centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

2024 Annual Report

The Centre for Population Genomics values diversity in our team and our work. We believe that including all human diversity in genomic research will empower medical care that benefits everyone.

We pay our respect to all Aboriginal and Torres Strait Islander cultures, and to their Elders past and present. We gratefully accept the invitation in the Uluru Statement from the Heart “to walk with us in a movement of the Australian people for a better future”.

Table of Contents

We are currently living through the most exciting period in the history of biology and medicine, driven by rapid advances in genomic technology that are already transforming the prediction, diagnosis, and treatment of many human diseases. However, unless we can rapidly build more representative resources for genomic research, the benefits of these technologies will not be accessible to everyone - and the nascent field of genomic medicine runs the risk of exacerbating already substantial inequities in health outcomes between different communities.

The Centre for Population Genomics (CPG) exists to tackle this urgent challenge. We launched in mid-2020 with long-term committed support from two remarkable Australian research institutions, the Garvan Institute of Medical Research (Garvan) in Sydney and Murdoch Children's Research Institute (MCRI) in Melbourne. Four years later, thanks to that unwavering support, we have grown to a team of over 45, and successfully raised over $26M in industry and grant funding to expand our mission.

Our mission is an ambitious one, but this year we are closer than ever to making equitable genomic medicine a reality.

The CPG is, by necessity and design, a deeply unusual organisation. We operate as a single, geographically distributed, remote-first team, with staff at multiple locations across Eastern Australia and New Zealand. Our team is composed primarily of professional staff rather than traditional academics, and our software engineering and project management teams have built infrastructure that allows us to manage the largest collection of genomic data in Australia. We are intensely collaborative, with a network of research, clinical, and industry partners including many key global players in genomics and advanced analytics; and we choose our projects based on their path to long-term, real-world impact on the equitable prediction, diagnosis, and treatment of human disease.

The last year has been an exciting one for the CPG's scientific programs. After a long period of planning and deep community engagement, we launched recruitment for the OurDNA program, which will build the

first genomics repository for many Australian communities currently under-represented in genomics, creating a long-term resource for scientific and clinical impact spanning over 10,000 diverse Australians. We began analysis of a newly-generated collection of genome and single cell RNA sequencing from over 1,800 individuals - the largest in the world - which will lead to deeper understanding of the function of human genes. And our rare disease program developed a new automated approach to diagnosing patients affected by severe genetic disorders, resulting in over 185 new genetic diagnoses for Australian families.

This annual report lays out the major accomplishments of the last year, and is also an opportunity to reflect on what's coming next. One exciting new avenue is the incorporation of new artificial intelligence tools into the diagnosis of genetic disease - an effort that was substantially boosted this year by an $8M federal grant that supports CPG's leadership of a new national consortium, bolstering our already productive collaborations with industry partner Microsoft Research.

I am extremely proud of the accomplishments of the CPG team over the last 12 months, and excited by the year ahead. My sincere thanks to everyone in the team for their hard work; to our institutional partners at Garvan and MCRI for their critical financial, operational, and intellectual support; and to our new funders and supporters. Our mission is an ambitious one, but this year we are closer than ever to making equitable genomic medicine a reality.

Daniel MacArthur

DIRECTOR, CENTRE FOR POPULATION GENOMICS

We find ourselves at an exhilarating juncture in genomic medicine, where cutting-edge technologies are poised to make unprecedented clinical impacts. Australia's unique position, with its robust healthcare infrastructure and research capabilities, offers immense opportunities to harness these technological advancements for the benefit of our population.

A cornerstone of our success has been the highly collaborative model that underpins CPG as a partnership between two of Australia's leading institutions in genomics - the Murdoch Children's Research Institute and the Garvan Institute of Medical Research. Collaboration is woven into the very DNA of CPG, enabling us to achieve breakthroughs that would be impossible in isolation.

Our commitment to improving clinical care from the outset has been pivotal, and has already yielded significant benefits for patients with rare disease. Investment in data science has also been a critical driver of our progress. We are particularly excited about the new funding secured for a national consortium in artificial intelligence that will be led by CPG. This investment underscores our belief in the transformative potential of AI to revolutionise genomic medicine, paving the way for more precise and personalised healthcare solutions.

As we look to the future, we remain committed to leveraging the power of genomics to improve health outcomes. Our focus on collaboration, clinical impact, and data science will continue to guide our efforts, ensuring that the benefits of genomic medicine are realised for all.

Kathryn North

EXECUTIVE DIRECTOR, MURDOCH CHILDREN'S RESEARCH INSTITUTE

The Garvan Institute of Medical Research has a long and proud history of genomics research, and this is reinforced in our new strategic plan. The Centre for Population Genomics is a central part of that strategy.

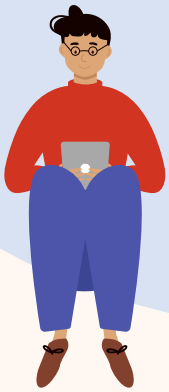
The opportunity to partner with MCRI in CPG is tremendously exciting. It brings together Garvan's large-scale genomics and data science capabilities with a global leader in child health and the clinical translation of genomics.

CPG reflects our shared view that the future of medical research will increasingly require institutional collaborations built around shared interests, shared values, and complementary expertise.

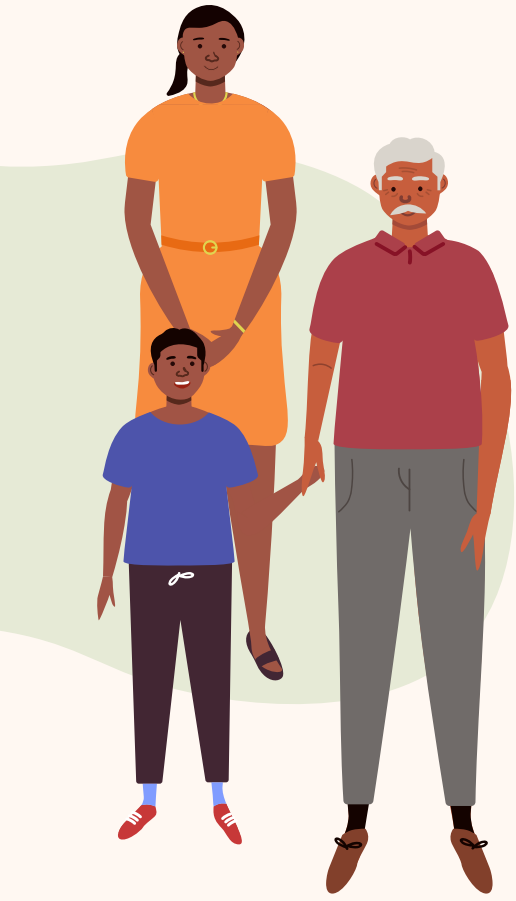
I am excited about many aspects of CPG's work, but the investment in building equitable resources for clinical impact is especially critical. I look forward to the progress with the OurDNA program and the opportunity to ensure that the tools for diagnosis, prediction, and treatment of disease will benefit all Australians.

Benjamin Kile

EXECUTIVE DIRECTOR, GARVAN INSTITUTE OF MEDICAL RESEARCH



Introduction



Our team's expertise is not limited to scientific and technical domains. We have professionals skilled in community engagement, policy development, and advocacy, ensuring that our research is conducted ethically and inclusively.

The Centre for Population Genomics was established in 2020 as a joint initiative between two well-established Australian medical research institutes: Garvan and MCRI. The Centre's formation was driven by a shared vision: to ensure that the incredible opportunities created by the exponential rise of genomic medicine lead to better healthcare outcomes for all Australians.

The Centre's inaugural Director, Daniel MacArthur, returned to Australia to take on this role after 12 years driving world-leading genomics projects in the UK and USA, including the development of ExAC and gnomAD, the world's largest and most widely-used resources of human genetic variation. Over the last four years the Centre has grown to a team of over 45, with expertise spanning genomics, data science, software engineering, statistical genetics, community engagement, participant recruitment, operations and project management, law, policy, advocacy, and other fields.

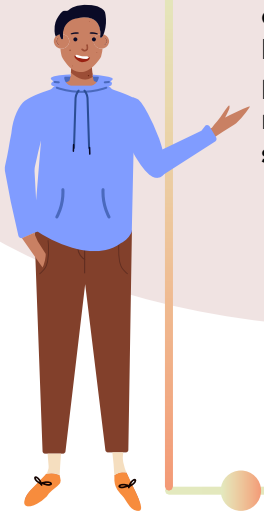
The team is structured along two axes (page 9). Firstly, we have three program teams, reflecting our major areas of scientific focus: the OurDNA team drives our major project on increasing the representation of diverse Australian communities in genomic research; the Population Analysis team builds and deploys pipelines for analysing large-scale genomic datasets; and the Rare Disease team works with clinicians and researchers across Australia to apply new genomic and analysis methods to improve the genetic diagnosis of families with severe genetic disorders.

Secondly, the work of the programs is supported by five cross-cutting foundational teams providing critical infrastructure and processes: Operations, Software Platforms, Project Management, Scientific Affairs, and Equity & Advocacy.

The Centre's operating model is fundamentally open and collaborative, reflecting its origins as a deep partnership between two independent research institutes: CPG projects span more than 80 collaborators including researchers, clinical labs, and industry partners across Australia and around the world (page 14). These collaborations, and the infrastructure built by our foundational teams, allow the Centre to operate at an unusual level of scale: CPG resources now span the largest human genomic datasets in the country.

The Centre's team and capabilities have allowed us to already deliver substantial scientific and clinical impact, and to attract over $20M in competitive federal grants and more than $5.5M in industry funding. This report describes our current programs and achievements, which lay the foundations for the much larger projects that will be needed to ensure a truly equitable future for genomic medicine.

Retrospective



Over four years of operation the Centre has hit many critical project and funding milestones - here are some of them

