# QMB Abstracts Genomics satellite

## G1: Beyond one T2T human genome: What's next?

Arang Rhie[1]

[1] Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

In 2022, the Telomere-to-Telomere (T2T) consortium released the first complete human genome from a human cell line, CHM13hTERT (CHM13)[1]. Having been derived from a complete hydatidiform mole, CHM13 has a 46,XX karyotype but is almost entirely homozygous. This simplified assembly of its genome, but prevented assembly of a Y chromosome. To add the missing Y sequence, we chose another human cell line, HG002, to assemble the X and Y chromosomes. HG002 has been well characterized for its genome and is widely used as a benchmark dataset by the Genome in a Bottle consortium. We combined the highly curated and annotated Y chromosome with the prior T2T-CHM13 assembly to form a new, complete reference for all human chromosomes, T2T-CHM13v2[2].
In my talk, I will present the latest resources released for utilizing T2T-CHM13v2 as a reference, discussing areas where T2T genomes enable new discoveries. Current best practices for sequencing and generating automated diploid T2T genome assemblies will be discussed, along with the remaining challenges. At the end, current plans and efforts in human and non-human T2T projects will be shared, concluding with opportunities in the forthcoming T2T pan-genome era.

1. Nurk S, Koren S, Rhie A, Rautiainen M, et al. (2022) *The complete sequence of a human genome*. Science.
2. Rhie A, Nurk S, Cechova M, Hoyt SJ, Taylor DJ, et al. (2022) *The complete sequence of a human Y chromosome*. bioRxiv.

## G2: Detection and Mitigation of Microbiome DNA Presence in Saliva-Derived Whole Genome Sequence Data

Halliday, B.J.[1], Markie, D.[2], Goodin, E.[1], Black, M.[3], Sporle, A.[4,5], Watson, H.[6], Wilcox, P.[7], Robertson, S.P.[1]

[1]Department of Women's and Children's Health, University of Otago, Dunedin, NZ, [2]Department of Pathology, University of Otago, Dunedin, NZ, [3]Department of Biochemistry, University of Otago, Dunedin, NZ, [4]iNZight Analytics Ltd, Auckland, NZ, [5]Department of Statistics, University of Auckland, Auckland, NZ, [6]Ngāti Porou Hauora, Te Puia Springs, NZ, [7]Department of Mathematics and Statistics, University of Otago, Dunedin, NZ.

DNA extracted from saliva is suitable for high-throughput sequencing studies and crucial in situations where taking blood samples is not practical or appropriate. However, using saliva samples risks inclusion of the oral microbiome, which may compromise downstream applications of the sequencing data. As part of the Genomics Aotearoa Variome project, a subset of anonymised whole-genome sequence data derived from either participants' saliva (*n*=263) or blood (*n*=9) samples were assessed. The Variome project aims to assist genomic diagnostics in Aotearoa by generating a population specific genomic catalogue to more accurately represent allele diversity in Māori.

Preliminary cohort screening revealed oral microbiome presence in saliva samples, with a median of 8.5% non-human reads (ranging between 2.2-67.9%) in participant sequences. Misalignment of bacterial sequence to the human reference genome was observed in affected samples, inducing false positive variant calls and other quality anomalies. The presence of bacterial derived variation presents a considerable risk to the goals of the Variome project, as it appears novel when compared to other human allele frequency databases and occurs consistently across affected samples.

To ensure a robust and representative Variome resource, a mitigation strategy was developed using the taxonomic classification tool Kraken2 to partition sequence data and quarantine of non-human sequences. Consistent with expectation, data derived from blood samples were found to have at least 98% of sequences classified as human. For saliva samples, non-human sequences were found to originate primarily from oral microbiome bacteria, showing taxonomic consistency between affected samples.

Classification of sequences using Kraken2 and subsequent alignment of human and unclassified reads led to a median decrease in erroneous misalignment of 66%. This method is theoretically portable to any target species for the mitigation of non-target species contamination. However, the additional computation required for sequence classification prior to alignment may be impractical for some sequencing studies.

## G3: MapQsee—Improving Visualisation of Sequence Read Mapping

Astra Heywood[1], Tim Millar[1], Usman Rashid[2], Emily Koot[3], Andrew Catanach[1], Chen Wu[2], Cecilia Deng[2] & Susan Thomson[1]

[1]The New Zealand Institute for Plant and Food Research Limited, Lincoln, New Zealand
[2]The New Zealand Institute for Plant and Food Research Limited, Auckland, New Zealand
3 The New Zealand Institute for Plant and Food Research Limited, Palmerston North, New Zealand

Many plant families have polyploid genomes (*Asteraceae, Actinidiaceae, Convolvulaceae, Ericaceae, Poaceae, Rosaceae, Rubiaceae, Solanaceae*) that represent economically important food, fodder, and ornamental crops. In New Zealand, we have a special interest in *Actinidia* (Kiwifruit) where genetic analysis outcomes are used for important breeding decisions that may have great financial consequences. *Actinidia* species' genomes can range from diploid (2n) to octoploid (8n), with the majority of commercially available kiwifruit being either tetraploid (4n) or hexaploid (6n). Downstream analyses of genetic reads such as population dynamics and phenotype/genotype correlations rely on accurate reference mapping and variant calling. However, existing toolkits are not optimised for polyploid genomes so features such as allelic patterns of inheritance can be lost during processing. Here we present MapQsee, an app that visualises the quality of reference mapping that can be used to: 1. Analyse the quality and depth of read mapping from sequence data, from whole genome sequencing down to targeted marker site quality; 2. Visualise relationships between a population of samples, such as experimental impact; 3. Generate matrices and bed files that identify reliable targeted regions (marker loci) that may be used in downstream analysis. We demonstrate a use-case of MapQsee by exploring an example set of genetic markers in kiwifruit that are used in selection decisions. We also highlight that read-quality assessment with MapQsee has uses in downstream applications such as polyploid identification and allele calling.

# G4: COVID-19 risk variant associations with chromatin remodelling, DNA maintenance and surfactant genes are infection dependent in the lung

Jaros, R.K.[1], Golovina, E.[1], O'Sullivan, J.M.[1,2,3,4,5*]

[1]The Liggins Institute, The University of Auckland, Auckland 1023, New Zealand, [2]Maurice Wilkins Centre for Molecular Biodiscovery, The University of Auckland, Auckland 1010, New Zealand, [3]MRC Lifecourse Epidemiology Unit, University of Southampton, United Kingdom, [4]Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore, [5]Australian Parkinson's Mission, Garvan Institute of Medical Research, Sydney, New South Wales, Australia

During viral infection the structure of host chromatin is modified. It is generally assumed that these chromatin modifications will affect variant-gene mapping, and therefore gene expression. What is not clear is how limitations imposed by host germline risk affect the expression changes that occur with infection induced chromatin remodelling. Critically, this lack of information extends to how germline variants associated with severe SARS-CoV-2 impact on tissue-specific gene expression changes in response to infection-induced chromatin conformation changes. Here we combined temporal chromatin conformation data from SARS-CoV-2 stimulated cells with a lung spatial-eQTL gene expression analysis to contextualise the functional effects and contributions of germline risk on the severe phenotypes observed in SARS-CoV-2. We identify changes in lung-specific SARS-CoV-2 risk variant-gene mapping across the infection time course. Our results provide evidence for infection-induced chromatin remodelling that impacts the regulation of genes associated with the severity of SARS-CoV-2 infection. The gene targets we identified are functionally involved in host chromatin modifications and maintenance and the expression of these genes is amplified by SARS-CoV-2-induced epigenetic remodelling. The effect of this remodelling includes transcriptional changes to gene targets such as *SMARCA4*, *NCOR1*, *DNMT1*, *DNMT3a*, *DAXX*, and *PIAS4,* all critical components of epigenetic control mechanisms and SARS-CoV-2 antiviral activity, along with several genes involved in surfactant metabolism. We show how severe-phenotype-associated eQTLs form and break in an infection time-course-dependent manner that mimics positive feedback loops connecting germline variation with the process of viral infection and replication. Our results provide a novel bridge between existing COVID-19 epigenetic research and demonstrate the critical role of epigenomics in understanding SARS-CoV-2-risk-associated gene regulation in the lung.

