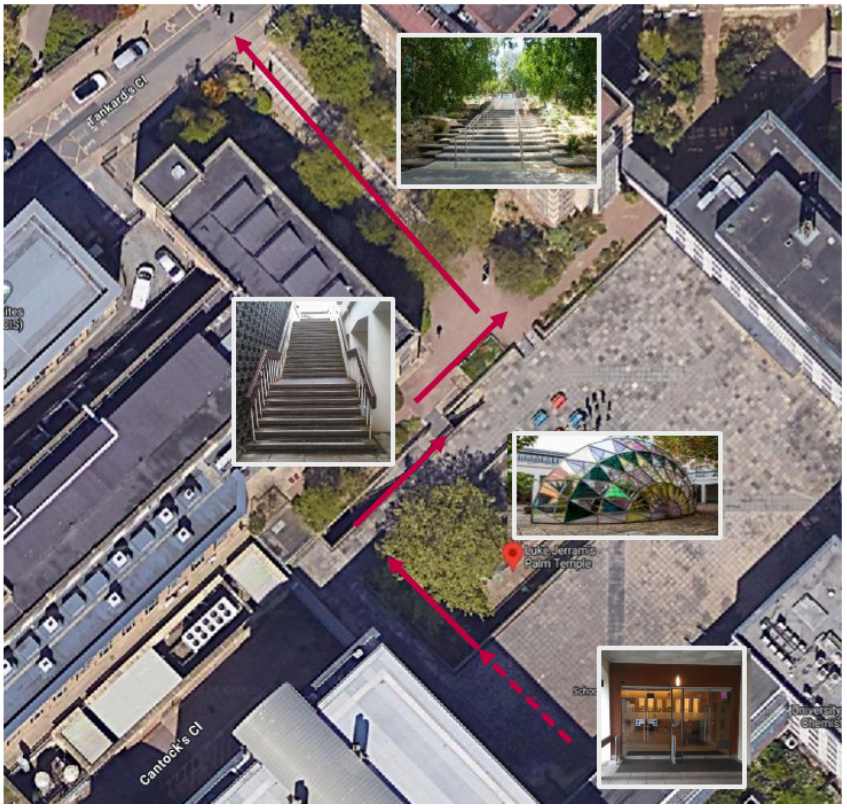
GENOME SCIENCE 2024

Welcome Pack and Abstract Booklet

Fire Safety



- Exit areas (Main entry)
- The Fire Assembly point is University Walk, at the top of Cantock's Steps from the patio area.
- Raise the alarm by breaking the glass on a RED fire alarm call point.
- Leave immediately and do not return to collect personal belongings.

Floorplan

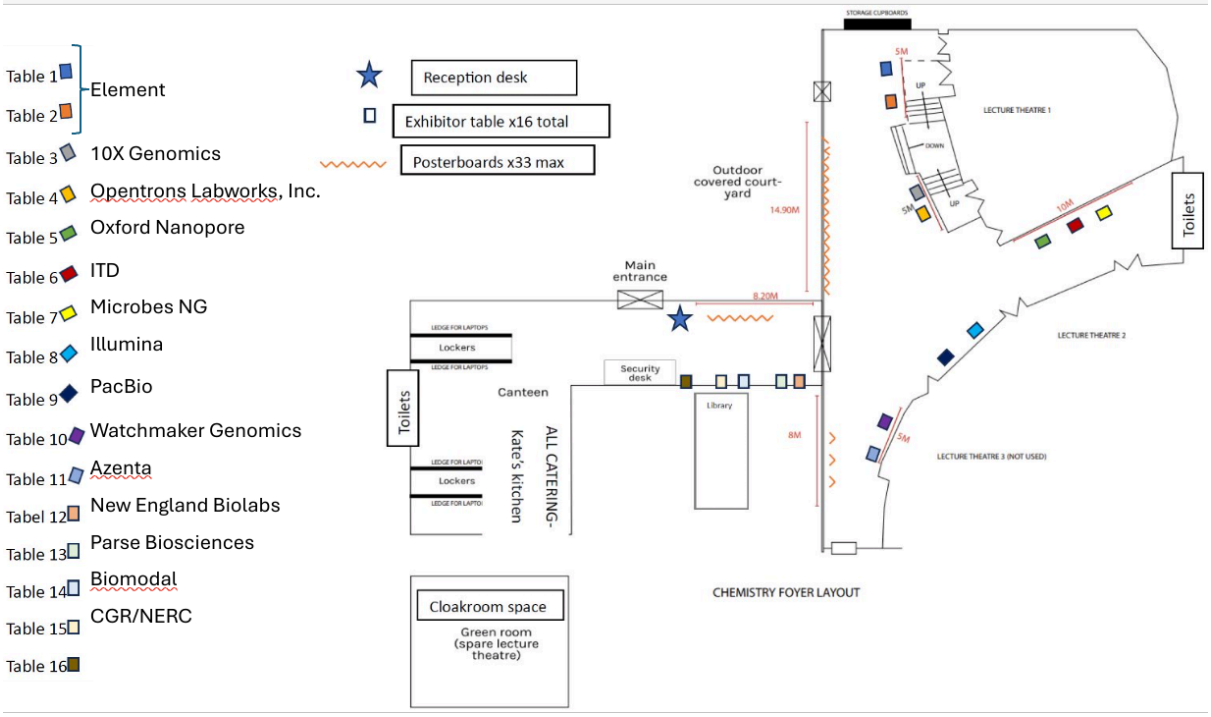


Table 1
Table 2
Element
Table 3
10X Genomics
Table 4
Opentrons Labworks, Inc.
Table 5
Oxford Nanopore
Table 6
ITD
Table 7
Microbes NG
Table 8
Illumina
Table 9
PacBio
Table 10
Watchmaker Genomics
Table 11
Azenta
Tabel 12
New England Biolabs
Table 13
Parse Biosciences
Table 14
Biomodal
Table 15
CGR/NERC
Table 16
Reception desk
Exhibitor table x16 total
Posterboards x33 max
Outdoor covered court-yard
14.90M
Main entrance
8.20M
Toilets
Lockers
Canteen
Security desk
Library
8M
Kate's kitchen
ALL CATERING-
Lockers
Cloakroom space
Green room (spare lecture theatre)
LECTURE THEATRE 1
LECTURE THEATRE 2
LECTURE THEATRE 3 (NOT USED)
5M
10M
UP
DOWN
CHEMISTRY FOYER LAYOUT

Sponsors

We are indebted to our sponsors. They very much make this event possible and drive the technology behind our science. Please take some time to visit their exhibits during the breaks.

Gold Sponsor

Element Biosciences	**Element Biosciences**

Silver Sponsors

NEW ENGLAND Biolabs	**New England Biolabs**
	Opentrons
PacBio	**PacBio**
Parse BIOSCIENCES	**Parse Biosciences**

Bronze Sponsors

10x GENOMICS	**10x Genomics**
AZENTA LIFE SCIENCES	**Azenta**
biomodal	**Biomodal**
UNIVERSITY OF LIVERPOOL	**Centre for Genomic Research**
illumina	**Illumina**
IDT INTEGRATED DNA TECHNOLOGIES	**IDT**

microbes NG	**Microbes NG**
Oxford NANOPORE Technologies	**Oxford Nanopore**
WATCHMAKER GENOMICS	**Watchmaker Genomics**

Programme

https://bristol.genomescience.org.uk/programme/

Schedule: Tuesday 16th July

Tuesday 16th July 2024 - LT1 (Afternoon)			
Start (GMT)	Finish (GMT)	Presenter	Presentation Details
12:00	13:30	Registration, Lunch and Networking	
13:30	**14:30**	**Keynotes**	**Chair: Nick Loman, University of Bristol**
13:30	14:00	Lucy Burkitt-Gray (UK Biobank)	Massive-scale whole genome sequencing: Insights from the UK Biobank
14:00	14:30	Richard Durbin (University of Cambridge)	Insights from high quality genome sequencing across the tree of life
14:30	**16:00**	**Evolving Technologies 1**	**Chair: Michael Quail, Wellcome Sanger Institute**
14:30	15:00	Shawn Levy (Element)	Integrated Multiomic and Genomic Profiling for Precision Biology on the AVITI24 platform
15:00	15:30	Jason Betley (Illumina)	Illumina Technical and Product Roadmap Update
15:30	16:00	Neil Ward (PacBio)	Platinum Genomes
16:00	16:30	Coffee Break	
16:30	**18:00**	**Evolving Technologies 2**	**Chair: Michael Quail, Wellcome Sanger Institute**
16:30	17:00	Andy Larrea (Ultima)	Ultima Genomics UG100: A flexible, low-cost, scalable sequencing solution
17:00	17:30	Graham Hall (Oxford Nanopore)	ONT Technology Update
17:30	18:00	Nicola Cahill (10x Genomics)	The application of single cell & spatial technologies to study the tumor microenvironment
18:00	19:30	Drinks Reception and Poster Session 1	

Schedule: Wednesday 17th July - Parallel Session A

Wednesday 17th July 2024			
Start (GMT)	Finish (GMT)		
08:30	09:30	Coffee and Pastries	

Parallel Session A - LT2			
Start (GMT)	Finish (GMT)	Presenter	Presentation Details
09:30	**11:00**	**Plant and Animal Genomics**	**Chair: Al Darby, University of Liverpool**
09:30	09:55	Alex Cagan (University of Cambridge)	Somatic evolution across the tree of life
09:55	10:20	Gary Barker (University of Bristol)	Unlocking hidden genetic diversity in hexaploid bread wheat
10:20	10:35	Francisco Rodriguez-Algarra (Queen Mary, University of London)	Ribosomal DNA Copy Number Variation Associates with Hematological Profiles and Renal Function in the UK Biobank
10:35	10:50	Carla Canedo-Ribeiro (University of Kent)	New method to detect chromosomal reciprocal translocations using long-read sequencing
10:50	11:00	Silver Sponsor Talk	New England Biolabs: UltraExpress™: Streamlined library prep for RNA and DNA samples from New England Biolabs
11:00	11:30	Coffee Break	
11:30	**13:05**	**Human & Clinical Genomics**	**Chair: Jonathan Coxhead, Newcastle University**
11:30	11:55	Sandi Deans (University of Edinburgh)	Ensuring a high quality NGS clinical service
11:55	12:20	Mike Hubank (Institute of Cancer Research, London)	Circulating tumour DNA diagnostics in the NHS
12:20	12:35	Steven Hair (Newcastle University)	Utilising long-read sequencing approaches to detect structural variants of clinical significance in childhood cancer
12:35	12:50	Ania Piskorz (CRUK Cambridge Institute, University of Cambridge)	Unlocking the potential of FFPE cancer specimens in predicting response and effect of cytotoxic therapy
12:50	13:05	Nicholas Timpson (University of Bristol)	Exome Sequencing of UK Birth Cohorts
13:05	14:00	Photo, Lunch and Poster Session II	

14:00	**15:30**	**Epigenetics/Human & Clinical Genomics**	**Chair: Ania Piskorz, CRUK Cambridge Institute University of Cambridge**
14:00	14:25	Areeba Patel (University of Heidelberg)	Methylation based classification of diagnostically relevant tumour classes
14:25	14:50	Alex de Mendoza (Queen Mary University of London)	Early Origins of Eukaryotic DNA Methylation Pathways
14:50	15:05	Hannah Trivett (University of Liverpool)	Clinical metagenomics can resolve pathogens to strain-level identification direct from stool
15:05	15:20	Aimee Hanson (University of Bristol)	Considering confounding in rare variant genome wide association studies
15:20	15:30	Silver Sponsor Talk	Pacific Biosciences: The power of high accuracy sequencing. Game-changing sequencing capabilities
15:30	16:00	Coffee Break	
16:00	**17:30**	**Bioinformatics & Machine Learning in Genomics**	**Chair: Matt Loose, University of Nottingham**
16:00	16:25	Rhydian Windsor (Oxford Nanopore)	Generative Modelling of Nanopore Signals
16:25	16:50	Leonid Chindelevitch (Imperial College London)	Insights into AMR from large-scale genotypic and phenotypic data analysis
16:50	17:05	Tim Downing (Pirbright Institute)	Pangenome variation graph analysis reveals insights into livestock poxviruses
17:05	17:20	Rory Munro (University of Nottingham)	Adaptive Sampling on PromethION: Soaring Close to the Sun with Icarust and Readfish
17:20	17:30	Silver Sponsor Talk	Opentrons: Accessible, touch-screen enabled automation for hands-off NGS library preparation

