bio- & medical informatics research network Antwerp

biomina

RESEARCH DAY

18 DECEMBER 2023 - ANTWERP

University of Antwerp

biomina biomedical informatics research network Antwerp **Research Day – 18/12/2023**

PROGRAM

9:00 | Registrations

9:30 | **Prof. dr. Kris Laukens**: Welcome & introduction to biomina

Morning session

10:00 | **Erin Sauve** - ITM: *Buzzworthy: Novel HIVE sequencing technology makes single-cell sequencing possible for malaria field isolates*

10:15 | **Dr. Wim Cuypers** - Adrem Data Lab: *A Leap Forward in Revolutionizing Infectious Disease Surveillance with Nanopore Sequencing*

10:30 | **Dr. Jamie Lustermans** - MST Centre of Excellence: *Extracellular electron transfer by candidate flocking bacteria in cable bacteria sediment*

10:45 | **Romi Vandoren** - Adrem Data Lab: *Unraveling the Complex T Cell Receptor-Microbiome Interaction Network in the Colon Through Advanced Computational Analysis*

11:00 | **Anwar Hiralal** - MST Centre of Excellence: *Closing the genome of unculturable cable bacteria using metagenomic hybrid sequencing*

Coffee & poster session (11:15-11:45)

11:45 | **Sebastiaan Valkiers** - Adrem Data Lab: *Detection of convergent immune responses through quantification of the T-cell receptor similarity space*

12:00 | **Julia Camacho** - EVECO: *Identification of candidate genes for adaptation to contrasting light environments along the water column in deep-water cichlid fishes*

12:15 | **Vincent Van Deuren** - Adrem Data Lab: *RapTCR: Rapid Exploration and Visualisation of T-cell repertoires*

12:30 | **Nicky De Vrij** - ITM/Adrem Data Lab: *Persistent T cell unresponsiveness associated with chronic visceral leishmaniasis in HIV-coinfected patients*

12:45 | **Prof. Hannes Svardal** - EVECO: *Sex, selection, and (barriers to) gene flow: the complicated evolutionary histories of large chromosomal inversions in benthic Malawi cichlids*

Lunch break & poster session (13:00-14:00)

Afternoon session

14:00 | **Nishkala Sattanathan** - MEDGEN: *WiNGS - Federated approach for genomics data sharing and analysis*

14:15 | **Dr. Kerry Mullan** - Adrem Data Lab: *STEGO.R for (easy) interrogation of combined scTCR repertoire and scRNA-seq data*

14:30 | **Christophe Verwimp** - MEDGEN: *Development and validation of a CNV analysis and visualization pipeline for WES data*

14:45 | **Charlotte Adams** - Adrem Data Lab: *Fragment ion intensity prediction improves the identification rate of non-tryptic peptides in timsTOF*

15:00 | **Dr. Tim Van De Looverbosch** - Laboratory of Cell Biology and Histology: *Accurate Nuclei Detection in 3D Cell Systems Through Centroid Prediction*

15:15 | **Ceder Dens** - Adrem Data Lab: *The pitfalls of negative data bias for the T-cell epitope specificity challenge*

Keynote speaker

15:30 | **Prof. dr. David Martens** - Antwerp Center on Responsible AI (ACRAI): *Responsible AI: Balancing Opportunities with Ethical Risks*

16:15 | Closing remarks

16:30 | Estimated end

Posters

Nicky de Vrij – ITM/Adrem Data Lab: *Human Leukocyte Antigen alleles as risk factors for visceral leishmaniasis development in HIV co-infected individuals*

Ana Regina de Abreu – MEDGEN**:** *Implications of genome-wide DNA methylation sequencing methods in the pursuit of cancer biomarker discovery*

Elise Coopman – VIB CMN: *Methylmap: visualization of nucleotide modifications for large cohorts.*

Sara Alidadiani - VIB CMN: *Transcriptomic analyses as part of the international FTLD-FUS consortium highlights role for mitochondrial dysfunction in disease.*

Dr. Geert Vandeweyer – UZA: *Accelerating Clinical Whole Exome Sequencing Analysis: Leveraging WDL and Cromwell on AWS Batch*

Janne Heirman – Adrem Data Lab: *Improved spectrum clustering increases interpretability of molecular networks*

Sarah De Beuckeleer – Lab. Of Cell Biology and Histology: *Unbiased identification of cell identity in dense mixed neural cultures based on nucleocentric phenotyping*

Dr. Juan Sebastian Piedrahita Giraldo – Adrem Data Lab: *Metabolite Identification and Analog Discovery through Deep Learning Modeling of Mass Spectrometry Data*

Thomas Vanpoucke – MEDGEN: *Scalable, flexible, and customizable GUI based workflow for multiple methylome data types*

Tijs Watzeels – VIB CMN: *Pipeline for cell-specific isoform calling from long-read single-nuclei sequencing*

Farhan Ul Haq - MEDGEN: *Differentiating ROS1 kinase states targeted by tyrosine kinase inhibitors*

Baukje Bijnens – VIB CMN: *Hexanucleotide repeat expansions in C9orf72 alter microglial responses and prevent a coordinated glial reaction in ALS*

Mahdi Safarpour – FAMPOP: *Unravelling the contribution of human mobility in malaria parasite persistence: current state and challenges towards elimination*

Discover our socials:

biomina
www.biomina.be

BIOMINA

biominabe.slack.com

@biominaBE

biomina@uantwerpen.be

Sign up for our Social Event later today (via WhatsApp)!

Share your contributions with #biominaRD2023 !

O.1: Erin Sauve (Institute of Tropical Medicine Antwerp)

Buzzworthy: Novel HIVE sequencing technology makes single-cell sequencing possible for malaria field isolates

Erin Sauve[1], Pieter Guetens[1], Pieter Monsieurs[1], Anna Rosanas-Urgell[1]
1: Malariology Unit, Institute of Tropical Medicine Antwerp, Belgium

Aim of the Study: Until now, single-cell sequencing of *Plasmodium* parasites has been limited to the lab due to resource requirements and a lack of preservation solutions. Here we present 3 parasite enrichment techniques which take advantage of well-known methods used for *Plasmodium* including Plasmodipur filters, MACS columns, and Percoll and Nycodenz gradients and combine them with HIVE single-cell capture to make single-cell sequencing possible for both lab cultures and whole blood from patients in the field.

Methods: HIVE single-cell sequencing utilizes small devices (HIVEs) to capture cells in a honeycomb-like pico-well system loaded by gravity or with a centrifuge. These devices can be stored and shipped worldwide to a laboratory for library preparation. Each HIVE is filled with one sample, making this system ideal for collecting samples as patients present to clinics or multiple time points where the ability to process them simultaneously can prove cost-effective.

Results: Our results from a *P. knowlesi* mock field sample, show that the regardless of enrichment methods used with HIVE single-cell sequencing, reproducible UMAPS include all intraerythrocytic life stages, with some variation in life stage composition due to the inherent differences between methods. We recovered between 2300 and 8800 high quality cells, after filtering based on thresholds of 300 genes and 600 transcripts per cell, with all samples covering at least 94% of *P. knowlesi* genes. When mapping the expression of housekeeping gene, pktubulin, across the UMAP, we found consistent expression across the life cycle and for pkDBP, a gene known to be involved red blood cell invasion, we found specific, high expression in the schizont stage, as expected.

Conclusions: By combining parasite enrichment with HIVE preservation, single-cell sequencing is possible in resource-limited settings. Our next step is to apply HIVE technology to capture *P. vivax* malaria isolates to answer biological questions including invasion and drug resistance.

O.2: Dr. Wim Cuypers (University of Antwerp)

A Leap Forward in Revolutionizing Infectious Disease Surveillance with Nanopore Sequencing

Wim L. Cuypers, Julia M. Gauglitz, Wout Bittremieux, Kris Laukens
Adrem Data Lab, University of Antwerp

Aim of the Study:

To tackle infectious diseases in low-and middle-income countries using nanopore sequencing for improved surveillance of zoonotic and vector-borne infections, our study focuses on two main objectives: (1) establishing adaptive nanopore sequencing as a benchmark for efficient pathogen DNA enrichment, and (2) creating 'squiDBase', a database of raw nanopore data, to facilitate the development of algorithms enhancing this sequencing method.

Methods:

Our proof-of-concept study employed adaptive nanopore sequencing for direct *Plasmodium* sequencing from blood samples, eliminating the need for prior enrichment. The focus was on enriching *Plasmodium* falciparum DNA in both controlled benchmark samples and patient samples with various parasitemia levels.

Utilizing PostgreSQL and Milvus, we have created a prototype database that is currently being populated with raw nanopore sequencing data of various pathogens, with an initial focus on arboviruses.

Results:

Adaptive nanopore sequencing significantly enriches *P. falciparum* DNA in patient samples, bypassing the need for time-consuming laboratory enrichment procedures. Key findings include:

- A 3-5 fold enrichment of *P. falciparum* bases in samples with 0.1-8.4% *P. falciparum* DNA.

- Successful sequencing of patient samples with parasitemias of 0.1%, 0.2%, and 0.6%, covering at least 97% of the *P. falciparum* reference genome.

- High concordance with Sanger sequencing in 38 drug resistance loci, suggesting the quality of sequencing data is adequate for clinical research in patients with parasitemias of 0.1% and higher.

