GENOME SCIENCE 2025

Genome Science UK 2025
9-11th July 2025
Newcastle University

Book of Abstracts

Contents

GENOME SCIENCE 2025

Thank you to sponsors:

GOLD

Element Biosciences

PacBio

Roche

SILVER

10x Genomics

Agilent Trusted Answers

biomodal

DNE Diagnostics North East

Dovetail Genomics Part of Cantata Bio, LLC

Earlham Institute Decoding Living Systems

illumina

New England Biolabs

Oxford Nanopore Technologies

Parse Biosciences

TaKaRa

Twist Bioscience

Yourgene Health

BRONZE

Watchmaker Genomics

SCALE biosciences

Vazyme

GENOME SCIENCE 2025

Genome Science 2025
Keynotes, Invited Speakers and Sponsor Abstracts

A Longer Look: Benchmarking Long-Read Sequencing for Single-Cell Transcriptomics

Dr Anita Scoones – Earlham

Authors: Anita L. Ahlert Scoones1, Yuxuan Lan1, Charlotte Utting1, Lydia Pouncey2, Ashleigh Lister1, Sofia Kudasheva1, Neelam Mehta3, Naomi Irish1, David Swarbreck1, Karim Gharbi1, Wilfried Haerty1, Adam P Cribbs3, 4, David J. Wright1†, Iain C. Macaulay1†

† Equal Contribution

Affiliations: 1 Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK 2 University of East Anglia, Norwich, UK 3 Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, National Institute of Health Research Oxford Biomedical Research Unit (BRU), University of Oxford, Oxford, UK 4 Oxford Centre for Translational Myeloma Research University of Oxford, Oxford, UK

Abstract

Single-cell RNA-seq has enabled the profiling of millions of cells, driving large-scale cell atlas projects. Yet by relying on 3’ or 5’ counting with short-read sequencing, many studies overlook critical biology, including alternative splicing, gene fusions, and single-nucleotide variants. Long-read (LR) sequencing technologies (Oxford Nanopore and PacBio) can recover full-length transcripts in bulk and single-cell contexts, and recent advances in throughput and accuracy have made LR scRNA-seq increasingly practical. However, most benchmarking studies demonstrating its utility predate these technological improvements.

Here, we present a comparative analysis of current commercial LR scRNA-seq approaches, performing parallel analyses of matched 10X cDNA on Illumina short-read, PacBio, and ONT platforms. We comprehensively assess platform-specific performance, evaluating sequencing throughput, accuracy, clustering, and cell-type identification concordance across datasets.

We demonstrate the utility of LR scRNA-seq for isoform-resolved analyses, enabling direct immunoglobulin chain reconstruction and detection of alternative splicing in immune-relevant genes such as CD45. We also evaluate the impact of CRISPR-based depletion of highly expressed transcripts, showing that redistributing sequencing capacity away from uninformative transcripts enhances gene and isoform coverage, enabling deeper, more interpretable full-length transcript profiling in single cells. Finally, we highlight key workflow limitations related to cDNA synthesis biases and inadequate read filtering common across LR platforms.

Overall, our benchmarking provides a practical framework for researchers seeking to integrate LR approaches into their single-cell transcriptomic studies, enabling richer biological insights beyond gene counts alone.

Accurate detection of small non-coding RNAs using NEBNext Low-bias Small RNA Library Prep Kit

Louise Williams – New England Biolabs

Abstract

Changes in small non-coding RNA (sncRNA) expression have been implicated in the development and progression of cancers, neurological, and cardiovascular diseases. High-throughput sequencing is a powerful tool for the characterization of sncRNA sequences, however, library preparation methods often limit the accuracy and sensitivity of detection.

Here we introduce a ligation-based small RNA library preparation method that is characterized by reduced bias in addition to increased detection of sncRNA. The NEBNext Low-bias Small RNA Library Prep kit uses a broad input range of 0.5 ng to 1000 ng total RNA and generates libraries that can be made in a single day using a streamlined protocol with bead-based size-selections and cleanups. Even representation of sncRNAs was confirmed using a pool of synthetic miRNAs where approximately 90% of miRNAs were within 2-fold of the expected number, compared to less than 30% with competitor methods. Additionally, miRNA detection was consistent using 0.5 ng to 1,000 ng of total RNA from human brain. This low-bias method also detects 2'O-methylated sncRNAs, such as piRNAs and plant miRNAs, without any protocol modifications. This method shows a robust capability to generate high quality libraries with increased confidence in the detection of sncRNAs.

Empowering Discovery: Agilent's Integrated Genomics Solutions for the Future of Precision Science

Rahna Ayub – Agilent

Abstract

Agilent Technologies delivers end-to-end genomics solutions that accelerate discovery and improve data quality. This talk will spotlight innovations in target enrichment, NGS library preparation, and automation—highlighting how platforms like SureSelect and Magnis are enabling efficient, high-confidence genomic analysis. In an era where genomics is transforming medicine, agriculture, and environmental science, Agilent Technologies stands at the forefront of innovation. Visit our stand to get updated on our newest launch, AVIDA for liquid biopsy target enrichment.

Enabling Spatial Omics at true single-cell resolution

Matthieu Pesant - Takara

Abstract

We stand at the early stages of a revolutionary leap in biology. Spatial multiomics is lighting the way to an advanced understanding of the inner workings of human health and disease. Meet Trekker, the first true single-cell mapping kit. Designed as a simple reagent kit, Trekker integrates seamlessly with existing single-cell sequencing workflows and preserves the high molecular sensitivity of single-cell data. You can now transform standard single-cell genomics data into spatial data through a simple, instrument-free and straightforward workflow. We will discuss how Trekker simplifies spatial omics and how it integrates seamlessly with major single-cell transcriptomics and multiome workflows for biomarker discovery.

Leveraging single-cell long-read sequencing and machine learning for therapeutic target identification

Adam Cribbs - NDorms

Abstract

Single-cell long-read sequencing has opened new frontiers in resolving transcriptomic complexity, enabling isoform-level resolution, splicing analysis, and accurate variant detection. In this talk, I will present our application of scCOLOR-seq, a high-fidelity long-read single-cell RNA sequencing technology, to assess novel drug targets in multiple myeloma. By capturing full-length transcripts at single-cell resolution, we identify isoform-specific signatures and transcriptomic features invisible to short-read methods. Integrating this rich molecular data with a machine learning toolbox—built to prioritise therapeutic targets via graph-based and multimodal inference—we uncover both known and previously uncharacterised targets. This platform advances our ability to dissect disease biology and supports the development of more precise, isoform-informed therapeutic strategies in complex cancers such as multiple myeloma.

Low Cost, High Accuracy Next Generation Sequencing

Alex Watkins, Ultima Genomics

Abstract

Ultima Genomics is advancing the field of omics by enabling scalable, cost-efficient sequencing for multiple counting-based omics applications. Multimodal studies are increasingly reliant on low-cost sequencing, the ability to quantify gene and protein expression, as well as detection of genomic variants. At launch our UG 100™ sequencing platform was the first system to generate a $100 genome and in February we brought the price down to $80.

Mapping the Respiratory Microbiome at a Spatial Resolution

Josie Bryant – Sanger Institute

Abstract

Chronic lung diseases such as cystic fibrosis are driven by complex polymicrobial communities and host-pathogen interactions, yet our understanding of their spatial dynamics within lung tissue remains limited. Josie Bryant will present cutting-edge work from the Lung Microbiome Lab at the Wellcome Sanger Institute, integrating deep metagenomics and spatial transcriptomics to map bacterial populations and their interactions with host cells in situ. The talk will cover work on Mycobacterium abscessus and Pseudomonas aeruginosa and early results on how microbial localization and microenvironment influences virulence and immune response.

Keynote

Rapid intraoperative diagnostics of CSF tumours

Matt Loose – University of Birmingham

Abstract

Over the past few years, our team has been developing new approaches to improve the diagnosis of brain tumours using real-time genomic technologies. As genomic medicine continues to move closer to the point of care, we've focused on creating methods that can operate directly within the surgical setting—delivering meaningful results in time to inform clinical decisions during an operation.
In this talk, I will describe the development and application of ROBIN (Rapid nanopOre Brain intraoperatIve classificatioN), our integrated sequencing-based workflow designed for intraoperative and rapid post-operative classification of CNS tumours. Using nanopore technology, we're able to provide near-instant insights into a tumour's methylation profile while surgery is ongoing, and a more comprehensive molecular profile—including single nucleotide changes and structural variants—within 24 hours.
This work draws on adaptive sampling methods to enrich key genomic regions dynamically, allowing us to detect not just epigenetic signatures but also deletions, insertions, and other clinically relevant alterations in real time. We've combined these innovations with a suite of tailored machine learning classifiers—including Sturgeon and RapidCNS2—to ensure accurate tumour classification, even under the time and data constraints of the operating theatre.
To date, we've applied ROBIN in nearly 150 cases and have validated the approach across multiple centres. We are now working closely with NHS pathology teams to support the transition of this technology from research environments into routine clinical practice—bringing real-time genomic analysis into the hands of those making critical decisions at the time the patient needs it.
Note that all work presented has been for research use only.

Revealing mechanisms of chemo-resistance with next generation lineage tracing technologies

Cannell IG, Wild SA, Pearsall SM, Hemmer N, Spalding K, Hannon GJ and Sawicka K*

Abstract

Triple negative breast cancers (TNBCs) are characterised by high levels of intra-tumoural heterogeneity, which is thought to underly drug resistance. Until recently bulk sequencing approaches and phylogenetic reconstruction have been used to infer clonal evolution. Such methods are limited in their ability to provide mechanistic insight, and additional clonal characterisation, such as transcriptional profiling, is required to identify vulnerabilities of drug resistant clones. Single cell technologies have provided a critical advance in prospective lineage tracing. Our WILD-seq method utilises genetic labelling of individual tumour cells with inert, expressed barcodes. By coupling barcoding with single cell RNA-seq, WILD-seq simultaneously tracks clonal lineage and gene expression. We initially applied WILD-seq to two syngeneic TNBC mouse models and identified clonal drivers of sensitivity and resistance to taxane-based chemotherapy. Taxane-resistant clones shared a signature of high NRF2 signalling, a major anti-oxidant hub, which could be exploited as it leads to a dependence on non-essential amino acids (NEAAs). Combination treatment with L-asparaginase (to deplete extracellular asparagine) and taxanes significantly improved tumour growth suppression and survival in mice. We are currently exploring more broadly the ability of dietary restriction of NEAAs to promote response to therapy.

Although a powerful technology, WILD-seq only allows measurement of clonal lineage and transcriptome within a heterogeneous cell population. Further characterisation of individual clones and their downstream functional analysis is hindered by difficulties in isolating clones of interest. To overcome this, we are developing a second-generation technology called clone-TAMeR (Transcriptional Activation-Mediated Recovery) which has all the features of WILD-seq but with the additional capability of using CRISPRa to activate a fluorescent reporter in a specific clone of interest. This enables targeted retrieval by flow cytometry, and functional perturbation of isolated clones. In addition, we are adapting WILD-seq to be compatible with spatial transcriptomics, thus enabling the spatial organisation of clones and their relationship with the tumour microenvironment to be determined. Ultimately, we aim to provide the research community with a suite of technologies which allow investigations of clonal dynamics and behaviours within heterogeneous tumour models. We anticipate their application will provide new insights

into various aspects of tumour biology including chemoresistance and metastasis.

