red Blood Cells

Expression Quantitative Trait Locus QTL (eQTL) studies represent a key tool to understand the effects of genomic variation on gene expression levels. Here we present some preliminary

results of the eQTL analysis carried out within the frame of the PanCancer project, an international collaborative effort to annotate similarities and differences between 30 different cancer

types. Whole Genome Sequencing, with both germline and somatic calls, and matched tumour RNA-seq data from more than 1000 TCGA and ICGC cancer patients are available to this

results of the eQTL analysis carried out within the frame of the PanCancer project, an international collaborative effort to annotate similarities and differences between 30 different cancer

Expression Quantitative Trait Locus QTL (eQTL) studies represent a key tool to understand the effects of genomic variation on gene expression levels. Here we present some preliminary

results of the eQTL analysis carried out within the frame of the PanCancer project, an international collaborative effort to annotate similarities and differences between 30 different cancer

"Apricot and white wine are a great pair in food and wine matching. We recommend our customers to try this combination and share their feedback with us."
identifying the functional role of transcription factors by a novel method.

Recent advances have been focused on the identification of genetic factors in diseases. However, in the field of molecular biology, there is still no single method that is universally accepted for the analysis of genetic variations. The main focus of this study is to identify genetic factors that are associated with specific diseases, such as neurodegenerative diseases, cancer, and cardiovascular diseases. We have developed a novel bioinformatics pipeline that integrates large-scale genomic data from various sources to identify genetic factors associated with these diseases. This pipeline includes the identification of genetic variants, the association of these variants with specific diseases, and the functional validation of identified genetic factors.

The pipeline consists of three main steps: (1) genetic variant calling, (2) association analysis, and (3) functional validation. In the first step, we use a combination of short-read and long-read sequencing technologies to identify genetic variants in the human genome. We then use statistical methods to associate these variants with specific diseases. Finally, we use functional validation techniques to confirm the biological relevance of identified genetic factors.

Our pipeline has been applied to several large-scale genomic datasets, including the Exome Aggregation Consortium (ExAC), the Genome Aggregation Database (GnomAD), and the International Cancer Genome Consortium (ICGC). The results have shown that our pipeline is able to identify a large number of genetic factors associated with specific diseases. Furthermore, we have found that our pipeline is able to identify genetic factors that are not detected by other methods, such as deep sequencing and copy number variation analysis.

In conclusion, our pipeline is a powerful tool for identifying genetic factors associated with specific diseases. It has the potential to revolutionize the field of molecular biology by identifying new genetic factors that contribute to disease susceptibility and pathogenesis. We believe that our pipeline will be widely adopted by the scientific community and will greatly advance our understanding of genetic factors in diseases.
Transcription factor (TF) binding sites represent critical regulatory elements that drive gene expression. They are essential for coordinating cellular processes and are involved in various biological pathways, such as development, differentiation, and disease. TF binding sites can be identified using a variety of computational and experimental methods. One approach is through the use of machine-learning algorithms that can predict TF binding sites based on sequence features and functional annotations.

The No Promoter Left Behind (NPLB) method is an example of such an approach. NPLB is an unsupervised machine-learning method that partitions promoters into diverse architectures while identifying regulatory signals associated with any DNA-specified biological event reported at high-resolution. NPLB opens up avenues to learn new biology from high-throughput data, which is particularly useful in the era of next-generation sequencing.

In this study, we showcase the designed workflow on samples from the University Hospital Zurich. In collaboration with hospital oncologists, researchers at ETH Zurich, and the Swiss Tumor Board, we present a new approach to predict tumor trajectories based on TF binding sites. The method integrates clinical and genomic data to provide insights into the molecular mechanisms underlying cancer progression.
Predicting oligogenic effects using digenic association analyses

Oligogenic effects refer to the simultaneous involvement of several genetic loci in the etiology of a phenotype. While single-locus associations can be identified through genome-wide association studies (GWAS), the contribution of multiple loci is often underestimated due to the complex nature of many diseases. In this study, we aimed to predict oligogenic effects by analyzing digenic associations, which involve the simultaneous presence of two genetic loci in an individual.

We employed a statistical approach based on the calculation of genetic correlations and the estimation of disease risk. Our method takes into account the joint effect of two genetic variants and assesses their combined impact on disease susceptibility. We utilized a large cohort of patients with a specific disease, genotyped for a large number of polymorphisms, to identify significant digenic associations.

Our findings revealed several digenic associations that were strongly associated with the disease phenotype. These associations were confirmed through follow-up studies, and we observed that the combined effect of the two genetic variants was significantly greater than the sum of their individual effects. This suggests the presence of an epistatic effect, where the interaction between the two loci plays a crucial role in the disease development.

The identification of digenic associations can provide insights into the complex genetic architecture of diseases, highlighting the importance of considering multiple genetic loci in the interpretation of genetic studies. Our findings suggest that this approach could be a valuable tool for understanding the genetic basis of many complex diseases, where the contribution of multiple loci is crucial for accurate risk assessment and personalized medicine.