Conference Dinner - Bristol Harbour Hotel, The Sanisovo Room, Weds 17th [TICKETED]			
Start (GMT)	Finish (GMT)	Event	
18:30	19:30	Arrival Drinks	
19:30	00:00	Banquetting Menu Dinner	

Schedule: Wednesday 17th July - Parallel Session B

Parallel Session B - LT2			
Start (GMT)	Finish (GMT)	Presenter	Presentation Details
09:30	**11:05**	**Evolutionary Genomics I**	**Chair: Jordi Paps Montserrat (University of Bristol)**
09:30	09:55	Peter WH Holland (University of Oxford)	Darwin Tree of Life project: Evolutionary applications
09:55	10:10	Filipe Castro (CIIMAR/FCUP)	A Domino Effect? Of loss, duplication and novelty in a transition to the sea
10:10	10:35	Marta Farré-Belmonte (Kent)	The role of 3D chromosome folding in mammalian genome evolution
10:25	10:40	Tom Jenkins (Exeter)	Plastic versus adaptive responses to climate in barbastelle bats
10:40	11:00	Gold Sponsor Talk	Element Biosciences
11:00	11:30	Coffee Break	
11:30	**13:00**	**Microbes and Microbiomes I**	**Chair: Kate Baker, University of Cambridge**
11:30	11:55	Alexandre Almeida (University of Cambridge)	Meta-analysis of the uncultured gut microbiome reveals a novel keystone biomarker of hea
11:55	12:20	Lauren Cowley (University of Bath)	Using phylogeographical signal of Salmonella enterica serovar Enteritidis to train a hierarchical machine learning model to rapidly predict source attribution
12:20	12:35	Winnie Lee (Imperial College London)	Genomic population study of bloodstream Klebsiella spp. in 2020 in Southwest, UK
12:35	12:50	Eleanor Hayles (Quadram Institute, Norwich)	Genomic Epidemiology of SARS-CoV-2 in Norfolk, UK, March 2020- December 2022
12:50	13:00	Silver Sponsor Talk	Oxford Nanopore
13:00	14:00	Lunch and Poster Session II	
14:00	**15:35**	**Microbes and Microbiomes II**	**Chair: Nick Loman, University of Birmingham**
14:00	14:25	Alan Walker (University of Aberdeen)	Questioning the foetal microbiome illustrates pitfalls of low-biomass microbial studies
14:25	14:50	Edward Cunningham-Oakes (University of Liverpool)	Metatranscriptomics improves the laboratory diagnosis of infectious intestinal disease fron
14:50	15:05	Alice Nisbett (Quadram Institute, Norwich)	The genomic switcheroo: characterising genome rearrangements within typhoidal infections
15:05	15:20	Kirsty Sands (University of Oxford)	Providencia in the fly microbiome acting as a reservoir of blaNDM carriage: A threat in the dissemination of antimicrobial resistance and infection?
15:20	15:35	Steven Rudder (Quadram Institute, Norwich)	Genomic diversity of non-typhoidal Salmonella found within patients suffering from gastroenteritis in Norfolk, UK
15:35	16:00	Coffee Break	
16:00	**17:35**	**Evolutionary Genomics II**	**Chair: Sion Bayliss, University of Bristol**
16:00	16:25	Sandra Álvarez Carretero (University College London)	Bayesian methods to infer evolutionary timelines when deep divergences are present and large genomic datasets used
16:25	16:50	Jialin Wei (Bristol)	Convergent Genome Evolution In The Conquest Of Land By Animals
16:50	17:05	Chris Clarkson (QMUL/UCL)	Characterising structural variation in known pathogenic STRs across genetic ancestries, germ-line instability events and diseased cohorts
17:05	17:20	Sarah Quigley (Kent)	Investigating mechanisms of genomic rearrangement in the human genome

Conference Dinner - Bristol Harbour Hotel, The Sanisovo Room, Weds 17th [TICKETED]			
Start (GMT)	Finish (GMT)	Event	
18:30	19:30	Arrival Drinks	
19:30	00:00	Banquetting Menu Dinner	

Schedule: Thursday 18th July

Thursday 18th July 2024 - LT1 (Morning)			
Start (GMT)	Finish (GMT)	Presenter	Presentation Details
08:30	09:30	Coffee and Pastries	
09:30	**11:05**	**Single Cell Genomics**	**Chair: Lia Chappell, Wellcome Sanger Institute**
09:30	09:55	Jimmy Lee (Wellcome Trust, Sanger)	Shared molecular vulnerabilities of human cortical neurons in C9ORF72 Amyotrophic Lateral Sclerosis
09:55	10:20	Rebecca Berrens (University of Oxford)	Study transposable elements at single molecule level in single cells to understand their role in gene regulation.
10:20	10:35	Christos Proukakis (UCL)	The somatic CNV landscape of the Parkinson's disease brain at single cell resolution
10:35	10:50	Simon Cockell (Newcastle University)	Geographically weighted methods for spatial transcriptomics data analysis
10:50	11:00	Silver Sponsor Talk	Parse Biosciences: Smash the Limits of Single Cell Sequencing with Parse Biosciences
11:00	11:30	Coffee Break	
11:30	**12:30**	**Keynotes**	**Chair: Michael Quail, Wellcome Sanger Institute**
11:30	12:00	Deborah Williamson (UK Health Security Agency)	TBD
12:00	12:30	Matt Brown (Genomics England)	Improving Diagnostic Rates for Rare Diseases – the Genomics England program
12:30	THE END	Lunch and leaving	

Selected Speakers Abstracts

Plant and Animal Genomics

Ribosomal DNA Copy Number Variation Associates with Hematological Profiles and Renal Function in the UK Biobank

Francisco Rodriguez-Algarra
Queen Mary, University of London

Authors:
Francisco Rodriguez-Algarra, David M. Evans, Vardhman K Rakyan

Affiliations:
FRA and VKR: The Blizard Institute, School of Medicine and Dentistry, Queen Mary University of London, London, E1 2AT, United Kingdom.
DME: Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, 4072, Australia; Frazer Institute, The University of Queensland, Brisbane, Queensland, 4102, Australia; MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, United Kingdom.

Abstract:
The phenotypic impact of variation in repetitive genomic features, such as the multi-copy, multi-locus 47S ribosomal DNA (rDNA), is poorly understood. Despite its key role and displaying copy number (CN) variation in humans (~100-600 copies), strong evidence for rDNA-associated variation impacting human traits is still lacking.

We estimate rDNA CN on 157,227 whole genome sequences from white British UK Biobank (UKB) participants using a proxy method we develop and validate. Phenotypes in the blood â€œcountsâ€ and â€œbiochemistryâ€ categories appear enriched in a phenome-wide screen of rDNA CN associations. rDNA CN and systemic inflammation markers, including the neutrophil-to-lymphocyte ratio, strongly associate in targeted models. Further analyses suggest rDNA CN causally influences blood composition, showing no association between rDNA CN and factors known to alter blood composition measurements. Stringent population filtering retains the associations. Similar results replicate in other ethnicities and the second UKB sequencing release (further 209,681 participants). rDNA CN also associates significantly with renal function biomarkers. Estimates of Glomerular Filtration Rate derived from these associate negatively with rDNA CN, suggesting higher CN corresponds with worse renal function, leading to higher kidney failure risk. In summary, we find that rDNA CN associates with hematological profiles and renal function, highlighting a novel source of genetic variation influencing phenotype, and providing a broader explanation for the genetic basis of human trait variance.

New method to detect chromosomal reciprocal translocations using long-read sequencing

Carla Canedo-Ribeiro
University of Kent

Carla Canedo-Ribeiro (1), Craig R. G. Lewis (2), Justin Holl (3), Marta Farré (1)

1 School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NH, UK
2 PIC Europe, C/Pau Vila, 22 2o piso, 08174, Sant Cugat del Valles, Barcelona, Spain
3 Genus plc PIC, Hendersonville, TN, United States

Reciprocal translocations (RT) are one of the main chromosomal rearrangements (CR) that cause infertility. In species like pigs, RTs have an incidence between 0.5-1.5%, leading to smaller litter sizes. In the pig breeding industry, it is common to perform artificial insemination, which allows to use of sperm from a high genetic merit boar to inseminate several sows. If the boar has an undetected RT, this CR will be passed through generations and lead to a significant litter size reduction. RTs can be detected through routine cytogenetic techniques, such as Karyotyping and FISH. However, these techniques present limitations due to their resolution, i.e., cryptic RTs can be missed in Karyotyping, or the probes used for the FISH experiment might not fully hybridise the region affected by the RT. Moreover, in both cases, chromosome metaphase spreads from fresh samples are needed, limiting the shipping time.

To overcome these limitations, a new approach was developed to detect RTs and genotyping using ONT long-read sequencing. HMW gDNA was extracted from six pig frozen blood samples (four diagnosed as normal and two with an RT using FISH). Sequencing was performed in a minION device, obtaining 7.9x coverage. NanoSV was used to call structural variants, and a custom filtering bioinformatics pipeline was developed to identify reliable structural variants. Breakpoints for samples affected by an RT were then verified by PCR. Overall, we successfully identified all structural rearrangements, indicating that ONT long-read sequencing can be used as an alternative approach to detect RTs.

Human & Clinical Genomics

Utilising long-read sequencing approaches to detect structural variants of clinical significance in childhood cancer

Steven Hair (Newcastle University)

Steven Hair, Joe Fenwick, Eleanor G. Woodward, Richard Yim, Phoebe Snow, Letizia Marchetti, Lisa J. Russell, Sarra L. Ryan.
(Wolfson Childhood Cancer Research Centre, Newcastle University, UK.)

Background
Many childhood cancers are characterised by genetic abnormalities of clinical significance, including large structural variants (SVs). Many SVs are challenging to detect due to the variable repetitive sequences in which they are located. Long-read sequencing approaches allow for targeted sequencing and methylation profiling of difficult genomic regions such as the immunoglobulin loci, facilitating the detection of SVs and removing the need for more costly whole-genome sequencing.

Methods
Using a novel adaptive sampling (AS) ultra-long-sequencing approach we characterised and detected risk-defining SVs in seven cancer samples with known abnormalities for the first time. Samples were sequenced using the ONT ultra-long sequencing kit (ULK-114) with AS. The targeted region was a list of genes that have potential to cause risk-stratifying abnormalities in cancer. Data was processed through a custom bioinformatics pipeline, with alignment to T2T-CHM13v2.0 genome using Minimap2, SV calling using Sniffles2 and copy number variation calling using QDNAseq.

Results
The samples had an average library N50 of 57kb (44kb-75kb). SV calling identified an average of 2747 SVs (1287-7232) per sample. The following known risk-stratifying genetic abnormalities were detected in each respective sample: BCR::ABL1, IGH::MYC, IGH::DUX4, IGH::CCND1, IGL::MYC, iAMP21 and high hyperdiploidy. The breakpoints of translocations were resolved to base pair level in all samples and methylation profiling identified hypomethylation of DNA across translocated oncogenes.

Conclusions
This study demonstrates the potential of targeted ultra-long-sequencing as a tool for detecting and characterising complex SVs in cancer. The samples sequenced here represent challenging to characterise SVs, mostly involving the immunoglobulin loci that can be difficult to detect by standard testing. This study successfully detected all relevant SVs, along with additional variants that require further investigation.

Unlocking the potential of FFPE cancer specimens in predicting response and effect of cytotoxic therapy

Ania Piskorz
Cancer Research UK Cambridge Institute University of Cambridge

Ania Piskorz (1,2), Laura Madrid (2) , Joseph Thompson (2) , Alex Deamer (1), Johanna Barbieri (1), Rachel Barns (1), Emily Lythgoe (1), Marika Reinius (1) , Ashley Sawle (1), Abigail Edwards (1), Jason Skelton (1), Ugurcan Sal (1), Eleanor Denham (1), Tomiwa Adepetune (1), Julia Jones (1), Fay Skinner (1), Jo Heffer (1), Elsa Santos (1), Rita Carreira (1), Julia Ponte (1), Caroline Powell (1), Maria Vias (1) , Maria Escobar-Rey (3) , Wing-Kit Leung (1), Jamie Huckstup (1), Magdalena Sekowska (1), Ángel Fernández-Sanromán (1,3) , Filipe Correia Martins (1) , Jo Arnold (1), Florian Markowetz (1,2), Jason Yip (2) , Matthew Eldridge (1), James D Brenton (1,2,4,5,6), Geoff Macintyre (2,3)

1) Cancer Research UK Cambridge Institute University of Cambridge; 2) Tailor Bio Ltd UK; 3) Spanish National Cancer Research Centre (CNIO); 4) Department of Oncology, University of Cambridge; 5) Cambridge University Hospitals NHS Foundation Trust; 6) Cancer Research UK Major Centre – Cambridge University of Cambridge

Background
There are two main challenges in genomic biomarker discovery in high grade serous ovarian cancer (HGSOC). Firstly, because HGSOC is driven by TP53 mutations and extensive chromosomal instability (CIN) which cause genetic chaos, simply finding biomarkers in the noise to enable precision medicine, reduce toxic side effects of treatment, and improve overall response rates is highly demanding. Secondly, the most common archival tumour specimens collected in the clinic are formalin-fixed paraffin-embedded (FFPE) often with degraded DNA adding another layer of complexity during data analysis. In our study we tested whether we could use CIN signatures to predict patient response to platinum-based and doxorubicin chemotherapy in routinely collected diagnostic FFPE material of variable DNA quality.

