



Utrecht
Bioinformatics
Center

Utrecht Life Sciences

The 10th UBC Annual Symposium

Monday, October 7, 2024

6th Floor | Princess Máxima Center for Pediatric Oncology |
Heidelberglaan 25 | 3584 CS | Utrecht

[Programme](#)

[Keynote Speakers](#)

[Pitch Talks](#)

[Poster Presentations](#)



Programme

08:30 – 09:00 Registration

09:00 – 09:15 Welcome by the **UBC Executive Committee** and **Prof. dr. Isabel Arends** (Chair of the Life Sciences Deans)

09:15 – 10:00 Keynote talk [*"Detection of shared introns in ancient gene duplicates to shed light on the deep origins of eukaryotic genome complexity"*](#) by **Prof. dr. Berend Snel** (Utrecht University)

10:00 – 10:30 Coffee break and [poster presentations](#)

10:30 – 12:00 Pitch talks **"Sequencing and Bioinformatics"** chaired by Mark Bakker (UMCU)

- [*"Multiomic measurements of tRNAs and ribosome extension dynamics in single cells"*](#) by Mees van der Ent (Hubrecht Institute)
- [*"Searching for robust protein biomarkers in human plasma"*](#) by Isabel Houtkamp (UU)
- [*"NanoRCS: Multimodal tumor cell-free DNA profiling using nanopore-based consensus sequencing"*](#) by Li-Ting Chen (UMCU)
- [*"Comparative transcriptomics reveals a conserved immune response in six mushroom-forming fungi during interaction with their competitors"*](#) by Marieke van Maanen (UU)
- [*"Fungal biocontrol agents for Striga weed eradication"*](#) by Dr. Khyati Mheta Bhatt (Westerdijk Institute)
- [*"Bioinformatic approaches to decipher lettuce innate immunity"*](#) by Iñigo Bañales Belaunde (UU)

12:00 – 13:30 Walking lunch and [poster presentations](#)

13:30 – 14:15 Keynote talk [*"Tools and discoveries from long read transcriptome sequencing"*](#) by **Prof. dr. Ana Conesa** (Spanish National Research Council)

14:15 – 15:45 Pitch talks **"Modelling and Machine Learning"** chaired by Jip van Dinter (PMC)

- [*"Quantitative modelling of fate specification in the C. elegans postembryonic M lineage reveals a missing spatiotemporal signal"*](#) by Dr. Benjamin Planterose Jimenez (UU)
- [*"Decomposing plasmid evolution using a simple pangenomic language model"*](#) by Dr. Weizhen Xu (UMCU)
- [*"Methodology for Biomarker Discovery with Reproducibility in Microbiome Data using Machine Learning"*](#) by David Rojas-Velazquez (UU)
- [*"M&M: An RNA-seq based Pan-Cancer Classifier for Pediatric Tumors"*](#) by Fleur Wallis (PMC)
- [*"Exploring the diversity of fungal plant biomass conversion"*](#) by Dr. Jiajia Li (Westerdijk Institute)
- [*"RNA splicing as novel immune regulator"*](#) by Dr. Alejandra Bodelon (UMCU)

15:45 – 16:15 Break and [poster presentations](#)

16:15 – 17:00 Keynote talk [*"Why is it so hard to analyze single-cell RNA-seq data?"*](#) by **Prof. dr. Erik van Nimwegen** (University of Basel)

17:00 – 18:00 Closing ceremony followed by drinks

Keynote Speakers

Detection of shared introns in ancient gene duplicates to shed light on the deep origins of eukaryotic genome complexity

In this lecture we will discuss two things. First I will shortly reflect on the past ten years in bioinformatics with a focus on the Utrecht Science Park and the UBC. Subsequently we will discuss our recent investigations in utilizing spliceosomal introns to unlock one of the biggest questions in comparative genomics. Spliceosomal introns are a unique feature of eukaryotic genes. Previous studies have established that many introns were present in the protein-coding genes of the last eukaryotic common ancestor (LECA). Intron positions shared between genes that duplicated before LECA could in principle provide insight into the emergence of the first introns. In this study we use ancestral intron position reconstructions in two large sets of duplicated families to systematically identify these ancient paralogous intron positions. We found that 20-35% of introns inferred to have been present in LECA were shared between paralogs. These shared introns, which likely preceded ancient duplications, were wide spread across different functions, with the notable exception of nuclear transport. Since we observed a clear signal of pervasive intron loss prior to LECA, it is likely that substantially more introns were shared at the time of duplication than we can detect in LECA. The large extent of shared introns indicates an early origin of introns during eukaryogenesis and suggests an early origin of a nuclear structure, before most of the other complex eukaryotic features were established.



Prof. dr. Berend Snel

Group leader of the Evolutionary Genomics and Integrative Bioinformatics group, Theoretical Biology and Bioinformatics, Department of Biology, Utrecht University

Berend Snel studied biology at the university of Utrecht specialising in Theoretical Biology and Bioinformatics. He did his PhD project on comparative genome analysis and genome evolution in the group of Peer Bork at EMBL (Heidelberg,

Germany). Subsequently he first worked as a normal post-doc and then as a VENI post-doc in Martijn Huynen's Comparative Genomics group at the CMBI / NCMLS in Nijmegen, Holland. In 2006 he moved to Utrecht University as an Associate Professor. In Utrecht he leads the Evolutionary Genomics and Integrative Bioinformatics group as part of Theoretical Biology and Bioinformatics in the Department of Biology. As of June 2014, he was appointed Professor in Bioinformatics and chair of the Executive Committee of the Utrecht Bioinformatics Center. In 2016 Berend Snel was awarded a VICI grant. Berend has co-authored more than 110 peer reviewed papers which have been cited more than 15.000 times resulting in an H-index of 52.

Tools and discoveries from long read transcriptome sequencing

Long-read sequencing (LRS) technologies have firmly established themselves as viable alternatives to short-read methods in various applications. Over the past five years, LRS platforms such as Oxford Nanopore and Pacific Biosciences have significantly improved in accuracy and throughput, becoming realistic alternatives to Illumina RNA-seq for transcriptomics experiments. Our lab's pioneering work led to the creation of SQANTI, a widely used tool for the quality control and quantification of long-read transcript models. However, as technologies advance, the need for additional algorithms to support transcriptomics applications of LRS has grown. This includes not only accurate transcript identification and quantification but also analysis and correction of technological biases, experimental design considerations for large experiments, benchmarking resources, data visualization, and functional annotation of full-length transcript variants.

I will present the current suite of SQANTI tools, which provide a comprehensive solution to these data processing needs. For benchmarking, SQANTI-SIM is unique for simulating long-read and orthogonal data with precise control over transcript novelty. This enables the assessment of the identification of both annotated and novel transcript models. BUGSI is a set of universal single-isoform genes that can serve as an internal standard to troubleshoot RNA degradation and library preparation errors. The raw data quality in multi-sample experiments is assessed with SQANTI-reads, which identifies outliers, technology biases, and evaluates whether the data meets quality standards for discovery. SQANTI3 evaluates the performance of transcript reconstruction algorithms to accurately identify transcript models from long reads. The Filter, Rescue, and Requant modules curate transcript models to enhance transcriptome quality and quantification. Our methodologies reveal that lrrNA-seq data exhibit quantification biases distinct from those seen in short-read RNA-seq, necessitating specialized normalization methods. I will also discuss alternative approaches to defining joint transcriptomes in multi-sample experiments and their impact on transcript identification. Finally, I will introduce IsoAnnot, now integrated into the SQANTI framework, which distinguishes productive from unproductive transcripts and annotates them with functional labels to explore the biological impact of alternative splicing.



Prof. dr. Ana Conesa

Group leader of the Genomics of Gene Expression Lab (ConescaLab), Institute for Integrative Systems Biology (I2SysBio), Spanish National Research Council (CSIC) and University of Florida

Ana Conesa is a senior scientist in Bioinformatics and Computational Biology and a research professor at the Institute for Integrative Systems Biology (I2SysBio) at the Spanish National Research Council, leading the Genomics of Gene Expression Lab. She is a Professor in Bioinformatics at the University of Florida and co-founded the Biobam Bioinformatics, a spin-off company specialised in user-friendly software of biologists. She holds an Engineering degree from the Polytechnical University of Valencia (1993) and a PhD in Molecular Microbiology from Leiden University in The Netherlands (2001). Ana has built a scientific program in the development of algorithms and bioinformatics solutions for the analysis of Big Genomics Data, particularly gene expression and their impact on the phenotype. She has conceived and lead the creation of over a dozen of bioinformatics tools that apply to all kind of species, model and non-model organisms, plants, microbiomes, and humans. These include Blast2GO, maSigPro, NOISeq, Paintomics, SQANTI and TAPPAS. Her tools have tens of users world-wide and have received over 25,000 scientific citations. Ana has published over 135 scientific papers and lead international EU (STATegra, DEANN) and US (UF-TEDDY) projects.

Why is it so hard to analyze single-cell RNA-seq data?

Single-cell omics methods promise to revolutionize our understanding of the behavior of biological systems, measurement technologies have undergone spectacular improvements over the last decades, and the number of available multi-omics datasets is increasing rapidly.

However, to realize the promise of these multi-omics data requires rigorous computational methods that can unambiguously quantify what the raw measurements tell us about the underlying biophysical states of individual cells. I will argue that, in this regard, the current state of the single-cell omics analysis field unfortunately provides a rather depressing picture. Most tools subject the data to multiple layers of ad hoc and complex transformations and filters, with countless arbitrarily tunable parameters, and return results as abstract quantities that lack physical interpretation or error bars. Moreover, the fact that the number of proposed analysis methods increases almost as fast as the number of published datasets shows that any consensus on how to analyze such data is yet to emerge in this field.

In this talk I will discuss some of the reasons why rigorous analysis of single-cell omics data is so challenging, taking single-cell RNA-seq as an example, and make some proposals for a way forward that is arguably simpler, more transparent, and directly rooted in our understanding of the underlying biophysics and measurement processes.



Prof dr. Erik van Nimwegen

Group leader of the Switzerland
Genome Systems Biology (GSB)
Group, SIB Swiss Institute of
Bioinformatics, van Nimwegen Lab,
Biozentrum, University of Basel

After studying theoretical physics in Amsterdam, Erik van Nimwegen moved to the United States in 1995, performing his PhD studies at the Santa Fe Institute (SFI) in

Santa Fe, New Mexico, and receiving his PhD from the Faculty of Biology at Utrecht University in 1999. After a year of post-doc studies at the SFI, he spent three years as a fellow at the Center of Studies in Physics and Biology at the Rockefeller University, in New York. Since 2003 he has been a Professor of Computational Biology at the Biozentrum of the University of Basel, and group leader at the Swiss Institute of Bioinformatics since 2004. His main research topics concern genome evolution and the function and evolution of the regulatory networks by which cells control gene expression. He develops mathematical models for analyzing how regulatory networks evolve and function, and computational methods for the reconstruction of such networks from large biological data-sets.

Pitch Talks

Sequencing and Bioinformatics

Multiomic measurements of tRNAs and ribosome extension dynamics in single cells

Presenter: Mees van der Ent

Hubrecht Institute

The translation of mRNA into proteins plays a pivotal role in gene expression, and as such, it and its components are tightly regulated. Transfer RNAs (tRNAs) form the physical link between a codon and its cognate amino acid. Typically thought of as providing a housekeeping role, the genomes of eukaryotic cells instead contain different copies of tRNA genes with the same anticodon; expression differences between these isodecoders can lead to variations in tRNA pools. Additionally, tRNAs are post-transcriptionally modified, which is essential for their stability and function. Proper tRNA expression and modification are required for efficient translation. However, the extent to which variations in the tRNA pool influence translation remains unclear.

Here, we develop a multiomic method to simultaneously measure translation extension dynamics and tRNA abundances and modifications in the same cell. Comparisons of the measured tRNA abundances to nanopore tRNA sequencing were moderately correlated ($p=0.660$), validating the quantitative measurement of tRNA abundances in single cells. We then investigate the interplay between each codon's translational speed and the cognate anticodon's supply. Here, we see that the expression of only some tRNAs is well correlated to their demand. Next, we measure post-transcriptional modifications on tRNAs by quantifying reverse transcription misincorporations and using a novel approach to model the increased resistance of certain modifications to digestion. We show that we can detect Queuosine modifications on a subset of tRNAs and that, consistent with existing research, cells with lower levels of this modification translate this subset's cognate codons more slowly. This new technology makes it possible to measure the interplay between codons' translational speed and their cognate tRNAs' abundance and modifications. Further work will focus on applying this technology to systems with high variations in codon usage to investigate the role tRNA abundances and modifications have in translational regulation.