## G5: AssemblyQC: A reproducible best-practices NextFlow pipeline for assessment of assembly quality

Usman Rashid[1], Chen Wu[1], Jason Shiller[2], Ken Smith[1], Marcus Davy[2], Ross Crowhurst[1], Susan Thomson[3], Cecilia Deng[1,*]

* Correspondence: Cecilia.Deng@plantandfood.co.nz, [1]The New Zealand Institute for Plant and Food Research Limited, Auckland, New Zealand, [2]The New Zealand Institute for Plant and Food Research Limited, Te Puke, New Zealand, [3]The New Zealand Institute for Plant and Food Research Limited, Lincoln, New Zealand

As whole-genome sequencing technology improves and sequencing costs fall, assembling a genome has become a common task in genomics research. This is opening up opportunities to build pangenomes or access multiple reference genomes. The key is in estimating the quality of assemblies as they are being generated, or assessing the quality of the growing assemblies now publicly available for research. Current sequencing technologies and software are still limited, especially for complex genomes. Reconstructing genomes from DNA sequence reads produces errors. Thus a quick, reliable, and easy to use method to assess genome assembly quality is essential. We developed AssemblyQC, a reproducible best-practices NextFlow pipeline that evaluates assembly quality from a wide range of aspects, using open source tools mainly targeting chromosome-level assemblies. AssemblyQC incorporates file validation, contaminant checking, taxonomic classification and completeness checking. The pipeline also includes contact map and synteny analysis. We will present the pipeline for AssemblyQC, showcasing the benefits the pipeline offers for progressing and validating genome assemblies, providing publication-quality plots, and the presentation of results in a standardised HTML report. Overall, AssemblyQC streamlines the evaluation process by providing a standardized and integrated solution for assessing assembly quality, ensuring researchers can make informed decisions based on quantitative results. GitHub: https://github.com/PlantandFoodResearch/assembly_qc

## G6: Evolutionary dynamics of transposable elements in Actinidia species

Ting-Hsuan Chen, Cecilia Deng, Susan Thomson

The New Zealand Institute for Plant and Food Research Limited, New Zealand

Transposable elements (TE) are fundamental genomic components that have been identified in all kingdoms of life. They are a source of mutations that can cause various genetic and epigenetic impacts, such as exotic chromosomal recombination, accumulation of cis-regulatory element for stress response, generation of novel promoter that changes gene expression pattern, dynamic biogenesis of small RNA and long non-coding RNA, as well as rewiring chromosomal 3D structure compartmentation. These lead to temporal or tissue-specific regulation of gene expression, triggering diverse physiological responsiveness, and thus phenotypic variations that facilitate environmental adaptation or benefit breeding programs in plants and animals.

The genus Actinidia comprises important economic crop species with wide range of phenotypic variations. Interspecific comparison of TE landscape in Actinidia species can unveil traits of interest influenced by the evolutionary dynamics of TEs, providing knowledge that can be leveraged for breeding program. Here, we use the package EDTA (extensive de-novo TE annotator) to conduct pan-TE annotation across seven Actinidia species, resulting in a non-redundant TE library representing TEs of these species. The result shows that TEs occupy 40% to 50% of the examined genomes, with over half of the TE sequences derived from long terminal repeat (LTR) retrotransposon. The two major TE subclasses, LTR retrotransposon and terminal inverted repeat (TIR) DNA transposon, show distinct preference in genomic distribution. Over 6700 TE families were found as the core pan-TE families. We identified TE families exhibit various sequence divergence across species, suggesting different TE mobilization activity or evolutionary history in different species. By understanding the TE landscape and its influence on phenotypes, we may apply the knowledge in breeding of climate-resilient cultivars.

## G7: Phylogenetic analysis of single-cell RNA-seq expression data

Gavryushkina, A.[1], Pinkey, H.[2], Diermeier, S.[2], Gavryushkin, A.[1]

[1]School of Mathematics and Statistics, University of Canterbury, Christchurch, NZ, [2]Department of Biochemistry, University of Otago, Dunedin, NZ.

Phylogenetic relationship of cells within an organism can help us to understand how tumours develop in space and time, identify driver mutations and other evolutionary events that enable cancer growth and spread. Numerous studies have reconstructed phylogenies from single-cell DNA-seq data. Here we are looking into the problem of phylogenetic analysis of single-cell RNA-seq expression data, which is a cost-efficient alternative (or complementary) data source that integrates multiple sources of evolutionary information including point mutations, copy-number changes, and epimutations. Recent attempts to use such data although promising raised many methodological challenges. Here we present several computational approaches to these challenges including evolutionary analyses of Visium spatial transcriptomics data and other single-cell RNA-seq data with the aim of enabling a range of phylogenetics and population dynamics approaches such as phylogeographic analysis.

## G8: Nanopore methylation sequencing in a non-model plant species

Liau Y[1], Whibley A[1], Lizamore D[1]

[1]Grapevine Improvement Lab, Bragato Research Institute, Lincoln, NZ

Epigenetic regulation such as DNA methylation is an important feature of plant development and response to environmental stresses. Epigenetic modifications can impact gene expression, providing a route for adaptation to environmental cues that does not change the genetic sequence itself. Grapevine (*Vitis vinifera*) cultivars are typically clonally propagated, which results in the same genetic stock being grown in a diverse range of environments, and without conventional breeding, there are limited possibilities for recombination and reassortment of adaptive alleles. In this scenario, epigenetic diversities may be of key importance in shaping phenotypic variation. Our project aims to characterise and explore the features of whole genome DNA methylation in *Vitis vinifera*, cv. Sauvignon Blanc, a cultivar of value to the New Zealand wine industry.

Traditional methylation sequencing usually involves bisulphite treatment that has detrimental effect on DNA integrity and thus results in generally short DNA sequences. Recent advances in sequencing technology, such as that from Oxford Nanopore Technology (ONT), have enabled direct DNA methylation sequencing with the ability to differentiate the signal of methylated and non-methylated cytosines. This method can decipher DNA methylation status in reads of kilobases in length and has been widely studied in humans, where methylation occurs in the CG context, but less so in plants where methylation occurs in all C contexts.

We explore ONT whole genome DNA methylation on the latest flow cell (R10.4.1) and chemistry (SQK-LSK114) versions on PromethION sequencer. We compare and assess the accuracy of different DNA methylation calling tools including those developed by ONT (dorado, bonito, guppy) and by third parties (Deepmod2, f5c); and investigate the effects of different read depth on DNA methylation detection with this technology. This will provide a technical foundation to explore the contribution of DNA methylation differences in grapevine phenotypic traits.

## G9: Network analysis uncovers gene-regulatory intersections between juvenile arthritis and comorbid traits.

Pudjihartono, N.M.[1], Ho, D.S.W. [1], O'Sullivan, J.M.[1,2,3,4,5]

[1]Liggins Institute, University of Auckland, Auckland, New Zealand, [2]Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand, [3]MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom, [4]Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore, [5]Australian Parkinson's Mission, Garvan Institute of Medical Research, Sydney, NSW, Australia.

Juvenile idiopathic arthritis (JIA) is an inflammatory joint disease with complex genetic aetiology. As JIA progresses, patients commonly develop additional comorbid conditions, including other autoimmune diseases and cancer, leading to challenges in clinical management. However, the specific gene-regulatory mechanism linking these comorbid conditions remains unknown. To address this knowledge gap, we conducted a two-sample Mendelian Randomization analysis and identified 54 blood-expressed genes that have a causal role in JIA. By integrating data from expression quantitative trait loci (eQTL), 3D genome organization, and protein-protein interaction network, we identified sets of single-nucleotide polymorphisms (SNPs) that regulate the expression of these genes and their interaction partners. Querying these regulatory SNPs against a public database of genome-wide association studies revealed 89 comorbid traits sharing gene aberrations within this network. Notably, our analysis highlighted a set genes on chromosome 6p22.1 (*HLA-A, HCG4P5, HLA-T, MOG, TRIM26 , IFITM4PI*) involved in the association between JIA and specific autoimmune diseases, such as Crohn's disease, type 1 diabetes, asthma, and rheumatoid arthritis. Furthermore, we observed a direct protein-protein interaction connecting JIA causal gene *PSMB8* to *HLA-C,* a protein strongly implicated in psoriasis. This finding might explain why 5-9% of JIA patients fall into the "juvenile psoriatic arthritis" (JPsA) category. Among cancer traits, we found a distinct association between JIA and Hodgkin lymphoma through a set of genes in 6p21.3 (*FKBPL, PBX2, AGER*), as well as chronic lymphocytic leukaemia through the *BAK1* gene. Genes like *PBX2* and *BAK1* have been implicated in the regulation of cell cycle and apoptosis. Overall, our findings suggest that an individual's JIA phenotype is influenced in part by their genetic risk for co-occurring conditions. The identification of regulatory mechanisms linking JIA and comorbid traits enhances our understanding of disease origins and enables the identification of shared therapeutic targets, ultimately improving outcomes for patients with multimorbidity.

## G10: Finding new ways forward: Nanopore sequencing as a novel method of characterizing anorexia nervosa associated genetics.

Natasha Berthold[1,2,3], Patrick Anthony Akkari[1,2], Martin Kennedy[3]*
[1]University of Western Australia, Nedlands, Australia
[2]Perron Research Institute, Nedlands, Western Australia
[3]University of Otago Christchurch, Christchurch, New Zealand
*P. A. Akkari and M. Kennedy share equal final authorship.