Methods
We performed shallow whole genome sequencing (sWGS) using FFPE samples from 41 HGSOC patients treated with platinum, and doxorubicin following platinum. Across samples of variable FFPE quality, we compared different WGS methods, different RNA library prep methods and assessed the quality and challenges using 10x FFPE Visium.

Results
We successfully applied sWGS to FFPE with variable DNA quality across different WGS methods which enabled us to evaluate CIN signatures as a biomarker. In 41 HGSOC patients we validated our predictor of platinum sensitivity. We also discovered a new biomarker for doxorubicin sensitivity based on a CIN signature related to focal DNA amplification, validated in 26 patients treated with doxorubicin following platinum.

Conclusion
We showed that our CIN signatures single test could be applied to routinely collected FFPE material and guide therapy choices, with the potential to shift generalised chemotherapy treatment towards precision medicine. Our RNA quality matrix and method optimisation for FFPE enable us to extend this study to infer copy number alterations and transcriptional diversity in bulk RNA and spatial single cell resolution (10x Visium).

Exome Sequencing of UK Birth Cohorts

Nicholas Timpson
University of Bristol

Birth cohort studies involve repeated surveys of large numbers of individuals from birth and throughout their lives. They collect information useful for a wide range of life course research domains and biological samples which can be used to derive data from an increasing collection of omic technologies. This rich source of longitudinal data, when combined with genomic data, offers the scientific community valuable insights ranging from population genetics to applications across the social sciences. Here we present quality-controlled whole exome sequencing data from three UK birth cohorts: the Avon Longitudinal Study of Parents and Children (8,436 children and 3,215 parents), the Millenium Cohort Study (7,667 children and 6,925 parents) and Born in Bradford (8,784 children and 2,875 parents). The overall objective of this coordinated effort is to make the resulting high-quality data widely accessible to the global research community in a timely manner. We describe how the datasets were generated and subjected to quality control at the sample, variant and genotype level and offer preliminary analyses to illustrate the quality of the datasets and probe potential sources of bias. We introduce measures of ultra-rare variant burden to the variables available for researchers working on these cohorts and show that the exome-wide burden of deleterious protein-truncating variants, Shet burden, is associated with educational attainment and cognitive test scores. The whole exome sequence data from these birth cohorts (CRAM & VCF files) will be available through the European Genome-Phenome Archive and study access points.

Epigenetics/Human & Clinical Genomics

Clinical metagenomics can resolve pathogens to strain-level identification direct from stool

Hannah Trivett
University of Liverpool

Authors:
H. Trivett (1,2) , E. Cunningham-Oakes (1,2) , A.C. Darby (1,2)
1) Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool
2) NIHR Health Protection Research Unit in Gastrointestinal Infections

Abstract:
Metagenome sequencing (mNGS) is a promising method for rapid agnostic pathogen detection; however, it has yet to transition from research to routine clinical use. This investigation aimed to identify clinically relevant pathogens directly from faecal samples, providing antimicrobial resistance and virulence profiles for the pathogens of interest, compared to the gold standard culture-based identification and other molecular techniques circulating in clinical laboratories.

The performance of mNGS for clinical applications was tested by sequencing 15 clinical faecal samples with known aetiologies using PacBio HiFi sequencing. Twelve samples had pathogens that were previously detected using traditional molecular diagnostic methods, and three samples were determined to be pathogen-negative.

Of the 12 pathogen-positive samples, 10 metagenome-assembled genomes (MAGs) were produced for the pathogens of interest. Our results were comparable to culture-based molecular diagnostics and whole genome sequencing in clinical laboratories. Strain-level diversity was displayed across this dataset, where five samples contained more than two MAGs for E. coli. Metagenomic sequencing provided greater sensitivity than traditional methods of pathogen detection, identifying E. coli within Sample 20 that confer sequence types associated with intestinal infections, originally determined as a pathogen-negative sample. In addition, antimicrobial resistance profiles were produced using ABRicate of, which highlighted all 17 E. coli MAGs conferred resistance genes for cephalosporin and carbapenem, high-risk antimicrobial classes documented on WHO’s Bacterial Pathogen Priority List for AMR.

mNGS has the potential to be used as a culture-independent method that goes beyond current molecular and culture-based clinical diagnostics to identify clinically relevant pathogens at the strain level. It provides comprehensive AMR and virulence gene profiles that are efficient for guiding clinical treatment, highlighting its value within clinical diagnostic laboratories.

Considering confounding in rare variant genome wide association studies

Aimee Hanson
University of Bristol

Aimee Hanson (1), Daniel Lawson (1), George Davey Smith (1), Gibran Hemani (1)
(1) MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol

The interrogation of rare genetic variant associations with complex traits is now achievable in large-scale population cohorts of whole-exome or whole-genome sequenced individuals. This class of variant is enriched for mutations with large and predicted deleterious effects on biological function, making them informative in understanding molecular mechanisms of disease, and attractive instruments to proxy biological exposures in mendelian randomisation (MR) studies. The extent to which rare-variant associations are biased by inadequately controlled demographic (e.g. population stratification and assortative mating) and indirect (e.g. parental) genetic effects in population genome-wide association studies (GWAS) is currently unclear. Though rare variants reflect more recently emerging ancestral substructures in human populations, this is not completely captured by rare-variant derived principal components. Using a within-family model, in which such confounders are most successfully controlled, we tested for bias in published rare-variant associations with 166 quantitative traits in the UK Biobank. We saw little evidence of heterogeneity in rare SNP effect estimates from population GWAS and within-family models overall, but evidence for residual bias in some demographically stratified traits, including height and educational attainment, which subsequently biased MR causal estimates. Biased SNP effects captured a mixture of fine-grained population stratification and polygenic confounding, thorough which rare variants appeared to tag common-variant burden for polygenic phenotypes in carriers. Polygenic confounding also appeared to result in some false positive rare-variant associations with height, and may similarly impact associations with other highly polygenic traits. Future use of family studies to detect biased genetic effects will facilitate sound interpretation of rare-variant associations and improve accuracy in causal inference methodologies utilising rare genetic instruments.

Bioinformatics & Machine Learning in Genomics

Pangenome variation graph analysis reveals insights into livestock poxviruses

Tim Downing
Pirbright Institute

Authors: Chandana Tennakoon, Caroline Wright, Thibaut Freville, Tim Downing.
Affiliation: The Pirbright Institute, Surrey, UK.

Background: Poxviruses cause considerable livestock disease worldwide and seasonal epidemics in diverse foci. Major sources include goatpox virus, sheeppox virus and cattle-infecting lumpy skin disease virus (LSDV). Poxvirus have long double-stranded DNA genomes (~150 Kb) with a GC content of just 26%. Moreover, they possess high genetic diversity driven by high rates of mutation and recombination. Consequently, detecting mutations among poxvirus samples remains challenging and better approaches are needed.

Methods: Conventional approaches align reads from newly sequenced libraries to an existing reference genome. This biased approach works well for samples closely related to the reference, but less well for divergent, novel or recombinant isolates. Recent advances in pangenome variation graph (PVG) construction and read mapping facilitate a more effective approach that discovers more mutations. A PVG is a mosaic reference against which reads can be mapped.

Results: We demonstrate that more LSDV sample reads map to a PVG compared to mapping to a single genome and that this allows detection of new mutations, particularly low-frequency heterozygous mutations. We present a comprehensive Nextflow workflow to automate important processes: download, quality control, phylogenetic analysis, PVG construction, PVG analysis, PVG visualisation, PVG openness assessment, PVG diversity analyses, PVG community detection, PVG annotation, and traditional pangenome gene-based analysis. We use node- and k-mer-based methods to show that LSDV has a closed PVG, and that its conserved core genome comprises 98% of genomic bases (147,766 bp), with 2% remaining as the variable accessory genome.

Conclusion: Our work illustrates how PVG-based methods can be more effective for mutation detection, which impacts our understanding of diverse areas like evolutionary inference, transmission tracing, mechanisms of viral adaptation, vaccine design, and molecular diagnostics.

Evolutionary Genomics I

A Domino Effect? Of loss, duplication and novelty in a transition to the sea

Filipe Castro
CIIMAR/FCUP

Extreme ecological transitions between environments with contrasting osmolalities pose an immense physiological challenge for aquatic animals. A successful documented example of a lasting transition from freshwater to marine habitats is that of the stripped-eel marine catfish, Plotosus lineatus. To expose the genomic events underscoring the adaptation to a high salt environment, we assembled and screened the genome of P. lineatus. We deduce that genes involved in kidney ion reabsorption (e.g. Na+, Cl- amd K+) and glucose uptake have been targeted-deleted from the P. lineatus genome, unlike their freshwater relatives. These findings coexist with the phenotypic elimination of the urine dilution section of the glomeruli typically found in freshwater fish, and parallel a unique hyperosmotic urine production. Notably, the key anatomical novelty involved in salt secretion, the dendritic organ, is governed by a transcriptomic landscape typically found in the gill/kidney of marine teleosts (e.g. Na+,K+-ATPases), and the acquisition of tissue-specific gene expansions involved in the excretion of divalent ions (e.g. slc26a6). Our findings offer a exceptional insight into the molecular events leading to a singular conquering event of marine environments.

The role of 3D chromosome folding in mammalian genome evolution

Marta Farrā
University of Kent

Lucĩa Ãlvarez-Gonzālez (1), Cristina Arias-Sardā (2), Marta Farrā (2), Aurora Ruiz-Herrera (1)

1 Departament de Biologia Celālular, Fisiologia i Immunologia, Universitat Autānoma de Barcelona, Spain
2 School of Biosciences, University of Kent, UK

Understanding the evolution of chromatin conformation is fundamental for understanding the mechanisms responsible for the origin and plasticity of genome architecture. Ancestral genome reconstructions have shown that structural changes disrupting synteny preferentially cluster in regions that are prone to break and reorganize. It is also known that changes in gene expression caused by genome reshuffling may have a selective advantage through the development of new adaptive characters specific to different mammalian lineages. These data suggest that sequence composition is not alone in determining evolutionary plasticity but rather that the occurrence and subsequent fixation of genome rearrangements are multifaceted, involving (1) repetitive elements (i.e., making DNA more susceptible to chromosomal reorganization); (2) functional constrains (i.e., genes related to species-specific phenotypes); and (3) genome folding dynamics and its effect on gene regulation/function. This suggests that the permissiveness of some genomic regions to undergo genomic rearrangements, especially in germ cells, is influenced by chromatin 3D conformation.

Here, we describe the principles of 3D genome folding in vertebrates and show how lineage-specific patterns of genome reshuffling can result in different chromatin configurations. We (1) identified different patterns of chromosome folding across vertebrate species using HiC data of 7 mammalian species; (2) reconstructed ancestral marsupial and afrotherian genomes analysing whole-genome sequences of species representative of the major therian phylogroups; (3) detected lineage-specific chromosome rearrangements; and (4) identified the dynamics of the structural properties of genome reshuffling through therian evolution. We show that chromatin configurational changes in extant species result from ancestral inversions and fusions/fissions. Our results represent a comprehensive catalogue of the close interplay between chromatin higher-order organization dynamics and therian genome evolution.