[Go back to the programme](#)

Searching for robust protein biomarkers in human plasma

Presenter: Isabel Houtkamp

Utrecht University

High-throughput affinity-based assays enable the simultaneous quantification of thousands of proteins in a liquid biopsy sample, facilitating both the discovery of single protein biomarkers and the development of multi-marker predictive models for disease. Blood plasma is a powerful matrix for these targeted proteomics analyses because of its minimally invasive collection process. The handling of blood samples can however impact the structures of plasma proteins and, consequently, their accurate quantification. Such handling factors in the pre-analytical phase, from blood collection to storage, can vary between laboratories. Here we investigate the robustness of the targeted plasma proteome under pre-analytical variations in blood sample handling. This allows us 1) to prioritize novel biomarkers that are most robust against such variations for use in clinical practice, 2) to detect protein signatures of sample handling and 3) to characterize which properties make proteins robust against pre-analytical sample handling variations. This knowledge eases the path to clinical application of novel plasma protein biomarkers discovered through high-throughput affinity-based assays.

[Go back to the programme](#)

NanoRCS: Multimodal tumor cell-free DNA profiling using nanopore-based consensus sequencing

Presenter: Li-Ting Chen

University Medical Center Utrecht

Shallow genome-wide cell-free DNA (cfDNA) sequencing holds great promise for non-invasive cancer monitoring by providing reliable copy number alteration (CNA) and fragmentomic profiles. Single nucleotide variations (SNVs) are, however, much harder to identify with low sequencing depth due to sequencing errors. Here we present Nanopore Rolling Circle Amplification (RCA)-enhanced Consensus Sequencing (NanoRCS), which leverages RCA and consensus calling based on genome-wide long-read nanopore sequencing to enable simultaneous multimodal tumor fraction estimation through SNVs, CNAs, and fragmentomics. Efficacy of NanoRCS is tested on 18 cancer patient samples and seven healthy controls, demonstrating its ability to reliably detect tumor fractions as low as 0.24%. *In vitro* experiments confirm that SNV measurements are essential for detecting tumor fractions below 3%. NanoRCS provides the opportunity for cost-effective and rapid processing, which aligns well with clinical needs, particularly in settings where quick and accurate cancer monitoring is essential for personalized treatment strategies.

[Go back to the programme](#)

Comparative transcriptomics reveals a conserved immune response in six mushroom-forming fungi during interaction with their competitors

Presenter: Marieke van Maanen

Utrecht University

Mushrooms are prone to pests and diseases from a range of fungal and bacterial competitors, leading to substantial agricultural losses. While transcriptomic analyses have explored defence mechanisms of mushroom-forming fungi during interaction with a variety of microorganisms, nematodes and arachnids, little is known about whether these responses are conserved across species. To address this, we conducted a comparative transcriptomic analysis of six mushroom-forming fungi during interactions with fourteen different organisms. Genes encoding secreted proteins were consistently enriched across all interactions. These proteins can modulate interactions with competitors by inhibiting their growth and viability or by detoxifying harmful substances.

One of the conserved genes was a thaumatin-like protein (TLP), which is previously associated with plant defence responses. Therefore, we heterologously expressed this TLP using the *Pichia pastoris* expression system to explore its functional role. Interestingly, expression of this protein led to cell lysis and increased cell death in *P. pastoris*, indicating its potential antifungal activity. Through comparative transcriptomics, we identified conserved mechanisms in fungal defence, including secreted proteins, which expands our knowledge of the defence system of mushroom-forming fungi.

[Go back to the programme](#)

Fungal biocontrol agents for *Striga* weed eradication

Presenter: Dr. Khyati Mheta Bhatt

Westerdijk Fungal Biodiversity Institute

The root parasitic weed *Striga* causes yield losses of 30–100% in *Sorghum* and other cereal crops in sub-Saharan Africa. *Sorghum* is globally considered to be the fifth most important cereal crop. *Striga* infestation therefore negatively impacts the livelihood and food security of smallholder farmers. Our aim is to explore the potential of beneficial soil microbes as biocontrol agents to control the parasitic weed *Striga*. In a previous study, approximately 3,000 fungal strains have been isolated from Ethiopian field soils, *Sorghum* roots and *Striga* seeds. Out of 1262 fungal strains tested for their interaction with *Striga*, about 10% of them showed a negative interaction, including seed degradation, seed germination inhibition and haustoria degradation in *Striga*. In contrast, approximately 11% of the isolates induced *Striga* seed germination. In the present work, our aim is to characterize the mode of action of selected fungal strains using a comparative genomics approach to identify the molecular determinants responsible for the observed interactions. We sequenced genomes of a total of 36 fungal strains, belonging to diverse genera across fungal taxonomy. Based on the phylogenetic dereplication of genes encoding core biosynthetic enzymes and its correlation with the known phenotype specific to *Striga*-fungi interaction, we identified candidate biosynthetic pathways likely responsible for the inhibition of *Striga* seed germination. Further, we investigated potential of the fungal strains to produce agriculturally relevant six mycotoxins, based on the identification of orthologs of core genes involved in known mycotoxin pathways. This knowledge will aid the development of selected fungal species as biocontrol agents, hopefully leading to a better protection of *Sorghum* fields against *Striga* infestations.

[Go back to the programme](#)

Bioinformatic approaches to decipher lettuce innate immunity

Presenter: Iñigo Bañales Belaunde

Utrecht University

Cell surface receptors sensing molecules derived from or associated with microbial pathogen invasion (PAMPs) allow plants to mount an effective immune response (PAMP-triggered immunity, PTI). Outputs of PAMP recognition include reactive oxygen species (ROS) burst, mitogen-activated protein (MAP) kinase phosphorylation, and host transcriptional reprogramming. Surface immune receptors can confer broad-spectrum resistance, since PAMPs are often shared between distinct pathogens. PTI has been proposed as the basis for quantitative disease resistance (QDR), defined as an incomplete, durable resistance mediated by multiple loci. Importantly, it is becoming increasingly evident that PTI is also required for qualitative (or full) resistance conferred by intracellular immune receptors. Altogether, PAMP receptors represent a promising lead for resistance engineering in crops.

Despite being an attractive lead for breeding, there is limited knowledge on how PTI works and connects to other sectors of immunity in species other than *Arabidopsis thaliana*. Lettuce is an ideal candidate to fill this gap for *Asteraceae*, one of the largest families of flowering plants. Research on lettuce has advanced significantly in recent years with the establishment of efficient transformation protocols and high-quality genome resources. There is a strong interest to unravel PTI mechanisms in lettuce due to the increasing emergence of resistance-breaking pathogen isolates. One of the most severe diseases is caused by the oomycete *Bremia lactucae*. *Bremia* can quickly circumvent the qualitative effect of newly introgressed *Resistance (R)* genes, many of which map to loci encoding intracellular immune receptors. In this context, the use of PTI emerges as a promising approach to boost lettuce immunity. However, PAMP receptors in lettuce are virtually unknown and PTI in lettuce is not well understood.

Through this research, we aim to learn about conserved and distinctive features of PAMP-triggered immunity in lettuce, which is a model species for the large *Asteraceae* plant family and to inform breeding efforts for quantitative disease resistance. To this end, we combine a variety of bioinformatic and molecular techniques, including association mapping, comparative genomics, structure modelling, genetic complementation assays, and genome editing.

[Go back to the programme](#)

Modelling and Machine Learning

Quantitative modelling of fate specification in the *C. elegans* postembryonic M lineage reveals a missing spatiotemporal signal

Presenter: Dr. Benjamin Planterose Jimenez

Utrecht University

The invariant lineages of *C. elegans* provide tractable models to study how developing organisms robustly integrate spatiotemporal signals at the single-cell level via gene regulatory networks to reproducibly generate the right number of cells with the right fates at the right time and place. We here study the so-called postembryonic M lineage in hermaphrodites. This lineage begins from a single progenitor, giving rise to 18 cells with three distinct cell types – body wall muscles (BWMs), coelomocytes (CCs), and sex myoblasts (SMs); the latter proliferate at later stages to contribute to the nematode's reproductive organs. Extensive prior work has identified the key signaling pathways and transcription factors in the M lineage. The main symmetry breaking cues are: i) a ventrally restricted activation of the Notch pathway promoting SM over CC fate and ii) an asymmetric re-distribution of β -catenin and TCF among daughter cells along the anteroposterior (A-P) axis, i.e. the Wnt/ β -catenin asymmetry pathway. However, it remains unclear whether these pathways are sufficient to specify all cell fates accordingly or whether additional symmetry breaking cues are necessary. In this study, we use quantitative modelling to better understand fate specification in the postembryonic M lineage. Specifically, we modelled the anteroposterior symmetry break with increasingly complex models towards robustly reproducing fate specification in the wild type and mutants. This iterative process resulted in two alternative models that explain the experimental observations by either introducing an additional spatial (spatial symmetry break) or temporal cue (temporal symmetry break). Finally, we evaluate their plausibility and propose a series of experiments to provide support for the contesting models. Overall, our study highlights how a quantitative examination of mechanistic ideas can identify knowledge gaps and guide experimental follow-up.

[Go back to the programme](#)

Decomposing plasmid evolution using a simple pangenomic language model

Presenter: Dr. Weizhen Xu

University Medical Center Utrecht

Plasmids are autonomously replicating DNA molecules in bacteria separate from the chromosome, and play a key role in the intraspecies and interspecies horizontal transmission of antimicrobial resistance genes and virulence determinants. Studying plasmid transmission and evolution is thus critical in the control of infectious diseases, but is complicated by the hallmark characteristics of plasmids: 1. Plasmids are highly modular structures comprised of diverse subsequences from different ancestral origins, and 2. Plasmid sequence diversity results in the lack of a shared core genome to construct meaningful phylogenies from. A system of modularisation that can decompose a plasmid sequence into separate modules of divergent ancestry is thus necessary for typing/clustering large collections of plasmids into subgroups with meaningful and comparable evolutionary relationships.

Word2vec is a previous-generation natural language processing methodology using a shallow, two-layer neural network to encode dense vector representations of words based on their surrounding context. We repurposed word2vec into a plasmid modularisation system that recognizes synteny (i.e. regions of conserved gene order) within a plasmid sequence, which likely represent modules of different ancestry - we were able to identify recombination events and conserved transposons based on these syntenic patterns, and use the vector embeddings of these modules to search for similar modules in our plasmid dataset, with highly favourable runtime and low computational requirements.

[Go back to the programme](#)

Methodology for Biomarker Discovery with Reproducibility in Microbiome Data using Machine Learning

Presenter: David Rojas-Velazquez

Utrecht University

In recent years, human microbiome studies have gained significant attention due to their potential clinical applications. Advances in omics technologies and AI have led to promising outcomes in identifying potential biomarkers using machine learning tools. However, challenges such as small sample sizes, inconsistent results, and lack of uniform methodologies hinder reproducibility in biomedical research. We combined the DADA2 pipeline for 16s rRNA sequence processing with Recursive Ensemble Feature Selection (REFS) across multiple datasets to enhance reproducibility and achieve robust results. We conducted four experiments analyzing microbiome data from patients with Inflammatory Bowel Disease (IBD), Autism Spectrum Disorder (ASD), Type 2 Diabetes (T2D), and Parkinson's disease. Each experiment identified a biomarker signature in one dataset and validated it in at least two others. Our methodology addresses issues related to data dimensionality, inconsistent results, and validation across independent datasets. The findings from the four experiments, involving 13 different datasets, demonstrate that our approach provides robust and reliable results. While the results are promising, further research and experimentation are needed to establish the microbiome as an effective tool for medical purposes.

[Go back to the programme](#)

M&M: An RNA-seq based Pan-Cancer Classifier for Pediatric Tumors

Presenter: Fleur Wallis

Princess Máxima Center for pediatric oncology

Acquiring the correct diagnosis is a challenging but crucial process in pediatric oncology. Notably, rare tumors present a unique challenge given their infrequency and relative unfamiliarity among pathologists. As a result, these tumor entities tend to be affected by higher misclassification rates. Here we present M&M, a pan-cancer ensemble-based machine-learning algorithm specifically tailored towards inclusion of rare pediatric tumor (sub)types. The RNA-seq based algorithm can classify 52 different tumor types, plus the underlying 96 tumor subtypes. Furthermore, M&M encompasses samples from all tumor stages, treatment statuses and from several non-neoplastic tissues.