2020

Centre inception

2021

$3.2M Microsoft Research and Broad Institute collaboration begins

CPG is a partner on 6 grants awarded for rare disease diagnosis

First genomic data ingested and stored on the cloud

CPG and Garvan teams generate largest-ever Australian genomic dataset

2022

$10M grant awarded for the OurDNA program

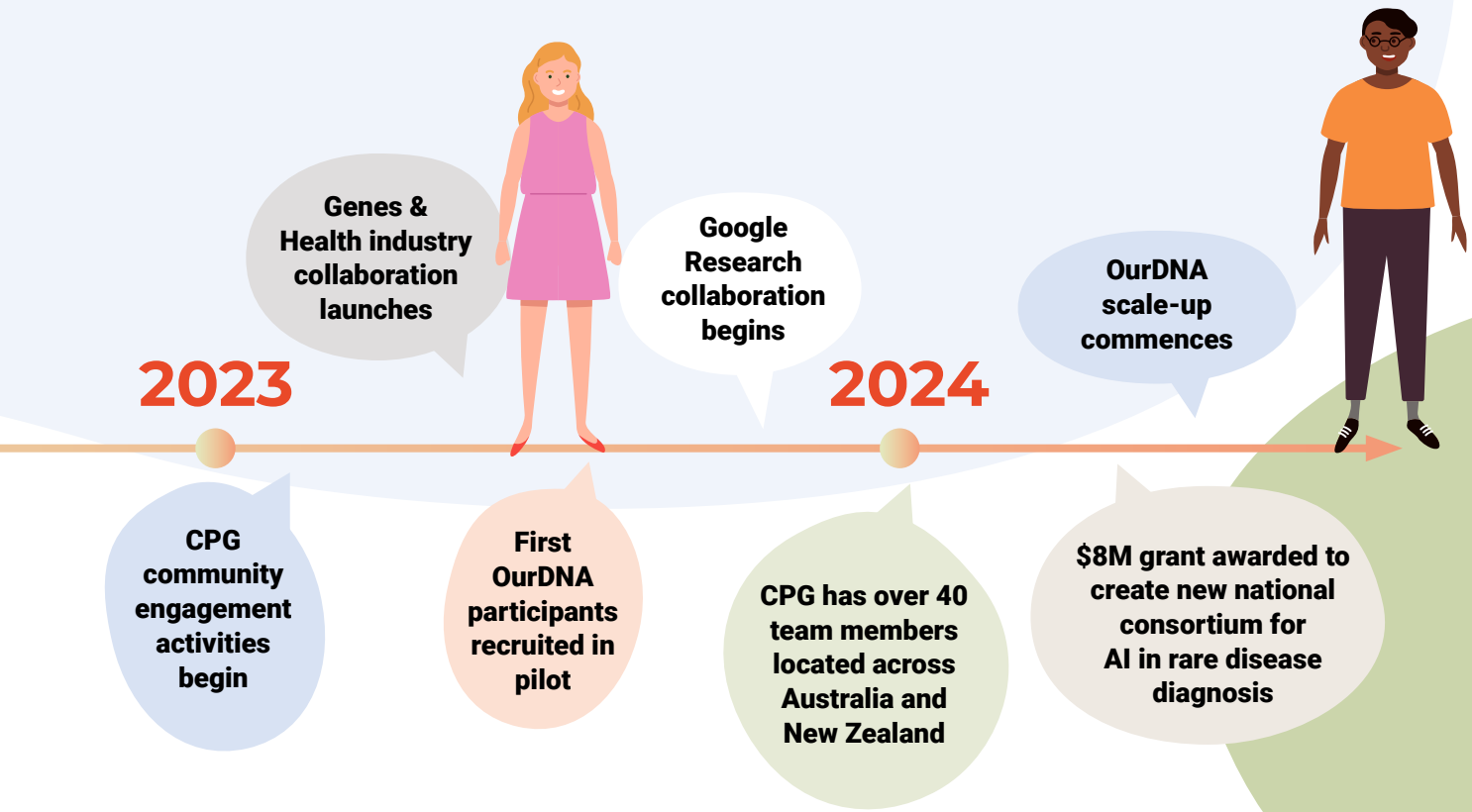
Open Targets industry collaboration launches

First diagnosis returned by Talos automated reanalysis pipeline

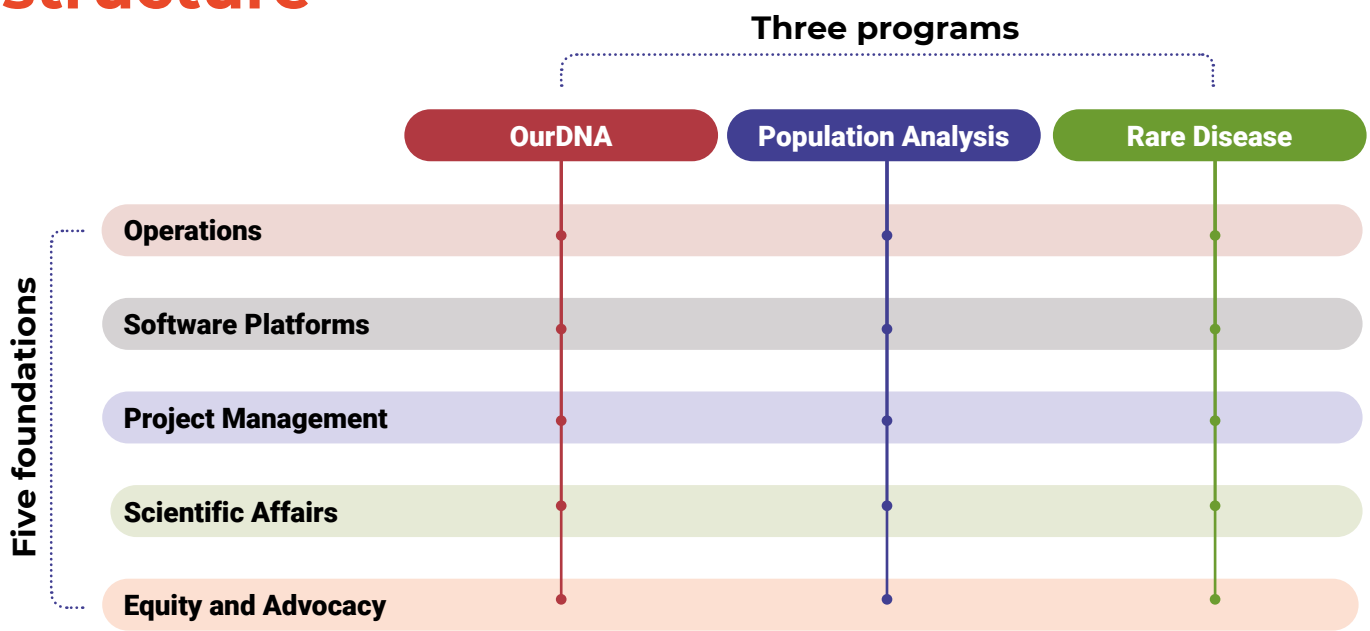
The CPG Mission

Existing global resources of genetic variation are completely missing many large Australian communities, spanning several million Australians. This has real consequences for healthcare: unless we can directly address this problem it will continue to exacerbate existing health inequities.

CPG is tackling this challenge by establishing respectful partnerships with diverse communities, collecting and analysing genomic data at transformative scale, and driving genomic discovery and equitable genomic medicine in Australia.



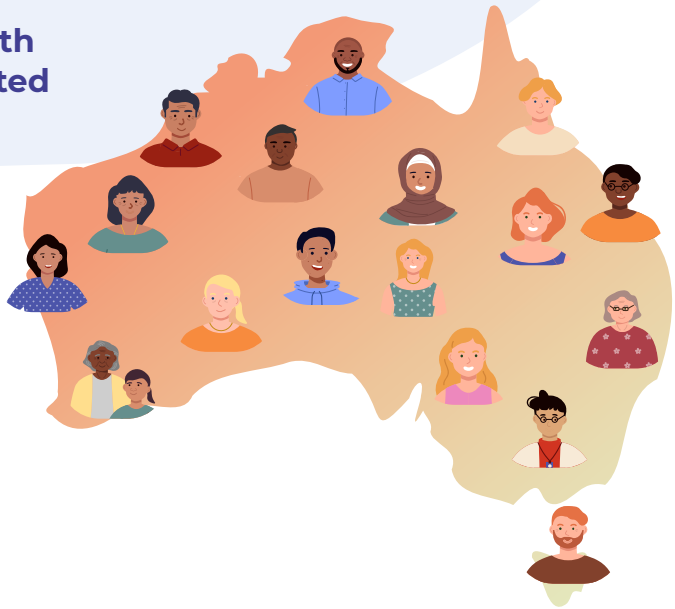
Organisational structure



Focus areas

The CPG's program of research builds on Director Daniel MacArthur's experience in the UK and US at the intersection of large-scale genomics, data science, and clinical translation, adapted to the unique opportunities in Australia, including our diverse population and streamlined path to clinical translation.

Map **genetic diversity** through deep engagement with under-represented communities



Understand the impact of variation on **gene function** through integrating genetic and cellular genomic data

Near-term clinical focus on **rare disease** diagnosis, long-term on risk prediction and therapeutic target discovery

Achievements

Genetic Diversity

Established strong community connections via meaningful outreach and engagement.

- **Over 40 community engagement** events and meetings held across four communities (Filipino, Vietnamese, Lebanese, Sudanese)
- **Participant information and education materials translated into 10 languages** (Arabic, Dari, Farsi, Fijian, Hazaragi, Samoan, Tagalog, Tongan, Urdu, and Vietnamese)

Successfully completed recruitment pilot in the Australian Filipino community

- **5 OurDNA** 'One-Stop Shop' events held
- **119 participants** recruited
- **Over 3,000 combined DNA,** plasma and blood cell aliquots in long term storage

Preparing for the first OurDNA browser release

- **5,900+ genomes** and **9,500+ exomes** ingested from existing Australian cohorts

Gene Function

Created the world's largest dataset of combined whole-genome sequencing and single cell RNA sequencing

- **Over 1,800 individuals** and **5 million cells**, spanning **28 blood cell types**
- **Unprecedented insights** into the effects of rare variation, structural variation, short tandem repeats and somatic variation on gene expression

Manually curated over 12,000 potential loss-of-function variants from population cohorts

- **Outputs are being fed into the drug development pipelines** of our pharmaceutical partners, supporting the future development of safe and effective therapeutics

Rare Disease

Built Australia's largest research-accessible genomics dataset from families affected by severe genetic disease

- **3,800+ genomes** and **2,250+ exomes** ingested from **31 active research cohorts**, including **14+ targeted disease areas**
- Over **280 Australian clinicians**, researchers and postgraduate students have used **CPG's CaRDinal analysis platform** to investigate the genomic data of rare disease patients

Real-world impact on the lives of patients and families

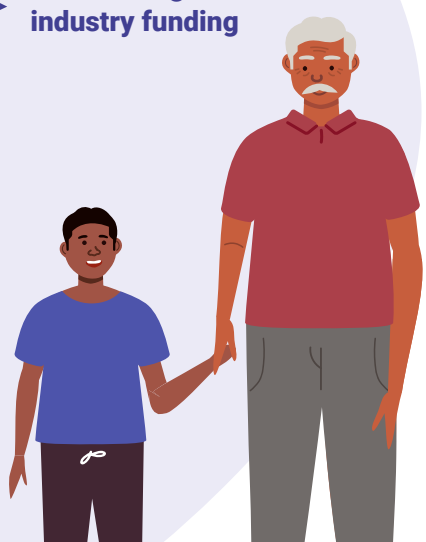
- **Developed a new model for routine reanalysis of undiagnosed patients**, now deployed monthly across thousands of patients in research and clinical cohorts
- **Diagnoses provided for over 380 Australian families** affected by rare genetic disorders

At a glance

828 terabytes of genomic data stored on secure cloud platform

$26.6M secured in grant and industry funding

57 papers published with CPG co-authors



Poised for Impact

Over the past four years the Centre has evolved from an idea into a mature, robust organisation that is already achieving substantial scientific and clinical impact across our key programs.

- → The OurDNA program has successfully engaged with diverse Australian communities, and begun large-scale recruitment to build a truly representative genomic resource for Australia
- → We have collaborated with national and global partners to build the largest cohort in the world combining whole genome sequencing and cellular genomic data, identifying hundreds of new genetic variants impacting immune cell biology
- → Our catalogue of curated gene-disrupting variants is already being used by pharmaceutical companies to design experiments to validate new drug targets
- → We have built the largest Australian collection of genomic data from families affected by severe genetic disease, developed a novel approach to automated reanalysis of research and clinical cohorts, and returned hundreds of new genetic diagnoses to families
- → We have secured over $26M in funding through competitive grant applications and industry agreements, allowing us to expand our scope and scale
- → We have contributed to 57 peer-reviewed scientific publications, with multiple high-impact CPG-led publications now in preparation

These successes are aligned with our commitment to driving projects where there is a clear path to real-world impact, and that move us meaningfully towards a more equitable future for genomic medicine. They also lay the foundation for critical, larger-scale work over the coming years.

Embedding equity in everything we do

Over the next decade, genomic medicine is set to impact every branch of healthcare. Australia is already well-advanced in clinical deployment of genomic technologies but faces a major challenge: Australia's diverse population includes over a quarter born overseas, yet at least 5 million Australians have ancestry from communities that are highly under-represented or entirely missing from global genomic resources. Many of these communities experience poorer health outcomes than other Australians, with higher rates of chronic conditions like heart disease, stroke, diabetes, and kidney disease[1]. Failing to rapidly address this critical gap will widen existing health disparities and result in less accurate genetic diagnosis, poorer risk predictions, and inequitable access to genomic medicine for many Australians.

We believe that for genomic advances to benefit everyone equally, genomics research requires a more equitable approach - one that actively seeks out the rich tapestry of human diversity. At the core of the Centre is our commitment to enhancing diagnosis, refining risk prediction and improving clinical delivery effectively and equitably. The Centre's focus on equity ensures that genomic benefits and advancements are fairly shared among all populations, promoting a more equitable healthcare system.