Further enrichment will be achievable with faster raw signal chunk classification through algorithmic advancements. Our in-development database squiDBase, enables efficient querying of raw nanopore data for this purpose, filling a current gap in the field.

Conclusions:

Our study establishes a new benchmark in pathogen DNA enrichment using adaptive nanopore sequencing. With squiDBase we set the stage for future algorithmic enhancements aimed at accelerating pathogen pathogen monitoring and pioneering a movement for preserving raw nanopore data.

O.3: Dr. Jamie Lustermans (University of Antwerp)

Extracellular electron transfer by candidate flocking bacteria in cable bacteria sediment

Jamie JM Lustermans[1,2], Kartik Aiyer[1], Mantas Sereika[3], Laurine DW Burdorf[1,2], Mads Albertsen[3], Ian PG Marshall[1], Andreas Schramm[1]

1: Center for Electromicrobiology (CEM), Institute for Microbiology, Department of Biology, Aarhus University, Denmark
2: Microbial Systems Technology Excellence Centre, Department of Biology, University of Antwerp, Wilrijk, Belgium
3: Center for Microbial Communities, Aalborg University, Denmark

Cable bacteria are filamentous sulfide oxidizers that separate sulfide oxidation from oxygen reduction by conducting electrons through wires in their filament. Freshwater cable bacteria, *Candidatus* Electronema aureum GS, exhibited a recently discovered phenomenon called 'flocking'. Flocking is described as hundreds of fast-swimming microbes that swim in a chemotactic fashion around electron-conducting cable bacteria filaments in a narrow (~20 µm) band under anoxic conditions. These flocks of bacteria appear only when cable bacteria are actively conducting electrons. Long-term observations of the flockers divulged that they are of diverse morphologies and likely of diverse phylogenies.

Flockers were hypothesized to donate electrons to cable bacteria via extracellular electron transfer (EET). Based on metatranscriptomics combined with metagenomics we have determined who these associated bacteria may be, and how they may exchange electrons with cable bacteria. We discovered 14 potential flocking bacteria of widely varying phylogeny belonging to Acidobacteriota, Armatimonadota, Campylobacterota, Desulfobacterota, Firmicutes, Gammaproteobacteria and Myxococcota. Additionally, we screened isolates from a suspected flocking origin with outermembrane cytochrome assays and are determining their electroactivity in bioelectrochemical systems. The combined outcomes showed that *Ca*. Electronema aureum GS has many potentially electrogenic associates of which a varying subset may be exchanging electrons at any given flocking instance.

O.4: Romi Vandoren (University of Antwerp)

Unraveling the Complex T Cell Receptor-Microbiome Interaction Network in the Colon Through Advanced Computational Analysis

Romi Vandoren[1,2,3], Peter Delputte[4], Benson Ogunjimi[2,5,6,7], Pieter Meysman[1,2,3] & Kris Laukens[1,2,3]
1: Adrem Data Lab, Department of Computer Science, University of Antwerp, Antwerp, Belgium
2: Antwerp Unit for Data Analysis and Computation in Immunology and Sequencing (AUDACIS), Antwerp, Belgium
3: Biomedical Informatics Research Network Antwerp (biomina), University of Antwerp, Antwerp, Belgium
4: Laboratory of Microbiology, Parasitology and Hygiene (LMPH) & Infla-Med Center of Excellence, University of Antwerp, Antwerp, Belgium
5: Antwerp Center for Translational Immunology and Virology (ACTIV), Vaccine and Infectious Disease Institute, University of Antwerp, Antwerp, Belgium
6: Centre for Health Economics Research and Modelling Infectious Diseases (CHERMID), Vaccine and Infectious Disease Institute, University of Antwerp, Antwerp, Belgium
7: Department of Paediatrics, Antwerp University Hospital, Antwerp, Belgium

Aim of the study:T cells and their T cell receptors (TCR) play a crucial role in the body's immune defense against a wide variety of pathogens. However, pinpointing the specific bacteria that activate T cells in complex environments such as the colon remains challenging due to the vast TCR and gut microbiome diversity. To address this challenge, we introduce a new computational framework known as Adaptive Immune Receptor Repertoire-Wide Association Study (AIRRWAS).

Methods: AIRRWAS uses advanced statistical analysis to link TCR sequences to their target microbial species, utilizing data from an extensive cohort of paired TCR-seq and 16S rRNA microbiome data. Highly correlated interactions between TCR clusters and bacterial species are extracted from the data, generating a complex TCR-microbiome network that predicts and visualizes TCR-bacterium interactions. By analyzing these TCR patterns, this innovative approach can provide novel insights into the complex and interconnected relationship between the microbiome and TCR repertoire, not only in the colon but at all immune-microbiome interfaces.

Results: Application of AIRRWAS on a cohort encompassing 8000 microbial amplicon sequence variants and 7.3 million TCRs across 74 individuals revealed 7000+ highly correlated TCR-microbiome interactions. These interactions are depicted in an interactive visual network, facilitating an in-depth analysis of TCR repertoire characteristics within the network. Interestingly, 284 TCR-microbiome interactions were shared by ≥3 individuals in the cohort. The AIRRWAS network also enabled a detailed exploration of TCR-microbiome interactions and their characteristics like clustering of different TCR chain types, analysis of CD4+ and CD8+ T cell distribution, identification of public TCR interactions and more. Emerging from our AIRRWAS network, experimental validation of identified interactions has been initiated.

Conclusions: This will provide deeper insights into which TCR-microbiome interactions are driving immune modulation in the colon, essential for the development of novel therapies that identify and target specific bacterial-TCR pairs. Therefore, the development of this AIRRWAS method represents a significant step forward in our understanding of the complex interplay between the microbiome and our immune system.

O.5: Anwar Hiralal (University of Antwerp)

Closing the genome of unculturable cable bacteria using metagenomic hybrid sequencing

Anwar Hiralal[1], Jeanine S. Geelhoed[1], Silvia Hidalgo-Martinez[1], Bent Smets[1], Jesper R. van Dijk[1], and Filip J.R. Meysman[1,2]

1: Microbial Systems Technology Centre, University of Antwerp, Antwerp, Belgium.

2: Department of Biotechnology, Delft University of Technology, Delft, The Netherlands.

Many environmentally relevant microorganisms cannot be cultured, and even with the latest metagenomic approaches, achieving complete genomes for specific target organisms of interest remains a challenge. Cable bacteria provide a prominent example of a microbial ecosystem engineer that is currently unculturable. They occur in low abundance in natural sediments, but due to their capability for long-distance electron transport, they exert a disproportionately large impact on the biogeochemistry of their environment. Current available genomes of marine cable bacteria are highly fragmented and incomplete, thus hampering the elucidation of their unique electrogenic physiology. Here, we present a metagenomic pipeline that combines Nanopore long-read and Illumina short-read shotgun sequencing. Starting from a clonal enrichment of a cable bacterium, we recovered a circular metagenome-assembled genome (5.09 Mbp in size), which represents a novel cable bacterium species with the proposed name *Candidatus* Electrothrix scaldis. The closed genome contains 1109 novel identified genes, including key metabolic enzymes not previously described in incomplete genomes of cable bacteria. We examined in detail the factors leading to genome closure. Foremost, native, non-amplified long reads are crucial to resolve the many repetitive regions within the genome of cable bacteria, and by analyzing the whole metagenomic assembly, we found that low strain diversity is key for achieving genome closure. The insights and approaches presented here could help achieve genome closure for other keystone microorganisms present in complex environmental samples at low abundance.

O.6: Sebastiaan Valkiers (University of Antwerp)

Detection of convergent immune responses through quantification of the T-cell receptor similarity space

Sebastiaan Valkiers[1], Koshlan Mayer-Blackwell[2], Andrew Fiore-Gartland[2], Kris Laukens[1], Pieter Meysman[1], Phil Bradley[2]
1: University of Antwerp, Antwerp, Belgium
2: Fred Hutch, Seattle (WA), USA

Aim of the study: The T-cell receptor (TCR) repertoire provides a detailed overview of past and present immune exposure of an individual. Yet, trying to pinpoint which T cells are responsible for a certain immunological condition is like finding a needle in a haystack. The TCR repertoire represents an enormously complex soup of hundreds of thousands of unique recombinations that each recognize a limited set of epitopes. Analysis of the TCR repertoire has shown a biased distribution of distinct T cell clones in the body. The probability of observing a T cell can be estimated based on several features of its sequence. Given this bias, some specific recombinations tend to be shared more often among different individuals. By quantifying the sequence similarity space in the TCR repertoire, we aimed to detect events of significant sequence convergence or significantly neighborhood enriched (SNE) TCRs.

Methods: An efficient indexing strategy was used to compute (TCRdist) distances between millions of TCRs. We developed a carefully constructed background model that preserves many of the biased properties in the source repertoire, including germline gene distribution and CDR3 sequence length distribution. SNE TCRs were defined as sequences that show a significant level of sequence neighbors as compared to the background model.

Results: A vectorized alternative to the TCRdist distance metric was developed that enables million-scale pairwise distance computation for TCRs. Using this framework, we show that SNE TCRs can be used as a proxy to identify vaccine-responding TCR clones. In addition, these high convergent TCRs have a higher tendency to be shared among multiple individuals. Furthermore, we observed many of the SNE TCR are associated to one or more HLA alleles.

Conclusions: The SNE TCR framework enables million-scale pairwise distance computation, and allows the quantification and prioritization of significant sequence convergence events in the TCR repertoire.