To facilitate infrequently occurring tumor (sub)type classifications, two different classifiers were created and integrated: a Minority classifier tailored towards correct classification of rare tumor (sub)types, and a Majority classifier with more predictive power for high frequency tumor entities. Classification took place on the tumor subtype level, from which the tumor type could be extrapolated. M&M could correctly classify the tumor type for 94.5% of the samples within the reference cohort. For high-confidence classifications (80% of samples), M&M could reach a precision of ~99%. For the low-confidence classifications, the correct tumor classification was often included in the three highest-scoring labels, leading to an overall accuracy of 98% within the top 3. For classes covered within the classifier, M&M's performance is comparable to existing class-restricted classifiers like the DKFZ methylation classifier for CNS tumors. An independent test cohort confirmed the robustness of M&M's performance.

Machine-learning algorithms for both adult and childhood cancer are increasingly used in the clinic, contributing towards increased patient survival. However, many tumor entities are currently missing from existing classifiers. Developing and introducing an extensive agnostic pan-cancer classifier in diagnostics has the potential to increase the diagnostic accuracy for many pediatric cancer cases, thereby contributing towards optimal patient survival and quality of life.

[Go back to the programme](#)

Exploring the diversity of fungal plant biomass conversion

Presenter: Dr. Jiajia Li

Westerdijk Fungal Biodiversity Institute

Plant biomass is one of the most abundant renewable resources on earth, and it serves as a primary raw material for conversion into value-added bioproducts. Filamentous fungi play a critical role in the global carbon cycle by degrading plant polysaccharides, as they are highly efficient degraders of plant polysaccharides. The conversion process of plant biomass involves several key cellular activities, including nutrient sensing, secretion of degrading enzymes (Carbohydrate-Active enZymes, CAZy), sugar uptake by transporters, intracellular metabolism, and the sophisticated transcription regulation system for precisely governing all aspects of the process. In our study, we focused primarily on the first three aspects: sugar metabolism, carbohydrate-active enzymes (CAZymes) involved in plant polysaccharides degradation, and sugar transporters. We investigated the diversity of these mechanisms across a range of fungal species, and aimed to elucidate the evolutionary adaptations underpinning these processes.

[Go back to the programme](#)

RNA splicing as novel immune regulator

Presenter: Dr. Alejandra Bodelón de Frutos

University Medical Center Utrecht

Upon encountering pathogens, effective immune responses demand rapid and precise changes of expressed proteins by immune cells. How this process is regulated remains incompletely understood. We hypothesized that the process of splicing is an overlooked layer of protein expression regulation that is directly involved in immune activation. We found that the expression of many genes involved in splicing was affected after activation in monocytes. To assess the entire splicing landscape, we used long-read RNA-seq in monocytes and assess the effect of activation on isoform expression and usage. We detected ~25.000 isoforms, of which 60% were novel not-annotated isoforms. In addition, we observed that activation modulated isoform-specific gene expression and that isoforms were differently used compared to the steady state. By assessing the effect of these changes, we detected that activation increased the usage of functional and productive isoforms. Taken together, these data suggest that RNA splicing is involved in the regulation of monocyte activation by the increase of functional isoform expression and usage. A better understanding of this process could pave the way for new strategies aimed at interfering with alternative RNA splicing for the treatment of immune-related diseases.

[Go back to the programme](#)

Poster Presentations

Posterboard	Presenter	Title
1	Roula Farag	Signatures of Breakage-Fusion-Bridge Cycles in Pediatric Cancer
2 (morning)	Ya Yuan Zhu	Revelations of Circulating Cell-free RNA in Atherosclerosis: Cell Type Deconvolution and Clinical Relevance
2 (afternoon)	Wilco Hermens	Resolving de novo Genome Assembly Graphs with Protein Embeddings
3	Pavel Sinitcyn	Global detection of human variants and isoforms by deep proteome sequencing
4 (morning)	Jan Paul van Meenen	How herd immunity mitigated a deadly second wave of COVID-19 in Manaus
4 (afternoon)	Weijuan Wang	Association of Antipsychotic Augmentation and Novel Drug Target Genes with Schizophrenia Risk: A Mendelian Randomization Study
5	Leron Kok	Integrating Ribosome Profiling and Proteomics to Validate Non-Canonical ORFs
6 (morning)	Ting Chia Liu	Machine Learning Analysis of Gut Microbiome Profiles in Infants from Different Feeding Practices
6 (afternoon)	Luuk van Vliet	Short-sightedness of HIV and its evolutionary extinction
7	Paula Gomez Zapata	Investigating the lifecycle switch in the fungal disease <i>Cronartium pini</i> through population genomic analysis

Posterboard	Presenter	Title
8 (morning)	Lorenzo Possanzini	MidMap: Integrating Molecular Atlases to Unlock Insights into the Midbrain
8 (afternoon)	Daan van Dam	Enhancing Colorectal Cancer Recurrence Prediction: A Multi-Feature Selection Approach and Transfer Learning Framework for Organoid Data
9	Ida de Vries	AlphaFill: enriching AlphaFold models with cofactors and ligands
10 (morning)	Faik Usul	Transcriptional Investigation of Cell Competition between Intestinal Adenocarcinoma and Wild Type Intestinal Cells in Mice
10 (afternoon)	Lorenzo Caranti	Modeling Yeast Metabolism in Plant Phyllosphere
11	Fleur Wallis	M&M: An RNA-seq based Pan-Cancer Classifier for Pediatric Tumors
12 (morning)	Laurens Janssen	Regulation of immune activation by transposable elements in monocytes
12 (afternoon)	Ilaria Coratella	Classifying glioblastoma cells based on developmental cell type resemblance using machine learning on single-cell RNA sequencing data
13	Dominika Tomczyk	Machine Learning Reveals New Gene Functions in Viral Respiratory Infections
14 (morning)	Gabe van den Hoeven	Hypothesis rejected: TdT does not conserve the D segment in T-cell receptor β-chain sequences

Posterboard	Presenter	Title
14 (afternoon)	Coen Micklei	<u>Microbes as cheese makers, from feast to famine</u>
15	Axel Rosendahl Huber	<u>Estimation of the number of cancer driver mutations induced in human tissues</u>
16 (morning)	Arne Kukkonen	<u>Comparative analysis of cardiomyocyte-fibroblast paracrine signaling involved in the myocardial ischemic tolerance of Acomys cahirinus</u>
16 (afternoon)	Ivo Borkus	<u>Technical factors influencing transcriptomic allelic distribution</u>
17	Saul Pierotti	<u>Extensive Gene x Environment interactions in the outbred vertebrate genetics of a standard physiological phenotype</u>
18 (morning)	Manuel Ayllongavilan	<u>Non-canonical peptides presented on MHC molecules as possible biomarkers for (immune)therapies</u>
18 (afternoon)	Nicola Greco	<u>Investigating the Evolutionary Origins of Winged-Helix Domains in the Dam1 and Ska Complexes of the Eukaryotic Kinetochore</u>
19	Raphaëlle Versini	<u>Lys716 in the transmembrane domain of yeast mitofusin Fzo1 modulates anchoring and fusion</u>
20 (morning)	Anne ten Vergert	<u>Exploring fundamental mutational patterns of recessive DNA repair syndromes using single-cell whole genome sequencing</u>

Posterboard	Presenter	Title
20 (afternoon)	Noémi Versluijs	Creating a hydraulic atlas of Sorghum bicolor to investigate the role of suberization during drought stress
21	Julian Paganini	Benchmarking pangenome graph construction methods for their application to foodborne pathogens
22 (morning)	Oscar Esteban	SNV detection and lineage tracing in single cells using long read RNA sequencing
22 (afternoon)	Rutger Ballieux	Single-Cell Classification and Disease Gene Identification in ALS and Alzheimer's Using Simple Classification Models
23	Laura Dijkhuizen	PhD courses on bioinformatics basics
24 (morning)	Sotiris Niarchos	Upsimpler: Data augmentation for ultra rare (sub)types in RNA-seq based pediatric tumor classification
24 (afternoon)	Vince During	Explaining unexpected deuterium labeling curves
25	Joanna von Berg	The Dutch childhood cancer genome project: Characterizing tumor drivers
26 (morning)	Miguel Sánchez Marín	Modelling of nanobody-antigen complexes using machine learning and information-driven docking
26 (afternoon)	Anna Delhaas	A discrete stochastic model on the dynamics of choanoflagellate chain colonies

Posterboard	Presenter	Title
27	Hidde Smits	Biologic therapies imprint epigenetic modifications in the skin-homing T cells of atopic dermatitis patients
28 (morning)	Maria Karaboeva	A detailed look into the intricate decision-making framework guiding cell-differentiation
28 (afternoon)	Judith Ahr	Computational design of coronavirus entry inhibitors
29	Emma Bernsen	Integrating Pharmacogenetics Upfront: Impact of Early Pharmacogenetic Screening in a Pediatric Oncology Cohort of 1,151 Subjects
30 (morning)	Mees Klunder	Analysing alternative polyadenylation in ALS using a predictive neural network
30 (afternoon)	Markel Benito	Leveraging deep learning to learn features of antibody recognition
31	Chrats Melkonian	Carbon-fixing and viral-infected Ferroplasma dominates acidic cave biofilm
32 (morning)	Marta Garcia Mondejar	Machine Learning-based analysis of plant phenomics data to better understand the growth-defense trade-off

[Go back to the programme](#)

Signatures of Breakage-Fusion-Bridge Cycles in Pediatric Cancer

Presenter: Roula Farag

Structural variants (SVs) are genomic rearrangements that lead to amplifications, deletions or translocations of chromosomal material of various sizes. In pediatric cancers, the prevalence and effect of complex SVs on chromosomal instability (CIN) remains unclear. Understanding the mechanisms and implications of CIN is crucial for unraveling the complexities of tumorigenesis in pediatric cancers. Breakage-fusion-bridge (BFB) is a prominent type of complex SV which can be linked to copy number changes in the genome. BFB cycles are characterized by repetitive cycles of breaks, followed by fusion of chromatin or chromosome ends, foldback inversions (FBIs) and mitotic bridge formation. BFB can manifest as oscillating copy number changes, amplifications and staircase-like gains associated with the presence of FBIs. Moreover, a sloping copy number signature consistent with successive BFB cycles and ongoing chromosomal instability was identified. We aim to fully characterize these different copy number signatures associated with BFB cycles in pediatric cancer. Using whole genome sequencing data from our pediatric cancer cohort, we categorize copy number alterations (CNA) associated with BFB. We classify them as paracentric or pericentric, gain or loss, and assign them to signatures including staircase, sloping or amplifications. We deployed GRIDSS for structural variant calling, PURPLE for ploidy scoring and LINX for identifying FBIs in BFB. So far, we identified BFB fold back inversions associated with sloping copy number signature in two osteosarcoma samples. Our findings highlight the various CNA signatures associated with BFB in our pediatric cancer cohort. We continue to identify BFB cycles in the remaining CNA signatures (ie staircase and amplifications) to establish their association and distribution within our pediatric cancer cohort. By combining these findings with long-read sequencing technologies, we aim to uncover larger FBIs and enhance our understanding of the prevalence and impact of BFB on CIN in pediatric cancer.

[Go back to the poster presentations overview](#)

Revelations of Circulating Cell-free RNA in Atherosclerosis: Cell Type Deconvolution and Clinical Relevance

Presenter: Ya Yuan Zhu

Atherosclerosis is characterized by the formation of plaques in the arterial walls, with their development and progression being primarily driven by various cell types. Commonly used methods are limited in capturing the cellular heterogeneity at a fine resolution and across large-scale studies. A potential proxy for measuring the heterogeneous cell type activity and the disease state is circulating cell-free RNA (ccfRNA). Utilizing deep learning-based cellular deconvolution on bulk RNA-sequencing data of ccfRNA, with single-cell RNA-sequencing data as a reference, the underlying cellular origins of the measured ccfRNA can be predicted. These cell type compositions are then analyzed for meaningful correlations with clinical characteristics of atherosclerotic patients, using general linear modeling and Cox regression. The composition of monocytes is expected to be associated with the occurrence of major adverse cardiovascular events. The outcomes of this study could provide an additional, minimally invasive tool for assessing disease status, or evaluating risk in real-time. Ultimately leading to the improvement of disease management and patient outcomes.