Size Matters - Enhance your long-read sequencing data and yields

Darren Marjenberg – Yourgene Health

Abstract

There are limited options for fragment length quality control (QC) of nucleic acid samples and libraries in long read sequencing workflows, such as PacBio workflows. Traditional direct current (DC) electrophoresis techniques are readily available but have limited utility for DNA fragments larger than 20 kb. A narrow set of pulsed field (PF) electrophoresis instrumentation exists for this purpose, but these are inaccessible or impractical for many laboratories due to a combination of cost and operating complexity. Motivated to address this gap, Yourgene Health designed and tested a low-cost modification to Yourgene Health's automated DC electrophoresis instrument (LightBench).

The presentation explains how now for research customers conducting long-read sequencing, who need both large fragment analytics (larger than 20 kb) and DNA size selection, the new LightBench® Discover is a high-precision 3-in-1 instrument that simplifies workflows by combining multiple functions into a single benchtop solution.

Enhancing research and accelerating the rate of discovery, the LightBench Discover is an integrated solution that replaces the need for multiple instruments, reducing costs, enhancing QC, simplifying workflows, and delivering high-accuracy analytics.

Why One Genome Just Isn't Enough: How Multiple Wheat Genomes Are Transforming Our Understanding of Plant Evolution

Anthony Hall – Earlham

Abstract

Recent advances in long-read sequencing have made it feasible, both economically and technically, to generate high-quality genomes at scale. Sequencing multiple individuals within a species is revealing unexpectedly complex patterns of genetic variation. In crop species like wheat, this variation underpins key phenotypic traits, including responses to abiotic and biotic stress, with important implications for breeding and agritech innovation. Using wheat as a case study, I will highlight how genome-scale data and improved annotations have uncovered extensive structural variation, shedding light on the role of introgression in shaping genome evolution and introducing beneficial diversity.

GENOME SCIENCE 2025

Genome Science 2025
Submitted Abstracts

2

High-Resolution Mapping of Bacterial Evolution Under Antibiotic Stress Using Single-Cell Whole-Genome Sequencing

Matt Bawn[1], Johana Hernandez[2], Gaetan Thilliez[3], Mark Webber[3], Eleftheria Trampari[3], Rob Kingsley[3], Neil Hall[2], Iain Macaulay[2]

[1]Newcastle University, Newcastle, United Kingdom. [2]Earlham Institute, Norwich, United Kingdom. [3]Quadram Institute, Norwich, United Kingdom

Abstract

Antimicrobial resistance (AMR) in *Salmonella enterica* poses a growing threat to public health, with resistance to fluoroquinolones such as ciprofloxacin reducing treatment efficacy. Traditional bulk sequencing often overlooks population heterogeneity, limiting our understanding of how resistance emerges and spreads. In this study, we applied single-cell whole-genome sequencing (scWGS) to investigate the evolutionary dynamics of *Salmonella* populations exposed to sub-inhibitory concentrations of ciprofloxacin.

Using Fluorescence Activated Cell Sorting (FACS), individual bacterial cells were isolated and subjected to Multiple Displacement Amplification (MDA) prior to Illumina sequencing. Genomic variants were identified with breseq and Bowtie2, while population structures were inferred using RAxML and RHierBaps. Principal component analysis (PCA) and rarefaction analyses quantified intra-population diversity and evolutionary trajectories.

Our data revealed that *Salmonella* populations can evolve resistance via distinct sub-lineages while retaining high levels of genetic diversity. Key resistance-associated mutations, such as those in *gyrA* and *acrR*, were detected at frequencies consistent with bulk sequencing but with enhanced resolution of sub-population structures. Notably, resistance evolved without significant genetic bottlenecks, indicating that weak antibiotic stress can promote parallel adaptive paths.

This study demonstrates the utility of scWGS in capturing fine-scale evolutionary dynamics in bacterial populations, offering insights missed by conventional methods. The ability to track resistance development at single-cell resolution provides a powerful tool for understanding AMR evolution and could inform strategies for monitoring and mitigating resistance in clinical and environmental contexts.

3

Artificial Lighting Induces Gene Expression Responses Linked to Oxidative Stress in Bats

Tom Jenkins[1,2], Thomas Foxley[2], Emma Stone[2], Orly Razgour[1]

[1]University of Exeter, Exeter, United Kingdom. [2]University of Bath, Bath, United Kingdom

Abstract

Artificial lighting associated with urbanisation is a key threat to biodiversity. Despite recent progress in understanding the impacts of artificial light on nocturnal animal behaviour, no studies have yet assessed responses at the genomic level. This study aims to assess the impact of artificial lighting on gene expression in a light-averse species, the lesser horseshoe bat (*Rhinolophus hipposideros*). A lighting experiment was carried out in which bats emerging from a roost were exposed to dark or light conditions. We collected blood samples from 20 bats, ten from each treatment, to extract and sequence total RNA. Blood was chosen to represent changes in gene expression as it is a non-lethal sampling method that offers a proven alternative to study changes in bat transcriptomes. We identified 251 genes that were upregulated in bats exposed to light. However, while all of the control bats consistently showed little or no changes in gene expression, the light-exposed bats showed variability in their levels of upregulation. Our gene enrichment analysis found biological pathways linked to oxidative stress responses, DNA damage responses and apoptosis, and shifts in metabolism, protein and organelle homeostasis, nutrient uptake and gas transport. This suggests that some bats are particularly sensitive to light and that exposure to artificial lighting can induce a strong response that exceeds normal cellular function. Hence, our work highlights that artificial lighting does not only induce behavioural responses in bats, but also induces pronounced physiological responses that can be observed at the genomic level.

4

ULTRA-LONG-READ OXFORD NANOPORE SEQUENCING ENABLES PRECISE CHARACTERISATION OF IGH::DUX4 REARRANGEMENTS AND OTHER COMPLEX SUBTYPE-DEFINING FEATURES IN CHILDHOOD LEUKAEMIA

Steven Hair, Joe Fenwick, Eleanor Woodward, Robyn Sherrard, Richard Yim, Jack Bakewell, Sarra Ryan

Newcastle University, Newcastle Upon Tyne, United Kingdom

Abstract

Understanding the genomic landscape of childhood leukaemia is vital for risk-stratification and guiding treatment. While many clinical drivers of leukaemogenesis are now well-characterised, certain structural variants remain difficult-to-detect, most notably *IGH::DUX4* rearrangements. Advances in sequencing technologies, particularly the ultra-long-read capabilities of Oxford Nanopore (ONT) sequencing, allow us to study these once-cryptic rearrangements for the first time.

We applied ONT ultra-long-read sequencing to diagnostic samples from 32 leukaemia patients, including seven with known *IGH::DUX4* rearrangements. Each sample was sequenced on a single PromethION R10 flow-cell using the ultra-long library kit (SQK-ULK114). Data were aligned to the T2T-CHM13V2.0 reference genome and SNPs, structural variants, allele haplotypes and CpG methylation profiled.

Sequencing produced a median library N50 of 82 kb (55-106 kb), with a median sequencing depth of 23.3x genome coverage (11.5-35.5x) and a mean read quality of 18.9±0.1. The longest single molecule read in the cohort was 1.46 Mb. All seven cases of *IGH::DUX4* were fully characterised using this method, elucidating the full length and location of the translocation to base-level resolution for the first time. Methylation profiling further revealed epigenomic-highjacking and a loss of 5mC methylation in translocated *DUX4* genes.

This study demonstrates that ultra-long ONT sequencing enables the detection and precise characterisation of *IGH::DUX4* rearrangements in leukaemia for the first time. These once-cryptic events were identified in all known cases, with precise breakpoint localisation. Other subtype-defining features such as high-hyperdiploidy, iAMP21 and *BCR::ABL1* were also detected using this method, highlighting its potential in both diagnostic and research settings.

5

A human phenotype ontology driven analysis of published clinical descriptions of rare neurodevelopmental conditions

Mosab Alsaedi[1,2], Peter Robinson[3], Alisdair McNeill[1,4]

[1]Division of Neuroscience & Neuroscience Institute, the University of Sheffield, Sheffield, United Kingdom. [2]Department of Genetic Medicine, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia. [3]Berlin Institute of Health at Charité (BIH), Rahel Hirsch Center for Translational Medicine, Berlin, Germany. [4]Sheffield Clinical Genetics Service, Sheffield Children's Hospital NHS Foundation Trust., Sheffield, United Kingdom

Abstract

Purpose

To identify where in the medical literature the most phenotypic information about rare, genetic neurodevelopmental conditions is found.

Methods

The Monarch annotator was used to annotate HPO terms in free text in medical research papers. We validated the Monarch annotator using the gold standard corpus of manually annotated abstracts. A curated list of genes associated with neurodevelopmental conditions was identified from the DDG2P gene list. A literature review was conducted to identify the initial disease-gene association papers for each gene, and papers describing follow-up case series for each gene. The Monarch annotator was used to annotate phenotypic information in research papers. The number of HPO terms reported in gene-disease association papers was compared to the number of HPO terms in follow-up case series statistically.

Results

The Monarch tool demonstrated an overall F1 score of 0.51, with a higher precision rate (0.55) than recall (0.49). This is comparable to other semi-automated annotation tools. The initial disease-gene association papers reported significantly fewer HPO terms than follow-up case series. The number of HPO terms per paper was linearly related to the number of people with the rare condition described in the paper. The follow-up case series frequently reported phenotypic information that was important to guide clinical management. The follow-up case series were published in journals with impact factors significantly lower than those of the gene-disease association papers.

Conclusions

The most in depth phenotyping for rare genetic neurodevelopmental conditions is reported in follow-up case series, in comparatively low impact journals.

6

Spike-in-controls for assessing the accuracy of methylation calling in EM-seq

Lucy Faulkner, Tamsin Robb, Henson Lee Yu, Harveer Dev

Early Cancer Institute, Cambridge, United Kingdom

Abstract

Enzymatic methylation sequencing (EM-seq) enables DNA methylation patterns to be uncovered through C>T conversion of unmethylated cytosines, via two enzymatic steps: TET2 oxidisation of 5mC/5hmC followed by APOBEC deamination of unprotected cytosines. Inefficiency in either enzymatic reaction can result in over- or under-conversion and resultant misclassification. Methylation controls assess the performance of methylation calling.

NEBNext® EM-seq (Twist Biosciences) was performed over multiple sets of samples. 0.1 pg of each methylation control was spiked into samples: enzymatically methylated pUC-19, a plasmid derived from *E. coli* (methylated control) and lambda bacteriophage DNA (unmethylated control).