O.7: Julia Camacho (Evolutionary Ecology Group, Department of Biology, University of Antwerp)

Identification of candidate genes for adaptation to contrasting light environments along the water column in deep-water cichlid fishes

Julia Camacho[1], Domino Joyce[2], Hannes Svardal[1]
1: University of Antwerp,
2: University of Hull (UK)

During their rapid adaptive radiations, cichlid fishes (Cichlidae) have conquered a variety of habitats and ecological niches, making them powerful model organisms to gain insight into the evolutionary mechanisms underlying ecological adaptation. In this study, we investigated the molecular basis of depth-specific adaptations in a group of deep-water dwelling cichlids from Lake Malawi, the clade Diplotaxodon, using a combination of genomics and transcriptomics. We conducted a genome-wide association (GWA) analysis for eye size, employed as a proxy for habitat depth, to identify genetic variation associated with contrasting depths. This analysis revealed multiple genomic regions enriched for Gene Ontology categories related to visual perception and photoresponse. Notably, the top 0.01% most significant variants were enriched for nonsynonymous substitutions, many of which map to vision-related genes. We also found signals of selection acting on these genes, making them intriguing candidates for depth-adaptation in this clade. These results are consistent with previous findings, which have linked the strong differences in light at different water depths with adaptations of the visual system. To further investigate changes in vision-related genes associated with the underwater environment, we performed a differential gene expression analysis using whole-eye transcriptomes of two species with marked differences in depth preferences. Among the highly differentially expressed genes there were multiple of the candidate vision-related genes highlighted by the genomic analysis, along with several additional genes with oxygen transport functions previously involved in depth-specific adaptation. The overlap between the two analyses provides strong evidence of a predominant role of vision for ecological specialization along a water depth gradient, where both coding and regulatory mechanisms are important. Overall, this study demonstrates the power of combining genomic and transcriptomic techniques in unravelling evolutionary adaptations.

O.8: Vincent Van Deuren (University of Antwerp)

RapTCR: Rapid Exploration and Visualisation of T-cell repertoires

Vincent Van Deuren[1], Sebastiaan Valkiers[1], Kris Laukens[1], Pieter Meysman[1]

1. Adrem Data Lab, Department of Computer Science, University of Antwerp, Antwerp, Belgium, AUDACIS, Antwerp Unit for Data Analysis and Computation in Immunology and Sequencing, University of Antwerp, Antwerp, Belgium

Aim of the Study

The acquisition of T-cell receptor (TCR) repertoire sequence data has become faster and cheaper due to advancements in high-throughput sequencing. However, fully exploiting the diagnostic and clinical potential within these TCR repertoires requires a thorough understanding of the inherent repertoire structure. Hence, visualizing the full space of TCR sequences could be a key step towards enabling exploratory analysis of TCR repertoire, driving their enhanced interrogation. Nonetheless, current methods remain limited to rough profiling of TCR V and J gene distributions. Addressing the need for improved TCR repertoire analysis, we developed RapTCR, a tool for rapid visualization and post-analysis of TCR repertoires

Methods

To overcome computational complexity, RapTCR introduces a novel, simple embedding strategy that represents TCR amino acid sequences as short vectors while retaining their pairwise alignment similarity. RapTCR then applies efficient algorithms for indexing these vectors and constructing their nearest neighbor network. It provides multiple visualization options to map and interactively explore a TCR network as a two-dimensional representation.

Results

Benchmarking analyses using epitope-annotated datasets demonstrate that these RapTCR visualizations capture TCR similarity features on a global level (e.g., J gene) and locally (e.g., epitope reactivity).

Conclusions

We introduce RapTCR, a Python toolkit designed to overcome the challenges of visualizing TCR repertoires, and demonstrate how it is an accessible and efficient solution for exploring and interpreting complex TCR sequencing datasets. RapTCR is available as a Python package, implementing an intuitive syntax to easily generate insightful, publication-ready figures for TCR repertoires of any size.

References

RapTCR preprint DOI: https://doi.org/10.1101/2023.09.13.557604

O.9: Nicky de Vrij (University of Antwerp / Institute of Tropical Medicine)

Persistent T cell unresponsiveness associated with chronic visceral leishmaniasis in HIV-coinfected patients

Nicky de Vrij[1,2], Julia Pollmann[3], Antonio M. Rezende[4], Ana V. Ibarra-Meneses[5], Thao-Thy Pham[1], Wasihun Hailemichael[6], Mekibib Kassa[7], Tadfe Bogale[7], Roma Melkamu[7], Arega Yeshanew[7], Rezika Mohammed[7], Ermias Diro[7], Ilse Maes[8], Malgorzata A. Domagalska[8], Hanne Landuyt[9], Florian Vogt[10,11,12], Saskia van Henten[12], Kris Laukens[2], Bart Cuypers[2], Pieter Meysman[2], Hailemariam Beyene[13], Kasaye Sisay[13], Aderajew Kibret[13], Dagnew Mersha[13], Koert Ritmeijer[14], Johan van Griensven[12], Wim Adriaensen[1]*

1: Clinical Immunology Unit, Department of Clinical Sciences, Institute of Tropical Medicine, 2000 Antwerp, Belgium
2: Adrem Data Lab, Department of Computer Science, University of Antwerp, 2020 Antwerp, Belgium
3: Department of Medical Oncology, University Hospital Heidelberg, National Center for Tumor Diseases (NCT) Heidelberg, 69120 Heidelberg, Germany
4: Department of Microbiology, Aggeu Magalhães Institute – FIOCRUZ/PE, Recife, Brazil
5: Département de pathologie et microbiologie, Faculté de médecine vétérinaire, Université de Montréal. Saint-Hyacinthe, QC, Canada.
6: Department of Immunology and Molecular Biology, Faculty of Biomedical Sciences, University of Gondar, Ethiopia
7: Leishmaniasis Research and Treatment Centre, University of Gondar, Gondar, Ethiopia
8: Molecular Parasitology Unit, Department of Biomedical Sciences, Institute of Tropical Medicine, 2000 Antwerp, Belgium
9: Clinical Trial Unit, Department of Clinical Sciences, Institute of Tropical Medicine, 2000 Antwerp, Belgium
10: National Centre for Epidemiology and Population Health, The Australian National University, Canberra 2601, Australia
11: The Kirby Institute, University of New South Wales, Sydney 2052, Australia
12: Unit of Neglected Tropical Diseases, Department of Clinical Sciences, Institute of Tropical Medicine, 2000 Antwerp, Belgium
13: Médecins Sans Frontières, Abdurafi, Ethiopia
14: Médecins Sans Frontières, Amsterdam, The Netherlands

A large proportion of HIV-coinfected visceral leishmaniasis (VL) patients exhibit chronic disease with frequent VL recurrence, despite viral suppression and initial parasitological cure. Due to a hard-to-reach population, knowledge on immunological determinants underlying this disease course is scarce. Therefore, we studied circulatory cellular immunity of a longitudinal HIV cohort in North-West Ethiopia, including asymptomatic *Leishmania*-infected individuals, and active VL developers. Compared to non-chronic VL-HIV patients, chronic VL-HIV patients exhibited persistently high levels of TIGIT and PD-1 on CD8+ and CD8- T cells throughout the disease course. Furthermore, these chronic VL-HIV patients showed a lower proportion of IFN-γ+TIGIT- CD8+/CD8- T cells and a higher proportion of IFN-γ-TIGIT+ CD8+/CD8- T cells, suggestive of impaired T cell functionality. At single-cell resolution, these chronic VL-HIV patients showed CD4+ T cell anergy throughout their anti-leishmanial treatment, characterised by a lack of T cell activation in the transcriptome, and a lack of a lymphoproliferative response at the clonal level. These findings suggest PD-1 and TIGIT play a pivotal in VL-HIV chronicity, and may be further explored for patient risk stratification. Our findings provide a strong rationale for adjunctive immunotherapy for the treatment of patients with chronic VL-HIV to break the recurrent VL disease episode cycle.

O.10: Prof. Hannes Svardal (University of Antwerp)

Sex, selection, and (barriers to) gene flow: the complicated evolutionary histories of large chromosomal inversions in benthic Malawi cichlids

Hannes Svardal
University of Antwerp

Chromosomal inversions – DNA sequences that are flipped end to end – play an important role in environmental adaptation, reproductive isolation, and sex determination in many organisms. However, there has been little evidence for their presence in large vertebrate adaptive radiations. Here we queried the genomes of 1,379 samples from the Malawi cichlid adaptive radiation for the presence of genomic regions with recombination-suppressed haplotypes – as expected for polymporphic chromosomal inversions. We discovered five such regions, each more than half a chromosome in length containing hundreds of genes, and used cytogenetics, long reads, Hi-C, and linked reads to confirm the presence of chromosomal inversions in four of these regions. Inversion haplotypes originated early in the benthic Malawi cichlid radiation, and contribute to species divergence, with some instances of haplotype transfer between ecologically similar but genetically divergent groups. We present evidence that at least three of these inversions contribute to sex determination in some extant species, but not in others. Despite their relatively high haplotypic divergence, the significance of these of these inversions for adaptation and diversification remains still largely mysterious as they remain polymorphic in some species without apparent correlations with phenotypes or sex.