[Go back to the poster presentations overview](#)

Resolving de novo Genome Assembly Graphs with Protein Embeddings

Presenter: Wilco Hermens

Genome assembly using short-read sequencing data presents significant challenges, particularly due to the presence of repetitive sequences that are longer than the reads themselves. These repeats create assembly graphs where the nodes representing the repeats have multiple incoming and outgoing edges. Our model can predict which pairs of incoming and outgoing edges are connected, greatly simplifying the assembly graph. This reduction in complexity improves the quality of contigs, enabling accurate and extended continuous sequences, which enhances research at the contig level.

[Go back to the poster presentations overview](#)

Global detection of human variants and isoforms by deep proteome sequencing

Presenter: Pavel Sinitcyn

An average shotgun proteomics experiment detects approximately 10,000 human proteins from a single sample. However, individual proteins are typically identified by peptide sequences representing a small fraction of their total amino acids. Hence, an average shotgun experiment fails to distinguish different protein variants and isoforms. Deeper proteome sequencing is therefore required for the global discovery of protein isoforms. Using six different human cell lines, six proteases, deep fractionation and three tandem mass spectrometry fragmentation methods, we identify a million unique peptides from 17,717 protein groups, with a median sequence coverage of approximately 80%. Direct comparison with RNA expression data provides evidence for the translation of most nonsynonymous variants. We have also hypothesized that undetected variants likely arise from mutation-induced protein instability. We further observe comparable detection rates for exon-exon junction peptides representing constitutive and alternative splicing events. Our dataset represents a resource for proteoform discovery and provides direct evidence that most frame-preserving alternatively spliced isoforms are translated.

[Go back to the poster presentations overview](#)

How herd immunity mitigated a deadly second wave of COVID-19 in Manaus

Presenter: Jan Paul van Meenen

Manaus, Brazil, experienced a severe second wave of COVID-19 deaths despite key studies suggesting high attack rates following the first wave. The novel gamma P.1 strain significantly increased mortality rates, especially in younger individuals. This study uses age-stratified transmission models that explicitly differentiate between infections and reinfections to investigate why the second wave was more lethal, and to assess the impact of previous immunity on epidemic dynamics. The model integrates data on severe acute respiratory disease in-hospital mortality, SARS-CoV-2 seroprevalence, and demographic information from Manaus (population 2.2 million). Fitting the model to the joint data, we were able to independently recover infection fatality rates for the first wave corresponding to age-dependent estimates found in meta-analysis studies. Heterogeneities in contact patterns led to differences in attack rates across age groups, with most younger individuals infected, while a significant proportion of older individuals remained susceptible after the first wave. Our findings reveal that while immunity provided by the first wave markedly reduced mortality risk, it did not halt the spread of the Gamma P.1 strain, resulting in increased deaths in the remaining susceptible population. Through counterfactual scenarios, we demonstrate that without the partial immunity from the first wave, the second wave would have been even deadlier. This study improves our understanding of how differences in contact rates, waning immunity followed by reinfection, and viral evolution influenced the severity of the COVID-19 pandemic in Manaus.

[Go back to the poster presentations overview](#)

Association of Antipsychotic Augmentation and Novel Drug Target Genes with Schizophrenia Risk: A Mendelian Randomization Study

Presenter: Weijuan Wang

Schizophrenia is a severe psychiatric disorder with significant societal burden. Antipsychotics remain the primary treatment, but due to their inefficacy, off-label agents are often prescribed as adjunctive therapy, especially for treatment-resistant schizophrenia (TRS). Reviews and meta-analyses of randomized clinical trials (RCTs) have examined the efficacy of various off-label agents, however, small sample sizes and high risks of bias of some RCTs limit their validity.

To triangulate current results, this study aims to investigate the following questions: Are the biological targets of medications with demonstrated effectiveness in schizophrenia associated with schizophrenia risk generally, and with TRS specifically? Do the genetic relationships between the biological targets and these two phenotypes as shown using MR align with RCT effect sizes?

Using eQTL and pQTL data as genetic instruments, we will conduct Mendelian randomization (MR) analysis with GWAS summary statistics as outcomes. Our goal is to synthesize evidence from RCTs, neurobiology, and genetics to better understand the effectiveness and biological plausibility of these agents in treating schizophrenia and TRS.

[Go back to the poster presentations overview](#)

Integrating Ribosome Profiling and Proteomics to Validate Non-Canonical ORFs

Presenter: Leron Kok

We and others have recently demonstrated the widespread translation of thousands of short non-canonical open reading frames (ncORFs). Translation of ncORFs is frequent in cancer and indicative of protein production that could drastically expand the canonical proteome. This "dark proteome" has gained interest as a source of new tumor-specific antigens for immunotherapeutic targeting. However, most ncORFs are currently not supported by high-confidence protein-level evidence. To address this, we established an international consortium. We reanalyzed 3.8 billion raw spectra of 414 different studies comprising 95,520 MS runs using the PeptideAtlas workflow for the presence of 7,264 ncORFs. Focusing on a subset of 118 immunopeptidomics datasets, we could confirm the peptide level presence of 1,785 out of 7,264 ncORFs. Of these ncORFs, 39% was supported by at least two unique HLA peptides. Strikingly, certain HLA alleles presented more ncORF peptides than others, indicating that cancer patients carrying these HLA alleles might be more susceptible to future immunotherapeutic strategies targeting non-canonical antigens. These findings improve our understanding of non-canonical translation and offer important insights for advancing immunotherapy development against this class of antigens.

[Go back to the poster presentations overview](#)

Machine Learning Analysis of Gut Microbiome Profiles in Infants from Different Feeding Practices

Presenter: Ting Chia Liu

Breast milk is widely recognized for its critical role in infant health, with emerging evidence indicating that early nutrition shapes the infant gut microbiome, immune system development, and long-term health. Feeding patterns, particularly the choice between human milk and formula, significantly impact the infant microbiota. In this study, we applied the Recursive Feature Selection (REFS) algorithm to identify taxonomic differences between breast milk and formula milk and their relationship to infant health. Using three microbiome datasets, we identified nine key bacterial taxa in the discovery dataset PRJNA633365. At the genus level, these include Cutibacterium, two Lactobacillus species, Bacteroides, two Bifidobacterium species, Bilophila, and Romboutsia, while Enterobacterales was identified at the order level. In the testing datasets DRA007599 and PRJNA515307, five and three of these taxa, respectively, were confirmed. The K-Nearest Neighbors algorithm achieved the best performance, with AUCs of 0.77 for PRJNA633365, 0.63 for DRA007599, and 0.75 for PRJNA515307. The PCoA plot highlights the effectiveness of the nine taxa in distinguishing breastfed from formula-fed infants. Our analysis revealed that breastfed infants exhibit a higher abundance of beneficial microbes, including Bifidobacterium and Lactobacillus, known to protect against allergies and infections. Future work will expand this analysis to additional datasets and further explore the link between the infant microbiome and health outcomes in relation to breast milk and formula.

[Go back to the poster presentations overview](#)

Short-sightedness of HIV and its evolutionary extinction

Presenter: Luuk van Vliet

The human immunodeficiency virus (HIV) is a clinically relevant virus and uncurable to this day. Despite this relevancy, little is known about the evolutionary trajectory of HIV. The virus spreads very slowly, while mutating relatively fast. This combination of properties indicates a mismatch of within-host and between-host evolutionary speed. This poster is about a multi-level model investigating this seemingly contradicting modus of evolution. It investigates the problem and how the disease can be rescued.

[Go back to the poster presentations overview](#)

Investigating the lifecycle switch in the fungal disease *Cronartium pini* through population genomic analysis

Presenter: Paula Gomez Zapata

Scots pine blister rust, caused by *Cronartium pini*, threatens pine forests in northern Fennoscandia, resulting in economic losses and ecosystem disturbances. This rust species exhibits two distinct lifecycles: heteroecious and autoecious. The coexistence of these different lifecycles within the same species raises questions about the evolutionary processes in rust fungi. The aim of this project is to identify genetic differences, patterns of gene flow between forms, and potential speciation events. We collected ten samples of each form from Swedish and Finnish pine forests and sequenced them using paired-end short reads. Additionally, long-read sequencing was performed on an extra sample to generate a high-quality reference genome assembly. Filtered short reads were then aligned to this reference genome to identify single-nucleotide variants (SNVs).

Preliminary population structure analyses indicate that autoecious and heteroecious samples are genetically distinct. All autoecious samples show a low proportion of heterozygous SNVs (<3%), suggesting a loss of sexual reproduction, whereas heteroecious samples exhibit a high proportion of heterozygous SNVs (>60%), indicating active sexual reproduction. Although phylogenetic analyses suggest that the lifecycle switch may have originated from a single event, we found some mating type alleles shared between individual samples of the autoecious and heteroecious forms, suggesting potential genetic exchange between these forms. This project will provide insights into the genomic regions that may restrict gene flow between forms. We anticipate that our findings will enhance the understanding of the evolutionary dynamics of *Cronartium pini* and offer practical applications for managing this disease in pine forests in northern Fennoscandia.

[Go back to the poster presentations overview](#)

MidMap: Integrating Molecular Atlases to Unlock Insights into the Midbrain

Presenter: Lorenzo Possanzini

The midbrain, involved in cognition and emotional regulation, displays a high level of cellular heterogeneity, consisting of various cell types. The precise composition of these cell populations remains unclear due to limitations in the size and complexity of existing datasets. Single-cell genomics, particularly single cell RNA sequencing (scRNAseq), has improved resolution in identifying distinct cell types. However, the lack of standardization in annotations across studies poses a significant challenge for comparative analysis. This project addresses that challenge by creating MidMap, a powerful tool designed to enhance the analysis of neural cells through integration and comparison of data from diverse sources.

We want to create MidMap, a comprehensive atlas that enables researchers to compare cells from different brain regions with existing published datasets. By providing this capability, MidMap helps researchers align their findings with those in the literature, even when labeling and classifications differ across studies. Furthermore, MidMap will allow for the merging of datasets, facilitating more comprehensive clustering and annotation of neural cell populations. A major focus of this study was the challenge of mitigating batch effects, which result from technical discrepancies during data processing. These effects were addressed by standardizing the use of alignment tools to the human reference genome. Additionally, CellBender was employed to remove technical artifacts such as ambient RNA contamination.

scVI will be utilized to capture biological signals and account for technical noise, laying the groundwork for a robust comparison across datasets. Future research will focus on performing differential gene expression (DGE) analysis to compare neurons from the substantia nigra (SN) and the ventral tegmental area (VTA). Additionally, the study will examine the differences between healthy VTA neurons and those impacted by Parkinson's disease. In conclusion, MidMap will be a valuable resource for midbrain research, enabling comparisons of neural cells across studies, integrating datasets, and refining classifications.

[Go back to the poster presentations overview](#)

Enhancing Colorectal Cancer Recurrence Prediction: A Multi-Feature Selection Approach and Transfer Learning Framework for Organoid Data

Presenter: Daan van Dam

This project investigates features driving recurrence in colorectal cancer by analyzing multiple data types, including sequence alterations, chromosomal variations, clinical data, count expression data and features derived from the count expression data. I aim to identify which features predict recurrence and apply these findings to organoid models of the same disease to test their generalizability. Additionally, self created gene sets that predict recurrence in organoids are evaluated for their ability to separate recurrence and recurrence-free groups in FFPE-preserved samples.

Preliminary results suggest significant challenges in transferring features that work well in FFPE-derived samples to organoid contexts. However, gene sets, for example, identified in organoids show more promising transferability to FFPE samples, indicating some level of shared signal between these systems. This suggests that while predictive biomarkers may be context-dependent, organoid-based findings could offer insights for FFPE-based clinical applications.

[Go back to the poster presentations overview](#)

AlphaFill: enriching AlphaFold models with cofactors and ligands

Presenter: Ida de Vries

With the public availability of the AlphaFold Protein Structure Database, structural biologists now have a valuable new tool for their research. However, biochemical interpretation of these structures is limited to amino acids only. Hence, co-factors, small molecules and/or metal ions, which are required for protein function or structural integrity are lacking. To address this limitation, we developed AlphaFill: an algorithm that enriches AlphaFold models with co-factors, small molecules and (metal) ions based on homologous structures available in the PDB-REDO databank. All AlphaFill models are freely available through alphafill.eu: a FAIR (Findable, Accessible, Interoperable and Reproducible) resource to support scientists interpreting the enriched AlphaFold models while creating new hypotheses and designing experiments.