In a cohort of prostate cancer cell-line samples, we demonstrated high fidelity in methylation calling (pUC-19: 99.4% methylated, lambda:99.7% unmethylated). However, flow-cytometry sorted prostate cancer cells saw lower efficiency in the methylated control; further analysis revealed an increase in unmethylated calls around ~1000 bp along the pUC-19 sequence, with a similar number of methylated calls throughout the plasmid.

Potential causes of this suboptimal methylation calling include inefficiency of TET2 oxidation, loss of methylation of pUC-19 or misalignment of reads to this region. We are exploring the possibility of *E. coli* contamination of samples through PCR of the input DNA, with primers designed to bind to this region.

Spiking controls into individual samples provides the opportunity to monitor performance at a sample level; however, it may be preferable to process controls as a separate reaction during laboratory processing (with both strategies suggested by the manufacturer). The results of a head-to-head comparison of these two approaches will be discussed.

7

Nanopore-based hybridization capture approaches for targeted viral metagenomics and whole-genome sequencing

Oscar Enrique Torres Montaguth[1], Sarah Buddle[1], Leysa Forrest[1], Tony Brooks[1], Helena Tutill[1], Rachel Williams[1], Sofia Morfopoulou[1], Judith Breuer[1,2]

[1]University College London, London, United Kingdom. [2]Great Ormond Street Hospital, London, United Kingdom

Abstract

Hybridization-capture approaches can significantly increase the sensitivity of metagenomics. However, their implementation with Oxford Nanopore Technologies (ONT) sequencing has been limited due to the lack of dedicated workflows and the need for an extra ligation-based library preparation step. We developed a universal four-primer PCR approach that allows the easy implementation of hybridization capture workflows with ONT without additional ligation-based steps.

We evaluated our protocol using targeted metagenomics and whole-genome sequencing (WGS) of viruses. For targeted metagenomics, we tested the Twist Bioscience Comprehensive Viral Research Panel (CVRP) on a dilution series from of a six-virus mock community in a human DNA+RNA background. The ONT-CVRP workflow increased sensitivity by 10 to 100-fold over untargeted ONT, making it suitable for the detection of low viral loads. Viral genome coverage was also increased compared to untargeted approaches, which could facilitate applications such as phylogenetic and epidemiological analysis. Real-time analysis allowed the detection of all viruses at high viral loads within two hours of sequencing, providing actionable results early in the run.

For WGS, we used Agilent SureSelect workflow using mock samples spiked with Cytomegalovirus (CMV) DNA and SARS-CoV-2 RNA and obtained high-coverage full genomes from both viruses.

Our protocol facilitates the use of hybridization-capture workflows with ONT, providing both rapid detection and whole-genome sequences where the pathogen is abundant and identification where pathogen loads are low. In future, it could be used not only in pathogen sequencing, but in a wide variety of clinical contexts, such as exome panels, pharmacogenomics and cancer panels.

10

Ultra-long Sequencing Reveals Structural Heterogeneity of Common Dicentric Chromosomes in Childhood Cancer

Eleanor Woodward, Steven Hair, Joe Fenwick, Richard Yim, Steven C. Clifford, Sarra L. Ryan

Newcastle University, Newcastle upon Tyne, United Kingdom

Abstract

Most cancers harbour chromosomal abnormalities, including ploidy changes and structural variants (SVs). Dicentric and isochromosomes, defined as chromosomes possessing two regions of centromeric sequence, are frequent in childhood cancer but poorly understood, largely due to the breakpoints falling within highly repetitive, complex regions of DNA. Ultra-long whole-genome sequencing (UL-WGS) by Oxford Nanopore Technologies enables the characterisation of complex SVs for the first time.

We performed UL-WGS on six childhood medulloblastoma (MB) cell lines, each harbouring a dicentric or isochromosome, confirmed by karyotyping and copy number array. Sequencing libraries were prepared using the ULK-SQK114 kit and sequenced on a PromethION with a single flow cell. Reads were aligned to the T2T-CHM13 reference genome and analysed using a custom bioinformatics pipeline designed to detect breakpoints within repetitive regions.

The median N50 and coverage were 69kb (range, 57-89kb) and 25x (range, 27-34x), respectively. Base-level resolution of dicentric and isochromosome breakpoints was achieved in all six samples, including centromeric breakpoints in three. Interestingly, in cases with the common isochromosome i17q, breakpoint heterogeneity was observed, and breakpoints diverged from those inferred by copy number array analysis.

The study demonstrates, for the first time, the utility of UL-WGS in resolving dicentric and isochromosomes in childhood cancer. The findings reveal structural heterogeneity and suggest variability between the genomic regions and genes involved in recurrent isochromosomes. These results offer novel insights into the mechanisms underpinning the formation and maintenance of dicentric and isochromosomes in cancer development.

11

Accurate detection of small non-coding RNAs using NEBNext Low-bias Small RNA Library Prep Kit

Maximilian J Fritsch Fritsch[1], Heather Raimer Young[2], Deyra Rodriguez[2], Louise Williams[2]

[1]New England Biolabs UK Ltd, Hitchin, United Kingdom. [2]New England Biolabs Inc, Ipswich, USA

Abstract

Changes in small non-coding RNA (sncRNA) expression have been implicated in the development and progression of cancers, neurological, and cardiovascular diseases. Use of sncRNAs as disease biomarkers requires their precise and sensitive detection. Therefore, high-throughput sequencing is instrumental in characterising sncRNAs for this purpose. However, library preparation methods often limit accuracy and sensitivity, and ligation-based methods introduce bias, thus obscuring true sncRNA composition. Consequently, improvements in library preparation methods are essential for using sncRNAs as clinical biomarkers.

We developed a novel, ligation-based small RNA library preparation method with reduced bias and increased sncRNA detection. Libraries can be prepared in a single day using a streamlined protocol with bead-based size-selections and cleanups. This method is robust, compatible with a broad input range, and effective with challenging samples like formalin-fixed paraffin-embedded (FFPE) RNA.

Using a pool of synthetic miRNAs, we confirmed even sncRNA representation with our low-bias method. Approximately 90% of miRNAs were within 2-fold of the expected number, compared to less than 30% with other methods. miRNA detection was consistent using 0.5 ng to 1,000 ng of total RNA from human brain. Additionally, this method detected 2'O-methylated sncRNAs, such as piRNAs and plant miRNAs, without protocol modifications. Libraries made from low-quality FFPE total RNA (1 ng to 100 ng) resulted in consistent yields and miRNA detection. Regardless of input or sample, this method generates high-quality libraries, increasing confidence in sncRNA detection and potential disease biomarker identification.

12

Barcoded ultra-long read sequencing for genetic exploration of cancer samples

Joe Fenwick, Steven Hair, Robbie Bain, Eleanor Woodward, Sarra Ryan

Newcastle University, Newcastle Upon Tyne, United Kingdom

Abstract

Ultra-long read sequencing has proven to be a valuable tool for creating highly contiguous telomere-to-telomere assemblies, which are providing insight into poorly explored genomic regions in human disease for the first time. Despite this, application to patient samples has been limited by material requirements. Here, we describe a method of barcoded ultra-long (BULK) read sequencing, which builds upon community driven efforts to optimise Oxford Nanopore library preparations and utilises adaptive sampling (AS) to enrich regions of interest, while reducing DNA requirements and cost by 66% and 64.2%, respectively, per sample.

BULK-AS was performed on nine haematological cancer samples, providing up to 30x read depth over targeted regions (median 23.1x; range, 17.6x-29.9x; n=9), with coverage comparable to standard ultra-long whole genome sequencing (UL-WGS) of the same regions of interest (ROI) (21.3x; 13.6x-35.2x ; n=8) (p = 0.306) and native barcode sequencing (NBD) with the same number of samples per library (19.8x; 11.5x-42.7x; n=9) (p = 0.101). However, the median N50 of BULK-AS (79kb; 64-106kb; n=9) was significantly greater than NBD (9kb; 8.3kb-9.7kb; n=9) (p = 2.6e-06) and similar to UL-WGS (70.5kb; 64.3kb-86.1kb; n=8) (p = 0.191). BULK produced reads up to 2.23Mb which helped to characterise complex and repetitive regions of the genome such as the entire D4Z4 array (T2T-CHM13, chr4:193,439,356- 193,553,170).

BULK-AS sequencing is a versatile, scalable tool that reduces sample input requirements and cost, while leveraging two of the most alluring features of the Oxford Nanopore platform; exceptional read length and preparation-free target enrichment.

15

Emerging Point-of-Care Testing (POCT) in Genomics

Shona Haston[1,2], James Elgey[1,2], Olushola Ewedairo[1,2], Beth Green[1,2], Alex Inskip[1,2], Katie Thompson[1,2], Andrew Mkwashi[1,2]

[1]Newcastle University, Newcastle upon Tyne, United Kingdom. [2]NIHR Innovation Observatory, Newcastle upon Tyne, United Kingdom

Abstract

As genetic testing becomes mainstream in standard-of-care settings, there may be in interest in employing genetic point-of-care tests (POCTs) that analyse genetic variation in humans rapidly and accurately. This horizon scan aimed to identify and characterise near-market (≤5 years) standalone POCTs that can identify variation(s) in the genome and epigenome that enable to use of genetic information to inform clinical decision making.

Searches conducted on Embase and MEDLINE identified a total of 348 unique technologies. The technologies were being developed for a range of indications; 18 broad conditions were represented. Cancer was the most common, accounting for 69.3% of the total technologies. Purposes were diagnosis (n=266), treatment decisions (n=20), prognosis (n=3), or a combination (n=55).

Technology readiness levels (TRLs) were used to determine the market readiness, alongside in-house tiers based on criteria including CE-mark, clinical trials, automation level, sample type, and clinical validation.

This scan identified 34 technologies designated 'true POCTs' by the above criteria, and 312 technologies requiring specialist operation or off-site analysis; some of these technologies may evolve into true POCTs within the five-year timeline to market, as defined in this horizon scan. Most of the technologies were at earlier stages of development (TRL 3 or TRL 4; n=266, 76.4%), with few at clinical trial stage (TRL 5 or TRL 6; n=4, 1.1%) or CE-marked (TRL 9, n=18, 5.2%).

Overall, a range of genomic POCTs were identified, though relatively few were at later stages of development and/or yet able to be classified as being 'true POCTs'.

16

Geminio: a Nextflow workflow to uncover bacterial copy number variations

Sarah Cameron, Andrew Preston

University of Bath, Bath, United Kingdom

Abstract

Copy number variations (CNVs) in bacteria are understudied due to their instability and complex nature. They arise during homologous recombination between repetitive sequences, such as transposable elements and rRNA genes, leading to the duplication of regions between them. Although maintaining extra regions of DNA is thought to incur fitness costs, in certain situations, CNVs are thought to confer advantages. The phenomenon known as heteroresistance, where numbers of antibiotic resistance genes are temporarily amplified, is frequently reported but difficult and labour intensive to comprehensively study and therefore it's contribution to the public health problem of antimicrobial resistance (AMR) remains largely unknown.