O.11: Nishkala Sattanathan (University of Antwerp)

WiNGS - Federated approach for genomics data sharing and analysis

Nishkala Sattanathan[1], Benjamin Huremagic[4], Geert Vandeweyer[1,2], Yves Moreau[3], Joris Vermeesch[4]
1: University of Antwerp
2: Antwerp University Hospital
3: ESAT, Catholic University of Leuven (KU Leuven)
4: Catholic University of Leuven (KU Leuven)

Aim of the study

Next-generation sequencing (NGS) has been increasingly used in a wide range of research communities, as well as in in routine clinical practice and leads to an ever increasing amount of sequencing data. The management of the expanding amount of data is challenging and, especially for human omics data, privacy protection is crucial. In contrast, unraveling the causes of rare diseases is critically dependent on data sharing, but progress is hampered by regulations and privacy concerns. To overcome the concerns associated with centralized human genomic data storage, we developed a federated analysis platform, referred to as Widely Integrated NGS (WiNGS). The presented approach enables data sharing and combined data-analysis of genomics data across a consortium without a centralized data store.

Methods

WiNGS was designed as a modular and multi-tier platform, with robustness , data privacy and flexibility in mind. The general implementation layout is developed with a WiNGS-UI(user interface) for data visualization and WiNGS Clients which store sensitive data and perform in-house analysis. WiNGS can be used to store variant data in a VCF(variant caller format), and to perform single sample analysis, trio-based analysis under different inheritance models, and cohort-based analysis.

Results

Currently, WiNGS has been deployed over six centers. Approximately 2500 samples are available across the different centers and they are used for in-house variant analysis and also perform cohort-based analysis in a privacy controlled manner.

Conclusions

WiNGS is a platform for next-generation sequencing data interpretation. One of the key features of WiNGS is its secure federated data storage, which ensures that sensitive data is only accessible to authorized users and is kept confidential. In addition, the platform offers full API functionality, allowing researchers to easily integrate the platform into their own analysis tools or pipelines. Cohort-based analysis allows researchers to aggregate data from multiple institutions in order to generate anonymous statistical insights. Overall, WiNGS is a powerful and flexible platform that provides researchers and clinicians with a secure and federated environment for collaborative next-generation sequencing data interpretation. Its range of analysis approaches, combined with its secure data storage and full API functionality, make it an essential tool for researchers working in the field of genomics.

O.12: Dr. Kerry Mullan (University of Antwerp)

STEGO.R for (easy) interrogation of combined scTCR repertoire and scRNA-seq data

Kerry Mullan[1,2,3], My Ha[2,4,5], Sebastiaan Valkiers[1,2,3], Benson Ogunjimi[2,4,5,6], Kris Laukens[1,2,3], and Pieter Meysman[1,2,3]
1: Adrem Data Lab, Department of Computer Science, University of Antwerp, Antwerp, Belgium
2: Antwerp Unit for Data Analysis and Computation in Immunology and Sequencing (AUDACIS), Antwerp, Belgium
3: Biomedical Informatics Research Network Antwerp (biomina), University of Antwerp, Antwerp, Belgium
4: Antwerp Center for Translational Immunology and Virology (ACTIV), Vaccine and Infectious Disease Institute, University of Antwerp, Antwerp, Belgium
5: Centre for Health Economics Research and Modelling Infectious Diseases (CHERMID), Vaccine and Infectious Disease Institute, University of Antwerp, Antwerp, Belgium
6: Department of Paediatrics, Antwerp University Hospital, Antwerp, Belgium

Introduction. The hypervariable T cell receptor (TCR), created through somatic recombination, allows for recognition of a diverse array of antigens. Single sequencing technologies allow capture of both the single cell expression data (scRNA-seq) with the paired single cell TCR sequencing (scTCR-seq). However, the current analytical pipelines have limited capacity to integrate both data levels. To overcome these limitations, we developed STEGO (Single cell TCR and Expression Grouped Ontologies) Shiny R application to facilitate the complex analysis required for understanding T cells role in various conditions.

Program parameters. STEGO.R application includes the Seurat quality control (QC) process, merging with Harmony, followed by semi-supervised cellular annotations with scGate. The scRNA-seq with scTCR-seq is broken down into four sections: top clonotype, expanded clonotypes, clustering (ClusTCR2) and target epitopes from TCRex predictions. The Shiny R interface also facilitates the program's accessibility to novice R coders.

Preliminary analysis. Out of 22 selected public datasets, 12 could be processed with STEGO.R. We re-interrogated the dataset concerning colon inflammations following melanoma therapies, as original studies did not integrate the scRNA-seq with scTCR-seq analysis. From one study, our novel process identified that the colitis expanded T cells were cytotoxic CD8+ T cells with over-represented transcripts including IFNG, GNLY, PFR1, GZMB, NKG7, HLA-DR, KLRD1 transcripts relative to both the non-expanded clonotypes, non-colitis cases and healthy colon donors. The analysis also identified a TRGV4 cluster associated with melanoma cases as well as two TRBV6-2 clusters specific to colitis.

Discussion. STEGO.R facilitates fast and reproducible analysis of complex scRNA-seq with TCR repertoire data. We have demonstrated its utility by extracting novel biologically relevant insights into T-cells. We anticipate this program will facilitate the identification of subtle T population differences and if these are specific to a TCR clone and/or the expanded repertoire.

O.13: Christophe Verwimp (UZA)

Development and validation of a CNV analysis and visualization pipeline for WES data

Christophe Verwimp[1], Katrien Janssens[1], Geert Vandeweyer[1]
1: UZA, centrum medische genetica

Aim of the Study

Development and validation of a CNV analysis and visualization pipeline for WES data.

Methods

CNVs are detected by read counts, the number of reads aligning to a genomic region. Sample counts are compared to the count distribution of a reference set, representing technical variation of a normal genotype. Deletions and duplications are expected to result in lower and higher counts respectively, compared to reference values.

We select a new set of samples from a rolling reference pool per analysis, considering gender, technical factors from library preparation and sequencing, and overall correlation of read counts between samples. CNV calling is performed by combining 3 CNV callers; Conifer, XHMM and Canoes. Consensus CNVs are determined by majority voting.

The pipeline is written in python and R, and can be run both on local and cloud infrastructure using WDL/cromwell. Results are visualized as a PDF report rendered with Rmarkdown, and interactively visualized in an R Shiny app.

Performance was validated on 262 non-benign, mutually covered CNVs called by SNP array in 216 samples.

Results

242/262 array-CNVs were confirmed by WES, yielding a sensitivity of 92.4%. We observed an effect of CNV length on sensitivity, which was ~50% for single-exon events (1 WES target), 99% for CNVs of at least 4 targets, and 100% for minimum 8 targets. 189/253 WES-CNVs were confirmed by array, yielding a precision of 74.7%. We also observed an effect of CNV length on precision, with most FPs covering under 30 WES targets, and the largest FP covering 53 targets. Thus, a threshold of 60 targets was set as a lower limit for orthogonal confirmation. Furthermore, SNP arrays are not a perfect golden standard, due to a lower resolution compared to WES. Hence, precision of WES-based CNV calling was potentially underestimated.

Conclusion

The presented WES-CNV pipeline, integrating 3 CNV callers, delivers high-quality results that are interpretable for molecular diagnostics.

O.14: Charlotte Adams (University of Antwerp)

Fragment ion intensity prediction improves the identification rate of non-tryptic peptides in timsTOF

Charlotte Adams[1,2], Wassim Gabriel[3], Kris Laukens[1], Mathias Wilhelm[3], Wout Bittremieux[1], Kurt Boonen[2,4]
1: Department of Computer Science, University of Antwerp, Antwerp, Belgium
2: Centre for Proteomics (CFP), University of Antwerp, Antwerp, Belgium
3: Computational Mass Spectrometry, Technical University of Munich, 85354 Freising, Germany
4: Sustainable Health Department, Flemish Institute for Technological Research (VITO), Antwerp, Belgium

Aim of the Study

The adaptive immune system can eradicate cancerous cells by recognising peptides bound to HLA-molecules present on the cell surfaces. In immunopeptidomics, these peptides—commonly termed immunopeptides—are isolated and characterized using mass spectrometry. To minimize false positives and improve spectrum annotation rates, peptide-spectrum match (PSM) rescoring can be used. This involves post-processing results from an unfiltered database search, during which multiple PSM features are used to distinguish between correct and incorrect PSMs. Recently, there has been significant interest in using additional features for PSM rescoring, including spectral features based on the similarity between experimental and predicted fragment ion intensities.

Because low abundant immunopeptides often occur, highly sensitive timsTOF instruments are increasingly gaining popularity. To improve PSM rescoring for immunopeptides measured using timsTOF instruments, we fine-tuned a deep learning-based fragment ion intensity prediction model.

Methods

Prosit is a deep neural network that can predict the fragment ion intensities for a given peptide sequence. To fine-tune Prosit, a dataset was generated by analyzing over 300,000 synthesized non-tryptic peptides on a timsTOF-Pro. This new model was used for PSM rescoring of public immunopeptidomics datasets.

Results

We achieved an up to 3-fold increase in the identification rate of immunopeptides after PSM rescoring compared to database searching. Furthermore, our approach increased detection of immunopeptides even from low input samples. Importantly, the immunopeptides identified after PSM rescoring are likely to bind HLA-molecules, as supported by motif analysis and binding affinity assessment.

Conclusion

By applying our new fragment ion intensity prediction model for PSM rescoring, we can drastically increase the detection of immunopeptides, which hold the potential to serve as valuable targets for immunotherapy.