Our mission to advance the equitable, population-scale implementation of genomic medicine in Australia and beyond is both a scientific and a moral imperative. Embedding equity in our core operations and future planning will ensure the lasting scientific and clinical impact of our efforts to bring the benefits of genomic medicine to underrepresented communities.

We believe that for genomic advances to benefit everyone equally, genomics research requires a more equitable approach - one that actively seeks out the rich tapestry of human diversity.

1. Australian Institute of Health and Welfare 2023, Chronic health conditions among culturally and linguistically diverse Australians, AIHW, Australian Government.

Collaborative network

Baynam

Davis

Ravenscroft, Laing

Kassahn

Amare

Benyamin, Lee

University of South Australia

Boughtwood

Best, Blombery

Southey, Lacaze

Goranitis

Mallett

Hewitt

Brown, Hermes

Geronimo

Vears
Savulescu
Wake
Christodoulou
Thorburn
White
Tan
Sinclair
Stark
Lunke
Delatycki
Sadedin

Yengo, Wray, Shah

Lundie

Galea

McNamara

Bennetts, Palmer, Ewans

Kirk, Oates

Irving, Karki, Daly, Thorpe

Medrum

McEwen

Kennerson

Figtree

Coiera, Berkovsky

Deveson
Ingles
Powell
Young
Kummerfeld
Siggs
Mallawaarachchi

International collaborators

Rehm, Talkowski, Daly, Samocha, Philippakis, Zhou

Dashnow

Rendon

Saleh

Martin

van Heel

Industry partners

Programs updates

Genetic Diversity: OurDNA

Existing global resources of genetic variation don't include many large Australian communities, spanning several million Australians. This results in inequity in access to accurate genetic diagnoses and future developments in genomic medicine. The OurDNA program is addressing this gap by building genomic resources for multicultural communities that are currently underrepresented or completely missing from genomics databases and are unable to fully benefit from genomic medicine.

We are taking a community-based participatory approach to:

- engage and partner with individual underrepresented multicultural communities to develop culturally appropriate recruitment, communications, and outreach strategies.
- recruit thousands of individuals from priority communities, primarily in Sydney and Melbourne, and collect DNA and cells to create new, enduring and more representative genomics resources (OurDNA Samples, Data and Browser).
- make data and samples available in a way that improves the diagnosis of patients from these communities and makes health and medical research that benefits Australian communities easier in the future, while respecting individual privacy.

We are beginning with recruitment in seven priority communities from five global regions that are currently underrepresented in genomics research: Filipino and Vietnamese Australians (South-East Asia), Lebanese Australians (Middle East), Sudanese Australians (East Africa) and Samoan, Fijian, and Tongan Australians (Oceania).

Our approach of engaging with multiple communities simultaneously is unique. In the Australian research setting, this has not been done at a population scale, in non-clinical settings, or for this purpose. We are balancing the need for huge scale - we are recruiting thousands of people - with the imperative to build trust in each community we wish to engage. In this first phase of **OurDNA**, we will build representative resources encompassing 10,000 participants from underrepresented communities. This will be the largest and most diverse genomic resource in Australia.

Samples
Blood and cell repository

- 10,000 participants from under-represented communities
- DNA and blood cells extracted
- Consented for research and recontact
- Basic health questionnaire, consent for data linkage

Data
Conrolled-access genomics datasets

- Whole genome sequencing for comprehensive variants
- Other assays (e.g. scRNA-seq, metabolomics)
- Opt-in validation and return of clinically actionable genetic findings

Browser
Open-access reference database

- Joint processing of newly recruited individuals with external population cohorts.
  Including 17,000 exomes and genomes from other Australian cohorts
- Public release of aggregate variant allele frequencies per community

Our team gained valuable practical experience and captured significant learnings, which we have incorporated as we launch into the next stage.

Achievements

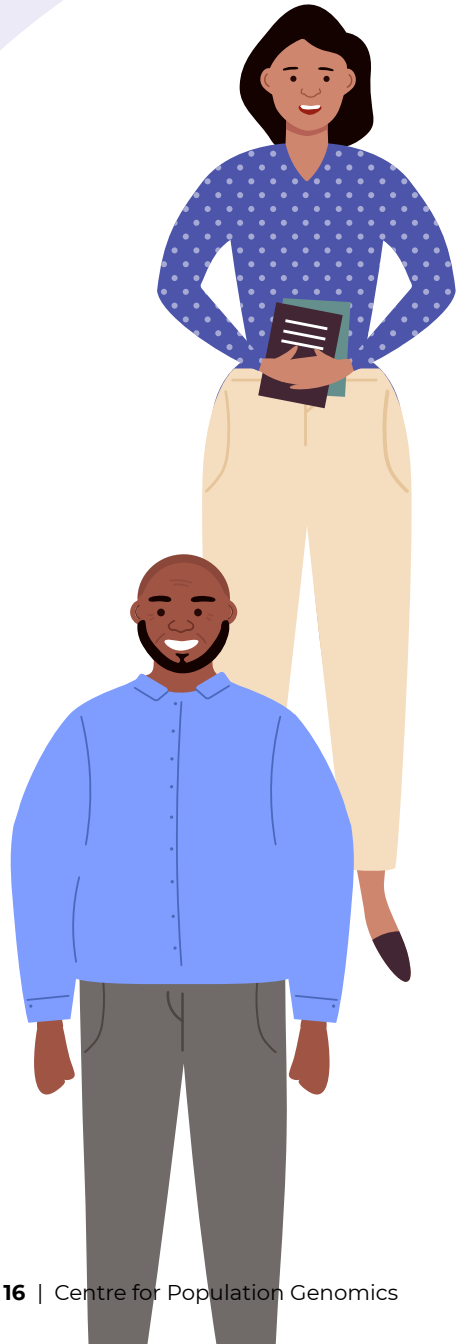
This year OurDNA completed a pilot with the Filipino community in Western Sydney, providing valuable lessons in advance of launching scaled recruitment in 2024. The pilot was launched in August 2023, with the process continuously adapted based on community feedback, and completed in June 2024 after sample collection and processing for 98 participants. Our team gained valuable practical experience and captured significant learnings, which we have incorporated as we launch into the next stage.

Establishing new logistics pipelines

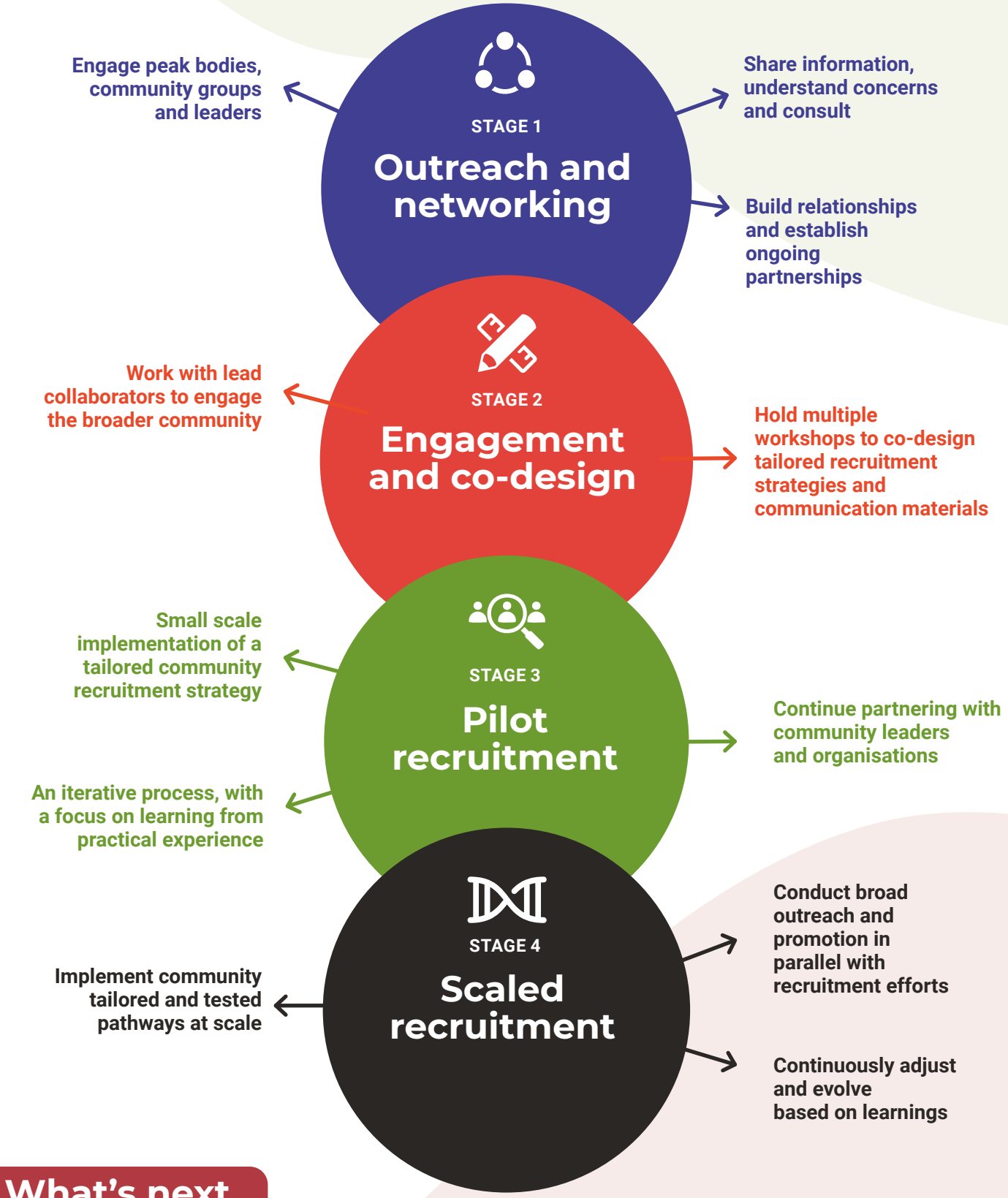
At the conception of OurDNA, we chose the ambitious approach of collecting blood instead of saliva to ensure that we could generate multiple high-quality outputs and maximise the potential for creating a wide variety of datasets in the future. Building a logistics pipeline for large-scale blood draws posed challenges due to the lack of existing pathways that could plug into our community-centred approach and deliver on our rapid sample processing goals. Working closely with our partners, we have successfully established a functional logistics pipeline that aligns with our recruitment strategies and ensures gold-standard sample processing within our time constraints. As we scale we will continue refining our logistics and processing pipeline with our partners.

A recruitment strategy for inclusive genomics

The OurDNA program's recruitment strategy prioritises engagement and responsiveness, and avoids the "set and forget" approach seen in many comparable studies. Leveraging insights from our pilot, we've developed the first iteration of a scalable, ethical, and sustainable recruitment model that is agile and rooted in community participation. This model ensures that recruitment pathways are designed in partnership with communities. As we continue to scale and refine our community-by-community approach, we will share this model as a framework for future genomics and medical research studies that is broadly applicable across the sector.



Recruitment Model



What's next

In 2024, **OurDNA** will transition from pilot to scale, launching outreach, engagement, and recruitment activities across at least three communities. These efforts will be driven by grassroots engagement and a broad outreach campaign. Concurrently, we will conduct community engagement activities with all priority communities, preparing for simultaneous recruitment throughout 2025.

The OurDNA 'One-Stop Shop': A community-led participant pathway and successful pilot cohort recruitment

In April 2024, we held our first **OurDNA** 'One-Stop Shop': an engaging recruitment event where community members could learn about the project, consent, and donate a blood sample, all at once. Driven by our partners, the Philippine Community Council of NSW (PCC-NSW), the one-stop shop has become a critical OurDNA participation pathway.

Initially, we assumed our primary recruitment pathway would centre on participants attending pathology collection clinics via an appointment-based system. After the pilot was launched in 2023, and despite capturing valuable learnings, we had not recruited the number of participants we had hoped for.