However, with whole genome sequencing being more accessible than ever, it may offer a solution by using read depth to estimate CNVs. Here, we have developed a Nextflow workflow, *Geminio*, to screen sequencing data for CNVs, cataloguing their location, length, gene content, annotations and estimation of the breakpoints. We report the first findings from screening 1705 Klebsiella pneumoniae isolates from the One Health SpARK study exploring the relationships between the genes in CNVs detected, their metadata and antibiotic susceptibility profiles. This pipeline will help provide novel information on CNVs in bacterial evolution and start to unpick their relationship with AMR.

17

Genomic epidemiology of plasmids carrying the *iuc5* locus and antimicrobial resistance genes in *Escherichia coli* and *Klebsiella pneumoniae* from One Health settings

Keira Cozens[1], Marjorie Gibbon[1], Natacha Couto[2], Tiffany Taylor[1], Matthew Avison[3], Edward Feil[1]

[1]University of Bath, Bath, United Kingdom. [2]University of Oxford, Oxford, United Kingdom. [3]University of Bristol, Bristol, United Kingdom

Abstract

Aerobactin, a plasmid-mediated siderophore encoded by the *iuc* locus, is a major contributor to virulence in *Klebsiella pneumoniae.* Whilst certain lineages of aerobactin are well studied, the epidemiology of *iuc5* remains poorly understood. Moreover, our previous work suggests that plasmids harbouring *iuc3* are acquiring antimicrobial resistance genes, which may pose a major threat to public health. Here we generate hybrid assemblies of the critical priority pathogens *E. coli* and *K. pneumoniae* from humans, animals, and the environment, to determine how frequently plasmids emerge that possess both *iuc5* and antimicrobial resistance genes.

We utilised *E. coli* and *K. pneumoniae* genome sequence data from a large One Health study conducted across Thailand, generated by the OH-DART consortium. We focused on strains carrying the plasmid-mediated virulence locus *iuc5*. Long-read sequencing was used to generate hybrid assemblies of 68 *E. coli* isolates and 2 *K. pneumoniae* isolates harbouring the *iuc5* locus. Resfinder and Kleborate were used to identify plasmids with acquired resistance genes and the *iuc5* locus, respectively.

In total we identified 230 circular plasmids from *E. coli* strains harbouring the *iuc5* locus. Of these plasmids, 70 carried *iuc5*, and 68 of these plasmids contained 1 or more antimicrobial resistance genes. These plasmids were isolated from human community samples, hospital samples, fresh markets and across several different chicken, duck, and fish farms. We also identified two *iuc5*-harbouring plasmids in *K. pneumoniae* isolated from hospital samples.

Here, we report that plasmids harbouring the *iuc5* locus and antimicrobial resistance genes are widespread across different ecological sources.

18

Cell typing using Nanostring's CosMX in a GBM context

Joshua Winter-Luke[1], Lucy Stead[1], Rachel Queen[2], Shoaib Ajaib[1], Steven Pollock[1], Michael Jenkinson[3,4], Andrew Brodbelt[3,4], Michael Cearns[3,4]

[1]University of Leeds, Leeds, United Kingdom. [2]Newcastle University, Newcastle, United Kingdom. [3]University of Liverpool, Liverpool, United Kingdom. [4]Walton Centre NHS Foundation Trust, Liverpool, United Kingdom

Abstract

Glioblastoma (GBM) is the most aggressive adult brain tumour. Despite current treatment of surgery and chemoradiotherapy median survival is only 15 months. The tumours highly heterogeneous nature, combined with its infiltrative properties means cancer cells will be left behind and these lead to tumour recurrence driven by therapy-resistant cells. This project, supervised by Dr. Lucy Stead and Dr. Rachel Queen, focuses on understanding the mechanisms of resistance using spatial transcriptomics from paired primary and recurrent GBM samples from 27 patients. These have been profiled using the NanoString CosMX platform with both 1,000 (1k) and 6,000 (6k) gene panels. This allows gene expression to be identified at a cellular level allowing us to characterise the tumour microenvironment (TME) and spatially distinct cell populations that may drive recurrence.

I have developed a pipeline to identify these cell types, including quality control, clustering (Leiden and non-negative matrix factorisation), cell typing, and copy number aberration analysis. This has allowed me to compare the 1k and 6k gene panels to identify how gene breadth affects cell type identification. By applying methods such as Jaccard indexing between cluster-specific marker genes and a curated cell type reference list, I can annotate clusters and evaluate the consistency between panels.

This groundwork will give me confidence in cell identification so that I can characterise transcriptional trajectories in primary vs recurrent GBMs, and ultimately identify personalised therapeutic targets based on spatial context and treatment-driven evolution.

19

Investigating mitochondrial DNA replication dynamics using thymidine analogues and ONT long read sequencing.

Katja E Menger, Thomas J J Nicholls

Biosciences Institute, Faculty of Medical Science, Newcastle University, Newcastle Upon Tyne, United Kingdom

Abstract

Mitochondria contain their own multicopy, circular genome (mtDNA) which is maintained independently of the nuclear genome. Each cell contains thousands of copies of mtDNA, and this number of molecules per cell is maintained by mtDNA replication. There are two canonical origins of mtDNA replication, with one origin for each strand.

We have developed a method that enables us to track the progression of mtDNA replication *in vivo* at single nucleotide resolution. Cells are treated with the thymidine analogue BrdU, which is transported into mitochondria and incorporated into replicating mtDNA. Native DNA is sequenced using ONT ultralong protocols, and sites of BrdU incorporation are identified using dedicated software.

By labelling mtDNA with BrdU for different periods we have been able to determine the proportion of replicating mtDNA molecules at each timepoint, and are able to identify sites of both replication initiation and termination.

To assess the perturbation of mtDNA replication we have labelled cells with BrdU following depletion of the topoisomerase TOP3A. This enzyme is responsible for decatenating mtDNA daughter molecules at the end of replication, and its loss is known to cause genome-wide mtDNA replication stalling. Using this method, we are able to detect and quantify this reduced rate of mtDNA replication in the absence of TOP3A.

Overall, our method is suitable to study mtDNA replication dynamics and through further development we aim to advance our understanding of how mtDNA is replicated under physiological and pathological conditions.

20

Performance assessment of a rapid library preparation kit for short-read RNA sequencing

Thomas Brabbs[1], Camilla Ryan[1], Gillian Reynolds[1], Tom Barker[1], Alex Durrant[1], Kendall Baker[1], Sacha Lucchini[1], Chris Watkings[1], Magnus Morgan[2], David Swarbreck[1], Salah Azzi[2], Scott Kuersten[3], Karim Gharbi[1]

[1]Earlham Institute, Norwich, United Kingdom. [2]Illumina, Cambridge, United Kingdom. [3]Illumina, Madison, USA

Abstract

RNA sequencing (RNA-seq) has become ubiquitous for transcriptome-wide analysis of gene expression in animals, plants, and microbes. The continued decline in sequencing costs have acclerated the demand for rapid and effective methods to carry out RNA-seq experiments at scale, and at low cost. Recent advances in RNA-seq library preparation have introduced significant improvements, including fewer steps, shorter incubation times, and reduced consumables usage. As part of our efforts to develop a robust low cost, low-input workflow for high-throughput RNA-seq library preparation, we assessed the performance of Turbo RNAseq® (Illumina), a new fast library preparation kit for RNA sequencing on short-read platforms. We tested the robustness, reproducibility, and sensitivity of Turbo RNAseq across a range of input amounts in both human controls and hexaploid wheat and evaluated its performance for the analysis of differential gene expression in wheat landraces against our current gold standard. These findings will inform our strategy for the miniaturisation and automation of RNA-seq library preparation.

21

The role of genome structural variations in mediating antimicrobial resistance in *Mycobacterium abscessus*.

Brian Mpofu[1,2], Sarah Cameron[2], Christopher Moon[1], Andrew Preston[2], Joanna Bacon[1]

[1]UK Health Security Agency, CDEP, Discovery Group, Salisbury, United Kingdom. [2]University of Bath, Department of Life Sciences, Bath, United Kingdom

Abstract

Over the past two decades, infections from non-tuberculous mycobacteria (NTMs) have increased significantly, with *Mycobacterium abscessus* emerging as a leading and highly drug-resistant pathogen. Its extensive intrinsic resistome has earned it the nickname “incurable nightmare” in clinical settings. While antimicrobial resistance in *M. abscessus* has been well studied at the gene level, the contribution of genomic structural variations (SVs) remains largely unexplored owing to previous sequencing limitations.

Recent advances in long-read sequencing, particularly Oxford Nanopore Technologies (ONT), now permit de novo genome assembly and enhanced detection of SVs in the GC-rich, repetitive genome of *M. abscessus*. This work aims to identify novel resistance-conferring mutations, including SVs. Using the broth serial dilution method in 2 mL deep-well blocks after 3 days of incubation at 37 °C, the minimum inhibitory concentrations (MICs) for amikacin and apramycin, a novel aminoglycoside, were determined to be 1–2 µM and ~0.5 µM, respectively, against *M. abscessus NCTC 14045*. Populations were isolated after 7 days of exposure at drug concentrations above these MICs, and subsequent serial passages revealed intermediate and high-level resistance phenotypes.

ONT sequencing identified known rrs mutations (A1408G, G1491T, and C1409T) in amikacin-resistant populations and, for the first time, the same mutations in apramycin-resistant isolates, indicating a shared resistance mechanism. Additionally, previously unreported SVs, including a 2.5 Mbp symmetric genomic inversion disrupting acetyl-CoA C-acetyltransferase and inner-membrane protein YhjD, were found in some amikacin-resistant strains. Further analyses and ongoing work aim to elucidate the functional roles of these SVs in aminoglycoside resistance in *M. abscessus*.

22

Testing the regulatory effects of indels and structural variants using massively parallel reporter assays

Lindsey Plenderleith, Rachel Owen, Timothy Connelley, Musa Hassan, Liam Morrison, James Prendergast

Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom

Abstract

Regulatory DNA sequences can vary substantially between individuals, and such variation drives a number of important phenotypes. As sequencing technologies improve it is becoming more feasible to move beyond single nucleotide regulatory variants (SNVs) and study the effects of larger changes such as insertion-deletion mutations (indels) and structural variants (SVs), which may often have larger regulatory impacts. We have used the Survey of Regulatory Effects (SuRE) approach, a genome-wide massively parallel reporter assay, to screen the cattle and human genomes to identify SNVs with regulatory effects, and are now leveraging this approach to study the effects of larger variants. The SuRE method, which tests the ability of individual genomic DNA fragments to initiate transcription in an otherwise promoterless plasmid, allows the effects of individual variants to be tested, considerably reducing the confounding impact of linkage disequilibrium. We have combined the experimental approach of SuRE with a novel graph genomics analysis pipeline to improve the detection of regulatory effects of indels and SVs. Using libraries generated from cattle we successfully tested almost 1.4 million indels and SVs, ranging in size from 1 bp to 1.5 kb, and identified around 13,000 with a significant effect on gene expression in primary cattle cells. Work is ongoing to characterise further these potential regulatory variants and their relevance to understanding how indels and SVs shape important phenotypes. These results validate our method as a new tool for evaluating the functional effects of longer variants.