O.15: Dr. Tim Van De Looverbosch (University of Antwerp)

Accurate Nuclei Detection in 3D Cell Systems Through Centroid Prediction

Tim Van De Looverbosch[1] , Sarah De Beuckeleer[1], Jan Sijbers[2], Winnok De Vos[1]
1: Laboratory of Cell Biology and Histology, University of Antwerp
2: Imec-Vision Lab, University of Antwerp

Aim of the study

In recent years, powerful deep learning (DL) algorithms were developed for segmenting cells and nuclei in 2D images. However, segmentation in 3D cell systems remains challenging due to high cell density, the heterogeneous resolution and contrast across the image volume, and the difficulty in generating reliable and sufficient ground truth for model training. Reasoning that many applications rely on nuclear segmentation but do not necessarily require an accurate delineation of their shapes, we implemented a 3D U-Net based method that rapidly predicts the position of their centroids. Finally, we tested if this model can serve as a prompt generator for automated 3D nuclei segmentation using a foundation model.

Methods

A 3D U-Net model was trained to predict 3D nuclei centroid probability maps from a 3D nuclear counterstain image. We pretrained the model using weak targets resulting from a conventional image processing pipeline on a large dataset. Next, we finetuned this model using supervised training with human annotations. Finally, we used the predicted centroids as prompts for nuclei segmentation using the Segment Anything Model (SAM).

Results

We show that our model outperforms existing methods in centroid prediction. In addition, we found that a pretrained + finetuned model significantly outperformed a model that was trained from scratch. Finally, we show that our model can serve as a lightweight prompt generator for predicting nuclei centroids as guides for automated nuclei segmentation based on SAM.

Conclusion

The nuclei centroid predictions of our model can be used easily integrated in downstream tasks, such as cell counting, cell type and phenotype prediction, and analyzing sample organization. Due to the ease and speed of centroid annotation compared to manual segmentation, specialized centroid prediction models can be developed more rapidly than segmentation models tailored to specific datasets. If segmentation is preferred, a lightweight model capable of predicting cell centroids could function as a specialized prompt generator for a more generalized image segmentation foundational model.

O.16: Ceder Dens (University of Antwerp)

The pitfalls of negative data bias for the T-cell epitope specificity challenge

Ceder Dens[1,2], Kris Laukens[1,2], Wout Bittremieux[1], Pieter Meysman[1]
1: Adrem Data Lab, Department of Computer Science, University of Antwerp, Middelheimlaan 1, 2020 Antwerpen, Belgium
2: AUDACIS consortium, University of Antwerp, Middelheimlaan 1, 2020 Antwerpen, Belgium

Aim of the Study

Recently, Gao et al. [1] introduced a combination of meta-learning and the neural Turing machine to tackle a very important but yet unsolved problem in immunology: the TCR–epitope binding prediction challenge for novel epitopes. They showed a big increase in performance compared to previous models and we investigated how they were able to achieve this.

Methods

We tested their model on an external dataset, containing negatives generated with a different technique.

Results

In this article, we describe how the technique used to create negative data for the TCR–epitope interaction prediction task can lead to a strong bias and makes that the performance drops to random when tested in a more realistic scenario.

Conclusions

All high-performing machine learning models can have problems when deployed in a real-world setting if the data used to train and test the model contains biases. This will result in unrealistic performance that is achieved due to shortcut learning.

References

1. Gao, Y. et al. Pan-Peptide Meta Learning for T-cell receptor–antigen binding recognition. Nat. Mach. Intell. 1–14 (2023) doi:10.1038/s42256-023-00619-3.

P.1: Nicky de Vrij (University of Antwerp - Institute of Tropical Medicine)

Human Leukocyte Antigen alleles as risk factors for visceral leishmaniasis development in HIV co-infected individuals

Nicky de Vrij[1,2], Romi Vandoren[2], Kadrie Ramadan[3], Anke Van Hul[1], Wim Cuypers[2], Bart Cuypers[2], Thao-Thy Pham[1], Johan van Griensven[4], Kris Laukens[2], Wim Adriaensen[1]*

1: Clinical Immunology Unit, Department of Clinical Sciences, Institute of Tropical Medicine, 2000 Antwerp, Belgium
2: Adrem Data Lab, Department of Computer Science, University of Antwerp, 2020 Antwerp, Belgium
3: Clinical Reference Laboratory, Department of Clinical Sciences, Institute of Tropical Medicine, 2000 Antwerp, Belgium
4: Unit of Neglected Tropical Diseases, Department of Clinical Sciences, Institute of Tropical Medicine, 2000 Antwerp, Belgium

HIV co-infection is one of the major challenges for visceral leishmaniasis (VL) disease control, as mortality rates are multifold higher in VL-HIV patients than in immunocompetent patients. Given this poor prognosis once infection has progressed to disease, tackling the infection in the asymptomatic stage before it progresses to disease would be a valuable asset to patient management. However, it is currently not fully known what predisposes individuals to progress to VL disease in HIV co-infection. In VL disease alone, the Human Leukocyte Antigen (HLA) gene region has already been associated with predisposition to VL development. In this work, we showcase the use and accuracy of a novel commercially available Oxford Nanopore Technologies sequencing-based HLA genotyping method to detect HLA associations with VL disease development in 124 asymptomatic Leishmania-infected HIV-coinfected individuals living in North-West Ethiopia. Analyses are still ongoing, but preliminary analyses show HLA-A*03:01 to be a risk factor for VL development in this Ethiopian HIV-positive cohort.

P.2: Ana Regina de Abreu (Center of Medical Genetics)

Implications of genome-wide DNA methylation sequencing methods in the pursuit of cancer biomarker discovery

Ana Regina de Abreu*[1,2], Joe Ibrahim*[1,2], Ligia Mateiu[1], Marc Peeters[2,3], Guy Van Camp[1,2], Ken Op de Beeck[1,2]
1: Center of Medical Genetics (CMG), University of Antwerp and Antwerp University Hospital, Edegem, Belgium
2: Center for Oncological Research (CORE), Integrated Personalized and Precision Oncology Network (IPPON), University of Antwerp and Antwerp University Hospital, Antwerp, Belgium
3: Department of Oncology, Antwerp University Hospital (UZA), Edegem, Belgium

Aim of the Study: DNA methylation is an important gene expression marker in human cancer and is extensively explored as a source of biomarkers. Numerous methods exist for profiling methylation, but many do not provide a comprehensive understanding of the disease. Despite its shortcomings, bisulfite sequencing remains the reference method. Although several methods for circumventing associated limitations have been proposed, there has yet to be a clear consensus on which is most appropriate for genome-wide methylation profiling.

Methods: We examined four different methods to find the strengths and weaknesses of each and make recommendations on which method is most suited for various profiling applications. We assessed DNA methylation profiles from one healthy blood genome and four human cancer genomes derived from fresh frozen tissue and cell lines. In addition, we sought to adapt to a clinical setting and checked the ability of each method to sequence FFPE-derived DNA. Genomic DNA was subjected to Whole-Genome Bisulfite Sequencing (WGBS), EPIC BeadChip, Enzymatic Methyl-Sequencing (EMSeq) and Oxford Nanopore Technologies (ONT).

Results: After rigorous quality assessment, we found high concordance between the methods, but differences in the efficiency of read mapping, CpG calling, and coverage were observed. ONT detected a significant number of sites inaccessible to short-read assays. Moreover, longer fragments benefited from fewer overlapping reads, which increased mapping efficiency and mean coverage per CpG. The methylation signals obtained by native ONT detection are highly comparable to short-read bisulfite- and enzymatic sequencing, with average Pearson correlation values of r = 0.93 for CpG methylation concordance. EPIC showed a slightly lower concordance with ONT sequencing (r = 0.92). Furthermore, due to structural interference with the nanopores, the current ONT chemistry does not permit native sequencing of FFPE-derived DNA. Moreover, despite using restoration methods EPIC only occasionally produces satisfactory results due to the fragmented state and poor quality of FFPE-derived DNA.

Conclusions: Our findings provide an exposition of the advantages, shortcomings, and performance of several profiling methods. Notably, when using FFPE samples, EPIC frequently fails to generate data, whereas native ONT sequencing always fails. Thus, WGBS or EMSeq are alternatives for FFPE sequencing. Overall, we provide a decision-making resource based on budget, DNA input, coverage, and performance for genome-wide DNA methylation profiling studies.

P.3: Elise Coopman (University of Antwerp)

Methylmap: visualization of nucleotide modifications for large cohorts.

Elise Coopman [1,2], Marleen Van den Broeck [1,2], Tim De Pooter [3,4], Geert Joris [3,4], Dennis W Dickson [5], Mojca Strazisar [3,4], Rosa Rademakers [1,2,5], Wouter De Coster [1,2]

1: Applied and Translational Neurogenomics Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium
2: Applied and Translational Neurogenomics Group, Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
3: Neuromics Support Facility, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium
4: Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
5: Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA

Aim of the Study: In recent years, epigenetics has become a crucial topic of interest for understanding biological functions such as regulating gene expression. Due to ongoing developments in the field of long-read sequencing technologies, they are now being applied in population-scale sequencing projects. Several tools for visualizing nucleotide modification patterns in one or a limited number of individuals are available; however, no software is tailored to large cohorts. Therefore, we developed methylmap, a tool for visualizing nucleotide modification frequencies per position scaling to hundreds of individuals or haplotypes in heatmap format for a genomic region of interest.