Plastic versus adaptive responses to climate in barbastelle bats

Tom Jenkins
University of Exeter

Understanding evolutionary responses to climate is crucial for assessing the vulnerability of biodiversity to climate change, especially for long-lived species, such as bats. In this study on wild western barbastelle bats (Barbastella barbastellus), we use transcriptomics and genomics to investigate (i) plastic responses to different in situ temperatures, and (ii) genetic adaptations to climate that may have evolved in the bats. Differential expression analysis revealed a number of upregulated genes in bats from lower temperatures (9.9-11.3oC), including genes reported to be important for non-shivering thermogenesis (e.g. UCP2). This suggests that these temperatures induce a cold stress plastic response and are likely near the thermal minima of active barbastelle bats. In contrast, few genes were differentially expressed in higher temperatures (15.0-16.4oC), which suggests that these temperatures are not near the thermal maxima experienced by active barbastelle bats across the range studied. Climate-associated loci identified by genotype-environment association analysis across the speciesâ€™ western range, revealed potential adaptations to maximum summer temperatures and annual precipitation levels. Of the variants located within gene regions, IGF1 and MAPKAPK2 have roles in thermogenesis and inflammatory responses. Genomic offset analysis highlighted an elevated risk of maladaptation to future climates under a high emissions scenario, but a relatively low risk under a global sustainability pathway. Our results provide evidence that barbastelle bats have evolved plastic mechanisms for acclimating to temperatures outside of their optimum range and genetic adaptations to warmer and drier climates.

Microbes and Microbiomes I

Genomic population study of bloodstream Klebsiella spp. in 2020 in Southwest, UK

Winnie Lee
Imperial College London

Authors:
Winnie Lee (1), Philip B Williams (2), Matthew B Avison (1)

1) School of Cellular and Molecular Medicine, University of Bristol, Bristol, BS8 1TD, UK
2) Bristol Royal Infirmary, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK

Background: Bloodstream infections (BSIs), caused by Klebsiella spp., particularly those harbouring blaNDM-1 are an increasing clinical challenge. Rapid empiric antimicrobial therapy is vital for management for BSIs. Understanding the genotypic basis of antimicrobial resistance (AMR) can help improve empiric choice and control spread.

Methods: We studied 261 Klebsiella spp. isolated from BSIs by a regional diagnostic laboratory serving 1.5 million people between January and December 2020. We examined the potential hospital transmission of Klebsiella pneumoniae using sequencing using Illumina-HiSeq-2500 (n=261) and the MinION (n=10). Sequencing data were analysed for AMR determinants and phylogenetic relationships. Concordance between phenotypic AMR and that prediction based on genotype was evaluated.

Results: K. pneumoniae ST307 was the predominantly identified multi-drug resistant sequence-type, harbouring determinants associated with fluoroquinolones, aminoglycosides, sulphonamides, phenicols and tetracycline resistance. ST23 and ST395 harboured most virulence genes amongst Klebsiella spp. Analysis of highly resistant ST15 K. pneumoniae containing blaNDM-1 indicated transmission to different patients within the same hospital ward. K. pneumoniae BSI isolates from two different patients showed 100% identity for AMR and heavy metal resistance on IncR and IncFIB plasmids. Concordance for genomic AMR prediction varied across bioinformatics tools for Klebsiella spp. Lowest critical errors were detected across all bioinformatics tools for cefotaxime. No critical errors were observed with Kleborate for ciprofloxacin.

Discussion/conclusions: High levels of resistance, presence of AMR and heavy metal resistance genes on mobile elements highlights the importance of continuous AMR monitoring and minimising delays in empiric antimicrobial therapy. WGS-based AMR prediction requires further optimisation prior to exclusive use for surveillance.

Genomic Epidemiology of SARS-CoV-2 in Norfolk, UK, March 2020- December 2022

Eleanor Hayles
Quadram Institute Bioscience

Authors:
Eleanor H. Hayles(1,2), Andrew J. Page(1,3), Robert A. Kingsley(1,2), Javier Guitian (4), The COVID-19 Genomics UK Consortium & Gemma C. Langridge(1)

(1) Quadram Institute Bioscience, Norwich Research Park, UK
(2) University of East Anglia, Norwich Research Park, UK
(3) Theiagen Genomics, Highlands Ranch, Colorado, USA
(4) Royal Veterinary College, 4 Royal College Street, London, UK

Background: In the UK, the COVID-19 Genomics UK Consortium (COG-UK) established real-time national genomic surveillance for SARS-CoV-2 during the pandemic. As a COG-UK partner, Quadram Institute Bioscience sequenced over 60,000 SARS-CoV-2 genomes, contributing to Norfolk becoming the most densely sequenced UK region. Retrospective SARS-CoV-2 investigation in this region will allow for understanding of lineage and variant diversity.

Methods: 29,406 SARS-CoV-2 whole genome sequences and metadata from Norfolk were extracted from the COG-UK dataset, dated March 2020 to December 2022, representing 9.9% of regional COVID-19 cases. Sequences were lineage typed using Pangolin, with subsequent analysis completed in R. Study approval was granted by the University of East Anglia (Ref:ETH2223-2684) and fell under COG-UK remit (PHE R&D Ref:NR0195).

Results: 401 global lineages were identified, with 280 appearing more than once and 125 over ten times. Temporal clustering identified a network of lineages based on first lineage emergence, split into 6 individual groups. Alpha, Delta, and Omicron variants of concern (VOC) accounted for 8.6%, 34.9% and 48.5% of sequences respectively. These formed four regional epidemic waves alongside the remaining lineages which appeared in the early pandemic prior to VOC designation and were labelled as ‘pre-VOC’ lineages. Regional comparison highlighted variance of VOC epidemic wave dates dependent on location.

Conclusion: This study is the first to assess SARS-CoV-2 diversity in Norfolk across a large timescale. SARS-CoV-2 was highly diverse and dynamic in its evolution in Norfolk between March 2020 – December 2022. The study highlights the utility of genomic methodology within pandemic response.

Microbes and Microbiomes II

The genomic switcheroo: characterising genome rearrangements within typhoidal infections

Alice M. Nisbet
Quadram Institute Bioscience

Authors:
Alice M. Nisbet[1], Emma V. Waters[1], Marie A. Chattaway[2], Gemma C. Langridge[1]
1.Quadram Institute Bioscience, Norwich Research Park, Norwich, NR4 7UQ, UK
2.Gastrointestinal Bacteria Reference Unit, United Kingdom Health Security Agency (UKHSA), London, NW9 5EQ, UK

Background: Enteric fever is the inclusive term for infections caused by a serovars Typhi, Paratyphi A and B(dTar-): the typhoidal . These pathogens can change their genomic structure (GS) via the rearrangement of large chromosomal fragments, leading to altered gene expression which may facilitate adaptation to their environment.

Methods: To determine GS variety and propensity for rearrangements within typhoidal , an array of UKHSA isolates collected between 2004-2019 was curated, corresponding to acute and carried infections (bacterial persistence >3 weeks). Oxford Nanopore Technologies long-read sequencing was used to generate genome assemblies from which GS was determined using socru.

Results: In S. Typhi, GSs were highly variable in both acute and carried infections, with 30 different GSs observed across the collection. Conversely, S. Paratyphi A and S. Paratyphi B were dominated by a single GS, GS1.1 and GS1.0 respectively. In S. Paratyphi A, alternatives to GS1.1 were almost exclusively observed in patients who had harboured the bacteria for >12 months, while GS1.0 was seen across all infection states of S. Paratyphi B. Interestingly, inversion of genomic fragment 1, encompassing half the genome, was common in S. Typhi and S. Paratyphi A. Within carriers, S. Typhi frequently rearranged over the course of infection, while the majority of S. Paratyphi maintained their GS.

Conclusions: Genomic rearrangements may play a role in S. Typhi persistence but is much less common in S. Paratyphi infections, highlighting a key difference between these human host-adapted bacteria that are often considered near identical.

Providencia in the fly microbiome acting as a reservoir of blaNDM carriage: A threat in the dissemination of antimicrobial resistance and infection?

Kirsty Sands
Ineos Oxford Institute for Antimicrobial Research,
Department of Biology, University of Oxford

Authors:
Kirsty Sands (1,2), Chioma Achi (1), Kate Cook (1), Shonnette Premchand-Branker (1), Mei Li (1), Edward Portal (1), Kenneth Iregbu (3), Seniyat Afegbua (4,11), Aminu Aliyu (5), Yahaya Mohammed (6), Ifenyinwa Nwafia (7), Oyinlola Oduyebo (8), Zainab Tanko (9), Ibrahim Abdukrasul (10), Timothy Walsh (1), Nigeria-AVIAR group (12)

1. Ineos-Oxford-Institute for Antimicrobial Research, Department of Biology, University of Oxford, United Kingdom
2. Infection and Immunity, Cardiff University, United Kingdom
3. Medical Microbiology, NHA, Abuja, Nigeria
4. Microbiology, Faculty of Life Sciences, ABU, Zaria, Nigeria
5. Medical Microbiology, AKTH, Kano, Nigeria
6. Medical Microbiology, UDUTH, Sokoto, Nigeria
7. Medical Microbiology, UNTHN, Enugu, Nigeria
8. Medical Microbiology, LUTH, Lagos, Nigeria
9. Medical Microbiology and Parasitology, KSA, Kaduna, Nigeria
10. Medical Microbiology, ABU, Zaria, Nigeria
11. Nigeria Defence Academy, Kaduna, Nigeria
12. AVIAR-Nigeria team

Abstract:
Houseflies and blowflies are often labelled as “insect-pests” and a pilot-study addressed the prevalence of clinically relevant antibiotic resistance genes (ARG) carried by flies circulating in hospitals as a precursor to an international surveillance-study. From 1,396 flies collected from eight Nigerian hospitals in 2022, 8% carried blaNDM carbapenemase ARG. Following whole genome sequencing, Providencia species were dominant (n=83/132). ARG were on plasmids for 75 and for eight Providencia vermicola, blaNDM-1 was chromosomal. 51 Providencia hangzhouensis of different sequence-types from multiple hospitals were carrying blaNDM-1 on non-mobilizable or conjugative predicted-motility plasmids, Five carried blaNDM-4 on IncC plasmids. Phenotypic evaluation revealed large minimum-inhibitory-concentration distribution against carbapenems. Providencia is a component of fly microbiome; however, these rates of blaNDM carriage is concerning as a mechanism to confer resistance to beta-lactam antibiotics, often relied upon in low- and middle-income hospitals. Genomic epidemiology of Providencia, plasmids and blaNDM variants will be analysed in context of global data to determine whether certain combinations are more likely to be transmissible.

Evolutionary Genomics II

Characterising structural variation in known pathogenic STRs across genetic ancestries, germ-line instability events and diseased cohorts

Chris Clarkson
QMUL/UCL

We present our analysis of ataxia causing repeat expansion loci (REs) using our novel tool, “Repeat Crawler”. This software supplements tools that call short tandem repeats (STRs) lengths by recording the presence and length of sub-repeat elements that both border and interrupt the disease STR in an allele-specific manner. By leveraging the 100,000 genomes project, we analysed a cohort of individuals from 5 separate genetic ancestries, and demonstrated a host of population-specific differences in the repeat sequences. We also demonstrate a distinct set of trends across “CAG” repeat genes. We show both the number of cis-modifier units that border and the number of the interruptions that occur within the repeat body robustly correlate with the overall length of the repeat. Beyond “CAG” repeats, we also studied the more complex repeats in FGF14 and RFC1. We uncover a variant of a known cis-modifier indel that associates with a distinct GAA STR size in the FGF14 locus. This variant is distinctly present in African genomes. Furthermore, we study repeat germline instability and show that the “AAGGG” pathogenic motif, in RFC1, only increases in size when a germline instability event occurs. Finally we studied the repeat structures in known ataxia patients and show that alleles with expanded NOP56 repeats tend to be missing a 3’ cis-modifying repeat element that is invariably present in the general population. Overall, our analysis illustrates the importance of structural variation in RE loci in diverse populations in order to understand their role in disease onset.

Investigating mechanisms of genomic rearrangement in the human genome

Sarah Quigley
University of Kent

Authors:
Sarah Quigley* and Marta Farrā
*School of Biosciences, University of Kent, Canterbury, England, CT2 7NZ

Abstract:
Understanding and identifying drivers of genomic rearrangements is crucial due to their role influencing changes in karyotype architecture and ability to influence speciation. Evolutionary breakpoint regions (EBRs) denote disrupted synteny among two or more species genomes. EBRs were previously identified in the human genome compared to 8 reconstructed ancestral genomes back to the mammalian ancestor. We investigated the genomic, transcriptomic, and epigenomic context of these EBRs, suggesting possible mechanisms for their formation. First, we identified housekeeping genes using gene expression data across 9 tissues in four species (human, macaque, cow, mouse) to define 841 ancestral, 263 primate-specific, and 935 human-specific housekeeping genes. With the inclusion of 5 other species (gibbon, chimp, pig, elephant, chicken) we detected genes under positive and negative selection in each branch leading to human. We then defined genomic regions located in the nuclear centre or closer to the lamina using published GPSeq data in humans. Overall, we found that EBRs are in gene dense regions, but not genes under negative selection nor housekeeping genes. EBRs correlated with TAD boundaries, are in genomic regions located inside the nucleus, and negatively associated with the nuclear lamina. Inversely, syntenic regions are enriched in both housekeeping genes and negatively selected genes. Supporting this, gene ontology analysis revealed EBRs are associated with adaptive responses, whereas syntenic regions associate with developmentally important terms involved in cellular maintenance. Our results suggest genomic rearrangements do not disrupt essential genes or regulatory blocks, and the rearranged regions tend to be positioned within the nuclear centre. Our results may provide an insight to the drivers of genome rearrangements and subsequently evolutionary changes.