Besides the AlphaFill databank containing “filled” AlphaFold models, we developed a webservice where users can upload their own protein structures and run the algorithm to enrich their structures of interest with ligands and cofactors. Here, we will present the latest features, applications and future developments that improve the interpretability of the protein-ligand complexes obtained by AlphaFill, and increase user flexibility while “filling” their macromolecular structures of interest.

[Go back to the poster presentations overview](#)

Transcriptional Investigation of Cell Competition between Intestinal Adenocarcinoma and Wild Type Intestinal Cells in Mice

Presenter: Faik Usal

Cell competition is one of the processes helping maintenance of the homeostasis where cells, which are otherwise viable, are eliminated by more competitive cells. There have been some molecules shown to be responsible from the competition status determination (Notum, Dpp) to apoptosis (dSparc, Flower, JNK genes) or cell extrusion, or to overproliferation (JNK signaling). Notably, it is observed in tissues with cancer cells in which cancer cells have been shown to proliferate at the expense of healthy cells in intestinal adenomas in fruit fly, in liver tumors in mouse, and in pancreatic adenocarcinoma in human. By identifying important players, intervention strategies can be developed to counteract the cell competition resulting in death of healthy intestinal cells. In this project, I would like to identify transcriptional responses of cell competition between intestinal adenocarcinoma and wild type intestinal cells in mice. The dataset came from a study by Garcia et al. in 2021 in which they used three-days-grown pure intestinal adenocarcinoma organoids, pure wild type intestinal organoids, and mixed organoids of wild type intestinal and intestinal adenocarcinoma cells. These cells were sequenced with T-ChIC method, and the resulting transcriptomics data is being used. Specifically, I would like to 1. identify cell type specific responses of cell competition by identifying differentially expressed genes between mixed and pure conditions and pathways/GO terms enriched in these, 2. carry out trajectory analysis in mixed wild type intestinal cells using RNA velocity to infer in which stage/cell type during differentiation from the stem cell the priming of cell competition occurred most dominantly and which genes contributed the most to this change. Additionally, along the trajectory, JNK and SCA1 expressions will be modeled to see whether there is a significant change in expression. So far, clustering was carried out which identified four clusters in cancer cells (except one cluster seemingly exclusive to mixed condition) which were unstable to parameter changes, and five stable clusters in wild type cells corresponding to niche cells (Goblet, endocrine, Paneth), stem cells (mostly composed of naive condition), transit amplifying cells (mostly composed of primed condition), primed differentiated epithelial cells, and naive differentiated epithelial cells.

[Go back to the poster presentations overview](#)

Modeling Yeast Metabolism in Plant Phyllosphere

Presenter: Lorenzo Caranti

The upper portion of plants, also known as the phyllosphere, represents an important environment where various microbial communities live. These communities are composed of different organisms that interact with each other as well as with the plant. The nature of these interactions can have significant consequences for the plant's health, as some of the organisms may be pathogens. *Fusarium graminearum* is one of the most problematic pathogens in wheat and inhabits the phyllosphere. Previous studies have identified certain yeasts from the genera *Metschnikowia*, *Aureobasidium*, and *Candida* that are capable of inhibiting the development of this pathogen.

To better understand the inhibition activity of these yeasts on *Fusarium*, we reconstructed multiple Genome Scale Metabolic models. This reason led us to test different automatic pipelines to create our models. We initially used the CarveMe, a fast automatic pipeline designed for prokaryotes. Subsequently, we used CarveFungi, an adaptation of CarveMe, that uses machine learning to make the compartmental prediction of the reactions and improve the quality of the outcome. Through parallelisation, we were able to reduce the time required for this step. The models were tested on different carbon sources to evaluate their prediction. The results are sufficient on glucose and poor on other carbon sources. Community simulation showed a correlation between the inhibition activity and the colonisation efficacy, confirming previous findings and opening to more specific questions.

[Go back to the poster presentations overview](#)

M&M: An RNA-seq based Pan-Cancer Classifier for Pediatric Tumors

Presenter: Fleur Wallis

Acquiring the correct diagnosis is a challenging but crucial process in pediatric oncology. Notably, rare tumors present a unique challenge given their infrequency and relative unfamiliarity among pathologists. As a result, these tumor entities tend to be affected by higher misclassification rates. Here we present M&M, a pan-cancer ensemble-based machine-learning algorithm specifically tailored towards inclusion of rare pediatric tumor (sub)types. The RNA-seq based algorithm can classify 52 different tumor types, plus the underlying 96 tumor subtypes. Furthermore, M&M encompasses samples from all tumor stages, treatment statuses and from several non-neoplastic tissues.

To facilitate infrequently occurring tumor (sub)type classifications, two different classifiers were created and integrated: a Minority classifier tailored towards correct classification of rare tumor (sub)types, and a Majority classifier with more predictive power for high frequency tumor entities. Classification took place on the tumor subtype level, from which the tumor type could be extrapolated. M&M could correctly classify the tumor type for 94.5% of the samples within the reference cohort. For high-confidence classifications (80% of samples), M&M could reach a precision of ~99%. For the low-confidence classifications, the correct tumor classification was often included in the three highest-scoring labels, leading to an overall accuracy of 98% within the top 3. For classes covered within the classifier, M&M's performance is comparable to existing class-restricted classifiers like the DKFZ methylation classifier for CNS tumors. An independent test cohort confirmed the robustness of M&M's performance.

Machine-learning algorithms for both adult and childhood cancer are increasingly used in the clinic, contributing towards increased patient survival. However, many tumor entities are currently missing from existing classifiers. Developing and introducing an extensive agnostic pan-cancer classifier in diagnostics has the potential to increase the diagnostic accuracy for many pediatric cancer cases, thereby contributing towards optimal patient survival and quality of life.

[Go back to the poster presentations overview](#)

Regulation of immune activation by transposable elements in monocytes

Presenter: Laurens Janssen

The proteome of immune cells needs to be quickly and precisely modified to result in an effective immune response. Transposable Elements (TEs) are DNA sequences that are able to move from one location to another in the host genome, and new evidence suggests that TEs significantly contribute to the gene regulation and splicing in immune cells. TEs are also induced in the response to viral infections and are involved in some autoimmune diseases. The main goal of this study is to characterize the contribution of TEs to the isoform diversity in activated and steady-state immune, taking data from the monocytes of patients with early arthritis, and applying 3rd generation ONT RNA-seq data.

[Go back to the poster presentations overview](#)

Classifying glioblastoma cells based on developmental cell type resemblance using machine learning on single-cell RNA sequencing data

Presenter: Ilaria Coratella

Tumors are heterogeneous entities exhibiting a variety of cell types in complex spatial configurations, complicating the analysis of tumor composition and origin. The genetic and phenotypic diversity of cancer cells hides the initial mutations and cellular behaviors that lead to tumorigenesis, making it challenging to trace their evolutionary paths. Single-cell RNA sequencing has shed light on this heterogeneity, demonstrating that tumor cells often mirror specific developmental lineages. Therefore, identifying resemblances between tumor cells and specific healthy developmental cell types provides a starting point for a better understanding of tumor cell characteristics, helping to navigate tumor heterogeneity and potentially enhancing our understanding of tumorigenic processes. Nevertheless, new large-scale single-cell datasets that capture gliogenesis have not yet been compared to glioblastomas.

This study aims to categorize glioblastoma cells based on the developmental cell types they most closely resemble. To achieve this, we constructed a single-cell RNA sequencing meta-atlas of the developing brain and compared these cell types to a glioblastoma single-cell RNA sequencing dataset. We employed two different methodologies: a machine learning approach using a random forest classifier, and a multidimensional distance-based approach utilizing the Mahalanobis distance. The analytical methods categorized tumor cells into different cell types. We are assessing whether these classifications are consistent across both methodologies, aiming to refine the techniques for more reliable predictions. This approach will enable us to determine whether tumor cells robustly resemble specific developmental cell types.

The newly acquired information on tumor composition might be used to investigate how the clinical properties of a tumor—such as its growth rate, response to treatments, and prognosis—and molecular properties like clonal structure correlate with its cellular composition. Furthermore, making this classification reliable and systematic may enhance our understanding of how these cells co-opt developmental pathways to foster tumor growth.

[Go back to the poster presentations overview](#)

Machine Learning Reveals New Gene Functions in Viral Respiratory Infections

Presenter: Dominika Tomczyk

Influenza virus, human metapneumovirus, rhinovirus, and respiratory enterovirus are the leading causes of acute respiratory infections worldwide. Seasonal epidemics caused by these viruses make it evident that some people are more sensitive to viral infections than others. This indicates the existence of yet-unidentified, otherwise healthy individuals, genetically predisposed to develop respiratory infections. As there are no vaccines against rhinovirus and human metapneumovirus, and the vaccination rates against influenza remain low, predisposed individuals may be constantly exposed to the contagious agents. Therefore, there is a high need for biomarkers capable of identifying such populations. In this study, we aimed to identify the overlapping differently expressed genes in patients infected with either influenza, enterovirus/rhinovirus, or human metapneumovirus, and investigate whether these genes are related to the host antiviral response or, if not, whether they can contribute to human genetic susceptibility to viral respiratory infections. Using the Recursive Ensemble Feature Selection (REFS), a machine learning-driven approach, we processed 133 RNAseq samples from BioProject PRJNA660611 (GEO accession: GSE157240), including 65 adults infected with influenza virus, 31 with enterovirus/rhinovirus, 17 with human metapneumovirus, and 20 asymptomatic controls. Out of 19,203 genes, we identified 7 top-ranked differently expressed genes common in all three infections, with a diagnostic accuracy of 0.76, and a standard deviation of 0.11. Five genes were grouped either as those related to antiviral response (two genes - IFI27 and CARD17P) or those possibly related to susceptibility to viral respiratory infections (three genes - NOG, MT2A, and ID3). Two genes remained of unknown function and could not be grouped. This way, we propose possible novel functions of NOG, MT2A, and ID3 genes that, if proven true, could become potential biomarkers of human susceptibility to viral respiratory infections.

[Go back to the poster presentations overview](#)

Hypothesis rejected: TdT does not conserve the D segment in T-cell receptor β -chain sequences

Presenter: Gabe van den Hoeven

T-cell receptor (TCR) diversity is fundamental to the immune system's ability to recognise foreign antigens. TCRs are put together in a semi-stochastic process called V(D)J-recombination, where the α -chain consists of a variable (V) and junction (J) gene segment, and the β -chain consists of a V, diversity (D), and J gene segment. During V(D)J-recombination one segment of each of these genes are recombined, with deletions and non-template nucleotide additions by a protein called Terminal deoxynucleotidyl Transferase (TdT) occurring at the junctions of these segments. Research has shown that some abundant β -chain sequences lack the D segment. We test the hypothesis that the absence of TdT, may allow for the deletion of the D segment. By comparing TdT knock-out with wild-type sequences in mice, we find abundant β -chain sequences without a D segment in both groups but no significant increase in the TdT knock-out group, suggesting that TdT does not protect against the deletion of the D segment. Additionally, our analyses revealed that V and J gene segment usage differs significantly between TdT knock-out and wild-type sequences, and that almost 60% of abundant wild-type sequences used either the TRBV1 or TRBV16 gene segment.

[Go back to the poster presentations overview](#)

Microbes as cheese makers, from feast to famine

Presenter: Coen Micklei

Research into microbial communities is important for a wide range of socio-economic activities, like the production of fermented foods like bread, wine or cheese. In this research, we focused on cheese production by microbes and the metabolic processes involved. A lot is known about the effect of microbes on the first step of the cheese making process and of the physiology of up to three species that make up the majority of the starter culture. Still, a lot of knowledge about the effect of the different species and strains on long term ripening and flavor development remains unknown.

In this project, we went beyond the conventional cheese-making microbial starter culture composed from three microbial species. We explored diverse microbial community compositions by introducing five additional species crucial for cheese ripening across various targeted experiments. We uncovered how the complex interplay of these increasingly diverse microbial communities can shape the flavor profile during cheese-making.