Methods & Results: Methylmap is developed in Python and builds on the following dependencies: pandas, numpy, modbam2bed (ONT), argparse, and plotly. Tabix is used for fast and efficient retrieval of the region of interest. The tool supports several input options: CRAM/BAM files with MM and ML tags, tab-separated files from the nanopolish methylation caller, and an overview tab-separated table with modification frequencies. Optionally, a gene/transcript annotation track can be added, supported by a GTF/GFF input file. Furthermore, hierarchical clustering on the modification frequencies can be performed. The figure output of the tool is in dynamic HTML format, allowing the end-user to further customize the view in any of the major browsers. Methylmap is available through PyPI and bioconda. The source code can be found at https://github.com/EliseCoopman/methylmap.

Conclusions: Over the past years, increased interest in epigenetic modifications resulted in extensive developments of technologies, making it possible to perform population-scale epigenetic studies. We developed methylmap, a *fast* tool for visualizing modification frequencies of large cohort sizes. Methylmap is a technology-agnostic tool supporting standardized BAM/CRAM, nanopolish, and a custom modification frequency table, which can, together with other features, be expanded in the future.

P.4: Sara Alidadiani (VIB)

Transcriptomic analyses as part of the international FTLD-FUS consortium highlights role for mitochondrial dysfunction in disease.

Sara Alidadiani[1,2], Wouter De Coster [1,2], Sarah Wynants [1,2], Marleen Van den Broeck [1,2], Cyril Pottier[1,2,3], Marka van Blitterswijk[3], Matthew C Baker[3], Mariely DeJesus-Hernandez[3], NiCole A Finch[3], Cristina Vicente [1,2], Tim De Pooter[1], Geert Joris[1], Mojca Strazisar[1,2], Dorothee Dormann[4], Saskia Hutten[4], Merel Mol[5], Harro Seelaar[5], Jeroen van Rooij[5], Rik Vandenberghe[6,7], Dietmar Thal[8,9,10], Elizabeth C Finger[11], Manuela Neumann[12], Ging-Yuek Robin Hsiung[13], Bernardino Ghetti[14], Nigel J Cairns[15], Carlos Cruchaga[16], Robert A. Rissman[17,18], Annie Hiniker[19], Julia Kofler[20], Charles L White [3rd21], Bret M Evers[21], Ang Lee Cyn[22], Glenda M Halliday[23,24], John B Kwok[23,24] Murray Grossman[25], Edward B Lee [26], EunRan Suh [26] Vivianna van Deerlin[26], M Marsel Mesulam[27,28], Sandra Weintraub[27,29], Margaret E Flanagan[27,30], Jon Rohrer[31], Claire Troakes[32], Safa Al-Sarraj[32,33], Joanna Biernacka[34], Melissa E. Murray[3], Leonard Petrucelli[3], Björn Oskarsson[35], Keith A. Josephs[36], Ronald C. Petersen[36], Bruce L Miller[37], William W Seeley[37,38] Neill R. Graff-Radford[35] Bradley F. Boeve[36], John van Swieten[5], Ian RA Mackenzie[13], Dennis W. Dickson[3], and Rosa Rademakers[1,2,3]

1: VIB Center for Molecular Neurology, Antwerp, Belgium
2: Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium
3: Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA
4: Johannes Gutenberg University (JGU) Mainz, Faculty of Biology, Institute of Molecular Physiology, Mainz, Germany; Institute of Molecular Biology (IMB) Mainz, Mainz, Germany.
5: Department of Neurology, Erasmus Medical Center, Rotterdam, The Netherlands.
6: Department of Neurology, UZ Leuven, Leuven, Belgium.
7: Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Leuven, Belgium.
8: Department of Pathology, UZ Leuven, Leuven, Belgium.
9: Leuven Brain Institute (LBI), KU Leuven, Leuven, Belgium.
10: Laboratory for Neuropathology, Department of Imaging and Pathology, KU Leuven, Leuven, Belgium.
11: Department of Clinical Neurological Sciences, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada.
12: German Center for Neurodegenerative Diseases (DZNE), Department of Neuropathology, University of Tübingen, Tübingen, Germany.
13: Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada.
14: Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA.
15: Department of Neurology, Knight Alzheimer Disease Research Center, Washington University School of Medicine, Saint Louis, MO, USA.
16: Department of Psychiatry, Knight Alzheimer Disease Research Center, Washington University School of Medicine, Saint Louis, MO, USA.
17: Department of Neurosciences, University of California, San Diego, La Jolla, CA, 92093, USA.
18: Veterans Affairs San Diego Healthcare System, San Diego, CA, 92161, USA.
19: Department of Pathology, University of California San Diego, La Jolla, CA, USA
20: Department of Pathology, University of Pittsburgh, Pittsburgh, PA, USA
21: Division of Neuropathology, University of Texas Southwestern Medical Center, Dallas, TX, USA.
22: Department of Pathology, London Health Sciences Centre, London, ON, Canada
23: Central Clinical School and Brain and Mind Centre, The University of Sydney, Sydney, Australia.
24: UNSW Medicine and NeuRA, Randwick, Australia.
25: Penn Frontotemporal Degeneration Center, Department of Neurology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.
26: Center for Neurodegenerative Disease Research, Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.
27: Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA.
28: Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA.
29: Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA.

30: Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA.
31: Dementia Research Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, London, UK.
32: London Neurodegenerative Diseases Brain Bank, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.
33: Department of Clinical Neuropathology, King's College Hospital NHS Foundation Trust, London, UK.
34: Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA.
35: Department of Neurology, Mayo Clinic, Jacksonville, FL, USA.
36: Department of Neurology, Mayo Clinic, Rochester, MN, USA.
37: Department of Neurology, Memory and Aging Center, University of California, San Francisco, CA, USA.
38: Department of Pathology, Memory and Aging Center, University of California, San Francisco, CA, USA.

Aim of the Study: Frontotemporal lobar degeneration with FUS pathology (FTLD-FUS) is a rare form of FTLD, explaining about 10% of patients. More than a decade after its initial description, we only have a limited understanding of the etiology of this disease subtype, thus severely hampering translational research efforts.

Methodology: We established the international consortium on FTLD-FUS and collected brain tissue samples and associated clinicopathological data from FTLD-FUS patients with the goal of deciphering its molecular underpinnings. Patients with atypical FTLD-U (aFTLD-U), basophilic inclusion body disease, and neurofilament inclusion body disease were collected, but analyses thus far focused on aFTLD-U. Short-read RNA sequencing data was generated from frontal-cortex tissue of 21 aFTLD-U and 77 FTLD-TDP patients and 20 controls (HiSeq4000, Illumina), and differential expression and weighted gene co-expression network analyses were performed.

Results: To date, we collected samples and clinical information from 96 pathologically confirmed FTLD-FUS cases from North America, UK, Europe, and Australia. Among aFTLD-U (n=96), the mean age at onset was 44.20 years (26-73), and the mean age at death was 50.85 years (33-77), with 68% male. Transcriptomic co-expression analysis revealed alterations in mitochondrial processes as well as nucleic acid binding when aFTLD-U was compared to controls. The dysregulation of mitochondrial processes in aFTLD-U was confirmed when another FTLD disease group (FTLD-TDP) was used for comparison.

Conclusion: In conclusion, we established an international FTLD-FUS consortium to elucidate the pathogenic mechanisms of FTLD-FUS. Brain transcriptomic analyses suggest the involvement of mitochondrial dysfunction in aFTLD-U, confirming an earlier report on FUS-induced impairment of mitochondria (1).

Reference: 1. Deng, J. et al. FUS interacts with ATP synthase beta subunit and induces mitochondrial unfolded protein response in cellular and animal models. Proc Natl Acad Sci U S A 115, E9678–E9686 (2018).

P.5: Dr. Geert Vandeweyer (Antwerp University Hospital)

Accelerating Clinical Whole Exome Sequencing Analysis: Leveraging WDL and Cromwell on AWS Batch

Philip Holmgren[1], Anne Mensink[1], Christophe Verwimp[1], Geert Vandeweyer[1]
Antwerp University Hospital

Aims of the Study: Enhance clinical WES data analysis efficiency using WDL and Cromwell on AWS Batch. Develop a robust, scalable pipeline for automated WES data processing, minimizing genomic analysis turnaround time.

Methods: The core WES analysis pipeline was implemented in the Workflow Definition Language (WDL), a domain-specific language for defining bioinformatics workflows. The pipeline encompasses read alignment, small variant calling, CNV calling, contamination detection and quality control. Cromwell, a workflow management system, was used to execute WDL workflows on AWS Batch. Leveraging AWS Batch elasticity and parallel processing capabilities, we optimized resource utilization for concurrent analysis by dynamically scaling up to 2,400 CPUs. Additional python-based managers embedded the WDL workflow in a a fully automated end-to-end flow linking sequencing infrastructure to clinical interpretation applications.

Results: Our implementation demonstrated a significant reduction in analysis time, compared to local analysis. AWS Batch enabled efficient scaling of computational resources, but full support was lacking in cromwell. We added retry strategies for spot instance preemption and memory-exhaustion, detailed tagging and support for the elastic file system (EFS). These enhancements significantly improved robustness, allowing routine analysis on lower-cost spot instances. Additional retry strategies to handle transient communication issues, inherent to cloud computing, allowed full on-premise monitoring, eliminating all costs related to idling cloud infrastructure.