Anorexia nervosa (AN) is a severe metabopsychiatric disorder with poorly understood aetiology. AN significantly impacts quality-of-life, is one of the most fatal psychiatric disorders, and has high relapse and partial remission rates. There is a recognised notable genetic contribution to AN aetiology. The landmark AN Genetics Initiative (ANGI) study associated eight genetic loci with AN and estimated single nucleotide polymorphism heritability contribution of ~11-17%. However, ~83-89% of heritability is still unaccounted for, and there exists no biologically targeted therapeutics. There is increasing evidence that structural variants (SVs) provide functional insights into complex traits, including neurological and psychiatric disorders, via a plethora of impacts on expression regulation. Moreover, recent advancements in genetic sequencing technologies present a heretofore unseen opportunity to interrogate complex genetics. Here, we propose using Nanopore sequencing to further the results of ANGI and perform a novel characterization of SVs in the ANGI AN-associated loci in a small pilot cohort of AN cases. The adaptive sampling capability was implemented to specifically target the eight ANGI loci and capture preliminary methylation data. Raw sequencing data were processed with established analytical bioinformatics tools Minimap2 and Sniffles2 and phased by WhatsHap for better resolution of potential structural variation. Publicly available highly resolved whole-genome sequencing data, such as the new Human Pangenome Consortium, were leveraged to validate our findings. SVs were prioritised by a multifaceted approach with a plethora of bioinformatics tools visualising data and estimating functional consequences to optimise detection of potentially informative SVs. This preliminary investigation is the first detailed characterisation of the genetic architecture potentially underlying AN and presents an exciting path forward in elucidating AN aetiology.

# G11: Changes in frequency of beneficial and deleterious variants in New Zealand dairy cattle

Deans, F.[1], Spelman, R. [1], Littlejohn, M.[1,2], Couldrey, C.[1], Harland, C.[1]

[1]Livestock Improvement Corporation, Hamilton, NZ

[2]Massey University, Manawatu, Palmerston North, New Zealand

Deleterious genetic variation in dairy cattle can result in abortion or lower producing animals, leading to economic losses through both loss of milk and loss of replacement calves. Testing for carriers of recessive genetic variants allows for selection against disease carriers and reduces the potential economic losses from these diseases. Some dominant genetic variants also benefit from genotyping, such as hornlessness in cattle. Horned cattle are the predominant phenotype in the New Zealand dairy herd, leading to widespread disbudding and de-horning practices to reduce risk to handlers and other cattle. Disbudding leads to poor phenotyping for genetically hornless cattle, meaning genotyping for hornlessness is often more accurate. Breeding for hornless, or polled, cattle is recommended as preferable to disbudding or dehorning due to animal welfare concerns and labour costs.

Two polled variants, four embryonic lethal variants and five production variants (causing lower production of milk solids) have been tested for in over 1.5 million dairy cattle born between 2012 and 2022. Selection against heterozygous and homozygous animals for the production variant known as 'small calf syndrome' has resulted in decreases in the frequency of this variant over time. Comparatively, the other production variants and the embryonic lethal variants did not have such active selection against them, resulting in very little change in the frequency of these variants over time. The beneficial polled variants have increased in frequency in the New Zealand dairy herd, likely aided by testing and selection for this trait by either farmers or breeding companies.

Using genotyping to test for carriers of beneficial and deleterious variants is an important tool for active selection for or against these variants. This active selection can impact changes in allele frequency over time, which may not occur naturally in difficult to phenotype or recessive conditions.

## G13: Pangenome Graph for Comprehensive Surveillance of Plasmids Conferring Antimicrobial Resistance in Aotearoa New Zealand

Zoe Yang[1], Hermes Pérez[1], Una Ren[1,] Kristin Dyet[1], Joep de Ligt[1.]

[1]Institute of Environmental Science and Research, Porirua, New Zealand

Antimicrobial resistance (AMR) is a critical global health concern and has been highlighted as a key area of interest for New Zealand (NZ) by the DPMC white paper. Bacteria acquire AMR primarily through plasmids, yet our understanding of plasmid frequency, bacterial species, and AMR genes is limited, hindering our ability to manage this threat. Our goal is to mitigate this situation by leveraging pangenome graph-based surveillance of plasmids in antibiotic-resistant bacteria. A pangenome comprises an integrated set of genomes from multiple individuals of a specific species. Pangenome graphs are data structures utilized to represent and analyze the genetic variation within a pangenome, offering a comprehensive view of its genetic diversity. Our study involves creating a surveillance platform, establishing a NZ pan-plasmid graph reference library through long-read nanopore sequencing, and associating AMR genes, plasmid types, and resistance profiles. We have developed the pangenome graph pipeline using the Pangenome graph build and the VG toolkit, which stands out due to its unbiased approach in treating all input genomes with equal importance. The plasmid signature and AMR profile of Carbapenemase-producing Enterobacterales (CPE) indicate substantial diversity. We used Nanopore long reads to sequence approximately 130 isolates of CPE, enabling us to assemble the chromosome and most plasmids. This contribution promises to strengthen our ability to manage AMR in NZ by enhancing our understanding, improving infection control measures, and facilitating long-term public health interventions through accessible, real-time surveillance.

## G14: Benchmarking workflow for evaluating the performance of structural variant callers on non-model species genomes

Bailey, S.[1], Rashid. U[1], Chagné, D.[2], Wu, C.[1], Thomson, S.[3]

[1]The New Zealand Institute for Plant and Food Research Limited (PFR), Auckland, NZ, [2]The New Zealand Institute for Plant and Food Research Limited (PFR), Palmerston North, NZ, [3]The New Zealand Institute for Plant and Food Research Limited (PFR), Lincoln, NZ

The characterization of structural variants (SVs) is crucial for understanding the genomic mechanisms that influence traits related to adaptation and domestication in both plants and animals. Current SV-calling methods are primarily developed using diploid human datasets and may perform differently in species with high repetitive sequence content, high heterozygosity and polyploidy, such as plants. They also tend to utilise a single linear reference genome which can introduce reference bias, as they do not encompass the total diversity within a species. Our aim is to identify SV-calling methods using Nanopore sequencing that are suitable for polyploid species and that can ultimately be used in conjunction with pan-graphs, to collate and curate SVs that may be underpinning agronomic traits. We report on SV benchmarking developments, specifically focusing on polyploid plants, incorporating a number of SV-calling tools and aligners. While a ground-truth SV set would be ideal for evaluating SV caller performance, such datasets are not available for non-model species. To overcome the absence of a ground-truth set, we have developed a workflow to simulate read sets with known SVs. Our benchmarking efforts aim to establish a higher standard for SV detection. To achieve this, we have used the workflow management system Nextflow, which allows for continuous evaluation and simplified tool addition. Our results advance the knowledge of SVs in non-model species and contribute to the understanding of the genomic mechanisms underlying phenotypic evolution.

## G15: Improving foundational genomic resources for informed conservation strategies in a critically endangered parrot

Wold, J.R.[1], and Steeves, T.E.[1]
[1] School of Biological Sciences, University of Canterbury, Christchurch, NZ

There is growing interest in the role of structural variants (SVs) as drivers of local adaptation, adaptive potential, and speciation. From a conservation genomics perspective, the characterisation of SVs in threatened species provides an exciting opportunity to complement existing approaches that use single nucleotide polymorphisms (SNPs) to detect adaptive and maladaptive loci and inform conservation management decisions. However, singular reference assemblies do not adequately capture the full breadth of genomic variation within a species or population. This impedes our ability to characterize large and/or complex variants, like SVs, and ultimately impacts research outcomes. Genome graphs offer an opportunity to characterize variation missing from a single linear reference and are better able to resolve complex variation. Nevertheless, the assembly of multiple high-quality and contiguous genomes is no small task for many species of conservation concern. Here, we leverage a chromosomally assembled reference genome, long-read sequencing, and chromosomic approaches to enhance foundational genomic resources for the critically endangered kākāpō (*Strigops habroptilus*). We explore the challenges associated with establishing a representative genome graph and present an exciting chromosomic approach to enhance genomic resources in a species of conservation concern.

## G16: Progress in pan-graph construction from large and complex plant genomes

Chen Wu[1], Sarah Bailey[1], Anby Zhou[2], Usman Rashid[1], Ken Smith[1], Ignacio Carvajal[1], Astra Heywood[3], Ting-Hsuan Chen[3], Ross Crowhurst[1], David Chagné[5], Cecilia Deng[1]* & Susan Thomson[3]*

* Correspondence: cecilia.deng@plantandfood.co.nz & susan.thomson@plantandfood.co.nz

[1]The New Zealand Institute for Plant and Food Research Limited, Auckland, New Zealand
[2]The Kiwifruit Breeding Centre, Auckland, New Zealand
[3]The New Zealand Institute for Plant and Food Research Limited, Lincoln, New Zealand
[4]The New Zealand Institute for Plant and Food Research Limited, Palmerston North, New Zealand

A single reference genome limits our ability to interpret the genetics underlying traits of interest due to reference bias, generally caused by the presence of large structural variants (SVs) that are undetectable using standardised read-mapping processes. Pan-genome graphs allow the integration of variants from multiple individual genomes into a single mapping reference, providing opportunities to analyse those genomic regions that are missing from a single reference genome. To enable pan-graph construction from whole-genome alignments, high-quality individual genome assemblies are required. Building numerous genomes therefore requires a degree of automation in the form of a pipeline. Our highly automated genome assembly and Hi-C scaffolding pipelines with several quality assessment steps embedded can efficiently produce high-quality chromosome-level genome assemblies. These assemblies are then subjected to whole-genome alignment and graph-building software. Currently, we are investigating methods to generate phased haplotypes in polyploid plants and the usability of long-read SV calling tools on polyploid hybrid crosses, as well as integrate SVs to reference genome. Our goal is to produce a reliable and informative pan-graph from both diploid and polyploid Actinidia sp. as a comprehensive mapping reference to largely inform breeding directions, the method of which will be used for other crop breeding programs. We will present our assembly pipeline with its application on assembling animal and plant genomes and preliminary results from polyploid phasing and SV calling.