Single Cell Genomics

The somatic CNV landscape of the Parkinson's disease brain at single cell resolution

Christos Proukakis
UCL Queen Square Institute of Neurology

Christos Proukakis (1, 2), Ester Kalef-Ezra (1, 2_, Zeliha Gozde Turan (1, 2), Dominic Horner (1, 2), Caoimhe Morley (1), Diego-Perez Rodriguez (1), Vanda Knitlhoffer (4), Iain Macaulay (4), Zane Jaunmuktane (1, 2), Jonas Demeulemeester (2,4), Fritz Sedlazeck (2, 5)

1 UCL Queen Square Institute of Neurology, London
2 Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, United States
3 Earlham Institute, Norwich Research Park, Norwich
4 KU Leuven, Department of Oncology, Leuven, Belgium
5 Baylor College of Medicine, Human Genome Sequencing Center, Houston, United States

Background
Parkinson's disease (PD), the second commonest neurodegenerative disorder, is rarely Mendelian, with heritability 30%. We have hypothesised a role for somatic mutations in its pathogenesis. We previously demonstrated somatic CNVs (gains) of the SNCA gene which encodes alpha-synuclein, the hallmark protein which aggregates in PD. As the evidence for large somatic CNVs in the human brain has increased, we now aim to elucidate the genome-wide somatic CNV landscape in PD brain.

Methods
We obtained post mortem human PD and control brains with ethics approval from Queen Square Brain Bank. We selected two brain regions: the cingulate cortex, which has prominent alpha-synuclein neuronal pathology, and the substantia nigra (SN), where dopaminergic (DA) neurons are heavily affected. We used fluorescent-activated nuclear sorting to select nuclei for whole genome amplification by PicoPLEX. After alignment to CHM13-T2T, we obtain CNV calls by Ginkgo.

Results
We selected 11 PD and 10 control brains, and sequenced a total of ~4,400 single cells . Interim analysis of cingulate cortex (2,100 cells) shows that 91% pass QC. There is no significant difference in the proportion of cells per donor containing at least one CNV in PD and control (10.2% v 12.3%). There are no significant associations of % of cells with CNV and age of onset or death. Further analysis for CNV characteristics is underway, and SN data will be integrated shortly.

Geographically weighted methods for spatial transcriptomics data analysis

Simon Cockell
Newcastle University

Authors:
Eleftherios Zormpas*, Antonis Giakountis†, Rachel Queen*, Alexis Comber‡, Simon J Cockell*,§

Affiliations:
* Biosciences Institute, Newcastle University, UK.
† Department of Biochemistry and Biotechnology, University of Thessaly, Greece.
‡ Leeds Institute for Data Analytics, University of Leeds, UK.
§ School of Biomedical, Nutritional and Sport Sciences, Newcastle University, UK.

Abstract:
Spatial transcriptomics (ST) enables the comprehensive analysis of the entire transcriptome with approaching single-cell resolution while preserving the spatial information of the tissue. The properties of spatial data give rise to methodological concerns that relate to the accommodation of three concepts: spatial autocorrelation, spatial heterogeneity, and the modifiable areal unit problem. These concepts are well understood in the geographical sciences, and analysis methods have been developed that take account of them. Here we discuss the application of such methods in the biological context using data from the Visium ST platform, specifically geographically-weighted principal component analysis (GWPCA) and fuzzy geographically-weighted clustering (FGWC). GWPCA performs multiple local PCAs using geographically weighted gene expression matrices. Each gene gets a loading score in each component at each location in the tissue. In this way, lists of genes that play significant roles in the expression variability at each location can be generated and mapped. Consideration of these rankings in a Visium dataset from prostate cancer tissue identifies gene signatures associated with disease state, severity, and prognosis. FGWC is a supervised clustering method that for each point in space assigns partial membership to multiple classes. The nature of some ST techniques – especially low-resolution techniques like Visium – means experimental spots will often derive from a mixed cell population. We show that FGWC incorporates this uncertainty to allow for more refined cluster definition. Geocomputational methods reveal the existing heterogeneity over space that must be accounted for in ST analysis.

Selected Poster Abstracts

CoronaHIT : a high-throughput sequencing method of SARS-CoV-2 genomes

Dave Baker
Quadram Institute Bioscience

The COVID-19 pandemic caused by the SARS-CoV-2 virus began late 2019 in Wuhan, China. Key to controlling the pandemic was understanding the epidemiological spread of the virus at global, national, and local scales. Rapid and accurate genomic surveillance was achieved by Whole-Genome Sequencing of the SARS-CoV-2 genome. We present ‘CoronaHIT’, a low-cost, high-throughput method for sequencing SARS-CoV-2 genomes on both Illumina and Oxford Nanopore platforms. Libraries were prepared using a modified Illumina DNA prep tagmentation approach with subsequent barcoding via PCR to uniquely label each sample for multiplexing. For the ONT platform, primers with a 3’ end compatible with the Nextera transposon insert and a 24 bp barcode at the 5′ end with a 7 bp spacer were used to PCR barcode the tagmented ARTIC PCR products. For the Illumina platform, PCR products were tagmented and barcoded using a similar custom 9bp barcoded primers with P5 and P7 ends multiplexed up to a theoretical 9216 samples. From March 2020 to March 2022 the Quadram Institute sequenced 87,000 Covid-19 samples using both Illumina and Nanopore via this method. The method can also be adapted for low-cost high-throughput sequencing of bacterial genomes.

Comparison of 3 polyA selection and rRNA depletion commercial RNA library preparation kits to interrogate the RNA contents of FFPE samples

Johanna Barbieri
Cancer Research UK - Cambridge Institute

Authors: Johanna Barbieri*, Emily Lythgoe*, Ugurcan Sal, Eleanor Denham, Tomiwa Adepetune, Ashley Sawle, James Brenton, Ania Piskorz

Affiliations: Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge, United Kingdom

Differential gene expression analysis is extremely important to understand the mechanism underlying cancer resistance to treatment, clonal selection during treatment and improve tumour classification. However, as the most common source of archival material for cancer diagnosis is formalin-fixed paraffin-embedded (FFPE) tissue, the quality of RNA is often compromised. We compare three RNA library preparation methods to identify the most suitable workflow for the FFPE tumour samples.

We selected 6 high-grade serous ovarian cancer (HGSCO) FFPE samples with variable quality one Fresh Frozen and one positive control UHRR (Roche, good quality).
We used 75ng of total RNA as input (3 technical replicates per sample) for three library preparation methods: 1) NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® mRNA (NEBNext mRNA), 2) NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® rRNA depletion (NEBNext rRNA depletion), 3) Watchmaker Genomics RNA Library Prep Kit with Polaris Depletion (Watchmaker Polaris).

As expected, the total gene counts for FFPE samples were lower with NEBNext mRNA compared to NEBNext depletion kits. The total gene counts for rRNA depletion workflows were similar with slightly higher uniformity with Watchmaker Polaris.
The percentages of reads aligned to the reference genome were similar between all rRNA depletion workflows. The percentage of reads assigned to ribosomal proteins in FFPE samples was similar between the rRNA depletions workflows but was much higher for NEBNext mRNA. We observed no insert size variability between all three methods.

Our study confirmed that polyA captured method is not feasible for tumour FFPE material. We obtained promising gene expression results for FFPE tumour tissues using both tested rRNA depletion methods with Watchmaker being more time efficient and easier to implement. We have now successfully applied Watchmaker Polaris library prep to larger FFPE HGSOC cohorts.

Metagenomic sequencing for diagnosis of Ventilator Associated Pneumonia

Jenna Callaway
University of Bristol

Background: Diagnosis of Ventilator Associated Pneumonia (VAP) is currently reliant on culture of a pathogenic organism. If sequencing were applied to samples from ventilated patients then the identification of a pathogenic organism, and subsequent diagnosis, would be possible substantially faster than the current 48-72 hours. The complexity of the microbiome present in the lungs complicates the identification of pathogens and introduces the need to distinguish infection from colonization. Methods: Endo-tracheal Aspirate (ETA) samples from ventilated patients suspected of having VAP were extracted and sequenced by both targeted 16S sequencing using Oxford Nanopore Technology (ONT) sequencing and Whole Genome Sequencing (WGS) using Illumina. The results obtained by these methods were compared to identify inherent biases in either method, and compared to clinical culture results to establish the benefits of either method for diagnosis of VAP.

Results: The results show that both sequencing methods are largely concordant with culture results, although both methods also have inherent biases. Conclusions: When biases are considered and the benefits of either method are factored in, this data shows that the use of 16S targeted sequencing using ONT gives results from which clinical diagnoses can be confidently made.

Hospital insect pests as carriers of pathogenic and AMR bacteria: a metagenomic approach

Kate Cook
University of Oxford

Kate Cook (1), Maryam Yusuf Ahmad (2,4), Firdausi Said Mohd (2), Aminu Aliyu (3), Abdulrazak Muhammad (4), Chioma Achi (1), Kenneth Iregbu (5), Timothy Walsh (1), Kirsty Sands (1)

(1) Ineos Oxford Institute for Antimicrobial Research, Department of Biology, University of Oxford, United Kingdom; (2) Murtala Muhammad Specialist Hospital, Kano, Nigeria; (3) Department of Medical Microbiology, Aminu Kano Teaching Hospital, Kano, Nigeria; (4) Department of Microbiology, Bayero University, Kano, Nigeria; (5) Department of Medical Microbiology, National Hospital Abuja, Nigeria

Insect infestation of hospitals is an infection prevention and control concern across low- and middle-income countries with tropical climates. Flies and cockroaches are known to carry antimicrobial-resistant (AMR) bacteria in Nigeria. The extent to which insects are able to disseminate these bacteria in hospitals, and therefore their threat to human health, is unknown.

Houseflies and cockroaches were collected from Murtala Mohammed Specialist Hospital in Kano, Nigeria. Flies were washed and dissected to produce pooled samples that represented bodily surface and gut microbiomes. DNA extraction and Oxford Nanopore Technologies shotgun metagenomic sequencing were conducted, followed by metagenomic analysis with BugSeq.

Relative abundances of bacterial species varied between insects and between gut compared to surface bacterial communities. Clinically relevant bacteria containing resistance markers against multiple classes of antibiotics were detected in houseflies and cockroaches, with blaCTX-M-15-positive Escherichia coli on their bodily surfaces. Additional extended-spectrum beta-lactamase genes were detected in both housefly samples, with blaDHA detected exclusively in Morganella morganii from the gut microbiome. Carbapenemase genes (blaNDM) were detected only on the housefly surface, with blaNDM-5 on an IncX3 plasmid. Aminoglycoside resistance markers were present across all samples - mostly in Serratia marcescens in cockroaches and Enterobacteriaceae species in houseflies. There were several instances of broad host-range plasmids containing multiple AMR genes that target beta-lactam and aminoglycoside antibiotics. This work highlights the potential of insect pests to contribute to the pathogenic and AMR landscapes of the clinical setting.

Genomic epidemiology of antimicrobial resistance and virulence plasmids in Escherichia coli and Klebsiella pneumoniae from One Health settings

Keira Cozens
University of Bath

Plasmids are major drivers of bacterial evolution, and recent advances in long-read sequencing make it possible to generate complete assemblies of plasmid genomes for large population samples. Here we apply this technology to the critical priority pathogens Escherichia coli and Klebsiella pneumoniae from humans, animals, and the environment, to determine how frequently convergent plasmids emerge that possess both antimicrobial resistance and virulence traits.

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, which is associated with E.coli, and used long-read sequencing to generate hybrid assemblies of 68 E. coli isolates and 2 K. pneumoniae isolates harbouring the iuc5 locus. Resfinder and Prokka 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, emphasising the importance of monitoring virulence alongside antimicrobial resistance surveillance within a One Health context.