Our PCC-NSW collaborators, however, emphasised that community event recruitment would resonate strongly with the Filipino community. This proved true when we promoted the **OurDNA** program at PCC-NSW's Pasko (Christmas) festival. Prospective participants approached us expecting to sign up and donate there and then. Grace Liston, our PCC-NSW co-lead, remarked that it was clear that Filipinos wanted a 'one-stop shop'.

At the start of 2024, our partners at PCC-NSW expressed their determination to reach our pilot recruitment goal and mobilised a 'Genomics Committee'. Working in partnership, we have organised a lively event with food, music, and dance that resulted in more participants recruited in one afternoon than multiple months of recruitment for the pilot. The event was not only a great success but also sparked community members to approach us to organise additional events.

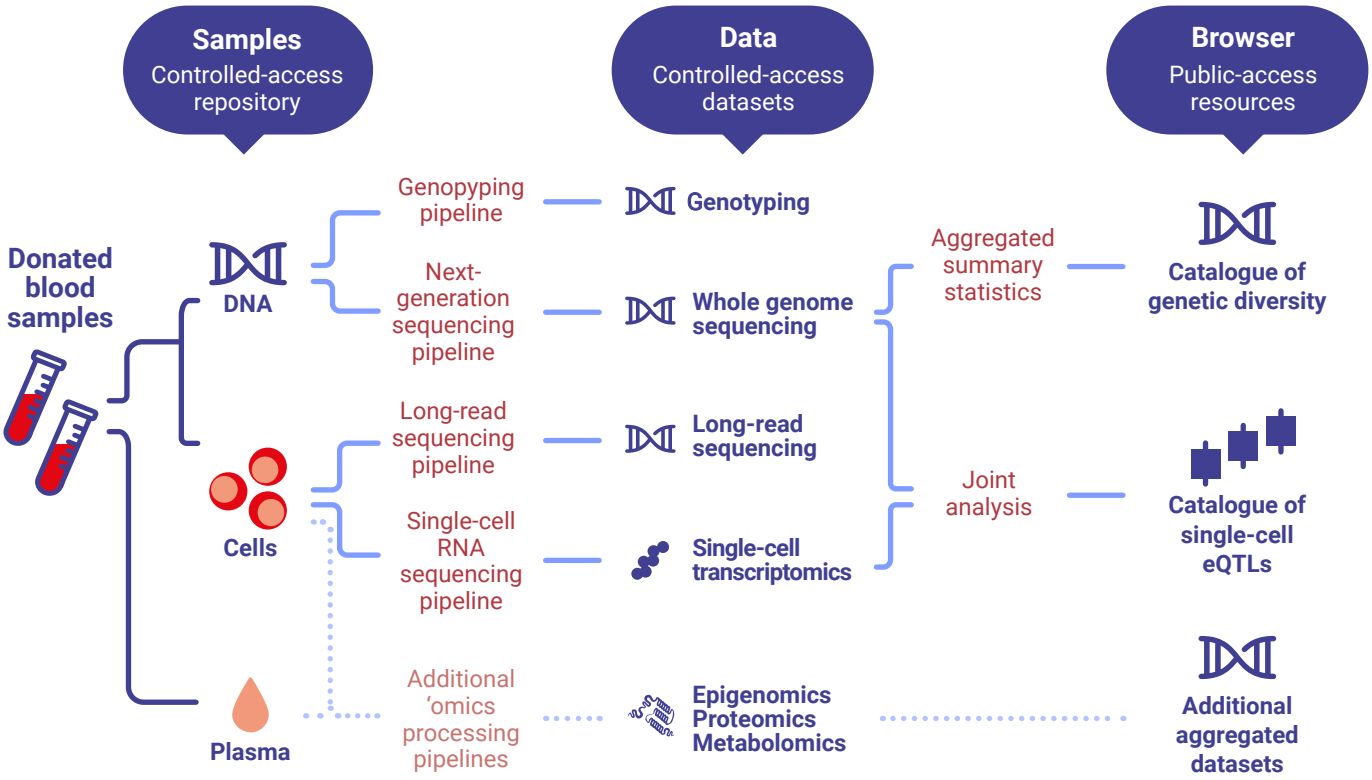
The success of the **OurDNA** 'one-stop shop' has reinforced the value of taking a community-based participatory approach to genomic research. It demonstrates the power of community ownership, drawing on community members' knowledge and expertise, and building partnerships with communities to make genomic medicine accessible to all.

Creating the OurDNA datasets

The OurDNA Samples repository contains DNA, cells and plasma, which provides the necessary raw material to profile our participants' samples across the complete biological spectrum from DNA, through RNA and protein, all the way to functioning metabolism. To capture this information, we are generating a range of datasets using the latest 'omics technologies, including short-read whole genome sequencing, long-read whole genome sequencing, and single cell RNA sequencing. In the future, we will expand this to include epigenomics, proteomics and metabolomics.

All these datasets will be made available to clinicians and researchers via OurDNA Data, a controlled-access system where they will need to apply to access specific data in order to perform research that aims to benefit the OurDNA communities.

CPG will also be analysing the data to produce a number of aggregate resources that summarise key information and can be safely made publicly available, ensuring the reach of OurDNA, and its potential for broad clinical and research impact, is as large as possible.



Establishing an inclusive genomics community of practice

Spearheaded by OurDNA, CPG is now a national leader in driving genomic research that engages underrepresented multicultural communities. We have established a strong network of collaborators across research and clinical genomics, multicultural and health community organisations, and industry, bringing together a community of practice that will enable us to address multicultural health challenges within genomic medicine.

The inaugural OurDNA Symposium, hosted by CPG with support from Australian Genomics, will be a pivotal event bringing together leaders from the genomics, government, industry and multicultural health sectors. As a satellite event of the GA4GH plenary in Melbourne in September 2024, we will attract a significant Australian and international audience. The symposium's primary objectives are to network, raise awareness, and develop strategies to tackle the critical issue of underrepresentation of diverse Australian communities, including Indigenous Australians peoples. This event marks a significant step in establishing CPG's presence and authority in the sector, and provides a platform to influence global discussions on multicultural health challenges.

The OurDNA browser

The first aggregate dataset CPG plans to release out of the OurDNA program is an open-access reference database that is freely available for download as well as explorable within our own OurDNA browser.

This will be produced by jointly processing genomic data from the newly recruited individuals with data from external population cohorts, performing realignment, variant calling, and stringent quality control in order to create a robust and harmonised dataset. The population-specific allele frequencies will be calculated to create a public database that provides clinicians and researchers with access to aggregate information about the genetic changes found in each community, similar to (and ultimately federated with) gnomAD.

In addition to the 119 Australian Filipino participants recruited into OurDNA to date, the first release of the browser will contain up to 17,000 whole genomes and exomes from a number of existing Australian cohorts, creating the largest resource of genetic variation ever assembled from Australian communities.

Tasmanian Ophthalmic Biobank

A biobank of blood and paired clinical information from Tasmanian participants with eye disease, as well as a control set with healthy eyes

- This cohort formed the basis of the initial OneK1K project, with single cell RNA-sequencing data available for all samples
- **Status**: whole genome sequencing for 1,055 samples has been generated and processed by CPG

BioHEART

A cohort of patients at risk of or suffering from coronary artery disease

- Together with the Tasmanian Ophthalmic Biobank, this cohort forms Phase 1 of the TenK10K project, with single cell RNA-sequencing data being generated for all samples
- **Status**: whole genome sequencing for 989 samples generated and processed by CPG (to date)

Mackenzie's Mission

A national project exploring the use of genomic testing for carrier screening in couples planning to have a child

- Particular effort was made to ensure this study enabled the participation of people from a broad spectrum of ethnicities, including Aboriginal and Torres Strait Islander people, people with limited English, those without computer/Internet access and those with disabilities
- Whole exome sequencing for 11,000 samples has been generated by the Mackenzie's Mission project
- **Status**: metadata and exomes from 9,671 participants have been ingested into CPG's systems, amounting to more than 100Tb of data

Medical Genome Reference Bank

A cohort of healthy Australians who are over 70 and have no reported history of cancer, cardiovascular disease or dementia

- 3,928 whole genomes have been sequenced by Garvan
- **Status**: initial alignment and variant calling has been performed as part of the NAGIM initiative; joint processing as part of OurDNA is planned for late 2024

OUR DNA
AUSTRALIAN GENETIC DIVERSITY DATABASE

Industry partnerships enabling community and scientific impact

Google Research

CPG has built productive relationships with industry partners that align with community interests and our long-term impact goals. This year, we formalised a research collaboration with Google which contributes to the OurDNA program and the broader CPG mission across multiple areas. Our ongoing collaboration with Google exemplifies how industry partnerships can help bridge the gap between research, community, industry, and clinical impact.

Prioritising community interests

As part of our collaboration, Google directly funds activities that support and expand the OurDNA team's community engagement efforts. Community engagement is a critical and often underfunded aspect of genomic research, and this support will allow us to expand our efforts rapidly. We are particularly excited to have an industry partner who recognises the importance of investing in community engagement.

Developing novel solutions

Working with underrepresented communities requires us to update our analysis approaches in a number of ways. One of these is in how we detect genetic variation, which typically relies on traditional reference genomes that were predominantly derived from a limited number of individuals, which did not fully reflect the breadth of human genetic variation. Using this biased reference for the OurDNA communities will likely result in us missing crucial population-specific variation. Therefore, Google is prioritising its contributions to improve the pangenome reference, which can better capture global genetic diversity. Partnering with the Genomics team at Google Research, we are working to refine and apply these tools together, pairing Google's software development strength with our community and population genomics focus to ensure they can be effectively applied to increasingly diverse datasets by researchers around the globe.

Leveraging cutting edge resources

Similarly, operating flexibly and at scale requires us to take new approaches to our compute infrastructure. To this end, CPG has been using cloud infrastructure, including Google's Cloud Platform, for genomic analysis, enabling us to run reproducible, secure, and scalable analyses on a future-proofed and adaptable infrastructure. This collaboration paves the way for future work with Google to make genomic resources and cloud-based analysis pipelines more broadly accessible.

The community engagement, co-design and recruitment activities of the OurDNA program are additionally supported by Google Australia's Digital Future Initiative.

OurDNA funding

The OurDNA program is supported by a 5-year $10M grant from the Medical Research Future Fund's Genomics Health Futures Mission that commenced in 2022. This funding supports a national consortium of 44 investigators from 24 institutions with expertise spanning community engagement, population studies, large-scale genomic sample acquisition, data generation and analysis, clinical and population genetics, ethics, economics and implementation science. The community engagement, co-design and recruitment activities of the **OurDNA** program are additionally supported by Google Australia's Digital Future Initiative. Providing culturally-aligned and language-appropriate information to prospective participants is essential for the success of the **OurDNA** program and other clinical and research genomics programs. To this end, in collaboration with Professor Mary-Anne Young and the Garvan's Clinical Translation and Engagement Platform and following national consultation and bi-lingual focus groups, we have developed a rich resource of written and audio materials in plain English and translated into ten languag-es. This participant information project was supported by a genom-ics implementation project grant from Australian Genomics that concluded in March 2024.

Gene Function

Genomic medicine promises to use knowledge of an individual's genome to inform healthcare: predicting the diseases each of us is most at risk for, making more accurate diagnoses for people currently suffering from disease, and designing new therapies and personalised approaches to medicine to ensure every patient is getting the best possible treatment. However, realising this potential requires an understanding of how genetic variation impacts health and disease.

In our Gene Function research domain, we aim to build up a deeper understanding of the biological impact of genetic variation, and the insights this can give us into the role of specific genes in the human body.

Understanding the impact of genetic variants on blood cell biology

The TenK10K Phase 1 project is a collaboration with Joseph Powell's lab at Garvan to combine whole genome sequencing and single cell RNA-sequencing data from over 1,800 individuals from two existing cohorts, the Tasmanian Ophthalmic Biobank and BioHEART studies. While these cohorts are largely of European descent, this project has allowed CPG to build new expertise in cellular genomics in advance of expanding this effort to include thousands of OurDNA participants from diverse communities.