## G17: Pangenome Graphs in Infectious Disease: A Promising and Practical Implementation for Comprehensive Genetic Variation Analysis

Zoe Yang[1], Andrea Guarracino[2,3], Patrick J. Biggs[4,5], Michael A Black[6], Nuzla Ismail[6], Jana Renee Wold[7], Tony R Merriman[6,8], Pjotr Prins[2], Erik Garrison[2], Joep de Ligt[1]*

[1]Institute of Environmental Science and Research, Porirua, New Zealand
[2]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Tennessee, USA
[3]Genomics Research Centre, Human Technopole, Milan, Italy
[4]Molecular Biosciences Group, School of Natural Sciences, Massey University, Palmerston North, New Zealand
[5]Molecular Epidemiology and Public Health Laboratory, School of Veterinary Science, Massey University, Palmerston North, New Zealand
[6]Department of Biochemistry, University of Otago, Dunedin, New Zealand
[7]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand
[8]Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Alabama, USA

Whole genome sequencing has revolutionized infectious disease surveillance for tracking and monitoring the spread and evolution of pathogens. However, using a linear reference genome for genomic analyses may introduce biases, especially when studies are conducted on highly variable bacterial genomes of the same species. Pangenome graphs provide an efficient model for representing and analyzing multiple genomes and their variants as a graph structure that includes all types of variations. In this study, we present a practical bioinformatics pipeline that employs the PanGenome Graph Builder and the Variation Graph toolkit to build pangenomes from assembled genomes, align whole genome sequencing data and call variants against a graph reference. The pangenome graph enables the identification of structural variants, rearrangements, and small variants (e.g., single nucleotide polymorphisms and insertions/deletions) simultaneously. We demonstrate that using a pangenome graph, instead of a single linear reference genome, improves mapping rates and variant calling for both simulated and real datasets of the pathogen *Neisseria meningitidis*. Overall, pangenome graphs offer a promising approach for comparative genomics and comprehensive genetic variation analysis in infectious disease.

## G18: Empowering Indigenous Communities in Genomics – the Australian Alliance for Indigenous Genomics (ALIGN).

Brown, A. [1,2] on behalf of the Australian Alliance for Indigenous Genomics (ALIGN).

[1]National Centre for Indigenous Genomics (NCIG), Australian National University, Canberra.

[2]Telethon Kids Institute, Adelaide, Australia.

Aboriginal and Torres Strait Islander Australians remain underrepresented in the governance of genomics sciences and remain virtually absent from databases of human genetic variation. Until such inequity is alleviated, the promise of genomics in Australia will remain unfulfilled. Guidelines for protecting the rights of Indigenous peoples in research exists, yet no guidance exists relating to the prioritisation, involvement, conduct, curation, analysis and sharing of genomic data for the explicit benefit of Indigenous Australians.

We seek to establish a national alliance that *synergistically builds and extends Indigenous leadership and involvement in genomic science, research, precision health care, data sciences, ethics and Indigenous knowledge systems to reduce health inequality among Australia's First Peoples*.

We have assembled individual, institutional and organisational proponents of genomics, bioethics, data science, and precision medicine, in partnership with Indigenous research and health care providers, community and consumers, to establish an alliance that will *articulate and prosecute a transformative approach to the development of responsible, culturally appropriate, nationally consistent and internationally relevant Indigenous genomics efforts*. The network will be led by Centres of Excellence across Australia. Each will support national priority programs in *Indigenous Governance; Policy; Capability Development and Data Systems* as well as leading Flagship Programs to maximise the realisation of benefit to Indigenous people from genomics.

Our entire national ecosystem will be predicated on consultation with, and leadership of, Indigenous consumers, communities and organisations, to clearly outline the *major priorities in genomics*, the necessary steps required to deliver *benefit from genomics,* and the articulation of a set of nationally informed and consistent *pathways to reduce inequality* experienced by Indigenous Australians.

## G19: Reference genomes, genome-wide heterozygosity, and historical demography of Hector's (*Cephalorhynchus hectori hectori*) and Māui (*Cephalorhynchus hectori maui*) dolphins.

Alvarez-Costes, S.[1], Baker, C.S. [2], Constantine, R.[3], Carroll, E.[3], Dutoit, L.[4], Ferreira, S.[1], Heimeier, D.[3], Gemmell, N.[1], Gillum, J.[1], Hamner, R.[2], Steel, D.[2]; Rayment, W.[5], Roe, W.[6], Urban, L.[1,7], Alexander, A.[1]

[1]Department of Anatomy, School of Biomedical Sciences, University of Otago, Dunedin, NZ.
[2]Marine Mammal Institute and Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Newport OR, USA
[3]School of Biological Sciences, University of Auckland, Auckland, NZ
[4]Department of Zoology, University of Otago, Dunedin, NZ
[5]Department of Marine Science, University of Otago, Dunedin, NZ
[6]School of Veterinary Science, Massey University, Palmerston North, NZ
[7]Helmholtz AI and Helmholtz Pioneer Campus, Helmholtz Munich, Neuherberg, Germany; Technical University of Munich, School of Life Sciences, Freising, Germany

Hector's and Māui dolphins are endemic to the waters of New Zealand. Characterized by small body size and coastal distribution, they face threats from human activities. There are an estimated 15,000 Hector's dolphins around Te Wai Pounamu (the South Island), classified as endangered by the IUCN. In contrast, the Māui dolphin, restricted to the West Coast of Te-Ika-a-Māui (the North Island) is among the most critically endangered marine mammals globally, with its population estimated at only 54 individuals. To understand the demographic history and evolutionary dynamics of both subspecies, we generated high-quality reference genomes using 10× Chromium-linked short reads and nanopore long reads. Additionally, we reconstructed the historical demography using PSMC and calculated genome-wide heterozygosity for both species. Combining *de novo* assembly tools with reference-guided techniques using chromosome-level genomes of closely related species, we assembled highly contiguous chromosome-level genomes for the Hector's and Māui dolphins (scaffold N50: 110 MB, scaffold L50: 9, and BUSCO completeness scores of 90.1% and 92.3%, respectively for the 2.3GB genome). Using these genomes, historical demography, reconstructed by PSMC, revealed that both Māui and Hector's populations have remained relatively stable at low numbers for thousands of years, with the divergence between the subspecies potentially due to the Last Glacial Period and associated environmental changes. Genome-wide heterozygosity analyses showed heterozygosity in Māui dolphins was 40% lower than Hector's and comparable to other cetacean species noted for reduced genetic diversity. Furthermore, the distribution of heterozygosity across chromosomes in Māui appeared to be more uneven, suggesting the presence of inbreeding within this subspecies. These high-quality chromosome-level genomes and the assessment of heterozygosity and historical demography contribute to a comprehensive understanding of the evolutionary history and genetic consequences of anthropogenic impacts on these dolphins, while also providing genomic resources that are facilitating ongoing conservation genomics research.

## G20 A high-quality genome assembly improves the detection of genetic variants in a wide-ranging marine teleost

Blommaert, J.[1], Sandoval-Castillo, J.[2], Wu, C.[1], Beheregaray, L.[2], Wellenreuther, M.[1,3]

[1]The New Zealand Institute for Plant and Food Research, Nelson, NZ., [2]Molecular Ecology Laboratory, College of Science and Engineering, Flinders University, Bedford Park, SA, Aus., [3]School of Biological Sciences, The University of Auckland, Auckland, NZ.