This study contributes to the growing body of knowledge on the genetic basis of diseases, emphasizing the importance of considering the joint effect of multiple genetic loci. Further research is needed to validate these findings and to explore the biological mechanisms underlying the identified digenic associations.
transcription factors – histones interplay in cellular disease regulation, modulating adaptive and stress responses. Recent studies have shed light on the molecular mechanisms by which histones and transcription factors interact, both directly and indirectly, influencing gene expression and cellular functions. These interactions play a critical role in various biological processes, such as cell differentiation, proliferation, and response to environmental stresses. Understanding the interplay between transcription factors and histones is crucial for elucidating disease mechanisms and developing therapeutic strategies.

DNA-Protein Binding Prediction

Transcription factors (TFs) are crucial regulators of transcription and gene expression. They bind to specific DNA sequences, called TF binding sites (TFBSs), to initiate gene expression. The identification of TFBSs is essential for understanding gene regulation and has significant implications for biomedical research. DNA-Protein Binding Prediction (DPBP) is an approach that aims to predict the binding sites of transcription factors from genomic sequences.

DNA-Protein Binding Prediction (DPBP)

DPBP is a computational method that predicts TFBSs by analyzing DNA sequences. It involves mapping the binding locations of TFs to DNA, which can be achieved through various experimental techniques such as chromatin immunoprecipitation (ChIP) and genomic footprints. The predicted TFBSs are then validated using biological assays and comparison with experimental data. This approach is widely used in genomics and bioinformatics for understanding the transcriptional regulation of genes and for identifying potential therapeutic targets.

TAPscan – An updated genome-wide transcription factor classification workflow

The TAPscan method, developed by Schmitz et al., is a powerful tool for identifying transcription factor binding sites in genomic sequences. TAPscan is an updated version of the genome-wide transcription factor classification workflow that has been widely used in the field. It includes improvements in the scoring function, increased accuracy, and enhanced ability to distinguish between different transcription factor families.

GeMoMa: A novel homology-based gene prediction method

GeMoMa is a novel method for predicting gene models in new genomes. It is based on a homology-based approach, comparing the novel genome to a related species to infer gene models. GeMoMa is particularly useful in predicting gene models in new genomes where existing functional annotations are lacking. It outperforms state-of-the-art methods, especially in challenging cases where the target genome has a low similarity to the reference genomes.

Network-based TSS prediction

The TSS prediction problem involves identifying the transcription start sites (TSSs) of genes. TSS prediction is a crucial step in understanding gene regulation and has implications for disease research. Various methods have been developed to predict TSSs, including machine learning approaches, graph-based methods, and network-based approaches. Network-based methods leverage the relationships between genes and their expression patterns to infer TSSs, providing a comprehensive view of the transcriptional regulatory network.

Next generation sequencing has led to a rapid increase in the number of sequenced genomes. Initial annotation of protein-coding genes in newly sequenced genomes is typically based on homology-based gene prediction methods. However, recent studies have shown that these methods may lack accuracy, especially in predicting genes that are not conserved between species. GeMoMa provides a novel solution to this challenge by incorporating additional information to improve gene prediction accuracy.

Functional genomics

Functional genomics is a field that focuses on understanding the functions of genes and regulatory elements in biological processes. It involves the use of high-throughput sequencing technologies to analyze genome-wide changes in gene expression, DNA methylation, and histone modifications. Functional genomics has played a crucial role in elucidating the molecular mechanisms underlying various diseases and has significant implications for personalized medicine.

Epigenetic control plays an important role in regulation of SM gene clusters. However, it is not yet shown if nucleosome positioning is the main mechanism underlying the early BMP4 response in breast cancer cell lines. Our study aims to uncover the early BMP4 regulatory target genes and adaptation. TSK maps sequence patterns to a high-dimensional feature space using the discriminative mismatch string kernel framework under SVM. Labeled examples from a source domain are mapped to the feature space and serve as a proxy for the target distribution, which would be more expensive and time-consuming to generate. In our study, we used a cross-organism setting, where we predicted the target domain-specific TSK scores.

Histone mark variations

Histone marks are modifications that occur on the histone proteins and provide additional information for DNA transcription. They can be classified into two main groups: histone modifications and DNA modifications. Histone modifications include methylation, acetylation, and deamination, which alter the accessibility of DNA to transcription factors. DNA modifications include cytosine methylation, which can lead to gene silencing.

Microorganism fermentation

Microorganisms that ferment the sucrose mainly into lactic acid, acetic acid, and ethanol. In this study, the species diversity of the water kefir microbiota was analysed using shotgun metagenomics. The results revealed a high diversity of microorganisms present in the samples, with Lactobacillus and Enterococcus being the most abundant genera. The availability of comprehensive genome databases and computational tools has substantially expanded our knowledge on metabolic interactions between HGM as well as interactions between the HGM and the host organism.

Study population unique in its genetic characteristics

The study population unique in its genetic characteristics, and represents an excellent opportunity to identify novel alleles regulating metabolite abundance in blood. We profiled 1,949 plasma metabolites.

In this study, the species diversity of the water kefir microbiota was analysed using shotgun metagenomics. The results revealed a high diversity of microorganisms present in the samples, with Lactobacillus and Enterococcus being the most abundant genera. The availability of comprehensive genome databases and computational tools has substantially expanded our knowledge on metabolic interactions between HGM as well as interactions between the HGM and the host organism.