Keira Cozens (1), Marjorie Gibbon (1), Natacha Couto (2), Tiffany Taylor (1), Edward Feil (1)

(1) University of Bath
(2) University of Oxford

Recommendations for k-mer usage in pathogen surveillance from the direct sequencing of over 1000 patient stool samples

Edward Cunningham-Oakes
University of Liverpool

Timely outbreak detection is crucial for controlling gastrointestinal infections, yet traditional laboratory-based surveillance is time-consuming and can cause delays. We compared culture-based and PCR diagnostics with sequencing data across 1,067 samples from patients with acute gastroenteritis in the INTEGRATE study. We aimed to determine if sequencing reads from clinical samples can accurately detect gastrointestinal pathogens at symptom onset.

We found strong positive correlations between metatranscriptomic and metagenomic data and gold-standard culture and PCR-based diagnostics for important community gastrointestinal pathogens, such as Campylobacter, Salmonella, and Rotavirus. We established a k-mer detection limit for pathogens based on microbiome complexity. We also recovered a complete transcriptomic profile of Salmonella Enteritidis from a patient later confirmed to have a Salmonella infection.

Our results suggest that directly sequencing clinical samples has diagnostic potential for pathogens of interest. Metatranscriptomics captures the broadest range of gastrointestinal pathogens and is comparable to gold-standard methods. However, non-read-based approaches are needed to accurately profile pathogenic E. coli.

AUTHORS AND AFFILIATIONS
Edward Cunningham-Oakes (1,2), Blanca Perez-Sepulveda3, Yan Li (3), Jay C. D. Hinton (3), Charlotte A. Nelson (4), K. Marie McIntyre (5), Maya Wardeh (6) , Sam Haldenby (4), Richard Gregory (4), Miren Iturriza-Gómara (3,7), Christiane Hertz-Fowler (4), Sarah J. O'Brien (5), Nigel A. Cunliffe (2,3), Alistair C. Darby (1,2,4*); on behalf of the INTEGRATE consortium

(1) Department of Infection Biology and Microbiomes, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK.
(2) NIHR Health Protection Research Unit in Gastrointestinal Infections, Liverpool, UK. Department of Clinical Infection, Microbiology and Immunology, Institute of Infection,
(3) Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK.
(4) Centre for Genomic Research, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK.
(5) School of Natural and Environmental Sciences, Newcastle University, Newcastle, UK.
(6) Department of Computer Science, University of Liverpool, Liverpool, UK.
(7) Centre for Vaccine Innovation and Access, Program for Appropriate Technology in Health (PATH), Geneva, 1218, Switzerland.

Exploring the effect of clinical FFPE sample RNA quality metrics and incubation time on spatial transcriptomics output.

Alex Deamer
Cancer Research UK Cambridge Institute

Background:
Formalin–fixed paraffin–embedded (FFPE) tissues are used for routine clinical cancer diagnosis and are very precious materials in research; however, the quality of RNA and tissue is variable. The degree of RNA degradation in FFPE samples as well as sample-slide adherence and its impact on 10X Visium performance is not well-known. Therefore, we tested whether any of the RNA quality metrics and incubation conditions can reliably determine the Visium FFPE performance.

Methods:
We performed 10X Visium FFPE CytAsisst on 28 high grade serous ovarian cancer (HGSOC) samples and 100 triple negative breast (TNB) cancer samples. All samples were sectioned at 5μM, sister slides were used to determine RNA quality via TapeStation HS RNA and 2 of the HGSOC samples were compared using 3- and 4-hour baking times. All libraries were sequenced to a depth of 35,000-70,000 reads per spot and analyzed using the 10X SpaceRanger platform.

Results:
We analysed the total genes detected alongside the median genes per spot and compared with the RIN/ DV50/ DV200 values using one way ANOVA for 41 of the HGSOC and 2 TNB. We observed a range of DV200 values (25-70%) but all (the lowest and highest) provided comparable total gene counts. 45 of the TNB samples and 18 of the HGSOC showed minor tissue detachment.
The 2 samples with modulated incubation times showed similar gene expression results and generated equal number of clusters.

Conclusions:
We did not observe any significant correlation between DV200/RIN score and Visium performance, but we observed sporadic correlation with DV50, which may be a better indication of assay success. We found that detachment was mainly observed in post-chemotherapy tissues with fat enriched components and that extended incubation time has no significant impact on gene expression but may reduce the tissue detachment and therefore improve Visium performance.

Exploration of the dynamic evolutionary rates of early animals.

Zhiqing Guo
University of Bristol

Around 530 million years ago during the Cambrian explosion, the nearly simultaneous emergence of most modern animal body plans (phyla) serves as strong evidence for a brief period of rapid phenotypic diversification. However, the precise rate and characteristics of this significant adaptive radiation remain topics of intense discussion. The evolutionary modes of the Cambrian explosion have been long characterized by debates between gradualism and punctuated equilibrium without reaching a definitive resolution. In this study, we utilize Bayesian morphological clock modelling to investigate the evolutionary dynamics of early animals, employing a dataset of 2,232 characters and molecular time-calibrated information. Our results demonstrate significant variability in evolutionary rates during the early Cambrian, particularly within the Protostomia clade. These morphological dynamics shed light on the mechanisms of evolutionary innovation and suggest complex patterns of evolutionary change that challenge traditional models. Our research enhances the understanding of early animal evolution by highlighting the intricate and variable nature of evolutionary rates during this pivotal period.

Utilising long-read sequencing approaches to detect structural variants of clinical significance in childhood cancer

Steven Hair
Newcastle University

Steven Hair, Joe Fenwick, Eleanor G. Woodward, Richard Yim, Phoebe Snow, Letizia Marchetti, Lisa J. Russell, Sarra L. Ryan.
Wolfson Childhood Cancer Research Centre, Newcastle University, UK.

Background
Many childhood cancers are characterised by genetic abnormalities of clinical significance, including large structural variants (SVs). Many SVs are challenging to detect due to the variable repetitive sequences in which they are located. Long-read sequencing approaches allow for targeted sequencing and methylation profiling of difficult genomic regions such as the immunoglobulin loci, facilitating the detection of SVs and removing the need for more costly whole-genome sequencing.

Methods
Using a novel adaptive sampling (AS) ultra-long-sequencing approach we characterised and detected risk-defining SVs in seven cancer samples with known abnormalities for the first time. Samples were sequenced using the ONT ultra-long sequencing kit (ULK-114) with AS. The targeted region was a list of genes that have potential to cause risk-stratifying abnormalities in cancer. Data was processed through a custom bioinformatics pipeline, with alignment to T2T-CHM13v2.0 genome using Minimap2, SV calling using Sniffles2 and copy number variation calling using QDNAseq.

Results
The samples had an average library N50 of 57kb (44kb-75kb). SV calling identified an average of 2747 SVs (1287-7232) per sample. The following known risk-stratifying genetic abnormalities were detected in each respective sample: BCR::ABL1, IGH::MYC, IGH::DUX4, IGH::CCND1, IGL::MYC, iAMP21 and high hyperdiploidy. The breakpoints of translocations were resolved to base pair level in all samples and methylation profiling identified hypomethylation of DNA across translocated oncogenes.

Conclusions
This study demonstrates the potential of targeted ultra-long-sequencing as a tool for detecting and characterising complex SVs in cancer. The samples sequenced here represent challenging to characterise SVs, mostly involving the immunoglobulin loci that can be difficult to detect by standard testing. This study successfully detected all relevant SVs, along with additional variants that require further investigation.

An integrated whole-genome sequencing and system biology approaches to predict antimicrobial resistance in the virulent bacterial strain of the Staphylococcus aureus

Md Imran Hasan
Texas A&M University San Antonio

Background
Staphylococcus aureus (S. aureus) is a significant pathogen in both community and hospital settings, causing substantial morbidity and mortality. This bacterium produces a vast array of secreted factors that contribute to its capacity to colonize various hosts, including mammals, humans, wildlife, and fish. The emergence of antibiotic-resistant strains of S. aureus has become a major public health concern.

Methods
To better understand the mechanisms underlying antibiotic resistance in S. aureus, we investigated 851 strains from the NCBI-Genome database. We analyzed the interaction network comprising 378 antimicrobial resistance (AMR) genes. Hub genes and functional enrichment analyses were assessed using AMR gene interaction networks. Furthermore, molecular docking was performed on the identified hub genes to validate their potential as therapeutic targets.

Results
Our analysis revealed that a majority of the genes in the network are involved in antibiotic inactivation, antibiotic efflux pumps, and resistance to common antibiotics such as tetracycline, beta-lactams, and fluoroquinolones. Notably, the genes clfA, fnbA, sdrC, and hly exhibited the highest frequency of relevant interactors in the network, identifying them as hub nodes. Molecular docking of these hub genes suggested their potential role in developing novel medications.

Discussions
Our findings provide valuable insights into the antimicrobial resistance mechanisms in S. aureus. The identification of key hub genes and their interaction networks enhances our understanding of the resistance pathways. Molecular docking further supports the potential of clfA, fnbA, sdrC, and hly as therapeutic targets. These insights can inform the development of new drugs to combat antibiotic-resistant S. aureus strains, addressing a critical and evolving public health challenge.

Development of a PCR-Tiling Whole Genome Sequencing Approach for Noroviruses in Wastewater

Richard Hill
Centre for Environment, Fisheries and Aquaculture Science (Cefas)

Authors: Scott, G., Ryder, D., Batista, F. and Hill, R.
Centre for Environment Fisheries and Aquaculture Science, Weymouth, Dorset, UK

Noroviruses (NoVs) are the leading cause of non-bacterial gastroenteritis and place considerable economic burden on society worldwide. Wastewater-based epidemiology (WBE) allows high-resolution investigation into genetic diversity of NoVs without mass clinical testing. Studies of NoVs in wastewater have focused on metabarcoding and sequencing of partial genomes, or target enrichment for metagenomics. However sensitivity of metagenomic approaches poses challenges as sample degradation and low viral loads hinder resolution of whole genomes. To resolve this a PCR-tiling approach for Nanopore whole genome sequencing of NoVs in wastewater was trialled. Primer assays generating both short- (≈400 bp) and long- (≈800 bp) amplicons were developed for four NoV types commonly detected in wastewater: GI.2[P2], GI.6[P6], GII.3[P12] and GII.4[P31]. These were tested with four high-fidelity polymerases and performance measured in amplification specificity and yield, sequencing depth, scheme coverage and single nucleotide polymorphism support. Size-selection methods were also compared in efficacy of reducing non-specific amplification, maximising target yield, cost-effectiveness and feasibility for high-throughput processing.

Amplification with Phusion polymerase and size-selection using SPRIselect and ExoSAP-IT™ demonstrated best performance. For genogroups GII.3[P12] and GII.4[P31] 100 % scheme coverage and >97 % reference genome coverage was achieved. However only >85 % reference coverage for GI.2[P2] and GI.6[P11] indicates further development and optimisation of the assay and bioinformatic pipelines is required before the method can be implemented into a quality assured WBE programme.
This study has shown potential for a PCR-tiling approach to study NoV diversity in matrices such as wastewater containing low and/or degraded viral loads. Utilising WBE could enable the discovery of novel types and improve NoV tracking throughout the population, aquaculture and recreational water settings.

Computationally Sorting Immune Cells From Single-Cell Epigenomics Data

Tayyibe Kansoy
University of Warwick

Cancer develops when the immune system fails to recognise cancer cells, or fails to act upon recognition of cancer cells. The immune response depends on sequential changes in the state of immune cells, such as activated state or exhausted state, for example. Single-cell sequencing of transcriptomes and/or epigenomes enables the analysis of all cell types in a tissue sample in a single assay at genomic scale. In order to understand the nature of the immune cell activation process and possible intermediate states, the first important step is to identify the immune cell types present in a tumour. However, detecting the state of the immune system is not simple because of the complexity of the immune system with its many cell types. In this research, the main aim is to identify the state of immune cells using single cell ATAC-seq data from blood as well as tumour samples. We devise a computational cell sorting approach for the recognition of immune cell types. This approach is relevant both to understanding the pathogenesis of cancer but also to improve understanding of how immune cells influence therapeutic outcome.