Teleost fishes represent the largest group of vertebrates, and this diversity is reflected in their genomes. Recent approaches assessing genome-wide genetic variants in tāmure/snapper (Chrysophrys auratus, family Sparidae) showed that structural variants (SVs) impact three times more basepair variation when compared to Single Nucleotide Variants (SNPs). However, some regions have been difficult to assess in this species, such as the region controlling sex. Here, we showcase 1) the utility of an improved high-quality genome in snapper using long read data, and 2) highlight how this assembly can enable the detection of an even larger catalogue of genetic variants with a focus on SVs, and 3) their interrelationships with genes and transposable elements (TEs). Here, we identified 36.5% more SVs than in the previous genome assembly, with most of the newly identified SVs being duplications, inversions, and translocations (as opposed to indels). Because larger SVs were more likely to be newly discovered in this long-read genome assembly, the proportion of the genome affected by SVs nearly doubled. Overall, 6.6% of annotated genes were affected by an SV. We also found that 25.71% of the sequence in this improved genome assembly could be identified as TEs, 3% more than previously identified. This is not outside the wide range of TE proportions in teleost fish genomes, from about 5% in pufferfish (Tetraodon nigroviridis) to 56% in zebrafish (Danio rerio). Importantly, most of the increase in identified TEs were TEs that could be classified. Most of the total TEs (7.2%) were DNA transposons, and remarkably few (1.45%) were identified as LTRs. This genome assembly, SV and TE catalogue form the basis of ongoing work to understand the environmental adaptation of wild populations and the genetic basis of important traits in aquaculture lines in this culturally, ecologically, and economically significant species.

## G21: Giving effect to Te Tiriti o Waitangi in the emerging field of chromosomics

Bramley, C.[1,2], Knight, J.[1,3], Davis T.[4,5], Wylie, M.[5,6], Roberts, J.[1], Steeves, T.[7]

[1]Flowjoanna Tapui Limited, 429 No 1 Line, Longburn, RD5, Palmerston North, Aotearoa-NZ.

[2]Ngāpuhi, Ngāti Kahu ki Whangaroa.

[3]Rangitāne.

[4]Kākāpō Recovery Programme, Department of Conservation, Invercargill, Murihiku, Aotearoa-NZ.

[5]Ngāi Tahu, Ngāti Māmoe, Waitaha.

[6]The New Zealand Institute for Plant & Food Research Limited, Box 5114, Port Nelson, Nelson, Aotearoa-NZ.

[7]School of Biological Sciences, University of Canterbury, Christchurch, Aotearoa-NZ.

The critically endangered kākāpō is a nocturnal, flightless parrot endemic to Aotearoa-New Zealand, and is considered a taonga (treasure) of Ngāi Tahu, a Māori iwi (tribe) of Te Wai Pounamu (the South Island of Aotearoa-New Zealand). Once widely distributed across the North and South Islands, kākāpō populations rapidly declined as a result of human disturbance and introduced mammalian predators. Kākāpō recovery is limited by low productivity–including early hatching failure–but intensive conservation management has grown the population from 51 birds in 1995 to 248 birds alive today. In an effort to determine the chromosomic basis of early hatching failure, we are part of a larger team combining new and existing genomic and cytogenetic resources for kākāpō. Whereas national and international publicly available scholarship regarding the genomics of culturally significant species is growing, no such scholarship exists for the emerging field of chromosomics. Here, we explore cultural perspectives and considerations associated with the creation and care of two primary cell lines made from tracheal tissue of two deceased kākāpō chicks, one female and one male. Beyond providing important context to our research, we are eager to advance early dialogue regarding the *in vitro* culture, use and storage of cells and tissues from culturally significant species to enhance conservation, both here in Aotearoa-New Zealand and beyond.

# G22: Early detection of lung cancer and identification of potential molecular targets for novel therapies via genetic and epigenetic screening

Weeks, R.[1], Almomani, S.N.[1], Macaulay, E.[1], Powell, R.[1], Bhat, B[2], Zuk, M.[3], Wasag, B.[3], Dziadziuszko, R[4]., Rzyman, W[5]., Eccles, M.R.[1*] and Ratajska, M.[1*]

[1]Department of Pathology, University of Otago, Dunedin School of Medicine, NZ,
[2]Department of Surgical Sciences, University of Otago, Dunedin School of Medicine, NZ,
[3]Department of Biology and Medical Genetics, Medical University of Gdansk, PL,
[4]Department of Department of Oncology and Radiotherapy, Medical University of Gdansk, Gdansk, PL, [5]Department of Thoracic Surgery, Medical University of Gdansk, PL,

[*] These authors contributed equally to the work.

Lung cancer (LC) is the third most commonly diagnosed cancer worldwide and is the leading cause of cancer-related morbidity and mortality, with more than a million deaths per year[1]. Most (75%) of LC patients are diagnosed in the advanced stages of the disease (III/IV), and despite significant advances in LC treatment, survival remains poor (a 5-year overall survival of 0% for clinical stage IVB)[2]. In contrast, recent studies demonstrated that LC is curable when early diagnosed (stage IA1), with a 5-year overall survival of 92% of patients[3]. Here we present the results of the successfully completed molecular profiling of 60 LC samples (stage IA-IB) and paired-matched normal lung tissue. We performed genome-scale DNA methylation analysis (using the 850K methylation array, Illumina), followed by targeted methylation sequencing and final validation of our results using QIAseq Targeted Methyl Custom Panel (Qiagen). We completed genetic profiling of all cancer specimens (Archer FusionPlex Lung assay), searching for LC actionable variants. The DNA methylation analysis revealed seven genomic regions as having the most significant differences in DNA methylation between LC and control samples. Some of the identified genes have previously been reported, e.g., hypermethylation of *SOX17* (which is known to antagonize Wnt/b-catenin signalling)[4].
In addition, we detected actionable somatic alterations in 46% of samples, with *KRAS* codon 12 alterations being the most common (28%).
Finally, we were able to reproduce our findings in cell free tumour DNA extracted from plasma of our lung cancer patients.

References:

1. Cancer Genome Atlas Research, N. (2014). *Comprehensive molecular profiling of lung adenocarcinoma.* Nature 511, 543-550.
2. Birring, S.S., and Peake, M.D. (2005). *Symptoms and the early diagnosis of lung cancer*. Thorax 60, 268-269.
3. Hoffman, P.C., Mauer, A.M., and Vokes, E.E. (2000). Lung cancer. Lancet 355, 479-485.
4. Yin, D., Jia, Y., Yu, Y., et al. *SOX17 methylation inhibits its antagonism of Wnt signaling pathway in lung cancer.* Discov Med. 2012;14(74):33-40.

## G23: Detecting non-coding genes in mushroom and truffle-like fungal genomes.

Chyou, T.[1], Fox, W.[1], Padilla, E.[2], Nilsen, A.[3], Mendoza, A.[4], Plett, J.[5], Summerfield, T.[3], Orlovich, D.[3] Brown, C.M.[1]

[1]Department of Biochemistry, University of Otago, Dunedin, NZ, [2]Centro do Ciencias Genomicas, Mexico City, Mexico, [3]Department of Botany, University of Otago, Dunedin, NZ, [4]Lincoln University, Christchurch, NZ, [5] Hawkesbury Institute for the Environment, University of Western Sydney, Sydney, Australia

We are interested mushroom development and the convergent evolution into truffle-like fungi. Non-coding RNA (ncRNA) genes are important for the development, but in complex fungal genomes these were often incompletely annotated, because the modelling focused mainly on a few model organisms like yeast. Frequently, only tRNAs and rRNAs were annotated. However, it is expected that core ncRNA genes, such as snRNA, snoRNA, SRP, telomerase RNA, RNase-P, and RNase-MRP, are present in most fungi. This shortcoming limited our understanding of fungal ncRNAs.

To address this issue, we retrained existing ncRNA models for fungi. Prior to retraining, we searched 384 representative fungal genomes from the NCBI for ncRNAs using Rfam covariation models (CMs). We initially observed an overall underprediction of core ncRNAs. For example, U1 was missing in 122, and RNase-P was missing in 103 out of the 384 genomes. We then taxonomically grouped the fungal genomes by class, utilised the Rfam CM to generate an aligned pool of predicted ncRNAs for each core ncRNA and each class, and then trained a new CM. This recursive process was repeated 4 times. Using retrained core ncRNA CMs improved the sensitivity significantly. For example, U1 was missing in only 12, and RNase-P was missing in only 16 out of the 384 genomes.

The retrained core ncRNA CMs were integrated with published rRNA and tRNA predictors, fungal snoRNA, and Rfam CMs, giving a comprehensive fungal ncRNA prediction pipeline. A total of 33,474 ncRNAs were predicted across the 384 fungal genomes. Of these, 31,134 ncRNAs did not overlap with exons on the same strand or were consistent with validated annotations, indicating a false-discovery rate of only 0.069. This pipeline serves as an improved tool for predicting essential fungal ncRNAs, that are potentially regulators of mushroom development.

## G24: Time to grow up: Validation of the timer-gene segmentation hypothesis in the honeybee

Delargy, E.[1], Dearden, P.K.[1,2]

[1] University of Otago Biochemistry Department, Dunedin NZ, [2]Genomics Aotearoa

*Apis mellifera*, the western honeybee, is essential to the environment and economy. They're an accessible model system for investigating the evolution of development. Segmentation, the division of the anterior-posterior axis into functional units, is a key developmental process that differs in morphology between insects. In recent years a hypothesis has been developed suggesting that these different morphologies are possible because of a flexible underlying gene regulatory network. Segmentation in honeybees is not well understood, and the morphology of segmentation is unusual, thus providing an opportunity to test the flexibility of the underlying GRN. Preliminary research into segmentation visualising gene expression suggests that honeybees segment progressively, similarly to the parasitic jewel wasp *Nasonia vitripennis*.