By analyzing metabolomics, microbial dynamics and genomics data in context of sensory data, we identified some key compounds that drive flavor diversification. Through an investigative approach using statistical analyses and dimensionality reduction we identified *Lactobacillus helveticus* as a key provider of amino acids and indirect inhibitor of galactose utilization. This is a prime example of how introducing certain 'roles' can hugely impact the final metabolome. Additionally, comparative genomics and genome scale metabolic modeling suggest that strain level variation can further drive the flavor profile in a certain direction. This research serves as a stepping stone towards employing complex microbial communities to change an environment in a desirable direction

[Go back to the poster presentations overview](#)

Estimation of the number of cancer driver mutations induced in human tissues

Presenter: Axel Rosendahl Huber

All cells in the body accumulate DNA mutations, and a fraction of these mutations have cancer driving properties. The billions of cells in our tissues and the several thousand mutations most cells accumulate during life suggest that a large number of cells acquire driver mutations during life. Knowledge on the number of driver mutations induced during life can help us to better understand the role of mutagenesis, and determine if mutagenesis is a rate-limiting factor for cancer progression. Here, we developed a method to estimate the number of driver mutations induced across tissues in the absence of selection. As input, we use mutation data of hematopoietic stem cells, lung basal epithelial cells, colon crypt stem cells, pancreas and skin (melanocytes, keratinocytes). For each tissue, we calculate neutral mutational rates for different mutational processes. These processes are represented as mutational signatures, and range from endogenous processes to exogenous processes, such as smoking and chemotherapy. In addition, the method allows for regional differences in mutational rates across genes. To extrapolate mutation rates to entire tissues, cell counts per tissue are obtained from the Human Cell TreeMap study. Calculations suggest every position across the genome is mutated multiple times across all studied tissues, with the exception being the hematopoietic system. We estimated in colorectal crypt stem cells the number of cells with two APC driver point mutations, driving the development of adenoma. In an average 55-year old male individual in 2-3 cells APC both alleles are estimated to be mutated. In patients with POLE germline mutations, which suffer from dramatically increased mutation loads, we estimate 65 cells with both APC alleles mutated. In future efforts, we aim to perform these estimates across all tissues and mutagenic exposures, enabling us to better understand the role of mutation induction in the development of cancer.

[Go back to the poster presentations overview](#)

Comparative analysis of cardiomyocyte-fibroblast paracrine signaling involved in the myocardial ischemic tolerance of *Acomys cahirinus*

Presenter: Arne Kukkonen

The pathological remodeling after myocardial ischemia is the leading direct cause of death worldwide. The propensity of the infarcted adult human heart to enter into a dangerous mode of months-long fibrosis, inflammation and cardiomyocyte death warrants research into regenerative medicine approaches. *Acomys cahirinus* is the first mammal reported to have the ability for highly adaptive, non-fibrotic regeneration of various tissues, including skin and ear cartilage - a trait lost in most mammals past the neonatal stage. Recently, it was demonstrated that *Acomys* has a higher tolerance to myocardial infarction (MI) than *Mus musculus*, characterized by reduced pathological remodeling and increased survival monitored up to 100 days post-MI. Studying cardiac ischemic tolerance in a mammalian model could potentially help to translate the findings into clinical context, which may be more limited in the case of other, more evolutionary distant regenerating organisms, such as Caudates (e.g. salamanders), flatworms and reptiles. This study focuses on the critical cardiomyocyte-fibroblast cross-talk of the post-MI heart, where direct-contact and paracrine signaling from the activated fibroblasts affect the function of cardiomyocytes, most strongly in the border zone surrounding the fibrotic scar area. The present study uses spatially-resolved cryosection RNA sequencing (tomo-seq) in a rodent model of myocardial infarction, combined with the bulk cell RNA-seq of fibroblasts from the animals, and hiPSC-derived cardiomyocytes treated with the fibroblast condition medium, to study how the differential ligand expression in animal fibroblasts could lead to the reduced inflammation markers observed in the condition-medium-treated cardiomyocytes. The hypotheses generated by the RNA-seq analysis are subsequently tested by in-vitro experiments.

[Go back to the poster presentations overview](#)

Technical factors influencing transcriptomic allelic distribution

Presenter: Ivo Borkus

Understanding the allelic distribution of heterozygous variants in transcriptomic data provides critical insights into the diverse phenotypic outcomes observed in human cardiomyopathy. Cardiomyopathies, such as hypertrophic cardiomyopathy (HCM), which often result from pathogenic variants in key genes, can present with widely variable clinical outcomes, even among individuals carrying the same mutation. This variability is hypothesized to be influenced by the allelic distribution of these variants, which may be shaped by a range of technical factors during data acquisition and analysis.

This study aims to explore the impact of technical factors on allelic distribution in transcriptomic data across multiple cohorts. We assess factors such as sequencing depth, variant quality, allelic concordance, variant concordance and mapping bias, evaluating their influence on the detection and quantification of allele-specific expression (ASE). Through an integrated analysis of transcriptomic data from various cohorts, we systematically dissect how these technical variables may skew allelic expression measurements and contribute to inconsistent interpretations of genotype-phenotype relationships.

By controlling for and understanding these technical influences, we aim to clarify the connection between heterozygous variants and their functional outcomes in cardiomyopathy. This work is essential in identifying reliable biomarkers and improving the accuracy of transcriptomic analyses, thus offering a deeper understanding of how allelic distribution shapes disease manifestations. Ultimately, these findings will guide future research and clinical approaches to the management of cardiomyopathy, contributing to the development of more personalized therapeutic strategies.

[Go back to the poster presentations overview](#)

Extensive Gene x Environment interactions in the outbred vertebrate genetics of a standard physiological phenotype

Presenter: Saul Pierotti

The collection, phenotyping, and genotyping of large numbers of unrelated humans for GWAS analysis have become routine in the last two decades. However, the paucity of environmental measures in human observational studies has meant that it is complex to estimate environment-specific genetic effects (G×E) in humans, despite extensive evidence of their existence in model organisms and farm animals. This inability means that we do not know what we are potentially missing as G×E effects, limiting the discoverability of biological mechanisms, and overlooking potentially modifiable environments.

The Japanese rice paddy fish, medaka, is an ideal model system for exploring this problem because wild medaka fish can be inbred to homozygosity, allowing for the establishment of a wild-derived genetic mapping panel. Here, we report the first systematic screening of G×E and G×G effects in vertebrates from wild-derived haplotypes. We focused on the heartbeat of medaka embryos at different temperatures. Using an advanced F2 cross design with 2209 F2 individuals, we discovered 16 loci, of which 10 (62.5%) had statistically clear G×E effects. We recapitulated the effects of the discovered loci using CRISPR techniques in at least four instances. Significant G×G and G×G×E effects were also observed.

We then deliberately mimicked the human scenario of each individual being measured only in one randomly assigned environment with varying levels of environmental measurement error. Using our discovered loci and effects as the ground truth, we could simulate arbitrary sample sizes and allele frequency distributions. We observed that increasing sample size can provide partial, but not complete, mitigation for a poor environmental measurement.

In summary, we show that the outbred vertebrate genetics of a standard physiological phenotype has substantial G×E effects, with consequences on the completeness of our knowledge of human genetics.

[Go back to the poster presentations overview](#)

Non-canonical peptides presented on MHC molecules as possible biomarkers for (immune)therapies

Presenter: Manuel Ayllongavilan

Recent advances in neoantigen research have accelerated the development and approval of tumor (immune)therapies, particularly antibody-based treatments for solid tumors. Neoantigens, arising from tumor-specific changes such as genomic mutations, dysregulated RNA splicing, and post-translational modifications, are newly formed antigens in tumor cells. The identification and prediction of these neoantigens have been facilitated by advancements in next-generation sequencing (NGS) and bioinformatics. They can be common between patients due to their independence of individual genetic mutations. These characteristics make them promising cancer-specific biomarkers that can be targeted in cancer treatment. However, there are several biological processes that need to take place in order to have a neoantigen on the cell surface, e.g. non-canonical translation and presentation by the major histocompatibility complex (MHC). In this study, we combine computational immunogenomics and large omics-data with machine learning (ML) to identify and validate non-canonical (cryptic) peptide candidates presented by class I MHC. As a result, several highly promising candidate biomarkers for cancer (immune)therapies were computationally identified.

[Go back to the poster presentations overview](#)

Investigating the Evolutionary Origins of Winged-Helix Domains in the Dam1 and Ska Complexes of the Eukaryotic Kinetochore

Presenter: Nicola Greco

The kinetochore is a vital macromolecular complex that ensures accurate chromosome segregation during eukaryotic cell division by mediating attachment to spindle microtubules. The Ska1 complex, found in metazoans, and the Dam1 complex, typically observed in yeast, reinforce this kinetochore-microtubule interface by tracking depolymerizing microtubule tips, which is essential for the precise movement of chromosomes during cell division. Despite their analogous functions, these complexes lack apparent homology, and recent phylogenetic evidence suggests that both were present in the last eukaryotic common ancestor (LECA), prompting questions about their evolutionary origins and functional redundancy.

Recent protein structure predictions have revealed that certain proteins within both Dam1 and Ska complexes contain winged-helix domains (WHDs), typically known for DNA binding but implicated here in microtubule interactions. The shared presence of WHDs in both complexes suggests either convergent evolution or underlying homology, making it essential to determine the evolutionary pathways that led to their incorporation into these kinetochore components. However, the small size and highly divergent nature of these domains have rendered traditional phylogenetic approaches ineffective in resolving their relationships.

To unravel the evolutionary history of the WHDs within the Dam1 and Ska complexes, this project employs a combination of highly sensitive homology search tools—including profile-based HHsearch and structure-based Foldseek—and structure prediction methods to identify and compare WHDs across a wide range of eukaryotic species. By constructing homology networks and applying clustering algorithms, the homology relationships between multiple proteins or orthogroups are visualized and analyzed. These networks provide a comprehensive view of the homology landscape and facilitate the identification of potential evolutionary groupings.

By elucidating the origins and evolutionary trajectories of these domains, this research seeks to enhance the understanding of kinetochore evolution and the mechanisms underlying chromosome segregation.

[Go back to the poster presentations overview](#)

Lys716 in the transmembrane domain of yeast mitofusin Fzo1 modulates anchoring and fusion

Presenter: Raphaëlle Versini

Outer mitochondrial membrane (OMM) fusion is an important process for the cell and organism survival, as its dysfunction is linked to neurodegenerative diseases and cancer. The OMM fusion is mediated by members of the dynamin-related protein (DRP) family, named mitofusins. The exact mechanism by which the mitofusins contribute to these diseases, as well as the exact molecular fusion mechanism mediated by mitofusin, remains elusive. We have performed extensive multiscale molecular dynamics simulations using both coarse-grained (Martini 2 and 3) and all-atom approaches to predict the dimerization of two transmembrane domain (TM) helices of the yeast mitofusin Fzo1. We identify specific residues, such as Lys716, that can modulate dimer stability. Comparison with a previous computational model reveals remarkable differences in helix crossing angles and interfacial contacts. Overall, however, the TM1-TM2 interface appears to be stable in the Martini and CHARMM force fields.

Replica-exchange simulations further tune a detailed atomistic model, as confirmed by a remarkable agreement with an independent prediction of the Fzo1-Ugo1 complex by AlphaFold2.

Functional implications, including a possible role of Lys716 that could affect membrane interactions during fusion, are suggested and consistent with experiments monitoring mitochondrial respiration of selected Fzo1 mutants.

[Go back to the poster presentations overview](#)

Exploring fundamental mutational patterns of recessive DNA repair syndromes using single-cell whole genome sequencing

Presenter: Anne ten Vergert

Recessive DNA repair syndromes encompass a group of rare genetic cancer predisposition disorders caused by biallelic pathogenic germline variants in genes related to DNA repair. Two aggressive examples, constitutional mismatch repair deficiency (CMMRD) and ataxia-telangiectasia (AT), often lead to development of tumors affecting the hematological system and brain. In this research study, we investigate the fundamental mutational patterns left by defective DNA repair in healthy bone marrow cells. We performed clonal expansion and primary template-directed amplification (PTA) followed by whole genome sequencing to obtain genetic information from single non-malignant hematopoietic stem and progenitor cells (HSPCs) of 2 patients with AT and 2 patients with CMMRD. Lineage trees displayed minimal sharing of variants between cells compared to the mutational load of the individual cells, showing no signs of clonal growth. One specific CMMRD patient had a very high mutational load with a mean of 6965 mutations, compared to a mean of 236 for AT patients and 605 for the other CMMRD patient. This can be attributed to receiving temozolomide treatment for primary glioblastoma before we obtained the healthy bone marrow sample. In a select number of HSPCs from patients with CMMRD we also observed the mutational signature related to temozolomide treatment (3/13). Possibly indicating that some cells are more likely to get affected by the treatment than others. Furthermore, we observed SBS26 which is a known MMRD signature also observed in tumors. Therefore, the fundamental mutational patterns of non-malignant HSPCs are comparable to that of tumors in CMMRD. On the contrary, most of the signatures we found in HSPCs from AT patients were not observed in their tumors. This could indicate that these cells might need additional alterations to obtain tumor cell characteristics.