This is already by far the largest dataset of this type globally, and gives us unprecedented insights into the effects of understudied variation classes on gene expression at single cell resolution. Newly explored genetic variants include rare variation, structural variation, short tandem repeats and somatic variation. To help with validation of these complex variants, we also have long-read (PacBio) sequencing available for a subset of 25 individuals.

At the American Society of Human Genetics 2024 meeting, two of CPG's trainees, Anna Cuomo and Hope Tanudisastro, will be presenting platform talks on their work to build quantitative trait loci catalogues from this dataset that link common, rare and complex variation to gene expression across 28 identified clusters of blood cells of various types and states.

We are now working with the Powell lab to scale this project to over 15,000 individuals, and integrate in information from a number of additional profiling technologies, including single-cell ATAC-seq, proteomics, metabolomics, as well as data from iPSC-derived cell lines.

This project has allowed CPG to build new expertise in cellular genomics in advance of expanding this effort to include thousands of OurDNA participants from diverse communities.

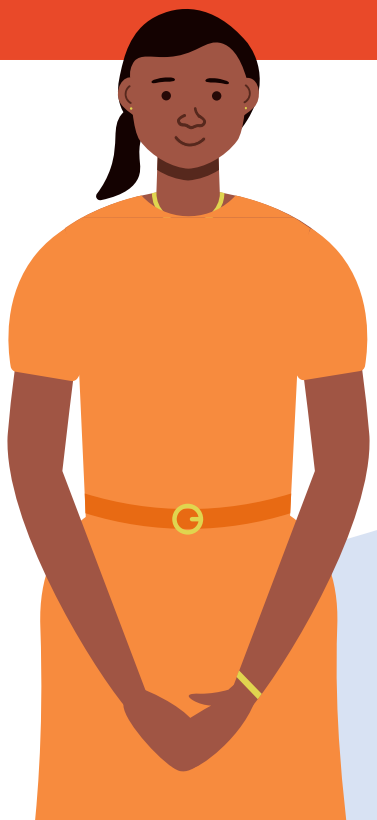
Using loss-of-function variants to identify new drug targets

Loss-of-function (LoF) variants are genetic changes that completely disrupt the function of a given gene. As such, they provide a powerful tool for understanding the biological function of the genes they affect, and the impact of ablating that function on human health and disease. For drug discovery, studies that carefully investigate the physical traits and health of individuals carrying specific LoF variants are a valuable resource when assessing the likely safety and efficacy of a novel drug target.

However, true LoF variants are generally uncommon, and, as such, predicted loss of function (pLoF) variants in healthy individuals from population cohorts are highly enriched for sequencing and annotation artefacts. This can make it difficult to draw conclusions based on pLoFs alone, and to prioritise specific pLoFs for follow-up studies.

We have therefore partnered with two large-scale industry consortia, Genes & Health and Open Targets, to manually curate approximately 20,000 variants that have been identified in broadly healthy population cohorts (gnomAD and Genes & Health) and are predicted to potentially result in loss of gene function. The output of this will be a highly curated set of likely loss of function variants, that can both be individually taken forward for careful follow-up study, and used as a whole as a training set for the enhancement of automated filtering approaches.

We have now developed an extensive standard operating procedure that can be used for the careful curation of loss-of-function variants, and have already used this to curate over 12,000 loss-of-function variants. These high-quality loss-of-function variant curations generated by CPG are being fed into the drug development pipelines of our pharmaceutical partners, supporting the future development of safe and effective therapeutics.



Rare Disease

Advances in genomics have revolutionised our ability to understand, diagnose and sometimes treat rare genetic disorders. However, despite these advances, a definitive genetic diagnosis - a critical step towards effective healthcare, and access to new therapeutics - is only achieved for around half of all affected Australian families. The CPG Rare Disease program delivers real-world impact to families impacted by rare diseases by developing new methods and approaches that provide faster and more accurate diagnoses.

The CaRDinal platform

The foundation of our rare disease activity is our CaRDinal collaborative rare disease analysis platform. The CaRDinal platform provides integrated, best-practice analysis and interrogation tools for genomic data from rare disease research cohorts. CaRDinal now supports short-read, long-read and RNAseq data types and provides comprehensive variant analysis including single nucleotide variant/indel, structural and copy number variants, short tandem repeat expansion analysis and mitochondrial variant analysis.

Through the CaRDinal platform, CPG provides the genomic analysis underpinning many of Australia's rare disease research projects of national significance. The CaRDinal platform currently hosts 31 national and state-based rare disease research cohorts from across Australia representing over 6,100 individuals from more than 3,600 families, Australia's largest rare disease research cohort. The harmonised genomic data and metadata provide CPG with a unique resource for both facilitating cross-cohort discovery (see page 26) and as training data for developing the next generation of analysis methods.

Development activity on the CaRDinal platform is currently focused on improving our support of analysis of long-read whole genome sequencing data as well as providing better integration of long and short-read datasets within the same family.

Talos automated reanalysis tool

Nearly half of all Australian genetic disease patients do not receive a genetic diagnosis from the first round of genomic analysis provided by clinical labs. One of the most impactful approaches to increase diagnosis rates is simply regularly reanalysing the genomic data incorporating changes in knowledge over time. Despite its effectiveness, systematic regular reanalysis has never been implemented at scale due to the substantial logistical and workforce requirements. Working with collaborators in the Victorian Clinical Genetics Service (VCGS), Microsoft Research, and the Broad Institute of MIT and Harvard, CPG has led the development of a novel automated approach to enable efficient periodic reanalysis at scale.

We have developed a new stand-alone analysis tool, Talos, designed to efficiently reanalyse variant calls from large cohorts of undiagnosed rare disease patients and identify variants that are highly likely to be causative of disease.

One of the most impactful approaches is systematically reanalysing the genomic data generated for the original test incorporating changes in knowledge over time

Key features of Talos include:

- Optimised for extreme specificity. Talos returns an average of less than one variant per trio, or two variants per singleton analysed.
- High sensitivity. Talos can identify over 80% of known diagnoses in our evaluation cohort.
- Integrated analysis of multiple variant types including single nucleotide variants, indels, structural and copy number variants.
- Automatic integration of the very latest gene-disease and pathogenic variant databases.

Talos is intended to be run periodically (i.e. monthly), highlighting candidate variants identified due to updated evidence such as new gene-disease relationships. Due to the combination of high specificity and real-time evidence updates, Talos enables effective continuous surveillance of cohorts of thousands of undiagnosed rare disease families to identify diagnoses based on changes in knowledge in days or weeks rather than years.

We have now used Talos to reanalyse the genomic data from ~3,200 clinical genomic tests from VCGS, resulting in over 185 novel diagnoses that have since been validated and reported back to patients by VCGS - representing nearly a 5% increase in diagnostic yield. This number is set to grow substantially as we continue to build on the Talos platform and apply it to thousands more undiagnosed Australian patients. We are also now refining the Talos software to make it easy to redeploy in other research and clinical labs around the world.

Harnessing Artificial Intelligence and Machine Learning to improve the diagnosis of Rare Disease.

Building from the success of our Talos reanalysis tool, the the CPG Rare Disease program is capitalising on recent advances in the field of AI to develop new approaches to improve our ability to diagnose rare disease. In June CPG was awarded an $8M MRFF grant that will allow us to invest heavily in this field.

The newly funded Australian Alliance for Secure Genomics and AI in Rare Disease (AASGARD) is a national program designed to accelerate the safe adoption of advanced analytics, including ML/AI, in genomic diagnostics. The program builds on established networks of academic and industry partners, spanning rare disease genomics experts, clinician-researchers, accredited diagnostic laboratories, experts in ML/AI, ethics, policy, and training, key industry partner Microsoft Research, and international partnerships with Genomics England and the Broad Institute of MIT and Harvard.

The AASGARD program aims to develop novel tools and pipelines, working with academic and industry partners to create new approaches for variant knowledge summarisation and clinical phenotype extraction, and integrating validated tools into portable open-source pipelines. These tools will be deployed at scale in both the CaRDinal research platform and at two leading diagnostic laboratories, resulting in new diagnoses for families, and real-world evaluation of outcomes.

Rare disease discovery enabled by CaRDinal platform data sharing

The CaRDinal platform and meta-cohort enabled CPG to contribute to a ground-breaking discovery that has rapidly developed in the last three months. Nicky Whiffin from Oxford's Center for Human Genetics recently observed mutations in the non-coding RNA RNU4-2 in undiagnosed individuals with neurodevelopmental disorders. Working with Nicky, CPG replicated this finding in the CaRDinal cohorts identifying six undiagnosed Australian patients with *de novo* mutations in this gene. We estimate that this single discovery may account for ~0.5% of all undiagnosed individuals with neurodevelopmental disorders.

This example demonstrates the power of CPG's CaRDinal platform. Consolidation of genomic data across such a wide range of of Australian rare disease cohorts provides a unique resource to identify novel causes of disease and efficient collaboration with the international rare disease community. Our manuscript describing this important discovery has been accepted for publication in Nature.

Rare Disease

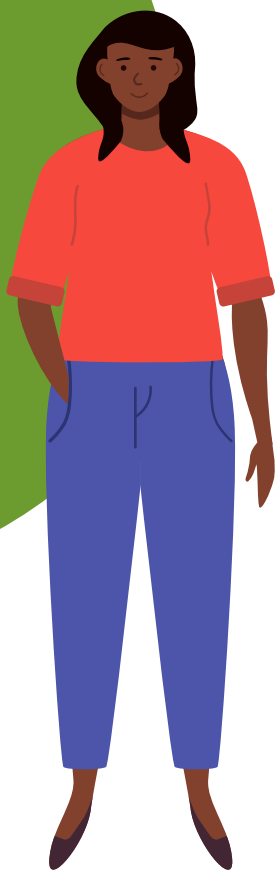
CASE REPORT

The impact of automated re-analysis with Talos

Reanalysis with Talos has delivered over 185 diagnoses to Australian patients to date. Many of these result from new gene-disease relationships, but Talos can also detect new variant types that were not detectable at the time of the original analysis.

For example, in 2018 Patient A presented with a phenotype consistent with Autosomal dominant polycystic kidney disease (ADPKD). They were recruited to the Australian Genomics KidGen flagship and underwent clinical whole genome sequencing with no diagnosis identified. As part of our systematic reanalysis of the KidGen flagship with Talos, we applied the advanced GATK-SV structural variant calling pipeline capable of detecting structural variant types that were not in scope for the previous generation pipelines.

Talos identified a balanced chromosomal translocation within the gene PKD2 that disrupts the function of the gene. Using Talos and the existing genomic data identified this diagnosis and provided the patient with a diagnostic report six years after the initial testing. Having this definitive genetic diagnosis will impact the clinical treatment of the patient and also inform important reproductive genetic counselling due to the specific consequences of balanced translocations.



Reanalysis with Talos has delivered over 185 diagnoses to Australian patients to date. Many of these result from new gene-disease relationships

Rare Disease funding

Many of the rare disease cohorts that form part of the CPG meta-cohort and are collaboratively analysed on CPG infrastructure are supported by grants from the Medical Research Future Fund (MRFF) Genomics Health Futures Mission and other MRFF funding schemes. These include initiatives in specific disease areas such as kidney disease (CIA Mallett, $3M), inherited bone marrow failure (CIA Blombery, $3M), ataxias and neuropathies (CIA Liang, $3M), mitochondrial disorders (CIA Thorburn, $3M), rhabdomyolysis (CIA Oates, $3M) as well as the Australian Undiagnosed Diseases Network (CIA Christodoulou, $3M). Other MRFF-funded rare disease collaborations focus on the development of novel tools and approaches to rare disease diagnosis including leveraging high-throughput functional assay data (CIA Rubin, $2.6M), deploying mass-spectrometry based functional genomics (CIA Stroud, $3M) and building a national long-read sequencing program (CIA Deveson, $3M). Our work to develop an automated approach to reanalysis of rare disease patient data is funded by both an MRFF GHFM grant (CIA Stark, $3M) and a research collaboration agreement with Microsoft Corporation and the Broad Institute of MIT and Harvard.