*Caudal*, *odd-paired* and *Dichaete* are key genes in the "timer-gene" segmentation hypothesis, first proposed in *Drosophila melanogaster*. Using *in-situ* hybridisation v.3 (Hybridisation-Chain Reaction), developmental expression series for *odd-paired* and *caudal* were made using *even-skipped* as a staging marker. Gene expression data indicates timer-genes may be acting similarly in honeybees as in *Drosophila*, implying the GRN is flexible enough to produce different morphological outcomes despite not changing appreciably in 300mya.

Gene expression information helps develop theories about how a GRN interacts but knockdowns using RNAi allows us to determine gene function by testing our proposed GRN. We are using RNAi to test the functions of the timer-genes in segmentation to dissect the regulatory interactions between each gene in the network. This aids understanding how a developmental system can evolve to produce the diversity of animal life.

## G25: Discovery and validation of large-effect recessive variants in New Zealand dairy cattle

Jivanji, S.[1], Davis, S.[1], Drury, L.[1], Ellison, R.[1], McCreesh, A.[1], Spelman, R.[1], Littlejohn, M.D.[1,2]

[1]Livestock Improvement Corporation, Newstead, New Zealand, [2]School of Agriculture and Environment, Massey University, Palmerston North, NZ.

Widespread use of artificial insemination in dairy breeding has seen a dramatic improvement in genetic gain over the decades, where a single elite bull may be used to inseminate hundreds of thousands of cows. If elite bulls carry rare, deleterious variants, this breeding strategy has the potential to drive these variants to problematic frequencies within a few generations. When this occurs, 'forward genetics' approaches can be applied to identify the causal variant where the observation of a disease phenotype precedes fine mapping of the variant. Reverse genetic screens that take advantage of growing genotypic resources in livestock can also be used to identify rare genetic diseases caused by recessive genetic effects that might otherwise have gone undetected within the New Zealand dairy population. In particular, the use of non-additive genome-wide association studies on quantitative traits such as growth and production have had success at identifying genetic diseases, where these traits can be used as proxies for whole-animal health status. Here, we report how a mixed-linear model approach implemented in BOLT-LMM was used to detect non-additive effects in over 165,000 cows using genotype data from a custom single nucleotide polymorphism (SNP) chip designed to include over 10,000 research markers predicted to influence gene function. From these analyses, we identified several novel associations that we predict to have deleterious effects on animal health and/or production, some segregating within our population at allele frequencies as low as 1%. Follow-up phenotypic characterisation of cows homozygous for these variants is currently underway to validate the predicted effect of these variants on animal health and/or production. By using these methods to identify rare, deleterious variants that cause recessive genetic disease, we can try to reduce their frequency within the population through sire selection and strategic mating to ensure our farmers are rearing healthy and efficient animals.

## G26: Evaluation of Loop Mediated Isothermal Amplification (LAMP) for the Identification of Endangered Species Common to the Illegal Wildlife Trade

Yugovich, $O.A.N.$[1,2], Bunce, M.[3], Cave, V.M.[4], Harbison, $SA.$[1,2]

[1] School of Chemical Sciences, University of Auckland, Auckland, NZ
[2] Institute of Environmental Science and Research, Auckland, NZ
[3] Department of Conservation, Wellington, NZ
[4] Department of Statistics, University of Auckland, Auckland NZ

Wildlife crime can be described as any unlawful activity associated with the exploitation and trade of wildlife and constitutes the third most lucrative global illegal business behind narcotics and weapons. Unlawful activities include harvesting, transporting, exchange (money or goods), and end use of wildlife or wildlife-derived products. Rare or endangered species are often sought after by collectors as these items are typically the most prestigious and valuable. Globally, the illegal wildlife trade is regulated by the Convention on the International Trade in Endangered Species of Flora and Fauna (CITES, 1973), and domestically in New Zealand by the Trade in Endangered Species Act (1989). Currently, species identification of suspicious items is conducted using morphology or DNA sequencing-based methods, however there is a demand for a rapid, portable, and cost-effective screening tool that could be used at the border by Customs staff. Loop mediated isothermal amplification (LAMP) is an amplification technique that occurs at a constant temperature, mitigating the need for expensive laboratory equipment. This study evaluates the use of a colorimetric LAMP assay as a presumptive screening tool for the detection of species prevalent in the ivory trade. Novel LAMP assays were developed to detect 8 CITES-listed species and 7 common adulterants in 30 minutes or less. The developed assays were tested for their specificity, sensitivity, and tolerance to common PCR inhibitors and high amounts of non-target DNA. The results of this study provides new insights into the utility of this amplification method as a rapid and simple presumptive screening tool for the identification of species common in the illegal wildlife trade.

## G27: Advancing aquacultural aspirations through genomic sequencing

Bailie, M.[1], Elliot, A.[2], Stephens, J.[2,3], Eason, C.[2-5], Kenny, N.J.[1]

[1]Department of Biochemistry Te Tari Matū Koiora, University of Otago, Dunedin, NZ, [2]Wakatū Incorporation, Nelson, NZ, [3]AuOra Ltd, Nelson, NZ, [4]CE Research Associates, Nelson NZ, [5]Faculty of Agricultural and Life Sciences, Lincoln University, Lincoln, NZ

Genomic sequencing is increasingly being used to address commercial problems in a variety of fields. It provides data that can be used for selective breeding, trait screening, and identifying and generating novel products. This information, however, has not always been generated in an equitable or culturally appropriate fashion, and benefits have not always been returned to source communities.

To aid Māori-lead aquaculture in New Zealand, we have adopted a tikanga-inspired approach to sampling and data sovereignty, combined with cutting-edge genomic sequencing techniques. This innovative methodology will be used to establish a New Zealand-centric framework for molecular biology and genome sequencing, cognisant of the need for partnership in this field.

Recent advancements in Oxford Nanopore sequencing technology have democratised access to genomic sequencing and have significantly enhanced the prospect of telomere-to-telomere assemblies at reasonable cost. We are sequencing the genome of kopakopa (the ribbed mussel *Aulacomya atra maoriana*), a species of commercial interest to Wakatū and the iwi of Te Tauihu, to reference standard. We will utilise this resource to underpin commercial aspirations while also providing an understanding that can be used for kaitiakitanga in the future.

Our approach will allow for the benefits of genomic sequencing to be returned to hapori Māori, while also providing data of broad interest to science and of utility to conservation, fostering a comprehensive understanding of this indigenous species, which is revered as a taonga, for the benefit of future generations.

## G28: Genome sequencing of New Zealand stick insects for understanding phylogeny and the evolution of asexual strategies

Collins, G.E.[1], Gandhi, T.[1], Palacios-Gimenez, O.M.[2], Blommaert, J.[3], Buckley, T.R.[1]

[1]Manaaki Whenua - Landcare Research, Auckland, NZ, [2]Friedrich-Schiller-Universität Jena, [3]The New Zealand Institute for Plant and Food Research

In New Zealand (NZ), there exist a total of 23 stick insect species which belong to 10 different genera. These species are exclusive to NZ, having migrated from New Caledonia ~20 million years ago. Since their arrival, these stick insects have undergone processes of radiation and diversification. A significant aspect of this radiation has been the emergence of asexual variants, particularly in alpine environments. Additionally, hybridisation events between two genera (*Acanthoxyla* and *Clitarchus*) have occurred resulting in the production of asexual offspring species. Some of these hybrid species possess triploid genomes with substantial sizes, reaching up to 5 Gb. Obtaining high-quality phased genome assemblies for these taxa poses considerable challenges, requiring long-read sequencing and high sequencing coverage to effectively resolve haplotypes. However, such genome assemblies are crucial for various investigative purposes. They would enable the identification of sex chromosomes, facilitate differential expression analyses, allow for detection of genome rearrangements and transposable element activity, and contribute to the estimation of divergence times among hybrid species. Here, we use Oxford Nanopore long-read sequencing to assemble the genome for *Acanthoxyla prasina*, an asexual hybrid species with a triploid genome size of 5 Gb. Additionally, we employ mitochondrial markers and BUSCO genes as reference for mapping Illumina short-read sequencing data, to reconstruct phylogenetic relationships for all NZ stick insects. Through the integration of these complementary approaches, we aim to elucidate the evolutionary history of NZ stick insects and establish a foundation for investigating the mechanisms underlying asexuality within these lineages.