[Go back to the poster presentations overview](#)

Creating a hydraulic atlas of *Sorghum bicolor* to investigate the role of suberization during drought stress

Presenter: Noémi Versluijs

Global warming requires increasing knowledge regarding crop protection to ensure food availability. An increasing issue in agriculture is drought, reducing crop yields. We therefore investigate how *Sorghum bicolor*, a drought-resistant crop, uses suberin to increase its drought resistance. Suberin is a hydrophobic barrier that can be produced in both the exo- and endodermis of roots. This research aims to investigate the effectiveness of suberin-based drought protection in different parts of the root while additionally researching potentially optimized approaches. To do this, we combine experimental data with computational models.

Experimental data provided insights into the developmental and anatomical aspects of the roots, as well as the location and intensity of the suberin lamellae in the root. This showed a clear gradient from suberin presence or absence and intensity in both the exo- and endodermis.

With these insights, we then parameterized a biophysical 2D-model of radial water flow. This model showed that under non-drought conditions, passage cells in the endodermis have an advantageous effect on flow rates, increasing water uptake from the soil into the xylem. However, under drought conditions, suberization is preferred to prevent a reverse flow back into the soil. Additionally, it was found that suberization in the exodermis does not substantially affect the flow rate. This highlights the importance of developmental suberization patterns and soil water potential for radial water flow. Additionally, we will use a 3D-model to simulate axial water flow through the root and soil-root interactions. With the combination of the parameterized 2D-model and the spatially explicit 3D-model, multiple anatomical and environmental scenarios can be simulated to investigate optimal suberization strategies under different environmental conditions.

[Go back to the poster presentations overview](#)

Benchmarking pangenome graph construction methods for their application to foodborne pathogens

Presenter: Julian Paganini

Background. Foodborne pathogens like Shiga-toxin producing *Escherichia coli* (STEC) exhibit significant genomic diversity. Accurately capturing this diversity is crucial for epidemiological and evolutionary studies.

Pangenome graphs are powerful tools for representing variation in microbial populations. Various methods exist for constructing these graphs, which use unitigs, clusters of orthologous genes (COGs), or multiple sequence alignments (MSA) to represent sequence (dis)similarity. Selecting the most appropriate method for downstream analyses poses a challenge.

Here, we compare the performance of multiple state-of-the-art pangenome inference methods using STEC genomes and examine the impact of using draft assemblies, complete genomes, or both.

Methods. Complete STEC genomes (n=175) and corresponding short reads were obtained from public databases. Short reads were assembled into draft genomes. Based on previous benchmarks and node representation method, six tools were selected: Bifrost and Cuttlefish (unitigs), PPanGGoLiN, Panaroo, and ggcaller (COGs), and Pangraph (MSA). Various proportions of complete genomes (0 - 100%) were mixed with draft assemblies and used as input for each tool. Network descriptive statistics compared the output graphs, including node number, length, degree distribution, edge number, and subgraph size. The accuracy of Shiga toxin (stx) gene cluster content determination was evaluated using recall and precision.

Results. Unitig-based methods (Bifrost, Cuttlefish) produced graphs with higher node and edge counts compared to COG (Panaroo, PPanGGoLiN, ggcaller) and MSA (Pangraph) methods, suggesting a more fine-grained representation of genomic diversity. However, Bifrost yielded more subgraphs (n=324) compared to Cuttlefish (n=4). Incorporating complete genomes decreased node and edge counts for unitig-based tools but increased them for COG-based tools, both explainable due to improved genomic context disambiguation. When the stx cluster was assembled into a single contig, all tools had a recall above 0.95. When fragmented, recall dropped below 0.15, but improved with complete genomes (Panaroo=0.7, ggcaller=1, PPanGGoLiN=0.5). ggcaller's precision decreased with complete genomes in both cases.

Conclusions. Unitig-based methods offer finer genomic resolution, but the increased number of nodes might complicate the construction of predictive machine learning models with limited datasets. Incorporating complete genomes enhanced most metrics explored in this work.

SNV detection and lineage tracing in single cells using long read RNA sequencing

Presenter: Oscar Esteban

Mutations play a critical role in cancer, as some of these will confer growth advantages over neighboring cells and drive tumor progression. This leads to selection and proliferation of specific clones, a process named clonal expansion. Using single cell RNA-sequencing (scRNA-seq) one can analyze the transcriptional heterogeneity of clones. However, the majority of scRNA-seq is processed using short-read sequencing (SRS) technologies such as Illumina. This method relies on reads of around 100bp into the 3' or 5' end of transcripts, providing an adequate coverage for transcriptomic analysis but limiting mutation detection. Therefore, linkage between clonality and transcriptomic heterogeneity remains an outstanding problem in the field.

Long-read sequencing (LRS) has recently been made compatible with scRNA-seq, supporting reads of up to 5000bp and substantially improving coverage. In this study, we are evaluating Oxford Nanopore Technologies (ONT), a type of LRS, comparing its performance to Illumina. To do so, we are working with a library coming from two Acute Lymphoblastic Leukemia (ALL) patients and sequenced with both technologies. First, we assessed ONT's performance at transcriptomic analysis. We then proceeded to do SNV calling using scNanoGPS, a tool specifically designed for single cell ONT data. With this, we obtained single nucleotide variations (SNVs) with single-cell resolution using ONT reads only. We are now working on linking the transcriptomic profiles of tumor cells to the subclones found in our samples, as well as obtaining a lineage tree of such cells.

With this study, we hope to bridge diverse sources of heterogeneity in cancer and gain further understanding of tumor progression.

[Go back to the poster presentations overview](#)

Single-Cell Classification and Disease Gene Identification in ALS and Alzheimer's Using Simple Classification Models

Presenter: Rutger Ballieux

This study investigates the use of Logistic Regression and Random Forest classifiers to analyse single-cell RNA sequencing (scRNA-seq) data from Alzheimer's disease (AD) and Amyotrophic Lateral Sclerosis (ALS) patients. The primary goals were to assess whether these models can accurately classify healthy and diseased cells despite challenges like high dimensionality and data sparsity, and to evaluate their ability to identify key genes potentially involved in disease mechanisms. Multiple methods of feature extraction, including random features, most variable features, principal components, Differential Expression, and WGCNA, were used and compared. Results suggest promising classification performance across different feature extraction methods, and the genes identified as important by the classifiers appear to be closely associated with known disease pathways.

[Go back to the poster presentations overview](#)

PhD courses on bioinformatics basics

Presenter: Laura Dijkhuizen

The Graduate School of Life Sciences (GSLs) and the UBC facilitate basic bioinformatics courses for all GSLs PhDs free of charge! In this poster, we present you an overview of these courses. Additionally, you can chat with the teachers to see if a course fits your goals and skill level.

[Go back to the poster presentations overview](#)

Upsimpler: Data augmentation for ultra rare (sub)types in RNA-seq based pediatric tumor classification

Presenter: Sotiris Niarchos

Accurate diagnosis of rare malignancies is crucial for optimal treatment and patient outcome in childhood cancer. However, diagnosticians often lack familiarity with the visual and molecular aspects of such tumor types due to their rarity. Machine learning (ML) offers a promising approach to develop tools that can assist pathologists in establishing these diagnoses. Nonetheless, training effective ML models typically calls for a substantial number of samples. Therefore, rare tumor (sub)types are often excluded due to their under-representation in existing datasets.

In this work, we introduce upsimpler, a data augmentation method for RNA-seq count data. The core concept behind the upsimpler algorithm is the generation of synthetic data points by subsampling the biological signal in the local neighborhood around a sample of interest.

Ultimately, upsimpler will be integrated into M&M, our in-house pan-cancer pediatric tumor classifier. The project's end goal is the inclusion of (sub)types with as few as 2 existing samples in M&M.

[Go back to the poster presentations overview](#)

Explaining unexpected deuterium labeling curves

Presenter: Vince During

DNA deuterium labeling is a powerful in vivo method for tracking the dynamics of populations of cells. Notwithstanding, the effectiveness of this method, interpreting the data can pose significant challenges, and there is a considerable debate on choosing the best mathematical models to interpret this data. We focus on explaining “strange” late data points. We often find a higher level of label in the data than any of these models predicts. Comparing models where either the lifespan, or the death rate, of cells follows a Gamma distribution, we find different mathematical solutions, but very similar fits to the data. Reassuringly, we find that a simple two-population model that is often used for fitting deuterium data, is perfectly capable of fitting data generated by both models based upon Gamma distributed parameters. The search for an explanation of the high late data points therefore continues.

[Go back to the poster presentations overview](#)

The Dutch childhood cancer genome project: Characterizing tumor drivers

Presenter: Joanne von Berg

Despite improvements in cure rates, cancer is still the leading cause of disease-related deaths among children in high-income countries. As childhood cancers are rare in comparison to adult cancers, concerted efforts are needed to advance research and improve patient care. In The Netherlands, all childhood cancer care and most research has been concentrated in a single national center, the Princess Máxima Center for Pediatric Oncology.

The institute's centralized position enables uniform generation of data from patients across the country. Whole-exome sequencing, whole-genome sequencing and RNA-sequencing are routinely performed for diagnostic and research purposes, leading to a representative national dataset with little batch effects. Currently this data collection consists of 1184 patient samples with rich and uniform clinical data. This collection is expected to grow to ~4000 patient samples by 2028.

We used this rich data resource to study the role of different types of somatic variation in tumor initiation. We called somatic single nucleotide variants, copy number alterations, and structural variants. Through integrative analyses we can identify tumor driver candidates that are likely missed by approaches that focus on one variant type. On the gene level, mutations seem to predominantly occur in certain primary groups (collections of similar tumor types). On the pathway level however, mutations seem to occur across all primary groups. Ultimately, we anticipate that the data collection presented here will further facilitate pediatric cancer research and provide an invaluable resource for precision oncology.

[Go back to the poster presentations overview](#)

Modelling of nanobody-antigen complexes using machine learning and information-driven docking

Presenter: Miguel Sánchez Marín

Nanobodies are small variable antigen-binding domains derived from camelid's heavy-chain antibodies (HCAbs). These proteins exhibit binding affinities of the same order than those of antibodies, which, along with their small size and unique structural characteristics makes them desirable for therapeutical and diagnosis applications. As for the heavy-chain variable region (VH) of antibodies, nanobodies are composed of four conserved framework regions (FR1/FR2/FR3/FR4) and three antigen binding hypervariable loops (CDR1/CDR2/CDR3). CDR3 in nanobodies is longer and more flexible than in antibodies, on average, and also exhibits a higher sequence variability than other regions. As a result, the CDR3 region of nanobodies can sample a very diverse loop conformational space and is considered the main contributor to antigen binding. This broad conformational space represents a sampling challenge for computational modelling methods. Moreover, the lack of coevolution signals in nanobody-antigen complexes also poses challenges to ML-based protein-protein interaction prediction methods. As such there is still an unmet need for accurately modelling this type of complexes. In this context, physics-based information-driven docking software like HADDOCK is likely to provide a solution as already demonstrated for antibody-antigen interaction modelling. In this project, we assembled a non-redundant nanobody-antigen crystal structure dataset and assessed the state-of-the-art of AI-based nanobody and nanobody-antigen structure prediction algorithms. Using different ensembles of nanobody predictions, we then benchmark different HADDOCK workflows using various information scenarios, with the aim of defining a nanobody-antigen prediction method from sequence that outperforms the best-to-date.