Equity and Advocacy

Participant Information Project – filling a critical gap in multilingual genomic resources

A major barrier for many diverse Australians in medical research is the paucity of culturally appropriate participant information and consent materials in community languages.

With funding support from Australian Genomics and the National Health and Medical Research Council to address this critical gap, between July 2022 to June 2024 the Centre's previous Inclusive Genomics team co-developed plain English materials translated into ten languages to support diverse Australian populations in genomics research participation and to promote inclusive, best practice approaches among researchers.

The in-language resources produced in multimedia formats throughout this Project contribute significantly to the Centre's real-world impact and ongoing efforts in:

enabling culturally and linguistically relevant mechanisms that drive equitable participation

addressing the chronic lack of inclusion and diversity in genomics research cohorts and reference datasets.

The Participant Information Project sets a best-practice benchmark for creating culturally aligned and language-appropriate participant information and education resources in clinical and population genomics research.

STEP 1

Drafting plain English materials

Topics selected and written in plain English by an interdisciplinary group of genetic counsellors, clinical researchers and educators.

- ***What is health research?***
- ***What is genetic and genomic research?***
- ***What are the benefits and risks of taking part in this research?***
- ***What happens to my information if I take part in research? (storage, data sharing, people)***

STEP 2

Expert consultation

Fifteen genomics researchers and cross-cultural communication experts invited to provide in-depth feedback on the materials.

STEP 3

Translation

Two translators per language group engaged to identify linguistic and cultural challenges with genomic terminology and ensure accuracy in translation.

STEP 4

Focus groups with community members

Ten focus groups with bilingual community members conducted for the ten priority language groups across Sydney and Melbourne, identified based on both clinical and population research needs.

Arabic	Dari	Farsi	Fijian	Hazaragi
Samoan	Tagalog	Tongan	Urdu	Vietnamese

STEP 5

Refining plain English and translations and establishing a best-practice model

Participant information and explanatory resources in 10 priority languages refined to offer

- Greater specificity for community members
- More culturally relevant and reflective definitions to the glossary
- More responsive translations based on participant needs and feedback

Resources to be made freely available to researchers under a Creative Commons license.

From Inclusive Genomics to Equity and Advocacy – looking ahead

Since March 2024, Inclusive Genomics has evolved to become a new function now known as Equity and Advocacy with a focus on developing and driving policy, advocacy and engagement strategies for CPG both externally and internally. Embedding advocacy and policy expertise in the Centre strengthens our strategic focus on building a multidisciplinary approach to advance equitable genomics.

We developed the Advocacy and Government Engagement Strategy 2024 – 2026 that serves as our forward roadmap for advocacy and government engagement efforts over the next two years – a critical period of scale and growth for the Centre. The Snapshot, found on page 29, provides an overview of our advocacy purpose, key stakeholder groups, current and planned actions that support our key focus areas to achieve impact.

Our continuous commitment to building cultural competence

In 2023-24, the Centre continued to embed cultural competence requirements and inclusion considerations into our operation. Investment in cultural competence training and professional development reflects best practices, fostering culturally safe environments in which our people operate and demonstrating our commitment to equity and inclusion both internally and externally.

CPG's advocacy against genetic discrimination

In June 2024, we formally requested Cabinet ministers to support legislation banning the use of genetic test results in life insurance underwriting to prevent genetic discrimination and safeguard the clinical potential of genomic medicine in Australia.

As our Advocacy and Government Engagement Strategy 2024 – 2026 progresses, we continue to advocate for sensible policymaking and ethical practices that eliminate access barriers to genomic research and benefit sharing.

Advocacy and Government Engagement Strategy

2024 – 2026 Snapshot

Purpose	Our vision to build a more equitable future for genomic medicine is at the forefront of our operation. This Strategy guides the development and implementation of our advocacy and government engagement priorities that (a) help strengthen CPG's capabilities, impact and commitment to equitable genomics at scale, (b) strategically expand on our successes.
Key stakeholders we work with	Industry, academic and clinical collaborators/partners who support our scientific pursuits Professional, peak organisations and consortia that bring relevant expertise and influence community standards for genomics Policy networks, government representatives and decision-makers in Australia that set relevant policy agendas and funding priorities Community members and leaders

Focus area 1: Equitable access and participation	Focus area 2: Development of sensible policies and responsible practices	Focus area 3: Adoption and integration of genomics innovation	Focus area 4: Open science and responsible data sharing	Focus area 5: Clinical translation of genomics research
We enable individuals from diverse backgrounds to participate in and benefit from genomics research.	We support policies and practices that increase equity in genomics research and implementation.	We lead the safe adoption and effective integration of advanced analytics to enhance disease diagnosis.	We pioneer open science and accessible research to deploy genomics broadly and equitably.	We are committed to translating genomic research into clinical practice to deliver real-world impact.
Key actions: • Drive equitable participation in genomic research through culturally appropriate engagement/ partnership with stakeholders that represent the community they serve • Develop and make readily available resources that facilitate community engagement at scale • Directly address critical gaps in the production/ sharing of diverse reference data to improve existing health disparities in priority populations	**Key actions:** • Contribute to policy making and advocacy efforts that tackle the ethical, legal and social barriers to equitable genomic advancements and benefit sharing • Foster interdisciplinary collaboration to build and advance best practices that drive health equity in genomics • Model appropriate use and scientific communication of genetic ancestry and population description labels	**Key actions:** • Facilitate the safe adoption/ integration of advanced analytics at scale, including AI and ML, to enable more accurate prediction and diagnosis of genetic disease • Reduce barriers to adoption and incorporation of these tools into national infrastructure for genomic information management, supported by evidence-based policy development and consultation with key government, research and clinical stakeholders	**Key actions:** • Model the release of publicly funded software as open-source code under a permissive license, and the early publication of scientific results as preprints at the time of journal submission to facilitate open access • Enable secure and responsible data sharing to maximise scientific and clinical impact, while respecting data sovereignty and community-specific access requirements and restrictions	**Key actions:** • Address unmet needs in rapid clinical decision-making through a highly accessible genetic database that can (1) deliver new diagnoses to underrepresented and diverse Australian families; (2) guide clinical interpretation/ therapeutic development • Translate project insights/ outcomes into compelling case studies and actionable recommendations to support investment in and clinical implementation of genomic medicine.

Team and Culture

The Centre's operating model was always destined to be unusual: our mission to build a more equitable and inclusive future for genomic medicine requires a non-profit model, but our focus on large-scale, long-term projects with clinical impact is non-standard for academia. We have thus built a culture and organisational structure that brings together aspects of both academic and startup models, benefiting from the flexibility and support of our two founding institutions, Garvan and MCRI.

The Centre now has over 45 staff members distributed across seven locations in Eastern Australia and New Zealand. We operate as a unified, remote-first organisation, with the vast majority of our work coordinated through the online communications platform Slack, and conducted on custom-built software platforms based on cloud infrastructure.

Our team members come from a wide variety of backgrounds, the majority from outside traditional academia, including software engineers, data scientists, project managers, operations specialists, and community engagement experts. To recruit and retain this diverse talent pool we have worked with Garvan and MCRI to develop new career paths, and to increase salaries for roles where we compete with the for-profit sector.

Our geographically distributed, remote-first model provides agility, and has helped us to effectively recruit talent in highly competitive areas such as software engineering; but it also creates challenges for effective communication and collaboration. To address this we have built a strong culture of online communication, processes for regular career check-ins, meetings, co-working days, and social events, and hold in-person whole-team gatherings 3-4 times a year to enhance collaborative work and social connections.

Given our mission, we see organisational innovation as every bit as critical as scientific and technical innovation. We experiment regularly with new processes and structures, and have built a culture of rapid feedback and continuous improvement fostered through regular project retrospectives that help us quickly identify what works, and fix what doesn't. This culture has allowed us to rapidly respond to challenges and new opportunities, such as the swift integration of new AI large language models (LLMs) into many areas of our work.

The next decade will see an overwhelming rate of change in genomics, data science, and therapeutics. Our team is well-placed to ensure we can respond quickly to that change, and take full advantage of the resulting opportunities to the benefit of the patients and communities we serve.

We see organisational innovation as every bit as critical as scientific and technical innovation

Our Mission

To establish respectful partnerships with diverse communities, collect and analyse genomic data at transformative scale, and drive genomic discovery and equitable genomic medicine in Australia

CPG Organisational Values

- We remember that our data come from people
- We do things that matter even if they are hard
- We share accountability
- We support each other
- We celebrate our differences

Near-term goals

Deliver scientific and clinical impact

We're generating and releasing new reference data from under-represented communities, new diagnoses for families affected by severe disease, and new understanding of the biological impact of genetic variation, while building the foundations for future scaling and impact across all of our programmes.

Better support our people

We're growing the team, bringing in new capacity and capabilities while maintaining high standards, and strengthening our processes around communications, feedback, management, and career development.

Strengthen and streamline our processes

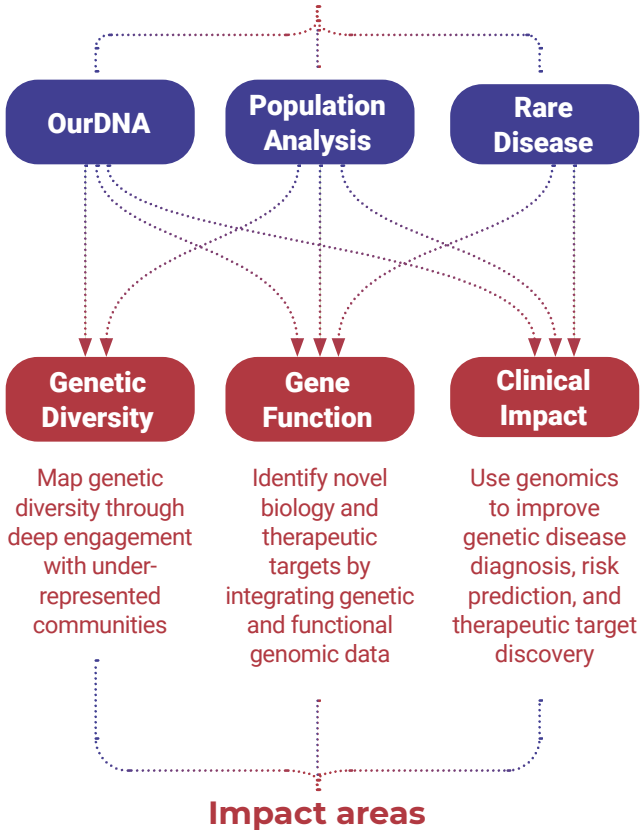
We're building a collective philosophy of continuous improvement: everyone plays a role in identifying inefficiencies and risks, and in strengthening our processes while maintaining speed and agility.

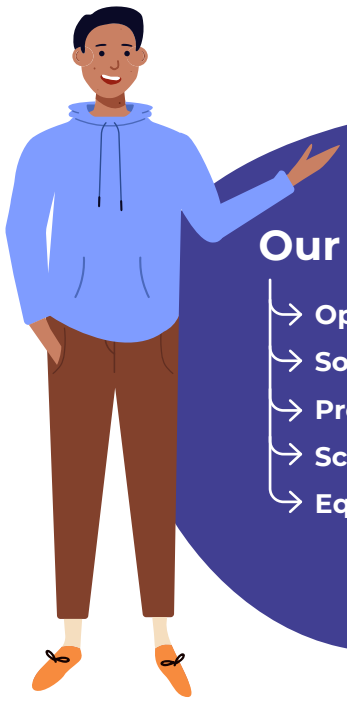
Build our external profile

We're increasing our external profile, establishing our leadership in genomic data management and analysis, community-led recruitment, and organisational innovation. We're sharing stories of our successes while remaining thoughtful and authentic.