# G29: Nanopore Sequencing of the Pharmacogene *CYP2C19* in a Ngāti Porou Hauora Cohort

Kerekere, T.[1], Hitchman, L.M.[1], Miller, A.L.[1], Wilcox, P.[2], Robertson, S.[3], Watson, H.[4], Kennedy, M.A.[1]
[1]Department of Pathology and Biomedical Science, University of Otago, Christchurch, NZ, [2]Department of Mathematics and Statistics, University of Otago, Dunedin, NZ, [3]Department of Women's and Children's Health, Dunedin, NZ, [4] Te Rangawairua o Paratene Ngata Research Centre, Gisborne, New Zealand

The *CYP2C19* pharmacogene encodes the principal metabolising enzyme for a variety of drugs including antiplatelet medicines, antidepressants, and anticancer compounds. Currently, there are 36 well-defined *CYP2C19* star alleles catalogued in the Pharmacogene Variation Consortium (PharmVar) database. An individual's star allele combination is used to assign their metaboliser status. Some star alleles result in impaired enzyme function, loss of function or increased expression. Only one increased expression allele (*CYP2C19*17*) is currently catalogued in PharmVar. However, a recently discovered CYP2C:TG haplotype, formed by two SNPs ~80kb upstream of *CYP2C19*, was shown to be associated with increased *CYP2C19* expression [1].

*CYP2C19* is highly polymorphic and much of the variation affecting this gene is unknown, especially in minority populations such as Māori. We hypothesise that there are novel variants in the Māori genome which may affect the function of CYP2C19. Therefore, in this project, we will catalogue the variation in the coding regions of *CYP2C19, and in the non-coding sites of CYP2C19*17 and the* CYP2C:TG haplotype, in a Māori cohort from Ngāti Porou Hauora (NPH), an iwi health provider in Te Tairāwhiti.

Large PCR amplicons encompassing the *CYP2C19* coding regions were designed and are being nanopore sequenced in our cohort (n=140). The functionality of novel variants will be assessed in silico, and those that potentially impact function will be Sanger sequenced to ensure they are not artifacts. In addition, we are also investigating the CYP2C:TG haplotype frequency using TaqMan assays. This study will paint a picture of the variability of the *CYP2C19* gene in the Māori population.

References

1. Braten, L. S., Haslemo, T., Jukic, M. M., Ivanov, M., Ingelman-Sundberg, M., Molden, E., & Kringen, M. K. (2021). A Novel CYP2C-Haplotype Associated With Ultrarapid Metabolism of Escitalopram. *Clin Pharmacol Ther*, *110*(3), 786-793.

## G30: Vascular endothelial growth factor A polymorphisms impacting patient survival

Meza-Alvarado, J.C. [1], Page R.A. [1], Mallard B.[1], Bromhead C.[1], Pilbrow, A.P.[2], Troughton, R.W.[2], Frampton, C.M.[2], Doughty, R.N.[3], Richards, A.M.[2], Palmer, B.R.[1]

[1]School of Health Sciences, College of Health, Massey University, Wellington, NZ

[2]Christchurch Heart Institute, Department of Medicine, University of Otago Christchurch, NZ

[3] Faculty of Medical and Health Sciences, University of Auckland, Auckland, NZ

Cardiovascular diseases (CVDs) are a leading cause of death worldwide. Accurate prognosis can be advanced by identifying specific biomarkers before the onset of critical adverse CVD outcomes. The vascular endothelial growth factor (VEGF-A) influences the cardiovascular system by regulating endothelial cell proliferation and vascular permeability. Increased plasma levels of VEGF-A have been observed in coronary heart disease and after acute myocardial infarction (MI). Genetic polymorphisms, such as rs6921438, have been associated with altered VEGF-A levels. High VEGF-A levels correlate with CVD risk factors such as metabolic syndrome and hypertension. VEGF-A impacts established pathways leading to CVD., Therefore circulating VEGF-A levels and SNPs, that regulate VEGF-A expression, represent potential novel prognostic biomarkers.

Genotype data for 30 SNPs from the same locus as the *VEGFA* gene were typed or imputed from a genome-wide scan of 1935 patient DNA samples from the Christchurch Coronary Disease Cohort Study (CDCS). Statistical analyses focused on assessing the association of genotypes with patient's cardiometabolic parameters and clinical outcome, including survival. The minor genotypes of a cluster of 5 SNPs located 173 kb downstream from *VEGFA* are strongly associated with elevated plasma levels of VEGF-A, elevated systolic blood pressure, body mass index (BMI) and amino terminal B-type natriuretic peptide (NT-proBNP). Following correction for other covariates (age, history of previous MI, BMI, NT-proBNP and creatinine) using Cox regression analysis we found patients carrying the high-risk genotypes of two of the imputed SNPs, rs6921438 and rs7767396, were independent predictors of survival. Patients with high-risk genotypes were respectively, 7.6 or 7.1 times more likely to die within 8 years of their index acute coronary event, than patients with the low-risk genotypes. Overall, rs6921438 and rs7767396 represent potential genetic markers for prognostic risk stratification after acute coronary events.

## G31: Transposable element copy variation in cultivated apples, and their wild apple relatives

Venon, A.[1], Chen, X.[1], Alix, K.[1], Cornille, A.[1]

1. Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190 Gif-sur-Yvette, France.

Transposable elements (TE) are repetitive regions of the genome known to play a role in key agronomic traits in crops. TEs can also be involved in population adaptation to new environments. The role of TEs in perennial crop domestication, including fruit trees, still needs to be studied. We estimated copy number of TEs among populations of the cultivated apple (*Malus domestica*) and its wild apple relatives, sequenced in short reads. We built an original bioinformatic pipeline combining short-reads and TE tool predictors (RepeatExplorer[1] and REPdenovo[2]) to obtain an exhaustive database of TE consensus. We then mapped the raw short reads onto this TE database to get the copy number of each TE consensus. Statistical analyses revealed TE copy number variations among cultivated and wild apples and within the cultivated apples depending on their uses, with LTR-retrotransposons class I being the most represented TEs. Further analyses will help understand which specific TEs can explain the observed variation. Our study is the first step to unravel the role of TEs in the domestication of fruit tree crops and their role in evolution.

1. Novak, P., Neumann, P., Pech, J., Steinhaisl, J. & Macas, J. *RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics* **29**, 792–793 (2013).
2. Chu, C., Nielsen, R. & Wu, Y. *REPdenovo: Inferring De Novo Repeat Motifs from Short Sequence Reads. PLOS ONE* **11**, e0150719 (2016).

## G32: Structural Variation in NZ sheep – identification to implementation

Clarke, R.M.[1], Hess, A.[2], Caulton, A.[1], Brauning, R.[1], McRae, K.M.[1], Chen, A.[3], Dodds, K.G.[1], Clarke, S.M.[1]

[1]Agresearch, Invermay, Mosgiel, NZ, [2]University of Nevada, Reno, USA, [3]University of Otago, Dunedin, NZ.

The challenge to provide energy-rich protein to the world while reducing the environmental impact is one of the largest challenges facing the animal industry today. To address this challenge more complex variation such as structural variations and epigenetics need to be utilised when selected animals to breed from. Structural variations (SV) play a major role in genetic diversity and phenotypic variations; however, they remain largely unexplored in most domesticated animals.

SNP arrays can be used to determine SV by utilizing the intensity values derived from each sample. The International Sheep Genomics Consortium Ovine High-density chip contains 606,006 SNPs and has been used in the NZ Sheep industry for several years with thousands of animals genotyped. This population also contains phenotypic data associated with breeding traits providing an excellent opportunity to identify SV present in the NZ sheep population.

A major hurdle in utilizing structural variations lies in the development of tools that can seamlessly capture multiple types of information in a high-throughput and cost-effective manner. We propose Oxford Nanopore Technologies (ONT) adaptive sampling as a potential solution well-suited for this purpose.

Adaptive sampling is a software-controlled enrichment unique to nanopore sequencing platform which enables targeted sequencing of specific regions of interest at higher coverage. This approach enables simultaneous capture of multiple sources of information such as methylation, mutations, and structural variances, in a single run.

We have utilised the SNP arrays to identify SV present in NZ sheep and aim to validate these regions using ONT adaptive sampling. We have trialled this technique to evaluate the possible use of adaptive sampling as a high-throughput cost-effective tool for the animal industry. This presentation will discuss the SVs identified in NZ sheep and the potential benefits and challenges currently being explored to integrate adaptive sampling into the animal industry.

## G33: Prenatal cannabis exposure alters offspring DNA methylation at genes involved in schizophrenia and neurodevelopment, across the life course

Noble, A.J.[1], Adams. A.[1], Satsangi, J.[1], Boden, J.M.[2], Osborne, A.J.[3]

[1]Translational Gastroenterology Unit, Nuffield Department of Experimental Medicine, University of Oxford, Oxford, UK, [2]Christchurch Health and Development Study, Department of Psychological Medicine, University of Otago Christchurch, Christchurch, NZ, [3]School of Biological Sciences, University of Canterbury, Christchurch, NZ.