25

Hitting the Mark: Comparing Targeted, PCR-Free, Long Read Sequencing Techniques for the Verification of Transgenics

Aurélie Reijnen[1,2], Daniel Snell[1], Jessica Olsen[1], Graham Duddy[1], Xiang Li[1], Sunita Varsani-Brown[1], Stephanie Frost[1], Katherine Courtis[1], Robert Goldstone[1], Ben Davies[1]

[1]The Francis Crick Institute, London, United Kingdom. [2]University of Glasgow, Glasgow, United Kingdom

Abstract

Targeted PCR-free sequencing methods have a wide range of applications including clinical diagnostics, pharmacogenomics and the verification of transgenic organisms. Enzymatic targeted PCR-free sequencing methods include Nanopore Cas9-Targeted Sequencing (nCATS) which dephosphorylates DNA ends and subsequently utilises Cas9 to induce double-stranded breaks flanking the region of interest. The ligation of Nanopore sequencing adaptors requires phosphorylated DNA ends, meaning that most of the adaptor ligation and subsequent sequencing occurs at the targeted region. Computational targeted PCR-free sequencing methods include adaptive sampling approaches. By providing the genomic regions of interest, the sequencer can align the reads to these regions in real time and can eject any off-target reads from the flowcell's pores. This results in the reduced sequencing of reads that do not align to the region of interest and allows flowcell pores to be available for sequencing on-target reads. We aimed to compare the use of nCATS, adaptive sampling, and whole genome sequencing to evaluate if targeted sequencing methods represent an alternative, potentially more cost-efficient, route for the PCR-free validation of modified alleles. By analysing multiple case studies of transgenic mouse lines, we evaluated the use of targeted PCR-free long read sequencing methods the validation of genetically modified organisms. Additionally, we explored multiplexing targeted sequencing methods to provide a more cost-effective approach. Targeted approaches were able to identify unexpected features such as large duplications and regions of plasmid integration, demonstrating their capabilities for validating genetically modified organisms.

26

Benchmarking Long-Read scRNA-Seq Approaches and Optimising Transcript Resolution

Anita Scoones[1], Yuxuan Lan[1], Charlotte Utting[1], Lydia Pouncey[2], Ashleigh Lister[1], Sofia Kudasheva[1], Neelam Mehta[3], Naomi Irish[1], David Swarbreck[1], Karim Gharbi[1], Wilfried Haerty[1], Adam P Cribbs[3,4], David Wright[1], Iain C Macaulay[1]

[1]Earlham Institute, Norwich, United Kingdom. [2]University of East Anglia, Norwich, United Kingdom. [3]otnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, National Institute of Health Research Oxford Biomedical Research Unit (BRU), University of Oxford, Oxford, United Kingdom. [4]Oxford University, Oxford, United Kingdom

Abstract

Single-cell transcriptomics has enabled the profiling of millions of cells, fuelling large-scale cell atlas efforts. However, most scRNA-seq methods rely on 3’ or 5’ end counting to maximise throughput, limiting resolution of biologically meaningful events including alternative splicing, fusion genes, and single-nucleotide variants (SNVs).

Long-read sequencing enables isoform-level expression profiling, improving detection of splicing, transcriptional diversity, and novel regulatory isoforms. Technologies such as Oxford Nanopore (ONT) and PacBio (PB) now support full-length transcript sequencing from both bulk and single-cell cDNA. Recent advances in library preparation and sequencing throughput have brought long-read scRNA-seq closer to the mainstream, though many benchmarking studies predate these developments.

Here, we present a comparative analysis of commercial long-read approaches for scRNA-seq. Using cDNA generated from the 10X Genomics platform, we performed parallel sequencing across Illumina short-read, and PB and ONT long-read platforms. We also evaluated CRISPR-based depletion of abundant transcripts prior to long-read sequencing. Single-source cDNA enabled direct cross-platform comparisons of performance metrics, clustering, and cell type identification. While all approaches yielded informative data, common limitations—particularly inefficiencies in cDNA generation and read filtering—reduced the proportion of usable reads for downstream analysis.

To address these challenges, we are developing targeted enrichment strategies to selectively amplify genes and isoforms that define cell types and states—many of which are lowly expressed and often missed in whole-transcriptome profiling. This enables deeper coverage of full-length transcripts, supporting robust detection of SNVs and gene fusions beyond 5’ or 3’ ends, and improving the biological interpretability of long-read single-cell datasets.

27

Spatiotemporal and Single-cell Sequencing of chicken trachea reveals differences in host Responses after Vaccination with Pathogenic and Attenuated Infectious Bronchitis Virus (IBV)

Graham Freimanis[1], Ge Wu[1], Francesca Young[2], Ke Yuan[2], David Robertson[2], sarah keep[1], Erica Bickerton[1]

[1]The Pirbright Institute, Woking, United Kingdom. [2]University of Glasgow, Glasgow, United Kingdom

Abstract

Infectious Bronchitis Virus (IBV), an avian coronavirus, poses a major threat to the global poultry industry. Live-attenuated vaccines (LAVs) are currently the primary defence against IBV infections, but these vaccines, produced by serial passage of a virulent field isolate in embryonated hen eggs, have been shown to potentially revert to virulence, leading to outbreaks. The mechanisms behind this reversion and overall vaccine stability remain poorly understood.

To better understand virus and host responses to LAVs and how these impact upon LAV stability, we employed single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics (ST) to profile cell types and interrogate their spatial organisation within both healthy and IBV-infected chicken trachea. A pathogenic strain of IBV, M41-CK was used to produce a LAV in ovo. Tracheal samples were collected at days 1, 4, and 14 post-infections from chickens infected with either the pathogenic (M41-CK) or attenuated (M41-SK) virus, with a mock-infected group as a control.

Our analysis revealed distinct transcriptomic profiles that were dependent on the viral phenotype (pathogenic vs. attenuated). Using integrated computational analysis, we identified host transcriptomic markers linked to IBV infection and localised the IBV genome to specific cell types within the infected trachea. Combined with genomic data from viral genomes and Immunofluorescent staining, we are integrating multi-layered datasets to better understand the evolution of these LAVs after vaccination.

This study provides novel insights into the mechanisms of avian coronavirus infection and evolutionary dynamics of LAVs post-vaccination, offering valuable information to enhance the rational design of more effective vaccines.

28

Characterising the role and evolution of temperate bacteriophages in chronic respiratory infections of Pseudomonas aeruginosa in Cystic Fibrosis (CF) patients using their first isolate.

MARIKLAIRI KIOURKOU[1], Mohammad Adnan Tariq[1], Roger Levesque[2], Antony de Soyza[3], Malcolm Brodlie[4], John Perry[3], Audrey Perry[3], Emma Mars[3], Andrew Nelson[1], Darren Smith[1]

[1]Northumbria Univeristy, Newcastle Upon Tyne, United Kingdom. [2]University of Laval, Quebec, Canada. [3]Freeman Hospital, Newcastle Upon Tyne, United Kingdom. [4]Royal Infarmary, Newcastle Upon Tyne, United Kingdom

Abstract

Cystic fibrosis (CF) is an autosomal recessive genetic disorder marked by impaired mucociliary clearance, that may lead to chronic bacterial infections and persistent lung inflammation. *Pseudomonas aeruginosa* (Pa) infection is associated with increased morbidity and mortality in CF patients due to its ability to adapt and evolve in the lung microenvironment. Temperate phages, play a functional role in bacterial adaptation and evolution in the CF lung longitudinally . This study aims to characterize the arms race evolution between temperate phages and bacteria by investigating how they influence *P. aeruginosa* physiology, and metabolism and examining their carriage at the onset of chronic infection and longitudinally. Phage evolution could provide insights into the pathophysiology of CF lung infections, particularly during the transition from early to chronic infection stages. Using a set of clinical isolates (n=315) provided by the Freeman Hospital, Newcastle, sequenced using a 384-indexing strategy, using the PacBio Revio platform. Analysis confirms the use of long read sequencing at high plexity to both assemble Pa genomes and mine for prophage carriage. . This research seeks to highlight the differences in bacteriophage genome complexity carriage associated with late-stage lung infections in CF, contributing to a deeper understanding of bacterial adaptation and chronic infection in this patient population.

29

Viral and Bacterial Impact Preceding Disease Onset in Preterm Infants: a Cross NICU Study in the UK

Rui Santos[1], Andrew Nelson[1], Gregory Young[2], Mohammad Tariq[1], Wen Yew[1], Jasmine Jessop[1], Nicholas Embleton[2], Christopher Stewart[2], Janet Berrington[2], Darren Smith[1]

[1]Northumbria University, Newcastle upon Tyne, United Kingdom. [2]Newcastle University, Newcastle upon Tyne, United Kingdom

Abstract

Very low birthweight preterm infants with less than 32 weeks of gestational age are more susceptible to develop diseases such as necrotising-enterocolitis (NEC) and development of late-onset sepsis (LOS). In the UK, 1 in 12 babies are born prematurely. In addition, 1 in 20 of these preterm infants (< 32 weeks of gestational age) are at risk of developing NEC and have an associated higher mortality rate of >20%. Development of a stable gut bacterial microbiome plays an important role in early child health, however the role of the gut virome in preterm infant's health and disease early in life remains underexplored. We characterised the gut virome (NovaSeq 6000, Illumina) and microbiome (full-length 16S rRNA PacBio Kinnex) of 39 preterm infants that developed NEC and/or LOS, and 57 healthy preterm infants matched for gestational age and neonatal intensive care unit (NICU). This study importantly does not only focus on dsDNA bacteriophages, but also ssDNA bacteriophages from preterm infants admitted to 13 different NICU sites across England. A smaller subset (74 stool samples of 392) involving 7 preterm infants that developed disease, and their matched controls was used to target ssDNA bacteriophages. The data and expansion of dsDNA phage diversity illustrates that the viral compartment can be used as an indicator of the development of NEC.

30

QuantumScale RNA: A scalable single-cell transcriptomics platform

Tushar Shah[1], Herbert Gong[2], Dmitry Pokholok[1], Bryce Alves[1], Nate Chapin[1], Fan Zhang[1], Jerushah Thomas[1], Nicholas Pervolarakis[1]

[1]Scale Biosciences, San Diego, USA. [2]Scale Bio, San Diego, USA

Abstract

Single-cell RNA sequencing (scRNA-seq) has revolutionized transcriptomics but cost and throughput limitations hinder wider adoption. Scale Bio has developed an innovative, highly parallelized workflow combining ScalePlex and QuantumScale technologies to overcome these challenges. ScalePlex increases throughput by multiplexing samples during fixation, minimizing batch effects and preserving sample integrity. QuantumScale then facilitates transcriptomic profiling at an unprecedented scale, enabling the analysis of up to 4 million cells in a single experiment.