Ensure long-term financial sustainability

We're securing new and diverse funding opportunities through grants, philanthropy, and industry engagement. We're working towards having a clear understanding of our future financial needs, and a well-defined strategy for meeting them.





Our foundations

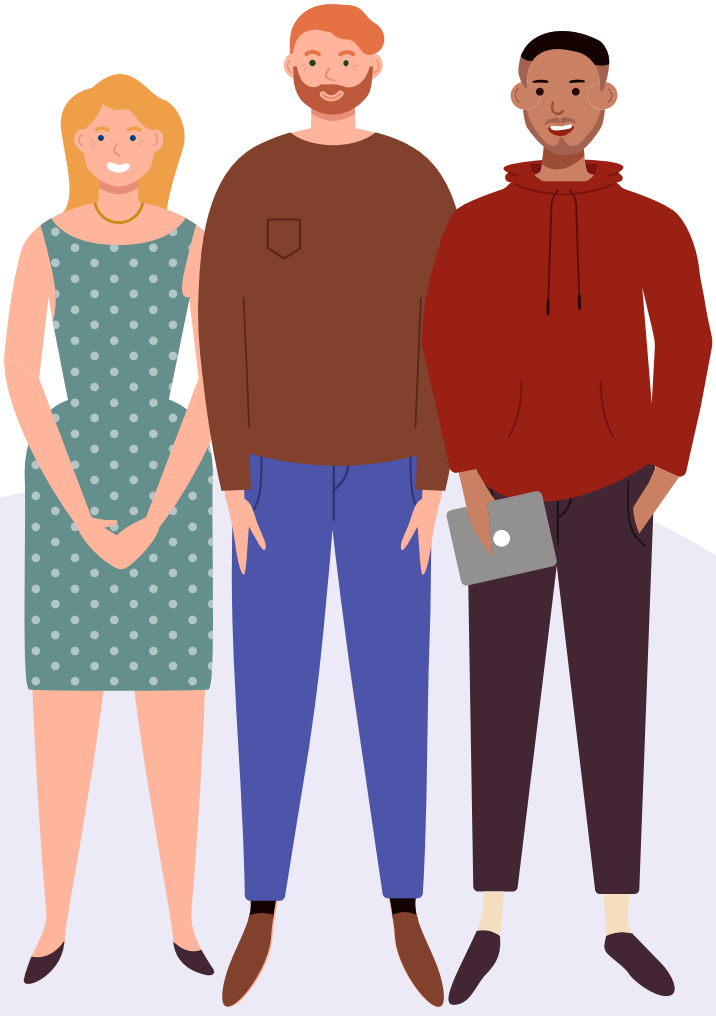
- Operations
- Software Platforms
- Project Management
- Scientific Affairs
- Equity & Advocacy

Our External Presence

Publications and conference presentations

Among the 20 publications contributed to by CPG staff in the past year, were research manuscripts published in Nature, Nature Genetics, New England Journal of Medicine and the American Journal of Human Genetics. Two of these were publications stemming from gnomAD work led initially by Daniel MacArthur in his previous role at the Broad Institute of MIT and Harvard, which were published simultaneously Nature and Nature Genetics with Daniel as second-last author, reflecting his initial leadership role and ongoing contributions to the gnomAD program. Daniel also contributed to an opinion piece on genomic newborn screening, led by Zornitza Stark in collaboration with UK-based colleagues, which was published in Nature Medicine. Members of the CPG rare disease analysis team collaborated with Marina Kennerson and Gina Ravenscroft on a manuscript published in Journal of the Peripheral Nervous System, which reports on several diagnoses of Charcot–Marie–Tooth neuropathy that were enabled by CPG's seqr platform and the automated reanalysis pipeline, Talos.

The work of the CPG was presented at research conferences in both oral and poster format including the American Society of Human Genetics Annual Meeting 2023, the National Multicultural Health and Wellbeing Conference 2023, the 2023 Australasian Biospecimen Network Association Conference and the European Society of Human Genetics Annual Meeting 2024.



External engagement opportunities

Hearts & Minds Investment Leaders Conference 2023

Hearts & Minds is a philanthropic group committed to supporting Australian medical research. Every year the Sohn Hearts & Minds Investment Leaders Conference brings together leading fund managers to present their investment ideas. Each conference also features two Australian research leaders, who are invited to share their insights on emerging areas of medical research. Alongside Misty Jenkins (WEHI), Daniel MacArthur featured in the 2023 conference at the Opera House in Sydney, where he presented on the future of genomic medicine.

David Danks Oration 2024

Daniel was invited to present the David Danks Oration at the University of Melbourne in August 2024, on "Human Diversity in the age of genomic medicine". This event, open to the public, attracted over 150 attendees from a wide variety of backgrounds.

SXSW Sydney 2024 conference panel

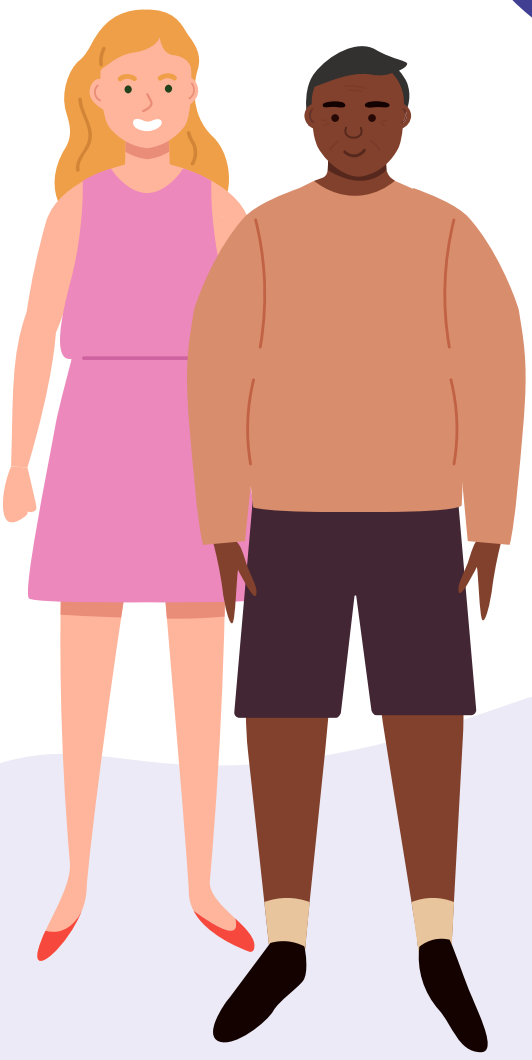
The genomic transformation of medicine

SXSW, which has a long history in the US as a forum for the gathering of thought leaders, launched an Australian partner event in 2023, bringing together local and international innovators in various domains from technology and gaming to music and film. SXSW Sydney 2024 will take place from 14-18 October and will feature a panel discussion proposed by the CPG. This session will explore the explosive change currently underway in the field of medicine, driven by exponential advances in genomics, data science, and therapeutics.

The panel will involve world-leading experts in each of these three fields, chaired by genomic medicine and child health advocate Sarah Murdoch: genomics expert, and Director of the CPG, Daniel MacArthur, DeepMind AI researcher Clare Bycroft, and stem cell therapeutics developer Enzo Porrello, covering current state and future opportunities for these transformative technologies, and grappling with the impending challenges of cost, data privacy, and equity in the genomic era.

"... if Australia can make the right decisions - investing in the right people, at the right scale - we have a real opportunity to be among the global leaders in this field, creating longer, healthier lives for everyone"

- Daniel MacArthur.

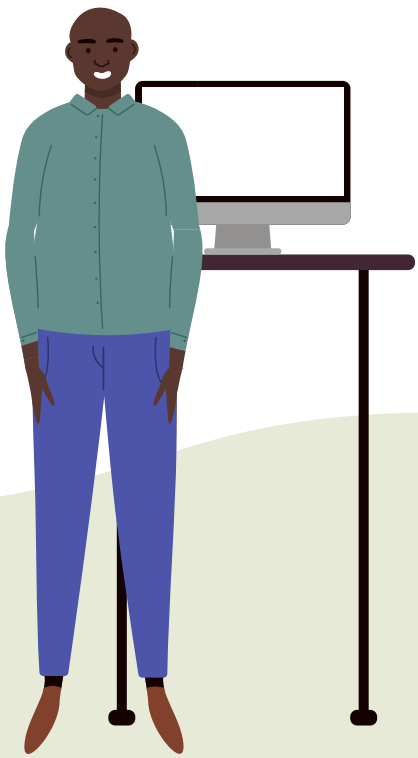


Funding

New funding secured
July 2023 - June 2024

Google partnership

Funding awarded over 5 years to accelerate the community engagement activities of the OurDNA program, expediting participant recruitment and ultimately the creation of the bioresource that will drive the future impact of the OurDNA program. This partnership agreement also formalises CPG's ongoing collaboration with Google Research, which aims to overcome reference biases to improve our ability to call population-specific variation in under-represented populations. We are working to build population-specific reference genomes using samples from OurDNA participants. Collaborators in the Genomics team at Google Research are then developing a variant caller that can effectively leverage these references to improve variant calling in our whole genome sequencing data.



2023 Medical Research Future Fund Genomics Health Futures Mission grant

A national platform for evaluation and integration of advanced analytics in the diagnosis of genetic disease

$8M awarded over 4 years for the establishment of the Australian Alliance for Secure Genomics and AI in Rare Disease (AASGARD), a program involving 32 investigators across Australia and internationally, designed to accelerate the safe adoption of advanced analytics, including ML/AI, in genomic diagnostics.

- **CPG funding $4.7M**
- **Clinical partners VCGS and SA Pathology $2.2M**
- **AI and data $0.6M**
- **Ethics and education $0.4M**

Cumulative funding totals from inception to June 2024

Government grants

CPG awarded to date: $21M

- National Health and Medical Research Council (2000001, 2009982)
- Medical Research Future Fund (2008820, 2007567, 2008249, 2007959, 2007548, 2007681, 2015969, 2016030, 2025138, 2025450, 2032931)
- NIH (Broad Institute of MIT and Harvard subawards 5U01HG011755, 1U24HG011450)

Success rates

- CPG-led success 100% (3/3; $21.4M)
- Other success 71% (27/38; $95.5M)

Industry agreements

CPG awarded to date: $5.6M (including $696k infrastructure support)

- Microsoft Corporation, Broad Institute of MIT and Harvard subaward
- Google Digital Future Initiative
- Open Targets
- Genes & Health

The Future

The Centre is still at the beginning of two journeys, both moving at dizzying speed. Firstly, genomic medicine as a field is rapidly evolving from disconnected genomic, computational, and healthcare technologies into an ecosystem that will ultimately provide longer, healthier lives for everyone. Secondly, over the last four years our Centre has moved from an idea into an established, highly functional team of over 45 expert staff; but we have much further to go to achieve our goals.

Here are some of our major focus areas over the next 3-5 years:

Building the foundations for equitable genomic medicine. Over the next three years, the first phase of the OurDNA program - collection of DNA, cells, and genomic data from 10,000 diverse Australians - will rapidly improve our ability to diagnose families with genetic disease from currently underrepresented communities, and provide insight into the impact of genetic variation across Australian communities. However, this is only the first step in a much more ambitious program to lay the groundwork for equitable healthcare in Australia. Over the next decade, we need to build much larger resources of genomic and clinical data, ultimately spanning hundreds of thousands of Australians, to ensure the next generation of genomic and therapeutic technologies are designed to benefit all of our communities.