Prenatal cannabis exposure (PCE) is of increasing concern globally, due to the potential effects on offspring neurodevelopment. However, our understanding of its molecular impact remains unknown, meaning that we do not know whether the impact of PCE on neurodevelopment is causative, or simply correlational. In order to begin to understand whether PCE can be considered (at least partly) causative for adverse neurodevelopmental outcomes in exposed children, we undertook epigenome-wide association studies at multiple timepoints, examining the effect of PCE and co-exposure with tobacco using two longitudinal studies; the Avon Longitudinal Study of Parents and Children (ALPSPAC) and the Christchurch Health and Development Study (CHDS) at birth, age 7 and age 15 (ALSPAC), and age 27 (CHDS). Our findings reveal genome-wide significant DNA methylation differences in offspring at birth, 7 years, 15 years and 27 years associated with PCE alone, and co-exposure with tobacco. Importantly, we identified CpG sites within the genes *LZTS2*, *NPSR1*, *NT5E*, *CRP2*, *DOCK8*, *COQ5* and *LPAR5* that are differentially methylated at multiple time points throughout development in offspring. Notably, functional pathway analysis showed enrichment for differential DNA methylation in neurodevelopmental and schizophrenia pathways, and this was consistent across all time points in both cohorts. Given the increasing volume of epidemiological evidence that suggests a link between PCE and adverse neurodevelopmental outcomes in exposed offspring, this work highlights the need for further investigation into PCE, particularly in larger cohorts.

## G34: Network Analysis Links Melanoma Germline Risk Loci to Somatic Driver Genes Through Novel Genetic Pathways

Michael Pudjihartono[1,2], Justin O'Sullivan[1,2,3,4], William Schierding[1,2]

[1]Liggins Institute, The University of Auckland, Auckland, NZ, [2]The Maurice Wilkins Centre, The University of Auckland, Auckland, NZ, [3]Australian Parkinson's Mission, Garvan Institute of Medical Research, Sydney, New South Wales, Australia, [4]MRC Lifecourse Epidemiology Unit, University of Southampton, United Kingdom.

Understanding the underlying biological mechanisms driving melanoma is crucial to develop targeted treatments. We have previously identified 151 target genes, across three distinct skin tissues, whose expression levels are altered by inherited single nucleotide polymorphisms (SNPs) in 42 melanoma risk loci. These germline susceptibility target genes participate in larger co-regulatory mechanisms that involve gene regulatory networks (GRNs) and protein-protein interaction networks (PPINs).

In this study, we constructed PPINs using the proteins encoded by the 151 target genes as the foundation, expanding the network to multiple levels. We observed a significant enrichment of interactions between the proteins encoded by melanoma target genes and the protein products of known melanoma high-penetrance germline and somatic driver genes. Notably, coregulatory connections to key somatic driver genes *TERT, BRAF, NRAS, CDKN2A, MAP2K1,* and *MITF* were captured within the first two levels of the constructed network.

Furthermore, we integrated the 151 target genes into GRNs and discovered that within skin cells these genes are involved in shared gene regulatory processes that link melanoma to many of its primary comorbid phenotypes, including skin pigmentation, immune response and other cancers.

Our results are consistent with melanoma as a systems-level network phenomenon, where germline susceptibility target genes impose risk, in part, from alterations to these larger regulatory networks. Our study presents a new way of scrutinising the biological implications of genetic variants associated with melanoma and provides a starting point for further experimental validation.

## Summary of Abstracts for the Poster Session Template

| No. | Title | Presenter | Institutions |
|---|---|---|---|
| G1 | Beyond one T2T human genome: What's next? | Rhie, Dr. Arang | NHGRI / NIH |
| G2 | Detection and Mitigation of Microbiome DNA Presence in Saliva-Derived Whole Genome Sequence Data | Halliday, Dr. Benjamin | The University Of Otago |
| G3 | MapQsee—Improving Visualisation of Sequence Read Mapping | Heywood, Dr. Astra | Plant And Food Research New Zealand |
| G4 | COVID-19 risk variant associations with chromatin remodelling, DNA maintenance and surfactant genes are infection dependent in the lung | Jaros, Ms. Rachel | University Of Auckland |
| G5 | AssemblyQC: A reproducible best-practices NextFlow pipeline for assessment of assembly quality | Rashid, Dr. Usman | The New Zealand Institute for Plant and Food Research Limited |
| G6 | Evolutionary dynamics of transposable elements in Actinidia species | Chen, Dr. Ting-Hsuan | Plant And Food Research |
| G7 | Phylogenetic analysis of single-cell RNA-seq expression data | Gavryushkina, Dr. Alexandra | University Of Canterbury |
| G8 | Nanopore methylation sequencing in a non-model plant species | Liau, Dr. Yusmiati | Bragato Research Institute |
| G9 | Network analysis uncovers gene-regulatory intersections between juvenile arthritis and comorbid traits. | Pudjihartono, Mr. Nicholas | University Of Auckland |
| G10 | Finding new ways forward: Nanopore sequencing as a novel method of characterising anorexia nervosa associated genetics | Berthold, Ms. Natasha | University Of Otago Christchurch |
| G11 | Changes in frequency of beneficial and deleterious variants in New Zealand dairy cattle | Deans, Dr. Fenella | Livestock Improvement Corporation |
| G12 | Paediatric acute haematogenous osteomyelitis: Identification of bacterial genes and phenotype that predispose to adverse health outcomes | Joseph, Dr. Reece | The University Of Auckland |
| G13 | Pangenome Graph for Comprehensive Surveillance of | Perez, Mr. Hermes | ESR |

| | | | |
|---|---|---|---|
| | Plasmids Conferring Antimicrobial Resistance in Aotearoa New Zealand | | |
| G14 | Benchmarking workflow for evaluating the performance of structural variant callers on non-model species genomes | Bailey, Miss. Sarah | Plant & Food Research |
| G15 | Improving foundational genomic resources for informed conservation strategies in a critically endangered parrot | Wold, Dr. Jana | University Of Canterbury |
| G16 | Progress in pan-graph construction from large and complex plant genomes | Wu, Dr. Chen | Plant And Food Research |
| G17 | Pangenome Graphs in Infectious Disease: A Promising and Practical Implementation for Comprehensive Genetic Variation Analysis | Yang, Dr. Zoe | ESR |
| G18 | Empowering Indigenous Communities in Genomics – the Australian Alliance for Indigenous Genomics (ALIGN). | Brown, Prof. Alex | ANU \| TKI |
| G19 | Reference genomes, genome-wide heterozygosity, and historical demography of Hector's (Cephalorhynchus hectori hectori) and Māui (Cephalorhynchus hectori maui) dolphins. | Alvarez-Costes, Mr. Sebastian | University Of Otago |
| G20 | A high-quality genome assembly improves the detection of genetic variants in a wide-ranging marine teleost | Blommaert, Dr. Julie | Plant And Food Research |
| G21 | Giving effect to Te Tiriti o Waitangi in the emerging field of chromosomics | Bramley, Ms. Claudia | Flowjoanna |
| G22 | Early detection of lung cancer and identification of potential molecular targets for novel therapies via genetic and epigenetic screening | Ratajska, Dr. Magdalena | University Of Otago |
| G23 | Detecting non-coding genes in mushroom and truffle-like fungal genomes | Chyou, Dr. Te-yuan | University Of Otago |
| G24 | Time to grow up: Validation of the timer-gene hypothesis in the honeybee | Delargy, Miss. Erin | University Of Otago |
| G25 | Discovery and validation of large-effect recessive variants in New Zealand dairy cattle | Jivanji, Dr. Swati | Livestock Improvement Corporation (LIC) |

| | | | |
|---|---|---|---|
| G26 | Evaluation of Loop Mediated Isothermal Amplification (LAMP) for the Identification of Endangered Species Common to the Illegal Wildlife Trade | Yugovich, Ms. Olivia | ESR and UoA |
| G27 | Advancing aquacultural aspirations through genomic sequencing | Bailie, Mr. Marc | The University Of Otago |
| G28 | Genome sequencing of New Zealand stick insects for understanding phylogeny and the evolution of asexual strategies | Collins, Dr. Gemma | Manaaki Whenua – Landcare Research |
| G29 | Nanopore Sequencing of the Pharmacogene CYP2C19 in a Ngāti Porou Hauora Cohort | Kerekere, Miss. Te Whetu Aarahi | University Of Otago |
| G30 | Vascular endothelial growth factor A polymorphisms impacting patient survival | Meza Alvarado, Mr. Juan Carlos | Massey University |
| G31 | Transposable element copy variation in cultivated apples, and their wild apple relatives | Venon, Dr. Anthony | Université Paris Saclay |
| G32 | Structural Variation in NZ sheep – identification to implementation | Clarke, Dr. Rebecca | Agresearch |
| G33 | Prenatal cannabis exposure alters offspring DNA methylation at genes involved in schizophrenia and neurodevelopment, across the life course | Osborne, Dr. Amy | University Of Canterbury |
| G34 | Network Analysis Links Melanoma Germline Risk Loci to Somatic Driver Genes Through Novel Genetic Pathways | Pudjihartono, Mr. Michael Alexander | University Of Auckland |