To demonstrate this capability, we processed 4 million nuclei from eight brain samples across 768 ScalePlex wells with unique sample barcodes. Nuclei were isolated, fixed, and pooled using ScalePlex, achieving a high cell capture rate. Deep sequencing was performed on 1 million nuclei, with shallow sequencing on the remaining 3 million. The resulting data demonstrated robust cell recovery, high sensitivity comparable to previous smaller-scale studies, and minimal batch effects across all samples. This study validates the feasibility and reproducibility of ultra-high-throughput scRNA-seq with ScalePlex and QuantumScale. This powerful combination unlocks unparalleled scalability for scRNA-seq, paving the way for groundbreaking discoveries by enabling comprehensive cellular analysis in complex biological systems.

31

Entwined Cellular Mechanisms and Hypomorphicity Could Explain the Clinical Variability in Monogenic Diseases Associated Mutations in TGFβ Signaling Pathway

Bassam R. Ali, Nesrin Gariballa, Fadia Abu-Sailik, Besan Alsaafeen

Department of Genetics and Genomics, College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, UAE

Abstract

Several autosomal-dominant monogenic disorders affecting the vascular system have been conclusively associated with mutations in major components of the Transforming Growth Factor-β (TGFβ) signaling pathway including mutations in Endoglin, TGFBR1, TGFBR2, ALK1 and BMPR2. The clinical presentations of the disease-causing alleles in these genes are highly variable which warranted detailed studies to delineate their underlying cellular mechanisms. We demonstrated the involvement of the stringent Endoplasmic Reticulum (ER) protein folding quality control mechanisms such as ERAD in the loss of function caused by a significant number of mutations in these proteins. More recently, we have demonstrated the involvement of dominant negative effects of some Endoglin ER-retained mutants on the wild type allele (1). The combination of haploinsufficiency and dominant negative effects may explain some of the variability observed in their clinical presentations. In addition, we speculate that hypomorphicity is contributing significantly to this variability. Understanding the detailed cellular mechanisms is expected to aid in developing novel therapies (2) and enhance our understanding of functional implications of variation in the human genome.

(1) Gariballa, N., Badawi, S., and Ali, B. R. (2024). Endoglin mutants retained in the endoplasmic reticulum exacerbate loss of function in hereditary hemorrhagic telangiectasia type 1 (HHT1) by exerting dominant negative effects on the wild type allele. Traffic, 25(1):e12928.

(2) Gariballa, N., Mohamed, F., Badawi, S., and Ali, B. R. (2024). The double whammy of ER-retention and dominant-negative effects in numerous autosomal dominant diseases: significance in disease mechanisms and therapy. Journal of Biomedical Science, 31(1):64.

33

Whole genome sequencing of Cyprinid herpesvirus 3 isolates collected in the UK over the last twenty years using an amplification free long read sequencing approach

David Ryder, George Scott, Gareth Wood, Jacqueline Savage, Richard Paley, Frederico Batista

Cefas, Weymouth, United Kingdom

Abstract

Cyprinid herpesvirus 3 (CyHV-3), also known as Cyvirus cyprinidallo3, is the cause of a highly virulent and contagious disease in common carp (*Cyprinus carpio*) that emerged in the 1990s, causing significant damage to the valuable aquaculture and angling industry worldwide. Genomic resources for CyHV-3 are limited, with only 22 genomes published so far using different methodological approaches. Although the CyHV-3 genome is the largest of any known herpesvirus, the layout of the genome largely follows that of other class A herpesviruses like Human herpesvirus 6, with a ~250kbp unique, non-repetitive coding region flanked by direct terminal repeats ~22kbp in length. Our study had the following aims: (1) develop an amplification free method to sequence DNA from viruses grown on a cell line using long read sequencing, (2) determine the genetic diversity of approximately 60 isolates of CyHV-3 collected in the UK between 2003 and 2024; and (3) identify highly polymorphic genomic regions for development of phylogenetic markers. Initial results for the first 28 isolates yielded a low level of host contamination (≤ 40%), long CyHV-3 read lengths (n50: ~22kbp) and even coverage across the length of the genome (coefficient of variation: ~20%). Initial assembly of reads resulted in the direct terminal repeats being erroneously identified as evidence of a circular plasmid, however, after confirming the number of repeats and their similarity through alignment and variant calling, the misassembly was manually corrected. Following assembly, phylogenetic analysis identified isolates from three different lineages, and genomic regions with high nucleotide diversity.

34

Evolving Landscape of scRNA-seq

Alex Deamer, Maria Vias, Marina Golotiuk, Maribel Valenzuela, Jason Skelton, Eleanor Denham, Ugurcan Sal, Isabelle Oliver, Johanna Barbieri, Ashley Sawle, Matthew Eldridge, James Brenton, Ania Piskorz

CRUK CI, Cambridge, United Kingdom

Abstract

Background

Single-cell RNA Sequencing (sc-RNAseq) is an established and powerful methodology offering new insights into individual cancer cell biology, response to treatment, tissue/tumour microenvironments, and their variable functional transcriptional states. The scRNA-seq technology landscape is rapidly evolving with variety of methods based on Drop-seq, PIP-seq, Combinatorial barcoding and kinetic confinement. We tested 7 commercially available sc-RNAseq workflows using PEO1 and PEO4 high grade serous ovarian cancer cell line models with known and quantified characteristics to accurately determine the efficacy of each technology.

Methods

PEO1 and PEO4 were trypsinised, washed and counted using hemocytometer. We used between 2 and 4 technical replicates for each cell line per each tested method and maximum recommended cell number input for each technology, with 1 million cell input for all fixation-based protocols (10X Flex, Parse, Scale Biosciences). We performed in parallel: 10x 3' V4/ 10x 3' OCM/ 10x GEMX Flex/ Parse WT v3/ Illumina T10 (Fluent) / SimpleCell (CS Genetics) and QuantumScale (Scale Biosciences) All libraries were sequenced using Illumina NovaSeqX and analyzed using their respective pipelines.

Results and Conclusions

We will present the comparison results between different technologies focused on: data quality (cell capture rate, total and median gene detection rate, UMI counts), reproducibility (assessed using technical replicates), experiment workflow timeline, cost and the feasibility of each workflow where number of cells varies from very low (hundreds of cells) to high (millions of cells). We will generate a guide to support effective and efficient experimental design for cancer translational studies.

35

Differential gene expression: bulk mRNA or total RNA?

Eleanor Denham, Johanna Barbieri, Shalini Rao, Ashley Sawle, Isabelle Oliver, Ugurcan Sal, Matthew Eldridge, Jason Carroll, Ania Piskorz

Cancer Research UK Cambridge Institute, Cambridge, United Kingdom

Abstract

Differential gene expression analysis using bulk RNA sequencing is a key tool to track transcriptional changes during disease progression in many cancer models. mRNA rather than total RNA methods are often selected due to cost effectiveness. We wanted to investigate if gene counts are similar between those two methods, as there is limited data available comparing them. We also compared the performance of two mRNA library preps, as producing mRNA libraries at scale can be significantly improved by using more streamlined workflows.

We used pancreatic tissue samples from KC Mice with an activating mutation in KRAS, with 4 biological replicates at 2 time points, and 3 technical replicates per sample. We compared 3 different library preparation methods: Illumina stranded mRNA, Watchmaker Genomics mRNA, Watchmaker Genomics RNA with Polaris depletion.

Both Illumina and Watchmaker mRNA library preps showed consistent total read counts, low ribosomal content, similar percentages of aligned reads and consistent count density. However, Watchmaker mRNA library prep has a shorter workflow, allowing more time efficient library generation and tolerates lower total RNA input.

The total gene counts mapping to coding regions for the rRNA depletion method was lower, due to a higher percentage of reads mapping to intergenic, intronic and untranslated regions. Our data showed that a read depth of between 30-40M for the depletion library was sufficient to reach the same level of gene counts as mRNA library preparation sequenced at 20M reads. An in-depth analysis of mRNA vs total RNA will be summarised in this poster presentation.

36

Understanding gene expression involved in immune response to SARS-CoV-2 Infection

Lee Murphy, Audrey Coutts, Giovanna Weykopf, Wendy Bickmore, Simon Biddie

University of Edinburgh, Edinburgh, United Kingdom

Abstract

Genome Wide Association Studies have identified 49 genetic loci involved in critical illness due to COVID-19. This has led to an understanding of the structural consequences of protein-coding variants and has identified druggable targets. The loci associated with severe COVID-19 harbour multiple plausible genes that may contribute to pathophysiology. In addition, genes are known to undergo alternative exon usage to generate multiple transcript isoforms, with expression which can be cell-type and context- specific. These transcript isoforms can be challenging to resolve using short-read sequencing. In this work we have investigated transcript isoforms across a severe COVID-19 risk locus, including transcripts that are lowly-expressed, and are cell-type specific. We have utilised Twist custom probes to enrich for a panel of genes combined with the power of long-read single-molecule sequencing by Oxford Nanopore to enable deep long-read sequencing of transcripts. We resolve the transcript isoform complexity across the locus, including unannotated gene isoforms, and full-length transcript structures that cannot be resolved using short read sequencing. This methodology can be used to investigate variants that alter splicing or isoform expression, and provides a powerful method for deep, quantitative, characterisation of transcript isoforms.

37

PyamilySeq: Transparent and interpretable gene (re)clustering and pangenomic inference highlights the fragility of conventional methods.

Nicholas J Dimonaco

School of Biological Sciences, Queen's University Belfast, Belfast, United Kingdom. Department of Computer Sciece, Aberystwyth, Aberystwyth, United Kingdom

Abstract

Understanding the genetic diversity and functional potential of microorganisms depends on grouping their genes by sequence similarity, function or structure. This is particularly important in highly interactive microbiomes, such as the human gut or rumen, where microbial communities rely on intricate metabolic exchanges and interdependencies. However, commonly used gene clustering and pangenomic inference methods often lack transparency in their methodology and offer limited flexibility in parameterisation. These limitations introduce biases, potentially distorting the representation of microbial communities and their functional dynamics.

This study introduces PyamilySeq, a flexible and transparent framework designed to systematically identify challenges in gene clustering and pangenomic analysis, and support the development of practical solutions. PyamilySeq enables users to replicate core components of widely used pangenome tools while providing greater control over parameters and clearer insights into their impact. Through comparative analysis, I demonstrate how changes in parameters, such as sequence identity cut-offs, length thresholds, and even computational settings like CPU and memory allocation, can significantly alter gene family composition. Therefore, PyamilySeq highlights that many reported gene families in contemporary tools are artefacts of the algorithms themselves, rather than biologically meaningful groupings.