Evaluating genotyping methods and the need for advancing Cryptosporidium genomics

Deborah Blessing Oladele
Aberystwyth University

Cryptosporidium, a protozoan parasite of significant zoonotic importance, causes gastroenteritis in a wide range of hosts. Research has demonstrated varying host specificity among Cryptosporidium species, as well as differences in pathogenicity and infectivity within the same species. Genetic analyses have advanced our understanding of Cryptosporidium species identification, subtypes, and overall genomic structure. Nonetheless, substantial gaps remain in our knowledge, particularly regarding the disease dynamics influenced by the parasite's high genomic recombination capabilities. There is still no standardised, globally adopted multilocus genotyping scheme, and an over-reliance on sequencing a single gene, gp60. We have reassessed the accuracy of existing genotyping information in the NCBI repositories, as well as the effectiveness of genotyping methods used to obtain this information. Our analysis involved reviewing relevant recent papers (up to five years) containing new Cryptosporidium parvum and Cryptosporidium hominis gp60 sequences and genotypes; the most common causes of human cryptosporidiosis. We also analysed all 5523 Cryptosporidium gp60 sequences submitted to GenBank, using a script that queried the records for subtype information. A total of 1064 (19%) gp60 sequences were not allocated to a subtype by their authors. Other entries (~28%) were partially allocated, to a subtype family, not the full subtype. Additionally, multiple sequence alignments of these sequences have uncovered discrepancies in naming subtypes and their sequences. Our results highlight the need for a standardized, state-of-the-art nomenclature. For future work, we aim to further compare gp60 sequencing, Multiple-Locus Variable-Number Tandem Analysis (MLVA), and the more comprehensive but expensive Whole Genome Sequencing (WGS) for application to outbreak investigations and transmission dynamics.

Deborah Oladele and Dr Martin Swain
Department of Life Sciences, Aberystwyth University, Penglais, Aberystwyth, Ceredigion, SY23 3DA

Prof Rachel Chalmers and Dr Guy Robinson
Cryptosporidium Reference Unit, Public Health Wales Microbiology, Singleton Hospital Sgeti, Swansea, SA2 8QA

Funding acknowledgments: OneZoo CDT (BBSRC, MRC, NERC)

Decoding the mitogenome of Argopecten ventricosus (Bivalvia: Pectinidae)

Maria Isabel Ruiz-Ruiz
Universidad Peruana Cayetano Heredia

Maria Isabel Ruiz Ruiz[1], Edson Soto[1,2], Camila Zamora[1,2], Pedro Romero[2]
[1]Laboratory of Marine Biology, Faculty of Science and Philosophy, Peruvian University Cayetano Heredia, 15102, Lima, Peru
[2]Faculty of Biological Sciences, National University of San Marcos, 15081, Lima, Peru

Background: Argopecten ventricosus (G. S. Sowerby II, 1842) is a marine scallop that occurs in the Tropical Eastern Pacific. The oldest fossil records of A. ventricosus date back to the middle Miocene (14-11 Ma) in Cabo Falcón, Venezuela. Other records suggest its presence in Ecuador and California since the Pliocene (5.3 - 2.6 Ma). However, a thorough taxonomic revision is required. Molecular phylogenies calibrated with fossils can clarify the divergence time of A. ventricosus. Thus, we characterized the first Peruvian mitogenome of A. ventricosus obtained from whole-genome sequencing.

Methods: Mitogenome assembly was conducted using GetOrganelle, and annotation was done with the MITOS Web Server. Phylogenetic reconstruction analysis was performed in BEAST 2.0 using concatenated protein-coding genes and rRNA genes with fossil calibration.

Results: The mitogenome of A. ventricosus spans 16,078 bp. The mitochondrial genome contains 13 protein-coding genes (PCGs), 23 tRNAs, and 2 ribosomal RNA (rRNA) genes, including the atp8 gene, the first official report of this gene for the Argopecten genus. Gene composition and arrangement are relatively conserved compared to A. irradians irradians and A. purpuratus, except for the position of the control region.

Discussion: The most recent common ancestor of A. ventricosus occurred around 4.5 Ma (HPD: 9.43 Ma - 1.79 Ma), prior to the definitive closure of the Panamanian Isthmus (~3 Ma). During the Mio-Pliocene (7-5 Ma) transition, the Argopecten fossil record increased in the Caribbean Sea, suggesting possible favorable conditions for its divergence. Collaborative research in the Tropical Eastern Pacific region is necessary to clarify the divergence time of A. ventricosus, consequently enhancing the comprehension of its evolutionary history.

De novo assembly and annotation of the Patagonian toothfish (Dissostichus eleginoides) genome

David Ryder
Cefas

Patagonian toothfish (Dissostichus eleginoides) is an economically and ecologically important fish species in the family Nototheniidae. Juveniles occupy progressively deeper waters as they mature, and adults have been caught as deep as 2500 m, living on the southern shelves and slopes around the sub-Antarctic islands of the Southern Ocean. As apex predators, they are a key part of the food web, feeding on a variety of prey, including krill, squid, and other fish. Despite its importance, genomic sequence data, which could be used for more accurate dating of the divergence between Patagonian and Antarctic toothfish, for example, has so far been limited.

A high-quality D. eleginoides genome was generated using a combination of Illumina, PacBio and Omni-C sequencing. To aid the genome annotation, the transcriptome derived from a variety of toothfish tissues was generated using both short and long read sequencing methods. The final genome assembly was 797.8 Mb with a N50 scaffold length of 3.5 Mb. Approximately 31.7% of the genome consisted of repetitive elements. A total of 35,543 putative protein-coding regions were identified, of which 50% were functionally annotated. Transcriptomics analysis showed that approximately 64% of the predicted genes (22,617 genes) were expressed in the tissues sampled. Comparative genomics analysis revealed that the anti-freeze glycoprotein (AFGP) locus of D. eleginoides does not contain AFGP proteins, in contrast to the same locus in Antarctic toothfish (Dissostichus mawsoni). This corroborates previously published results looking at hybridization signals and confirms that Patagonian toothfish do not possess any trace of the AFGP genes..

We have published the Patagonian toothfish genome, which will provide a valuable genetic resource for ecological and evolutionary studies..

Any tissue samples used as part of this study were collected by Argos Froyanes from dead fish which had already been caught as part of their commercial fishing operations.

Thoughts and perspectives of metagenome sequencing as a diagnostic tool for infectious disease: an interpretive phenomenological study

Hannah Trivett
University of Liverpool

Effective infectious disease diagnostics (IDD) are vital for informing clinical decision-making regarding treating and managing gastrointestinal infections. Conventional clinical methods rely upon culture-dependent techniques, with little shift in accepting and integrating culture-independent sequencing methods into routine clinical IDD.

The project utilised semi-structured interviews to study stakeholder experiences within clinical settings and those at the forefront of microbial genomics. The interviews explored the decision-making process for deciding which diagnostics to use for identifying bacteria-causing infections and further identified drivers of developing and implementing metagenome sequencing and the current challenges of its implementation as an IDD. The data were analysed using thematic analysis, and an Interpretive Phenomenological approach was used throughout.

10 stakeholders were interviewed between July 2021 and October 2021, including Clinical scientists, consultants, and professors in academia. Five themes emerged: Diagnostic Choice, Barriers to implementation, open access and data sharing, COVID-19, and communication. Participants recognised the need for new diagnostics to be implemented to overcome the limitations of current diagnostic approaches but highlighted the barriers to integrating new diagnostics into clinical settings. However, participants felt that lessons could be learnt from using metagenomics in COVID-19 and how other diagnostic platforms have been integrated clinically over the last 20 years.

The study provided evidence for the knowledge gap in current literature and practice regarding developing and implementing metagenome sequencing as a potential infectious disease diagnostic. Knowledge of new genomic diagnostic testing is disproportionate around the UK, impacting the drive to integrate metagenome sequencing as an IDD. Access to new diagnostics must be increased to improve patient treatment and positively impact public health.

Authors:
H. Trivett (1,2), O. Oyebode (3), A.C. Darby (1,2)

1) Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool
2) Health Protection Research Unit in Gastrointestinal Infections (NIHR)
3) Wolfson Institute of Population Health, Queen Mary University of London

Genomic insights into clonal diversity in UK populations of the Potato aphid, Macrosiphum euphorbiae

Mark Whitehead
University of Liverpool

The potato aphid Macrosiphum euphorbiae is one of many polyphagous crop pests involved in the transmission of insect-vectored pathogens. While their North American counterparts reproduce via cyclical parthenogenesis, UK populations of M. euphorbiae appear to persist asexually, resulting in the maintenance of several genotypes, with some demonstrating genotype-specific traits; this includes innate resistance to parasitism from the hymenopterous parasitoid wasp Aphidius ervi. The genetic and molecular basis for genotype-specific traits is often unknown. Here we present a chromosome scale assembly for a parasitoid-resistant clonal line of M. euphorbiae and provide insights into the genotypic composition and distribution of UK potato aphid populations using microsatellite and whole-genome sequencing (WGS) techniques, focusing on geographically separated potato crops within two distinct areas of the UK (Merseyside and Tayside). We show that the genome consists of five chromosomal blocks, has a total size of 560 Mbp and is highly complete based on BUSCO (C: 95.6%). The sampled potato aphid populations were dominated by two genotypes, one of which is absent from commercial farm settings, suggesting either an intolerance to farming practices, such as insecticide use, or a broader host range. We suggest some putative gene functions using WGS data to explain the observed frequency of aphid genotypes. WGS data highlighted the asexual clonal lifestyle of M. euphorbiae genotypes in the UK, resolving individuals to a higher resolution than using microsatellite data. The work presented here will provide useful information for integrated pest management of potato aphids, elaborating on the relationship between genotype diversity and functional traits such as parasitism and insecticide resistance, and host plant use, as well as providing more resources for further comparative genomics studies within the Aphididae.

Authors:
Mark Whitehead (1), Alison Karley (2), Alistair Darby (1)

1) University of Liverpool, Liverpool, UK, L69 7ZB
2) James Hutton Institute, Dundee, UK, DD2 5DA

Investigating Genetic Variation in a Segmental Duplication Region Implicated in Developmental Disorders

Eleanor Woodward
Newcastle University

Authors:
Eleanor G. Woodward (1), Richard Yim (1), Steven C. Clifford (1) , Sarra L. Ryan (1)

1) Newcastle University Translational and Clinical Research Institute, Newcastle upon Tyne, UK

Segmental duplications (SDs) are reservoirs for human-specific genes and were pivotal in the evolution of humans from apes. Recurrent rearrangements of SDs are implicated in human disease and developmental disorders. SD length (>1kb), sequence identity (>90%) and structural diversity have hindered efforts to characterise these regions. Human SD haplotypes are complex meaning that next-generation sequencing techniques result in insufficient read lengths to resolve structural diversity and differences between alleles. This study aims to characterise a region of SDs located on chromosome 17, where structural variants contribute to Smith-Magenis Syndrome (SMS) and Potolocki-Lupski syndrome (PTLS).

The development of complete human reference genomes such as T2T-CHM13 and those of the Human Pangenome Reference Consortium provide valuable resources to delineate landscapes of SDs. SDs of the chromosome 17 region were identified in GRCh38 and T2T-CHM13 using the tool SEDEF. By aligning SD sequences to human reference genomes (N = 47) with minimap2, we observed variable SD copy number (2-8 copies) and orientation among healthy individuals. Overall, 20 different SD haplotypes were observed.

Notably, haplotypes were not significantly associated with an individual's sex or ethnic group ($p > 0.05$, paired t-test). Investigation of DNA methylation profiles across haplotypes suggests the presence of haplotype-specific features of gene regulation.

These preliminary findings provide the bases for a more comprehensive understanding of this dynamic region of the genome. The study contributes to the elucidation of mechanisms underlying recurrent rearrangements of SDs associated with human disease.

Comparing common strategies for ortholog selection used in phylogenomics

Mingzhu Yang
University of Bristol

Authors:
Mingzhu Yang, Mattia Giacomelli, Maria Eleonora Rossi, Jesus Lozano Fernandez, Phil Donoghue & Davide Pisani

University of Bristol

Abstract:
Phylogenomic studies rely on large superalignments including hundreds or thousands of concatenated orthologous gene families. A diversity of software has been developed to identify such orthologs from sets of proteomes. A strategy that has recently grown in popularity is extracting single-copy genes pre-identified in the BUSCO database from the species of interest (Seppey et al., 2019). A second commonly used strategy is using OrthoFinder (Emms et al., 2019) to define de novo orthogroups from which single gene families are then identified and extracted. Other similar strategies can be designed using different approaches (Li et al., 2003) to define the orthogroups (e.g. OrthoMCL,). One frequently overlooked problem in phylogenomics is whether there is correspondence between orthogroups inferred using different methods. While the expectation is that if these pipelines perform well, they should identify the same orthogroups, no evidence exists that this is the case, in fact available evidence suggests the opposite (Li et al., 2023).