Conclusions: Integration of WDL and Cromwell with AWS Batch and extensive local management routines, proved to be a powerful solution for expediting clinical WES data analysis. Our findings highlight the potential of cloud computing to streamline data analyis, but also highlighted the need for thorough code evaluation and optimization to keep unexpected costs in check.

P.6: Janne Heirman (University of Antwerp)

Improved spectrum clustering increases interpretability of molecular networks

Janne Heirman[1], Wout Bittremieux[1]
1: University of Antwerp

Tandem mass spectrometry (MS/MS) generates massive amounts of data, making data interpretation challenging, especially in the context of small molecules. Molecular networking addresses this problem by representing spectra as nodes and connecting them based on their spectral similarity, consequently finding groups of similar compounds without knowing their true identity. Clustering is often employed as a preprocessing step to reduce redundancy in datasets, increasing interpretability of the resulting network.

Aim of the study

This study aims to enhance spectrum clustering for small molecule molecular networking. Diverse clustering methods are evaluated and implemented into falcon. While falcon currently uses density-based clustering, Bittremieux et al. (2022) suggests that hierarchical clustering may outperform it. Our evaluation includes hierarchical clustering with complete, average, and single linkage, and DBSCAN.

Methods

After removing low-quality spectra and noise peaks, the spectra are clustered. Next, the molecular network is created, only preserving the medoid spectrum of each cluster. The clustering methods are evaluated on their completeness, proportion of clustered spectra and proportion of incorrectly clustered spectra. The molecular networks are evaluated on the number of clusters and the average degree. Prior to applying our method on a large, diverse dataset, we test our approach on a small toy dataset containing 1423 acylcarnitine spectra of which 621 were annotated as 22 unique compounds through library searching.

Results

Hierarchical clustering with complete linkage outperforms DBSCAN, clustering 93.11% of the spectra in the toy dataset at the expense of 0.66% incorrectly clustered spectra, achieving a completeness score of 0.9474. DBSCAN results in 83.84% clustered spectra with a completeness score of 0.8950 at the expense of 0.68% incorrectly clustered spectra. Hierarchical clustering reduces the nodes in the network to 216, with an average degree of 19.26, while DBSCAN yields 336 nodes, with an average degree of 31.49.

Conclusion

Pending further investigation, integration of hierarchical clustering into falcon is foreseen. Moreover, integration into GNPS would make these methods accessible to the scientific community. Nonetheless, our preliminary findings show great potential.

P.7: Sarah De Beuckeleer (University of Antwerp)

Unbiased identification of cell identity in dense mixed neural cultures based on nucleocentric phenotyping

Sarah De Beuckeleer[1], Tim Van De Looverbosch[1], Johanna Van Den Daele[1], Peter Ponsaerts[4], and Winnok H. De Vos[1,2,3,4]
1: Laboratory of Cell Biology and Histology, University of Antwerp
2: Antwerp Centre for Advanced Microscopy, University of Antwerp
3: µNeuro Research Centre of Excellence, University of Antwerp
4: Laboratory of Experimental Hematology, University of Antwerp

Aim of the Study

Induced pluripotent stem cell (iPSC) technology has revolutionized cell biology. However, the variability between iPSC lines and a lack of efficient technology to characterize downstream differentiation products hinder its adoption in routine screening settings. Our research aimed to facilitate the validation of iPSC-derived culture composition.

Methods

We have adapted a morphological assay based on cell painting to recognize cell types in dense and mixed cultures with high fidelity. To that end, we have employed UMAP dimension reduction and clustering, Random Forest and convolutional neural network classification to examine the morphological fingerprint of different cell types and states. We have benchmarked our approach using pure and mixed cultures of neuroblastoma and astrocytoma cell lines. In addition, we examined the robustness of our CNN prediction by iteratively limiting the input information.

Results

Our proof of concept using cell lines attained a classification accuracy above 96%. Through iterative data erosion we found that a reduced feature set based on the nuclear region of interest and its close environ (nucleocentric) achieved classification performance similar to that of whole cells and maintained accuracy in very dense cultures. CNN classification on more complex iPSC cultures revealed a performance of 96% for the identification of postmitotic neurons versus their neural progenitors in heterogenous in vitro cultures. Our cyclic cell-based approach significantly outperforms a classical condition-based CNN with an accuracy of 86%.

Conclusions

Our approach to nucleocentric morphological phenotyping provides a means to quantify cell composition in complex mixed cultures and holds promise for use in quality control of iPSC-derived cultures.

P.8: Dr. Juan Sebastian Piedrahita Giraldo (University of Antwerp)

Metabolite Identification and Analog Discovery through Deep Learning Modeling of Mass Spectrometry Data

Juan Sebastian Piedrahita Giraldo[1], Wout Bittremieux[1]
1: University of Antwerp

Aim of study: The use of mass spectrometry, and in particular tandem mass spectrometry (MS/MS), has proven to be a key tool for analyzing the composition of chemical compounds in medicine and other areas of the life sciences [1]. The patterns observed in the spectra allow researchers to identify the components of the target molecules along with possible similar compounds. Various spectral similarity metrics, like the cosine distance, are used during spectral library searching to match experimental MS/MS spectra to ground truth library spectra derived from known molecules [1]. However, these heuristic techniques do not fully correlate with the degree of similarity between the molecules under study. With the purpose of discovering novel structural analogs in metabolomics, we propose the use of Deep learning neural networks to model the relationship between mass spectrometry data and chemical structural information.

Methods: We leverage the availability of large scale mass spectrometry data in order to train our deep learning models. This data is sourced from both public library resources, such as GNPS, and private datasets. With the purpose of ensuring a variety of chemical compounds between training, validation and testing, we divide the data based on the Murcko Scaffold method which aims to group compounds based on shared molecule backbones.

Results: We successfully trained Deep Learning neural networks capable of modeling structural similarity, surpassing the performance achieved by previous similarity metrics. A Siamese Network based on Fully Connected Layers is able to predict the Tanimoto Similarity with an r2 score of 0.20 surpassing static methods such as modified cosine similarity. This approach enabled us to identify similar compounds for the target spectra derived from our datasets.

Conclusions:

This research has enabled us to model the relationship between mass spectrometry data and chemical structural information via Deep Learning Models, exploiting the availability of large scale datasets . We not only provide a more robust similarity metric for metabolite identification but also facilitates the discovery of analog compounds by exploiting the models trained. Given that structural similarity is not always reflected in classical spectral similarity, our deep learning algorithms promise to be a useful tool for detecting molecules with similar chemical characteristics.

References:

[1] Bittremieux, Wout, et al. "The critical role that spectral libraries play in capturing the metabolomics community knowledge." Metabolomics 18.12 (2022): 94.

P.9: Thomas Vanpoucke (University of Antwerp)

Scalable, flexible, and customizable GUI based workflow for multiple methylome data types

Vanpoucke Thomas[1,2], Annear Dale[3], Ibrahim Joe[1,2], Mateiu Ligia[1], Van Camp Guy[1,2], Kooy Frank[3], Op de Beeck Ken[1,2]
1: Centre of Medical Genetics, Human Molecular Genetics, University of Antwerp and Antwerp University Hospital, Edegem, Belgium
2: Centre for Oncological Research Antwerp (CORE), University of Antwerp and Antwerp University Hospital, Wilrijk, Belgium
3: Centre of Medical Genetics, Cognitive Genetics (COGNET), University of Antwerp and Antwerp University Hospital, Edegem, Belgium

Cancer is characterized by a variety of genetic and epigenetic changes that include DNA methylation modifications. Methylation aberrations occur across the entire genome and develop early in the disease, making the methylome a rich source of biomarkers. Four commonly used methylation detection techniques are whole genome bisulfite sequencing, enzymatic methyl sequencing, methylation arrays and nanopore sequencing; each technique comes with its own set of benefits and limitations.

While existing workflows for these types of data sets mostly consists of selected tools, this study introduces a highly adaptable pipeline that consolidates multiple standard tools for all four analysis techniques. Utilizing Snakemake, a Python-based workflow management system, we integrate tools for methylation data analysis and quality control specific for each data type. This way, data analysis is simplified, while retaining the innate flexibility of all tools. The output is standardized across tools, generating additional quality metrics, including coverage metrics and a bias plot for sequencing-based methods. All parameters are collected in one structured configuration file in JSON format, so they can easily and transparently be adapted. Additionally, a graphical user interface was designed, facilitating the creation of the configuration file . As a result, only a basic understanding of bioinformatics is required to use this tool, without compromising performance. The analysis is command-line based and can be run on various computing environments, but it is recommended to analyze large experiments on servers or cloud environments due to data size. Since this pipeline is based on Snakemake, advantages include easy parallelization, reproducibility, and readability of the code. Additionally, it empowers users with the flexibility to choose preferred tools and seamlessly integrates with downstream analyses through its unified output.

P.10: Tijs Watzeels (University of Antwerp - VIB Center for Molecular Neurology)

Pipeline for cell-specific isoform calling from long-read single-nuclei sequencing

Peter De Rijk, Tijs Watzeels, Fahri Küçükali, Jasper Van Dongen, Júlia Faura Llorens, Lara De Deyn, Lena Duchateau, Tim De Pooter, Geert Joris, Rosa Rademakers, Mojca Strazisar, Kristel Sleegers, Wouter De Coster
Sleegers lab, VIB Center for Molecular Neurology

Aim of the study: Long-read sequencing of single-cell libraries is emerging as a powerful technique to accurately study splicing patterns at a cell-type specific resolution. Classic methods require matched short-read sequencing experiments to accurately identify cell barcodes. With decreased error rate of Oxford Nanopore Technology sequencing, this need for short-reads is fading. However, current tools for short-read-free analysis of single-cell long-read data are lacking and leaves ample room for improvement. We developed Scywalker, a scalable tool for analyzing long-read single-cell and single-nuclei sequencing data, to address this need.