PyamilySeq reveals the fragility and tool-driven biases inherent to contemporary gene clustering and pangenome inference that too often goes unnoticed. This work aims for more biologically grounded insights into microbial diversity and function, critical for applications in antimicrobial resistance, pathogen surveillance, and microbial ecology, highlighting the importance of reducing reliance on rigid, one-size-fits-all tools to ensure a more precise representation of microbial diversity and function.

38

Beyond Taxonomic Composition: Metagenomic Evidence for Functional Niche Displacement in the Rumen Microbiome.

Katie Lawther[1], Nicholas Dimonaco[1], Abdulai Guinguina[2], Sophie Krizsan[3], Sharon Huws[1]

[1]Queen's University Belfast, Belfast, United Kingdom. [2]Natural Resources Institute Finland, Jokioinen, Finland. [3]Inland Norway University of Applied Sciences, Blæstad, Norway

Abstract

Metagenomics is now common practice in many microbiome studies, including those focused on rumen microbiology with the aim of methane mitigation. While metagenomic studies have greatly enhanced our ability to profile microbial communities, particularly in terms of taxonomic and functional composition, many analyses rely on static snapshots of these aspects. This study aims to unveil the dynamic interplay between taxonomy and function.

We demonstrate how shifts in taxonomic contributors to key functions, despite stable overall taxonomic profiles, reveal hidden layers of redundancy and resilience within the microbiome. Using methanogenesis in ruminant microbial communities as a model, we show that taxonomic stability can mask underlying changes in functional contributions to critical pathways, a phenomenon with major implications for methane mitigation strategies.

Our results highlight that while the total functional potential related to the direct methanogenesis pathway may appear constant, less obvious pathways, particularly “feeder” pathways, that supply methanogenesis precursors were significantly altered, contributing to methane reduction in the animals. We also identify shifts in the taxa responsible for these functions. This niche displacement suggests that microbiome function is more robust and resistant to manipulation than previously assumed. As a result, efforts to reduce methane emissions through microbiome-targeted interventions may underperform without detailed knowledge of the organisms contributing to function.

By tracking both functional shifts and the taxonomic origins of functional pathways longitudinally, our approach offers a more nuanced framework for interpreting microbiome data and highlights the value of linking function to taxonomy in microbial ecology.

39

Signatures of copy number alterations driving childhood cancer development

Robyn Sherrard, Richard Yim, Dean Thompson, Jack Goddard, Anya Fletcher, Jonathan Higgins, Christine Harrison, Steve Clifford, Anthony Moorman, Dan Williamson, Sarra Ryan

Newcastle University, Newcastle Upon Tyne, United Kingdom

Abstract

Aneuploidy is a hallmark of many childhood cancers, yet its genetic origins and the diversity of resulting chromosomal alterations remain poorly understood. This project aims to identify genetic drivers and mechanistic associations underlying specific aneuploidy profiles in paediatric acute lymphoblastic leukaemia (ALL) and medulloblastoma (MB) using a CNV signature-based framework.

Using *SigProfilerExtractor*, nine genome-wide copy number signatures were derived from SNP-array and whole-genome sequencing data across a large, well-characterised cohort (n = 3,374). These signatures capture recurrent, non-random patterns of structural variation, including whole-chromosome, arm-level, and focal events, and reflect chromosomal instability processes such as chromothripsis, haploidy, and whole-genome doubling. Six of the nine signatures appear specific to paediatric cancers when compared to adult CNV signatures described in Steele et al. (2022). Signature-defined clusters exhibit mixed but sometimes exclusive associations with ALL or MB.

These signatures are being analysed in relation to established molecular subtypes to assess whether CNV processes map onto or cut across known classifications. Integration with whole-genome and targeted sequencing aims to identify somatic mutations—particularly in mitotic, chromosomal segregation, and genome maintenance genes—enriched in specific CNV-defined groups. In MB, RNA-seq and methylation data will provide additional biological context and potential downstream effects. This is the first study to generate CNV signatures for investigating patterns of aneuploidy in paediatric cancer.

40

Characterising regionally-associated AMR lineages from a decade of genomic surveillance inSalmonella enterica serovar Enteritidis

Mike Nsubuga[1], Kristen Reyher[1], Andrew Dowsey[1], Lauren Cowley[2], Marie A. Chattaway[3,4,5], Sion Bayliss[1]

[1]University of Bristol, Bristol, United Kingdom. [2]University of Bath, Bath, United Kingdom. [3]UK Health Security Agency, London, United Kingdom. [4]NIHR Health Protection Research Unit in Genomics and Enabling Data, University of Warwick, Warwick, United Kingdom. [5]NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, Oxford, United Kingdom

Abstract

Background: *Salmonella enterica serovar* Enteritidis is a foodborne pathogen whose global spread is facilitated by international trade networks. The emergence of antimicrobial resistant (AMR) lineages complicates treatment efforts. We characterise regionally-associated S. Enteritidis lineages resistant to beta-lactams, tetracyclines, and fluoroquinolones and identify mobile genetic elements underlying resistance.

Methods: We analysed 22,366 S. Enteritidis genomes collected between 2014–2024 by UK Health Security Agency surveillance. Geographical associations were identified by integrating travel metadata with phylogenetic clustering. AMR genes were identified using AMRFinderPlus, with co-occurrence patterns assessed through correlation and network analyses.

Results and discussion:

AMR carriage was generally low, with quinolone resistance determining regions (QRDR) mutations in 26.86% of genomes, beta-lactamase genes (*blaTEM*) in 7.52%, tetracycline resistance genes (*tet*) in 7.05%, and aminoglycoside resistance genes (*aph*) in 1.81%. Among genomes with travel data (5,772), QRDR mutations were most prevalent in lineages from Eastern Asia (91.67%) and Northern Africa (49.9%). *tet* genes were prominent in Northern Africa (42.36%) and South-Eastern Asia (17.77%), while *blaTEM* occurred in 54.17% of Eastern Asian and 42.15% of Northern African genomes. *aph* genes were most prevalent in Eastern Asia (47.92%) and Western Africa (25.93%).

Co-occurrence analysis revealed chromosomal quinolone resistance (gyrA_S83Y) with plasmid-borne beta-lactam resistance (blaTEM) in 12.81% of genomes across multiple lineages, widespread across Western Asia and Eastern Europe. Transposons carrying tet(A), tet(B), and blaTEM-57 in 2.89% of genomes, predominantly from Northern Africa. Our findings highlight the role of mobile elements in AMR spread and underscore global risks from regionally emergent MDR lineages.

41

Integrating Methylomic and Fragmentomic Profiling in Soft Tissue Sarcoma: Unravelling Novel Epigenetic Signatures

Michael Wong[1], Maxime Tarabichi[2], Udayaraja GK[3], Anant Desai[4], Andrew Beggs[1]

[1]University of Birmingham, Birmingham, United Kingdom. [2]IRIBHM – J. E. Dumont, Brussels, United Kingdom. [3]University of Leeds, Leeds, Belgium. [4]Midlands Abdominal Retroperitoneal Sarcoma Unit (MARSU), Birmingham, United Kingdom

Abstract

Background

DNA sequencing and copy number variation (CNV) analysis have become central to genomic research. Methylomic and fragmentomic profiles are inherently present in cell-free DNA (cfDNA) and circulating tumour DNA (ctDNA). We integrated these epigenetic profiles in plasma-derived DNA from patients with liposarcoma (LPS) and leiomyosarcoma (LMS).

Method

Plasma-derived ctDNA samples were sequenced using the PromethION platform (Oxford Nanopore Technologies). Reference-based alignment enabled the detection of epigenetic modification. Methylation deconvolution and fragmentation analysis were performed using two validated computational tools.

Results

80 LPS and 24 LMS ctDNA were analysed. ONT data revealed a significantly reduced global methylation in LPS samples compared to LMS. Interestingly, locus gene methylation (RASSF1A) was generally higher in LMS compared to LPS. Cell type deconvolution revealed that granulocytes were the predominant cell origin in both LPS and LMS, with average cfDNA signals of 36.9% and 40.5%, respectively ($p = 0.2795$). The adipocyte and smooth muscle signals were 0.152% and 0.092% in LPS, in contrast to 0% and 0.05% in LMS, which reflects a low cell-type-specific tumour signal. A higher density of total whole genome DNA fragments was associated with a gained MDM2 copy number compared to a normal copy number in LPS.

Conclusions

Global hypomethylation is a potential hallmark of LPS. Hypermethylation of locus RASSF1A is associated with LMS. Methylation deconvolution may help detect the cell origin in cancer of unknown primary (CUP). Whole genome fragment size density appears to be more abundant in samples with a gained copy number of the MDM2 gene in LPS.

42

Assessing the utility of self-supervised contrastive learning in bacterial geographic source attribution

Caroline Schreiber[1], Mike Nsubuga[1], Lauren A. Cowley[2], Andrew W. Dowsey[1], Kristen K. Reyher[1], Sion C. Bayliss[1]

[1]University of Bristol, Bristol, United Kingdom. [2]University of Bath, Bath, United Kingdom

Abstract

The availability of bacterial whole genome sequencing (WGS) data has continued to increase in recent years, leading to improved surveillance capabilities and increased potential for predictive modelling. However, a significant challenge persists in that many publicly available genomes lack detailed metadata - such as geographic origin or antimicrobial susceptibility information - that is essential for some downstream analyses. Self-supervised machine learning approaches that learn patterns inherent in genomic data without requiring labels could have high utility for predicting such metadata to allow extraction of valuable insights from existing genomic databases.

We present the use of contrastive learning to predict the geographic source of *Salmonella enterica* serovar Enteritidis isolates based solely on their genomic content. Our model, built using a PyTorch deep learning framework, leverages genomic data from ~7500 previously published Salmonella isolates to learn discriminative features of geographic regions, utilising explicit labels which are present in only a third of the dataset.

The objective of this ongoing work is to demonstrate the utility of unsupervised methods to learn relevant information about genetic variation and genomic organisation and to assess the utility of contrastive learning as a method to account for inherent biases in datasets using only a fraction of data for which label information is available. This approach could significantly enhance the utility of existing genomic databases for both public health and fundamental science utility.

43

A streamlined EM-seq method for whole genome methylation detection

VK Chaithanya Ponnaluri, Vaishnavi Panchapakesa, Daniel Evanich, Matthew Campbell, Bradley Langhorst, Louise Williams

New England Biolabs, Ipswich, USA

Abstract

DNA methylation is an important epigenetic regulator of gene expression. In the human genome cytosines are methylated in the CpG context and are often clustered in CpG rich regions associated gene regulation. Traditionally, sodium bisulfite conversion was used to distinguish 5-methylcytosines and 5-hydroxymethylcytosines from cytosines. However, this method damages DNA and introduces significant sequencing bias. NEBNext Enzymatic Methyl-seq (EM-seq) is an enzymatic approach that minimizes DNA damage therefore enabling longer insert sizes, lower duplication rates and a more accurate quantification of methylation.