Here we identified 954 single gene families using BUSCO for a set of nine high-quality metazoan proteomes and three outgroups. For the same genomes, OrthoFinder identified 2937 single-gene families, 365 of which were present in all 12 taxa. We functionally annotated these families and looked at the pattern of correspondence between families inferred using BUSCO and OrthoFinder. We found that the BUSCO metazoan set is depleted in translational genes (KOG functional category J). Furthermore, there is no clear correspondence between orthogroups inferred using BUSCO and Orthofinder. We conclude that the use of alternative software for ortholog detection can potentially lead to the inference of different phylogenies, and it is thus crucial to better understand which of these approaches is more reliable.

Protocol optimisation to isolate exosomes and PBMC-derived RNA as the source of biomarkers for pancreatic cancer early detection

Nurul Sarifah
University College London

Full author list and affiliations:
1. Nurul Sarifah (Institute for Liver and Digestive Health, UCL)
2. Pilar Acedo (Institute for Liver and Digestive Health, UCL)
3. Kyra Fraser (Institute for Liver and Digestive Health, UCL)

Abstract:
Pancreatic cancer (PanCa) is an often-lethal malignancy, demanding improved early detection and reduced mortality through novel biomarkers. Exosomes and peripheral blood mononuclear cell (PBMC)-derived RNA are promising sources for PanCa biomarkers. While differential ultracentrifugation and commercial RNA extraction kits are commonly used to isolate exosomes and PBMC-derived RNA, respectively, they may yield differing qualities, impacting reliable discoveries. Thus, this study aimed to develop and optimise the isolation protocol based on these techniques. Exosomes from PanCa PANC-1 cells were isolated using an optimised differential ultracentrifugation technique and characterised via western blotting, nanoparticle tracking analysis (NTA), immunofluorescence, and flow cytometry, confirming pure and high-yield exosomes. PBMC-derived RNA from PanCa patients was isolated using a modified commercial RNA extraction kit, with NanoDrop spectrophotometry confirming high concentration and acceptable purity. To conclude, our protocol is feasible for obtaining high-quality exosomes and PBMC-derived RNA, paving the way for potential PanCa biomarker discoveries.

Matched fresh frozen and FFPE patient tissues reveal the enhanced sensitivity and data quality of a novel DNA library prep method

Maximilian Fritsch
New England Biolabs UK Ltd.

Margaret R. Heider (1), Adam Peltan (2), Maximilian J. Fritsch (2), Jian Sun (1), Adrian Reich (1), Brittany S. Sexton (1), Laura Blum (1), Bradley W. Langhorst (1), and Pingfang Liu (1)

1) New England Biolabs Inc., Ipswich, MA 01938, USA
2) New England Biolabs UK Ltd., Hitchin, SG4 0TY, UK

FFPE DNA poses many challenges for preparing NGS libraries, including low input amounts and highly variable damage from fixation, storage, and extraction. It is difficult to obtain libraries with sufficient coverage and sequencing artifacts arising from damaged DNA bases confound somatic variant detection.

We have developed a novel NGS library prep method compatible with both high quality and very low quality FFPE DNA samples, employing a novel enzymatic DNA repair mix, enzymatic fragmentation mix, and PCR master mix. This workflow was validated on real patient samples, using DNA extracted from matched tumor and normal tissue of various tissue types preserved by both fresh frozen and FFPE (DIN 1.5 – 6.8).

This new workflow reduced the false positive rate in somatic variant detection by repairing damage-derived mutations in FFPE DNA samples but also improved the library yield, library quality metrics (including mapping, chimeras, and properly paired reads), complexity, coverage depth, and hybrid capture library quality metrics. Comparing the variant calls from matched FFPE and frozen tissues revealed an improved sensitivity and accuracy of variant calling compared to mechanical shearing and other enzymatic fragmentation library prep approaches.

This new suite of enzyme mixes improves the overall library prep success rate from challenging FFPE samples, allowing even highly damaged FFPE samples to achieve high quality libraries with a greater sensitivity for somatic variant identification and coverage for CNV analysis. The workflow is robust and flexible, compatible with both FFPE DNA and matched high quality DNA samples as well as automation-friendly for convenience in sample processing.

This research was supported by New England Biolabs, Inc. - M.R.H, J.S., A.R., B.S.S., L.B., B.W.L., and P.L. performed the work as employees of New England Biolabs, Inc.

Evaluation and validation of metagenomics approaches for samples from non-sterile sites for diagnosis and surveillance

Luz Marina Martin-Bernal
University College London

Authors: Luz Marina Martin Bernal (1), Rachel Williams (1), Oscar Torres (2), Sergi Castellano (1), Judith Breuer (2).

1) Genetics & Genomic Medicine Department, UCL Genomics.
2) Infection, Immunity & Inflammation Department, UCL.

Clinical metagenomics (CMg) is a powerful genomic approach that enables comprehensive analysis of microbial populations in clinical samples. It surpasses conventional diagnostics by efficiently identifying known, novel, rare, and unexpected pathogens. The application of CMg in diagnostics, surveillance, and antimicrobial resistance (AMR) and virulence factors discovery are crucial for early detection of infectious diseases. Remarkable advancements have been made using CMg to detect pathogens in samples from sterile body sites, as shown by Great Ormond Street Hospital's clinical metagenomics service for brain tissue and CSF. We are now expanding this approach to samples from non-sterile sites, such as respiratory samples. Respiratory infections are common and highly transmissible, requiring early detection to prevent outbreaks. CMg could provide definitive diagnoses for these infections, but its routine clinical implementation faces challenges such as data quality variation, complex data analysis and high costs. Further, standardising CMg for respiratory infections requires defining clinical sample types, turnaround times, and notifiable pathogens, which impact the strategies used for the entire CMg pipeline from nucleic acid extraction to analysis. Our project aims to develop and standardise CMg workflows for non-sterile respiratory samples, starting with upper-respiratory tract samples like nasopharyngeal swabs. We will evaluate and compare existing methodologies, including long-read (Oxford Nanopore) and short-read (Illumina) sequencing approaches. The focus will be on nucleic acid extraction, sample preparation, sequencing, and rapid data analysis, aiming to create a comprehensive, rapid, and cost-effective CMg pipeline.

A Backbone Tree of Formicidae —— The First Step to Reveal Ant Evolution in the Cretaceous Terrestrial Revolution and Post K-Pg Extinction Era

Wentao Tao
University of Bristol

1. Wentao Tao, School of Biological Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK.
2. Davide Pisani, School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK. School of Biological Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK.
3. Philip Donoghue, School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK
4. Donato A. Grasso, Department of Chemistry, Life Sciences & Environmental Sustainability, University of Parma, 43124 Parma, Italy
5. Enrico Schifani, Department of Chemistry, Life Sciences & Environmental Sustainability, University of Parma, 43124 Parma, Italy

With over 15,000 species, colonising nearly all terrestrial environments, ants are one of the most successful insects on the earth. However, despite genome-scale data have been used to explore their internal relationship, there are still some inconsistencies between different researches. Mysteries, such as the relationship of the subfamilies Leptanillinae and Martialinae with other living ants, as well as the root age and divergence time of internal clades, are still waiting to be revealed. Here, we build up the largest genome-scale dataset, containing 91 ant species and 10 outgroups. After going through a series of cleaning pipeline to remove hidden paralogs and long branch species, we build up trees under the best fitting mixture model, then calibrate them with 16 fossils following the best practice of fossil calibration so as to investigate the evolutionary history of ants. Our phylogenetic analysis supports the hypothesis that the leptanillomorph clade (Leptanillinae and Martialinae) is the sister group to all other living ants. While the novel molecular clock analyses suggest that crown ants originated in the early-Cretaceous and diversified afterward with a younger age compared to the previous research. Therefore, in the future, by adding more molecular and fossil data, combining with the latest tree-building and fossil-calibrating method, we can better know about (1) the diversification of crown ants in the late-Cretaceous under the background of the angiosperm's explosion, and (2) how they occupied the niches previously belonging to stem group ants after the K-Pg event, finally achieved success today.

A Multi-omics Approach Reveals the Gut Resistome and Microbiome of Yellow Perch in Minnesota Lakes

Omar Jimenez Lopez
University of Minnesota

Authors: Omar Jimenez-Lopez (1), Tui Ray (1), Christopher Dean (1), Ilya Slizovskiy (1), Jessica Deere (1), Tiffany Wolf (1), Seth Moore (3), Alexander Primus (1), Jennifer Høy-Petersen (2), Silje Finstad (2), Jakob Mo (2), Henning Sorum (2), Noelle Noyes (1)

University of Minnesota (1), Grand Portage Band of Lake Superior Chippewa (3), Norwegian University of Life Sciences (2)

Wastewater of anthropogenic origin can significantly impact wildlife in natural water bodies, affecting the health, behavior, and reproduction of wild fish like yellow perch. These fish host various microorganisms, including bacteria that can develop antimicrobial resistance. Commensal bacteria can transfer resistance genes to pathogens, increasing health risks as infections become harder to treat.

This study aimed to analyze the gut microbiome and resistome (antimicrobial resistance gene profile) of yellow perch from two types of lakes under anthropogenic pressure: wastewater effluent-impacted lakes and undeveloped lakes. Yellow perch were sampled from five lakes in Minnesota, USA—three wastewater effluent-impacted and two undeveloped. DNA and RNA from these samples were extracted and sequenced for analysis.

Results indicate that the gut resistome and microbiome of yellow perch differ between lakes, likely due to varying anthropogenic pressure. The resistome predominantly consisted of macrolide resistance genes, particularly the MLS23S group, making up 53% of resistome sequences from effluent-impacted lakes and 73% from undeveloped lakes. The microbiome was dominated by the phyla Proteobacteria, Firmicutes, and Actinobacteria, with opportunistic pathogens Plesiomonas shigelloides and Aeromonas veronii more abundant in effluent-impacted lakes.

Ongoing construction of metagenome-assembled genomes (MAGs) may reveal novel components of the yellow perch microbiome and resistome. Metagenomic analysis of wild fish samples offers valuable insights into the effects of anthropogenic pressures on microbial communities in water bodies.

Investigating molecular response to therapy in high-grade serous ovarian cancer using long-read Nanopore sequencing

Justina Pangonyte
Cancer Research UK Cambridge Institute, University of Cambridge

Authors: Justina Pangonyte*, Emilie Trentesaux*, Thomas Bradley*, Gabriel Funingana, Maria Vias, James D. Brenton *First co-authors equal contribution
Affiliations: Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge, United Kingdom
Ethics approval: OV04 study approved by the local research ethics committee at Addenbrooke's Hospital, Cambridge, UK, (REC reference numbers: 07/Q0106/63; and NRES Committee East of England – Cambridge Central 03/018)

Background:
High-grade serous ovarian cancer (HGSOC) is a major cause of female cancer death, characterized by extreme chromosomal instability (CIN) and tumour heterogeneity. We have previously shown that copy number signatures and common oncogenic chromosomal aberrations remain stable between diagnosis and relapse, suggesting subtle structural variants or epigenetic changes may underly resistance mechanisms. Here, we utilise Oxford Nanopore (ONT) long-read sequencing and RNAseq to explore the impact of chemotherapy on large-scale structural variation (SV) and methylation.

Methods:
Matched ascites spheroid samples from five HGSOC patients were selected, capturing both pre- and post-recurrence states. Extraction was performed using the Qiagen AllPrep DNA/RNA Mini Kit. Libraries were prepared using an optimised ONT Ligation Sequencing V14 protocol on a PromethION P24 with R10.4.1 flow cells, targeting 40× coverage for tumour and 20× for germline samples over 72-hours. Data was analysed using the EPI2ME Labs Nextflow wf-somatic-variation and wf-human-variation workflows.

Results:
Various extraction methods for processing clinical ascites tumours were tested. The Qiagen AllPrep DNA/RNA Mini kit produced high purity DNA with a final yield of approximately 19ng and an N50 of 20kb. Additional purification using the NEB gDNA Purification kit, where necessary, improved DNA purity and sequencing performance without significantly impacting N50. Preliminary analysis from paired samples shows differential methylation and new SV at several gain/loss of function loci and Chr 17q. We are currently optimizing these analyses and will present detailed results on the poster.

Discussions:
We have developed robust methods for obtaining high-quality long read sequencing data from ascites spheroids enabling facile interrogation of the evolving genomic and epigenomic landscapes of HGSOC patients during treatment. ONT sequencing has strong potential for demonstrating SV and methylation differences from sequential clinical samples in complex CIN cancer genomes.