Methods: Scywalker was developed using a single-nuclei sample derived from fresh frozen human brain using a density gradient protocol1 and 10X Genomics droplet generation and barcoding. The 10X library was sequenced using both nanopore long-read and Illumina short-read sequencing. Scywalker works short-read-free, and its performance is on the same level as that of the standard short-read workflow, which uses Cell Ranger for the RNAseq processing. Scywalker starts from the raw sequencing data and runs up to the cell-type demultiplexed gene and isoform discovery and quantification.

Results: Scywalker outputs pseudobulk gene and isoform count tables, grouped per cell type. These can be used as input for standard differential gene expression and transcript usage analyses. Identification of non-empty droplets and percentage of cell type distributions show very high overlap, and we find a strong correlation in per-cell UMI count.

Conclusions: Scywalker is an important tool that allows long-read single-cell or single-nuclei analysis without the need for short-read sequencing. Other tools currently available are fall short in scalability and are not realistically applicable to the increasingly large single-cell data. Scywalker performs just as well as short-read standard barcode identification. Furthermore, it has the important advantage that it can identify and quantify (novel) transcript isoforms, which cannot be accurately determined in short-read data.

References

1. Habib, N., et al., 2016, Science 353 (6302), 925-8.

P.11: Farhan Ul Haq (University of Antwerp)

Differentiating ROS1 kinase states targeted by tyrosine kinase inhibitors

Farhan Ul Haq, Ken Op de Beeck, Guy Van Camp, Geert Vandeweyer
1: Center of Medical Genetics, University of Antwerp and Antwerp University Hospital, Prins Boudewijnlaan 43/6, 2650 Edegem, Belgium
2: Center for Oncological Research (CORE), University of Antwerp and Antwerp University Hospital, Universiteitsplein 1, 2610 Wilrijk, Belgium.

Abstract

Tyrosine kinases have emerged as promising targets for personalized anti-cancer therapy due to their role in regulating aberrant protein activity. Well-established targets such as epidermal growth factor receptor (EGFR) and anaplastic lymphoma kinase (ALK) have demonstrated significant success in specialized anti-cancer treatments. Recently, the oncogenic potential of c-ros oncogene 1 (ROS1) has been recognized, particularly in non-small cell lung cancer (NSCLC) patients with ROS1 gene-fusions, comprising 1~2% of cases globally. In these patients, the kinase domain remains conserved, leading to unregulated kinase activity, making ROS1 an attractive target for tyrosine kinase inhibitors (TKIs).

Aim of the Study

This study focuses on the molecular modeling and molecular docking of ROS1 kinase, shedding light on its conformational states targeted by type 1 and type 2 TKIs. Type 1 TKIs target the ATP-binding pocket, while type 2 TKIs interact with an allosteric pocket adjacent to the ATP-binding region, offering kinase-specific inhibition. Despite limited information on the native function and ligands of ROS1, computational modeling and simulation methods are employed to gain insights. Utilizing existing ROS1-specific data and homologous kinase information, such as ALK and the Src kinases, helps overcome the challenges posed by the lack of structural and experimental data on ROS1.

Methods

The homology modeling technique is applied using c-ABL kinase structure (PDB entry 1IEP) as a template to model ROS1 DFG-out structures as no such experimental information is available. The final selected models are employed to create docked complexes with Crizotinib and Foretinib. We also use all-atom molecular dynamics (MD) simulations to study protein-ligand dynamics using GROMACS.

Results

We successfully modeled the ROS1 kinase domain in the DFG-out state which provided structural basis for studying type 2 TKIs in the absence of experimental structural information. Furthermore, we also highlighted the key features distinguishing ROS1 states.

Conclusion

This study not only proposes parameters to distinguish between different states of WT ROS1, namely DFG-in and DFG-out states, but also delves into the evaluation of type 2 TKIs against ROS1. The findings from this computational exploration contribute valuable insights into the structural dynamics of ROS1 and its interactions with TKIs.

P.12: Baukje Bijnens (VIB CMN UA)

Hexanucleotide repeat expansions in C9orf72 alter microglial responses and prevent a coordinated glial reaction in ALS

Baukje Bijnens [1,2]#,Pegah Masrori [3,4,5]#, Kristofer Davie[6], Suresh Kumar Poovathingal[7], Annet Storm[4], Nicole Hersmus[4], Laura Fumagalli[1,2], Ludo Van Den Bosch[3,4], Mark Fiers[8], Dietmar Rudolf Thal[9,10], Renzo Mancuso[1,2]*, Philip Van Damme[3,4,5]

1. Microglia and Inflammation in Neurological Disorders (MIND) Lab, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium.
2. Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium.
3. KU Leuven - University of Leuven, Department of Neurosciences, Leuven Brain institute (LBI), Leuven, Belgium.
4. Laboratory of Neurobiology, VIB Center for Brain & Disease Research, Leuven, Belgium
5. University Hospitals Leuven, Department of Neurology, Leuven, Belgium.
6. Single Cell Bioinformatics Expertise Unit (CBD), VIB Center for Brain & Disease Research, Leuven, Belgium.
7. Single Cell Analytics & Microfluidics Core, VIB Center for Brain & Disease Research, Leuven, Belgium.
8. VIB Center for Brain & Disease Research, Leuven, Belgium.
9. Laboratory for Neuropathology, Department of Imaging and Pathology, Leuven Brain Institute
10. Department of Pathology, University Hospitals Leuven, Leuven, Belgium.

# These authors contributed equally
* These authors contributed equally

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease characterized by progressive motor neuron loss (MNs) in the motor cortex, brainstem, and spinal cord. ALS has a strong genetic background, a group of patients carrying a GGGGCC hexanucleotide repeat expansion (HRE) in the chromosome 9 open reading frame 72 gene (C9orf72). The pathogenicity of C9orf72 variants in ALS has been extensively studied and linked to loss of function (LOF) and gain of function (GOF) in neuronal and non-neuronal cells. Neuroinflammation and the high expression of C9orf72 in microglia point to a glial contribution, the mechanism by which they contribute to disease is not fully understood.

We isolated single nuclei of the spinal cord and motor cortex from sporadic (sALS) and C9orf72 patients and controls. We primarily investigated the transcriptomic profile of microglia and astrocytes. In the microglia we found a high expression of C9orf72, that was reduced in the HRE carriers. HRE-microglia are lower in reactive markers and display deficits in phagocytic and lysosomal transcriptional pathways, indicative of an impaired response in disease. Microglia from sALS cases displayed an increase in reactivity, indicating a different contribution of microglia in sporadic versus C9orf72 patients. Astrocytes presented with a dysregulated response in C9orf72 HRE, with cells remaining in a homeostatic state. This data provides evidence towards a coordinated glial response, ultimately contributing to ALS.

Overall we present a microglia loss-of function in C9orf72 HRE carriers, likely impairing microglial-astrocyte communication thereby preventing glial response.

P.13: Mahdi Safarpour (Global Health Institute (GHI))

Unravelling the contribution of human mobility in malaria parasite persistence: current state and challenges towards elimination

Mahdi Safarpour, Jean-Pierre Van Geertruyden, Christopher Delgado-Ratto
Malaria Research Group (MaRCH), Global Health Institute, Faculty of Medicine and Health Sciences, University of Antwerp, Belgium

Human mobility facilitates the importation of parasites from one region to another, potentially undermining efforts to eliminate malaria. Therefore, it is essential to quantify the role of human mobility in the spatial distribution and connectivity of malaria parasites in countries aiming for malaria elimination. Despite the availability of various statistical approaches for analyzing malaria transmission dynamics, there remains a significant gap in our knowledge regarding their ability to model the impact of imported malaria cases on the persistence of malaria transmission in endemic areas over time. Given this context, this systematic review aimed to aggregate existing evidence related to frequently employed statistical methods, their underlying assumptions, and limitations.

Out of 35 articles initially retrieved from PubMed and Scopus, 13 eligible studies published between 2007 and 2022 were identified after full-text screening. The findings underscored that admixture analysis, phylogenetic analysis, and discriminant analysis of principal components (DAPC) were the most used methods. These methods could identify genetic clusters among distinct parasite populations globally. Furthermore, some studies used within-infection fixation indices such as FST and FWS to discriminate between local and imported cases.

Despite the extensive use of these methods, they had significant limitations, the most important of which were: 1) Using a limited quantity of genetic data, which in practice reduces their capability to distinguish imported parasites in regions with closely related clones; 2) Analyzing the data cross-sectionally rather than longitudinally, limiting the applicability of these models for estimating the impact of human mobility on the persistence of malaria transmission over time. Given these limitations, we recommended to use alternative statistical methods, such as the Coalescent models or timed phylogenetic tree. Utilizing such methods, along with the parasites' whole-genome sequences and travel survey data, can provide more accurate information about the genetic changes within the parasite population introduced by human mobility.

Keywords: malaria elimination, imported malaria, molecular surveillance, statistical modeling