EM-seq has been optimized further. Here we present EM-seq v2 for DNA inputs ranging from 0.1 ng to 200 ng. EM-seq v2 is a streamlined protocol that reduces library preparation time via a reduction in end preparation and adaptor ligation time as well as by eliminating a cleanup step. Additionally, removal of the cleanup also results in savings through reduced plastics usage. Libraries typically had increased yield that will reduce sequencing duplications. Furthermore, all libraries displayed an even GC bias representation as well as consistent global methylation between the 0.1 ng to 200 ng DNA inputs. Coverage of genomic features such as CpG islands and enhancers was maintained to 1 ng but dropped as expected with the 0.1 ng inputs. EM-seq v2 is an enhanced, streamlined protocol that provides consistent methylation metrics for inputs ranging from 0.1 ng to 200 ng.

44

TET2 mutant-allele dosage as a determinant of 5'-azacitidine sensitivity: a pan-cancer approach

Zoë Hawking, Oliver Hanley, Azalia Benedicta, Hannah Melville, James Allan

Newcastle University Centre for Cancer, Newcastle upon Tyne, United Kingdom

Abstract

Introduction and hypothesis: Ten-eleven translocation 2 (TET2) operates as a key regulator of genomic methylation by catalysing the demethylation of 5-methylcytosine to 5-hydroxymethylcytosine. *TET2* loss of function mutations lead to global genomic hypermethylation, a phenomenon which is frequently observed in adult acute myeloid leukaemia (AML). *TET2* mutations are also reported in most other human cancers, including lymphoid malignancies and numerous non-haematological cancers such as melanoma and colorectal carcinoma. We have previously demonstrated that somatic biallelic *TET2* mutation in AML sensitises to 5'-azacitidine (Aza) monotherapy, and as such we hypothesise that lymphoid malignancies and solid tumours with biallelic *TET2* mutation will also be sensitive to Aza monotherapy.

Methods: This study aimed to test this hypothesis by using CRISPR-Cas9 to generate isogenic non-myeloid cancer cell line pairs with either *TET2* monoallelic or biallelic mutation. These cell models were then used to test sensitivity to Aza as a function of *TET2* mutant-allele dosage. RNAseq analysis was performed to investigate potential mechanisms of Aza sensitivity as a result of TET2 knockout, and thus as a function of *TET2* mutant-allele dosage.

Results: Biallelic loss of TET2 sensitised HEL AML cells and MHH-PREB-1 B-cell non-Hodgkin lymphoma cells to Aza monotherapy (p <0.0001), however this was not observed in TET2 null LS411N colorectal cancer cells (p = 0.98). RNAseq analysis showed defined clustering by cell type and *TET2* mutation status, and differential gene expression analysis identified several hallmark pathways commonly affected in TET2 null cell models, including E2F targets, G2M checkpoint, glycolysis, and oxidative phosphorylation.

45

Application of Genomics to Biosurveillance and One Health Approaches

Richard Hill, David Walker

Cefas, Weymouth, United Kingdom

Abstract

The COVID-19 pandemic underscored the need for effective national and international systems of surveillance for global health security and tackling emerging disease threats. A One Health approach to biosurveillance relies on multidisciplinary and cross-sectoral collaboration to detect and respond to emerging threats to human and animal health. Cefas demonstrated case studies of the utility of genomic sequencing, in particular within the wastewater sector, in surveillance of SARS-CoV-2, Norovirus and other foodborne pathogens, as part of a *One Sample, Many Analyses* approach. This can maximise the effectiveness of matrices such as wastewater and enable Wastewater-Based Epidemiology (WBE) to be most effective, and enhancing the precision of public health responses to emerging global threats.

There are many factors to consider in delivering a successful WBE surveillance programme, not least defining the surveillance questions and objectives to establishing sampling protocols, data governance, and quality assurance measures. Here we provide an overview of the applications of genomics to biosurveillance for health programmes and the considerations required to implement this effectively. For example efficient resource allocation, harmonisation and standardisation of laboratory techniques and analysis pipelines, and identification of capability gaps before a critical threat develops. Finally, we outline how formation of a new Environmental Biosurveillance Design Framework, developed as part of the UK Government National Biosurveillance Network, will offer comprehensive and structured guidance in this, be that for human, animal, plant or environmental health surveillance. When applied in well-structured surveillance programmes genomics is undoubtedly an invaluable tool to effectively support public health decision-making.

46

Near-complete bacterial genome insertion in a tick nuclear genome: evidence from tick cell lines and ticks

Jing Jing Khoo[1], Alexandra Beliavskaia[1], Mark Whitehead[1], Adriana Ludwig[1], Catherine Hartley[1], Alaa Al-Khafaji[1], Grace Ward[1], Stuart Armstrong[1], Germanus Bah[2], Maria Kazimirova[3], Alistair Darby[1], Lesley Bell-Sakyi[1], Benjamin Makepeace[1]

[1]University of Liverpool, Liverpool, United Kingdom. [2]Institut de Recherche Agricole pour le Développement, Centre de Recherche Agricole de Wakwa, Ngaoundéré, Cameroon. [3]Slovak Academy of Sciences, Bratislava, Slovakia

Abstract

In earlier screening for bacterial endosymbionts in tick cell lines at the Tick Cell Biobank, horizontal gene transfer (HGT) of bacterial genes into the tick genome was initially suspected when cell lines generated from the tropical bont tick *Amblyomma variegatum* yielded positive PCR amplification of genes from the pathogen *Rickettsia africae*, known to be transmitted by this tick, even though microscopic examination did not show any signs of bacterial infection in the cells. Here we describe the investigations leading to the discovery of an insertion of the nearly complete *R. africae* genome into the *A. variegatum* nuclear genome.

Next generation sequencing of *A. variegatum* cell line AVL/CTVM17 provided evidence for the insertion of a nearly complete *R. africae* chromosome (~1.2Mb) into the tick genome, although a ~54kb region containing genes essential for *Rickettsia* bacterial metabolism was absent. Sequence of the pRa plasmid, which is normally found in live *R. africae*, was also not detected. We conducted further analyses of the genome sequence coverage, single nucleotide polymorphism profile, and plasmid gene copy numbers, giving further evidence for the genetic differences between the *R. africae* genome insertion and *R. africae* bacteria infecting ticks.

These findings represent the first confirmation of a pathogen chromosome integrated into the genome of its vector and highlight the potential impact of bacterial genome insertion on pathogen or endosymbiont surveillance in arthropods. We also demonstrate the value of arthropod cell lines in facilitating research on HGT in arthropod genomes.

47

Making the most of microbiome biobanks: Enabling Cross-Cohort and Multi-Kingdom Analyses

Greg Young, Ella Bailey, Aryan Khirwadkar, Christopher Stewart, Christopher Lamb

Newcastle University, Newcastle Upon Tyne, United Kingdom

Abstract

Clinical microbiome studies require increasingly large cohorts due to inter-individual variability. As sequencing costs fall, sample collection rather than sequencing capacity becomes rate-limiting. Consequently, trials generate substantial, highly curated, resource-intensive biobanks which are frequently underexploited beyond primary endpoints.

Often, only a fraction of collected material is used. Moreover, sample handling tailored to specific analyses may compromise other investigations. For instance, stool preserved for bacterial DNA can hinder fungal or viral community profiling.

Cross-study integration is essential to improve the utility of existing biobanks, avoiding trial duplication and unnecessary cost. However, methodological differences in sample collection, storage, and processing remain major obstacles to data integration.

We address these barriers by validating cross-study and multi-kingdom compatibility using stool samples from UK microbiome trials in inflammatory bowel disease. We present:

A comparison of bacterial DNA extraction protocols - Preservation of individual microbiome profiles generated using two distinct nucleic acid extraction protocols (n=23 paired samples).

A modified virome protocol - Removing human and bacterial preserved DNA from samples collected for bacterial community profiling enables enrichment, extraction, and sequencing of virus-like particles.

By facilitating multi-kingdom analysis from single samples and demonstrating comparability across trials this research advances a more integrative and cost-effective approach to microbiome science. Robust profiling of combined biobank samples enhances statistical power, reduces study duplication, and promotes a systems-level understanding of host-microbe dynamics in complex diseases such as IBD.

49

Accurate bioinformatic host pangenome depletion using fast minimizer search

Bede Constantinides[1], John Lees[1], Derrick Crook[2], Nick Loman[1]

[1]University of Birmingham, Birmingham, United Kingdom. [2]University of Oxford, Oxford, United Kingdom

Abstract

Realising the value of large DNA sequence collections demands efficient search and extraction of sequences of interest. Search queries may vary in size from short gene sequences to multiple whole genomes that are too large to fit in computer memory. In microbial genomics, a routine search application involving both large queries and large collections is the removal of contaminating host genome sequences from microbial (meta)genomes. Where the host is human, *sensitive* classification and excision of host sequences is usually necessary to protect host genetic information. *Precise* classification is also critical in order to retain microbial sequences and permit accurate microbial genomic analysis. While human pangenomes have been shown to increase sensitivity of human sequence classification, existing bioinformatic host depletion approaches have either limited precision when used with metagenomes or large computing resource requirements.

We present a new sequence filter implementation for raw sequence files and streams. We demonstrate its leading accuracy for the task of bioinformatic host depletion using a fraction of the computing resources required by existing approaches approaches with equivalent accuracy. By querying a human pangenome index for minimizers contained in each input sequence, our open source approach is able to accurately classify human sequences in long reads at over 250Mbp/s with a commodity laptop. We present validation of sensitivity, specificity and speed with short and long reads for diverse catalogues of human, bacterial and viral genomes alongside existing methods. Beyond host depletion, Deacon is well suited to common sequence search and filtering applications involving large queries.

50

The Human Developmental Biology Resource

Aragorn Jones, Steven Lisgo, Deborah Henderson

Newcastle University, Newcastle upon Tyne, United Kingdom

Abstract

The Human Developmental Biology Resource (HDBR) is a MRC/Wellcome funded biobank providing embryonic, fetal and placental samples to support scientific research. Based at Newcastle University and University College London, material from 3 weeks of development is collected under an HTA research tissue bank licence with REC approval, meaning UK based researchers do not need to apply for separate ethics to use the material in their research.

Over the past 25 years, tissue has been supplied to over 800 individual research projects resulting in well over 500 publications. HDBR are a major tissue provider to the Human Developmental Cell Atlas and the Wellcome-funded Human Developmental Biology Initiative.

Registered projects receive fixed, frozen or fresh tissues that can be used to gain insights into the unique aspects of human development as well as validate findings from cell culture and animal studies. Bespoke dissections of organs or tissues of interest can also be requested, and the tissue prepared to users' unique requirements, including dissociated cells preparations; sectioned to microscope slides; or RNA, DNA or protein extractions.

Our website hosts a range of teaching and educational resources, from annotated human embryo histological sections and 3D models to gene and protein expression data.

We also offer experimental services utilising techniques such as multiplex and base specific in situ hybridization, simultaneous RNA and protein localisation, single cell RNA sequencing, and spatial transcriptomics.

Visit https://HDBR.org or https://HDBRAtlas.org to learn more about how to register a project and other HDBR resources.