[Go back to the poster presentations overview](#)

A discrete stochastic model on the dynamics of choanoflagellate chain colonies

Presenter: Anna Delhaas

The transition to complex multicellularity from unicellular organisms was a major step in the life history of animals, however how this transition happened is not completely understood. To better understand the evolution of multicellularity, researchers study organisms that exhibit simple multicellularity. A defining feature of these colonies is their size. Although colony size has been modeled in various species, no model has yet focused specifically on choanoflagellate colonies, which are the closest living relatives of animals. To gain deeper insight into the origins of multicellularity in animals, this thesis presents a simple discrete stochastic model to simulate choanoflagellate chain colonies. The model focusses on the interplay of division and breaking of cytokinetic bridges and how their underlying waiting time distributions affect the distribution of colony sizes. These events are modelled using various probability distributions and, through Maximum Likelihood Estimation (MLE) and fitting to experimental colony size data, determined that the waiting times for division are best represented by a gamma distribution, while those for bridge breaks follow an exponential distribution. Additionally, the model revealed that choanoflagellate chain colonies reach a steady state in the average number of cells per colony. We analyzed how the rates of cell division and bridge breakage influence this steady state, finding that the division rate is inversely related to colony size, whereas the bridge break rate follows a logarithmic relationship. Furthermore, a mathematical description to describe these steady states is established. These results offer new insights into the colonies of choanoflagellates, the closest living relatives of animals, and may therefore enhance our understanding of the origin of animal multicellularity.

[Go back to the poster presentations overview](#)

Biologic therapies imprint epigenetic modifications in the skin-homing T cells of atopic dermatitis patients

Presenter: Hidde Smits

Introduction. Atopic dermatitis (AD) is a chronic skin inflammatory disease marked by Th2 immune responses dominated by cytokines IL4, IL13, IL5, and IL22. New biologic therapies like dupilumab (an IL-4 receptor blocker) and upadacitinib (a JAK inhibitor) have transformed AD management. Long-term therapy is required to prevent flares. Dupilumab leads to prolonged remission after discontinuation, while upadacitinib is associated with faster relapse, suggesting dupilumab may have disease-modifying effects. DNA methylation, an epigenetic mechanism, has been studied in AD patients, particularly in CD4+CLA+ T cells. However, it's unclear if long-term treatment with biologics affects AD's epigenetic profile or induces new changes.

Approach. CD3+CD4+CLA+ T cells were isolated from AD patients at baseline and after 52 weeks of treatment with dupilumab (n=6) or upadacitinib (n=6), along with non-atopic patients (n=6). DNA methylation was assessed using Illumina EPIC bead array v2.0, and RNA sequencing was performed. Potential disease-modifying regions were identified using previously characterized SNPs and publicly available Hi-C data from CD4+ T cells and keratinocytes.

Results. We identified 303 significant differentially methylated regions (DMRs) between non-atopic and AD patients before treatment. These DMRs were linked to pathways such as cytokine receptor interactions, NF- κ B, and T-cell receptor signaling. After 52 weeks of dupilumab treatment, 7 DMRs shifted toward a non-atopic profile. Conversely, after upadacitinib treatment, 40 AD-related DMRs increased in the AD phenotype, and 214 non-AD-related DMRs emerged, indicating broader epigenetic changes.

Discussion. These findings highlight distinct epigenetic impacts of dupilumab and upadacitinib, suggesting dupilumab may have more favorable long-term effects. The different modification patterns could have clinical significance upon treatment discontinuation.

[Go back to the poster presentations overview](#)

A detailed look into the intricate decision-making framework guiding cell-differentiation

Presenter: Maria Karaboeva

Gastruloids are faithful models of embryonic development for mice which have been shown to form somites and showcase striking similarities to in vivo grown embryos in gene regulation (Susanne C. van den Brink, 2020). Here we employ single-cell RNAseq to obtain a deeper insight into the cell differentiation, cell types and genes involved in early development. Unperturbed development is being followed via the novel technique VASA-seq (full-length RNAseq which accounts for most types of RNA) of gastruloids in the stages 0h to 168h. This time frame allows us to closely follow the beginning of hematopoiesis and somitogenesis (Rossi, 2022) (Susanne C. van den Brink, 2020). Additional 10x Chromium RNAseq single-time-point samples in distinct types of differentiation media (HD, Hanging Drop (ESLIF –LIF +gelatin); EB, Embryonic Body (ESLIF –LIF gelatin) and RA, Retinoic Acid (FCS + LIF + retinoic acid)) are being assessed to show the effects of perturbation on cell fates. We utilize the wealth of information offered by single cell data (Lukas Heumos, 2023) using methods such as RNAvelocity (implemented in scVelo (Volker Bergen, 2020)), compositional analysis (scCODA (M. Büttner, 2021)), gene-set enrichment and inference of gene regulatory networks (GRNs with SCENIC (Sara Aibar, 2017)). We expect to find out the dynamic changes in transcriptomes including which genes contribute most to the speed of reaching a certain cell fate, how are cell fates and gene pathways altered in the listed conditions and suitable explanations of these effects in the context of gene regulation. The data preprocessing workflows for our data scChICflow (tchic protocol for VASAseq) and snakepit (for 10x Chromium) are openly available in GitHub under <https://github.com/bhardwaj-lab>.

[Go back to the poster presentations overview](#)

Computational design of coronavirus entry inhibitors

Presenter: Judith Ahr

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the COVID-19 pandemic have underscored the urgent need for strategies to mitigate the threat posed by zoonotic respiratory viruses. In response, computational design has become a key tool in the development of biologic therapies that target viral entry mechanisms. Recent advances in deep-learning models for protein design have shown remarkable potential in generating high-affinity binders for viral targets, such as those involved in influenza A hemagglutinin interactions. These models, coupled with high-resolution structural data of viral proteins and identified conserved epitopes, provide a powerful framework for designing inhibitors that block viral entry by disrupting key interactions between the virus and host cells.

By leveraging generative deep-learning techniques, it is possible to rapidly refine interactions at the molecular level and optimize viral entry inhibitors to increase their affinity, specificity, and breadth of action. This approach is particularly valuable for developing therapeutics capable of neutralizing multiple strains of betacoronaviruses, including SARS-CoV-2 variants. The integration of in silico design with structural biology enables the rapid generation and evaluation of antiviral therapeutics, laying the groundwork for future pandemic preparedness and the development of broadly acting antiviral agents.

[Go back to the poster presentations overview](#)

Integrating Pharmacogenetics Upfront: Impact of Early Pharmacogenetic Screening in a Pediatric Oncology Cohort of 1,151 Subjects

Presenter: Emma Bernsen

Pharmacogenetic guidelines are underutilized in pediatric oncology, primarily because most are derived from adult data and are not directly applicable to children. However, pharmacogenetic screening can facilitate upfront dose adjustments, thereby reducing the risk of drug-related toxicity. This study aimed to retrospectively assess the potential impact of pharmacogenetic screening in a national pediatric oncology cohort. We evaluated the eligibility for dose or treatment adjustments based on genetic profiles and drug use in 1,151 subjects. Nine genes and 13 drugs- or drug groups were selected for analysis. Due to the lack of evidence based pharmacogenetic guidelines for widely used anticancer drugs such as vincristine, doxorubicin, and methotrexate, we primarily included supportive care drugs, as well as thiopurines, irinotecan, and 5-fluorouracil.

Among the 1,151 subjects, 88% exhibited a non-normal phenotype for at least one gene, and 463 subjects had a potential actionable phenotype warranting clinical intervention. In total, for 167 first-use drug prescriptions drug or dose adjustments were available, with the majority of these interventions related to supportive care medications. Our findings suggest that pre-therapeutic pharmacogenetic testing in pediatric oncology could enable tailored drug dosing.

[Go back to the poster presentations overview](#)

Analysing alternative polyadenylation in ALS using a predictive neural network

Presenter: Mees Klunder

The predictive neural network model APARENT2 was used to analyse the effect of variants in amyotrophic lateral sclerosis (ALS) patients on alternative polyadenylation (APA). Variants from whole-genome sequencing data of both ALS patients and healthy controls (n=6538, n=2415, resp.) were scored with the APARENT2 model. No difference was found between the frequency of PAS-affecting variants in ALS patients versus healthy controls. A key limitation was the complex regulatory system surrounding pA, making it difficult to predict the effect up- or downregulation of a particular polyadenylation site (PAS) might have on the overall transcripts. This means we cannot exclude the absence of ALS associated PAS altering variants. Attempts to train an ALS specific APA-prediction model on TDP-43 knockdown PAS data did not result in a better-performing model with the training-data available. APARENT2 was used to predict the strength of newly identified PAS's in motor cortex tissue. The model scored these supposed brain-specific PAS's similarly to previously identified ones and the sequence around these PAS's showed enrichment for the CUX1 RNA binding site, which is a transcription factor mainly used in the brain. The outcomes of the APARENT2 model can be used to decide which brain-specific PAS's are to be used in further research on polyadenylation in the brain.

[Go back to the poster presentations overview](#)

Leveraging deep learning to learn features of antibody recognition

Presenter: Markel Benito

Recent advancements in deep learning techniques have demonstrated promise in a wide range of tasks relating to protein engineering. This study explores the potential of integrating deep learning techniques with immunology to enhance vaccine and antibody design. Traditional methods, such as ELISA, have provided insights into immune responses, but lack detailed understanding of antigen recognition mechanisms. By leveraging extensive experimental data on antibody-antigen interactions, we developed a novel in-silico pipeline relying on Protein Language Models (PLMs) which aims to optimise protein sequences for adjustable immune recognition. Our results suggest that pre trained and fine-tuned PLMs propose divergent mutations when supplied the same sequencing data derived from yeast surface display. Furthermore, changing the architecture of the underlying PLM impacted the number of proposed mutations and the corresponding location. Together, our findings demonstrate the ability of PLMs to learn biophysical properties of proteins features of proteins and highlight how PLMs can be used to suggest mutations relevant to protein engineering efforts.

[Go back to the poster presentations overview](#)

Carbon-fixing and viral-infected *Ferroplasma* dominates acidic cave biofilm

Presenter: Chrats Melkonian

Microbiologists have long sought to investigate the boundaries of life. With the aim to discover greenhouse-gas-eating microbes, we studied the microbial community in an extremely acidic biofilm of pH 1 in a volcanic cave in Romania, which was highly enriched in carbon dioxide and methane.

Previously, we enriched a methanotrophic *Mycobacterium* from this cave, expanding our knowledge of the possible types of metabolism in the Actinobacteria phylum. Here, we recovered twenty high-quality metagenome-assembled microbial genomes using metagenomics from caves biofilm. In addition to *Mycobacterium*, we found two new and dominant *Ferroplasma* Archaea species from different phylogenetic clades with distinct genetic content. Our analysis using pan-genomics suggests that both *Ferroplasma* share amino acid auxotrophies, pointing towards community interdependencies. A metaproteomics analysis revealed that the two strains have the potential to fix carbon dioxide for growth via the reductive tricarboxylic acid cycle and utilise the cave's available sulfur compounds. Indeed, proteins annotated as sulfur oxygenase/reductase and sulfide:quinone oxidoreductase were among the most abundant. Simultaneously, we found an unprecedentedly high abundance of *Ferroplasma* CRISPR proteins. Further virus-specific bioinformatics analyses identified novel viruses, many of which were predicted to infect *Ferroplasma*.

Overall, we characterised a diverse microbial community that includes two carbon-dioxide-fixing *Ferroplasma* species. It thrives in highly acidic conditions and it consumes greenhouse gases. Despite the predicted auxotrophies and signs of ongoing viral infection, both species of *Ferroplasma* were dominant community members within the cave's biofilm.

[Go back to the poster presentations overview](#)

Machine Learning-based analysis of plant phenomics data to better understand the growth-defense trade-off

Presenter: Marta Garcia Mondejar

The growth-defense trade-off is a biological process where the activation of immune responses inhibits growth and development. Despite its fundamental nature, the mechanisms underlying this trade-off remain poorly understood. Here, we used a lettuce diversity panel, immune response-activating chemicals, automated digital phenotyping tools, and genome-wide association studies (GWAS) to determine the genetic architecture of the growth-defense trade-off in cultivated lettuce.

Nearly 200.000 top-view RGB images were taken to estimate the growth and health of individual plants over time. Since available image analysis tools were unable to properly segment lettuce plants, owing to their diverse colors and shapes, we developed a custom analysis pipeline in R based on a K-Nearest Neighbours (KNN) algorithm. This supervised machine learning model was trained on a subset of manually labelled images to classify each pixel as either background or foreground (plant) based on color characteristics. This resulting binary mask allowed us to quantify the plant pixels and use this measurement as a proxy for plant size, enabling us to track growth across the different genotypes and treatment conditions.

The resulting data will be used for GWAS to uncover genetic loci associated with growth-defense responses. The insights gained from this research have the potential to guide the development of crops that are both high yielding and resilient to pests.

[Go back to the poster presentations overview](#)