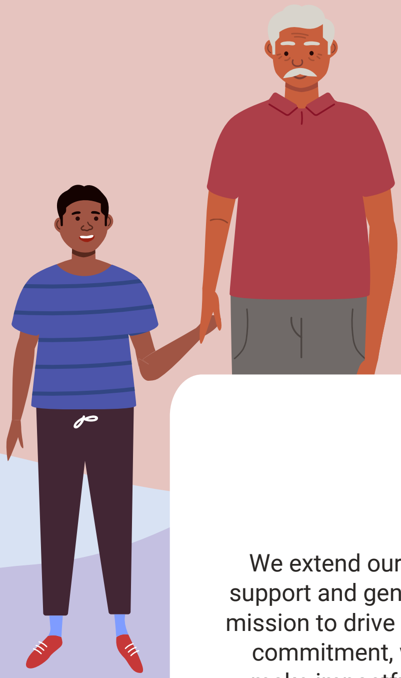
Integrating new technologies into research and clinical practice. The Centre is already driving large-scale research use of cutting-edge approaches, such as long-read whole-genome sequencing and artificial intelligence, to diagnose patients. Our well-established connections with clinical laboratories, and the newly-funded AASGARD consortium for AI-guided genomic diagnosis, provide powerful opportunities to move tools swiftly into clinical use. Our assembly of the OurDNA and CaRDinal cohorts means we can rapidly test new genomic technologies across diverse population cohorts and disease patients, and our cloud-based data platform, strong software expertise, and access to massive genomic datasets position us well for the deployment of new data science (including AI) approaches at enormous scale.

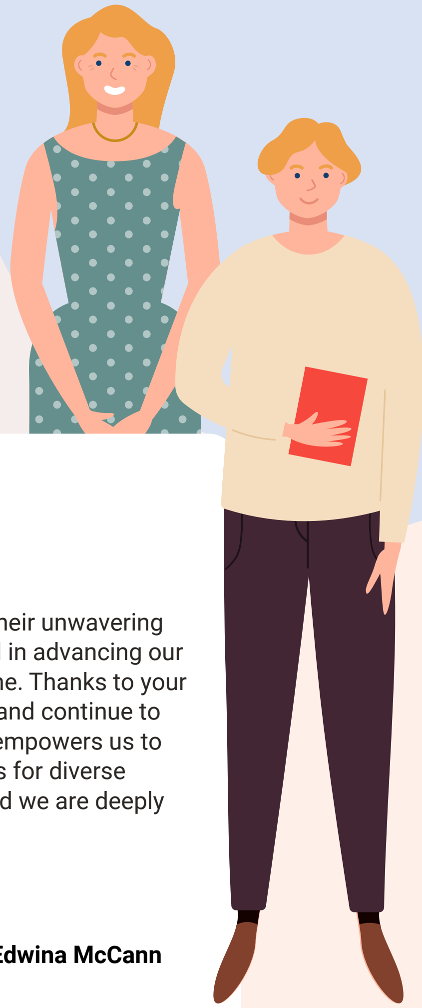
Achieving clinical impact in both diagnostics and therapeutics. Our clinical impact focus will move increasingly from the diagnosis to the treatment of severe genetic disease. The explosion of new therapeutic modalities for rare genetic diseases (such as RNA therapies and genome editing) will create new cures for many severe conditions, but we must ensure these treatments work across all ancestries. Our combination of population genomics and rare disease expertise positions us well to address this challenge, and this will be a major focus over the coming years. In parallel, the completion of the OurDNA and TenK10K resources will provide powerful platforms for identifying new drug targets and testing them at scale across living cells from diverse communities, and more partnerships with the biotech sector.

Continuing to innovate as an organisation

Living through an era of exponential change means that our success as an organisation will require responding quickly to shifting technological and social landscapes. We have built a modern, nimble organisational structure and a strong culture of continuous improvement, and we will continue to trial new technologies, processes, and operational models as we scale our team and projects, knitting together the best aspects of academic and startup culture.

We are well-placed to rapidly deploy the coming wave of genomic and data science innovations, and to ensure they result in improved clinical outcomes for all Australians.





Thank you

We extend our heartfelt gratitude to our funders and donors for their unwavering support and generosity. Your contributions have been instrumental in advancing our mission to drive genomic discovery and equitable genomic medicine. Thanks to your commitment, we have achieved significant milestones this year and continue to make impactful strides in our research. Your belief in our vision empowers us to push the boundaries of science and improve health outcomes for diverse communities. Together, we are making a profound difference, and we are deeply grateful for your partnership and trust.

Allie & Dan Rosen

Dominic & Emma Stevens via the Leslie Stevens Fund for Newborn Care

Edwina McCann

Gillon & Laura Mclachlan

Hearts and Minds Investments Limited

Jack & Sharon Cowin

Jack Zhang

John Brown Cook Foundation

Joseph Palmer Foundation

Lysia O'Keefe

Marc Freeman & Camilla Freeman-Topper

Miffany & James Blythe

Niall Lenahan

P&S Bassat Foundation

Rajeev Natarajan

River Capital Foundation

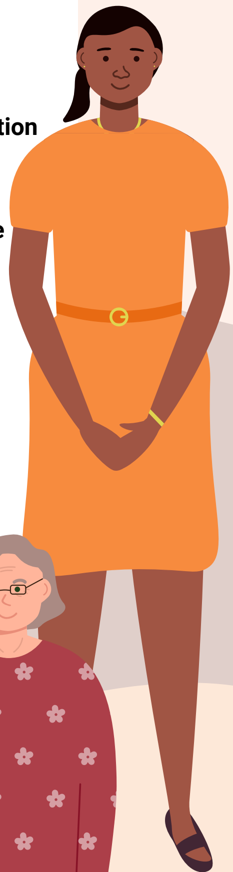
Robby & Sarah Ingham

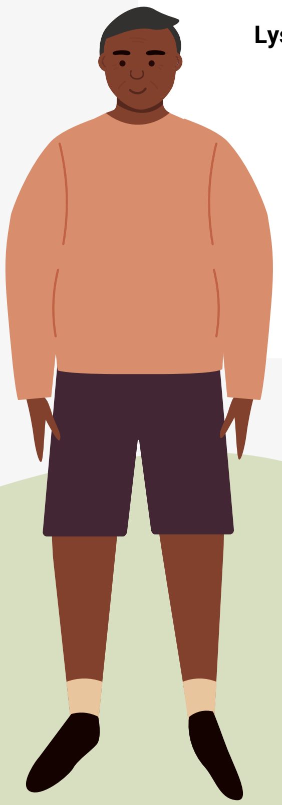
Sally White OAM

The V & C Comino Foundation

Valeska Bloch

Vas Katos & Nicole Georgelos





APPENDIX

Publications

* Names of CPG authors are underlined

Beerepoot S, Verbeke JIML, Plantinga M, Nierkens S, Pouwels PJW, Wolf NI, Simons C, van der Knaap MS. **Leukoencephalopathy with calcifications, developmental brain abnormalities and skeletal dysplasia due to homozygosity for a hypomorphic CSF1R variant: A report of three siblings**. Am J Med Genet A. 2024 Jun 27:e63800. doi: 10.1002/ajmg.a.63800. PMID: 38934054.

Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD, Poterba T, Wilson MW, Tarasova Y, Phu W, Grant R, Yohannes MT, Koenig Z, Farjoun Y, Banks E, Donnelly S, Gabriel S, Gupta N, Ferriera S, Tolonen C, Novod S, Bergelson L, Roazen D, Ruano-Rubio V, Covarrubias M, Llanwarne C, Petrillo N, Wade G, Jeandet T, Munshi R, Tibbetts K; Genome Aggregation Database Consortium; O'Donnell-Luria A, Solomonson M, Seed C, Martin AR, Talkowski ME, Rehm HL, Daly MJ, Tiao G, Neale BM, MacArthur DG, Karczewski KJ. **A genomic mutational constraint map using variation in 76,156 human genomes**. Nature. 2024 Jan;625(7993):92-100. doi: 10.1038/s41586-023-06045-0. PMID: 38057664.

Cuomo ASE, Nathan A, Raychaudhuri S, MacArthur DG, Powell JE. **Single-cell genomics meets human genetics**. Nat Rev Genet. 2023 Aug;24(8):535-549. doi: 10.1038/s41576-023-00599-5. PMID: 37085594.

Farbehi N, Neavin DR, Cuomo ASE, Studer L, MacArthur DG, Powell JE. **Integrating population genetics, stem cell biology and cellular genomics to study complex human diseases**. Nat Genet. 2024 May;56(5):758-766. doi: 10.1038/s41588-024-01731-9. PMID: 38741017.

Grosz BR, Parmar JM, Ellis M, Bryen S, Simons C, Reis ALM, Stevanovski I, Deveson IW, Nicholson G, Laing N, Wallis M, Ravenscroft G, Kumar KR, Vucic S, Kennerson ML. **A deep intronic variant in MME causes autosomal recessive Charcot-Marie-Tooth neuropathy through aberrant splicing**. J Peripher Nerv Syst. 2024 Jun;29(2):262-274. doi: 10.1111/jns.12637. PMID: 38860315.

Guo MH, Francioli LC, Stenton SL, Goodrich JK, Watts NA, Singer-Berk M, Groopman E, Darnowsky PW, Solomonson M, Baxter S; gnomAD Project Consortium; Tiao G, Neale BM, Hirschhorn JN, Rehm HL, Daly MJ, O'Donnell-Luria A, Karczewski KJ, MacArthur DG, Samocha KE. **Inferring compound heterozygosity from large-scale exome sequencing data**. Nat Genet. 2024 Jan;56(1):152-161. doi: 10.1038/s41588-023-01608-3. PMID: 38057443.

Hort Y, Sullivan P, Wedd L, Fowles L, Stevanovski I, Deveson I, Simons C, Mallett A, Patel C, Furlong T, Cowley MJ, Shine J, Mallawaarachchi A. **Atypical splicing variants in PKD1 explain most undiagnosed typical familial ADPKD**. NPJ Genom Med. 2023 Jul 7;8(1):16. doi: 10.1038/s41525-023-00362-z. PMID: 37419908.

Hudson R, Abeysekera N, Wolski P, Simons C, Francis L, Farnsworth E, Bennetts B, Patel C, Spijker S, Mallett A. **De novo HNF4A-associated atypical Fanconi renal tubulopathy syndrome**. J Nephrol. 2024 Jan;37(1):191-197. doi: 10.1007/s40620-023-01666-0. PMID: 37308774.

Ingles J, MacArthur DG. **The moral and practical urgency of increasing diversity in genomics**. Eur Heart J. 2023 Dec 21;44(48):5157-5159. doi: 10.1093/eurheartj/ehad365. PMID: 37377076.

APPENDIX

Publications

Kipkemoi P, Kim HA, Christ B, O'Heir E, Allen J, Austin-Tse C, Baxter S, Brand H, Bryant S, Buser N, de Menil V, Eastman E, Murugasen S, Galvin A, Kombe M, Ngombo A, Mkubwa B, Mwangi P, Kipkoech C, Lovgren A, MacArthur DG, Melly B, Mwangasha K, Martin A, Nkambule LL, Sanchis-Juan A, Singer-Berk M, Talkowski ME, VanNoy G, van der Merwe C; NeuroDev Project; Newton C, O'Donnell-Luria A, Abubakar A, Donald KA, Robinson EB. **Phenotype and genetic analysis of data collected within the first year of NeuroDev**. Neuron. 2023 Sep 20;111(18):2800-2810.e5. doi: 10.1016/j.neuron.2023.06.010. PMID: 37463579.

Lemire G, Sanchis-Juan A, Russell K, Baxter S, Chao KR, Singer-Berk M, Groopman E, Wong I, England E, Goodrich J, Pais L, Austin-Tse C, DiTroia S, O'Heir E, Ganesh VS, Wojcik MH, Evangelista E, Snow H, Osei-Owusu I, Fu J, Singh M, Mostovoy Y, Huang S, Garimella K, Kirkham SL, Neil JE, Shao DD, Walsh CA, Argilli E, Le C, Sherr EH, Gleeson JG, Shril S, Schneider R, Hildebrandt F, Sankaran VG, Madden JA, Genetti CA, Beggs AH, Agrawal PB, Bujakowska KM, Place E, Pierce EA, Donkervoort S, Bönnemann CG, Gallacher L, Stark Z, Tan TY, White SM, Töpf A, Straub V, Fleming MD, Pollak MR, Õunap K, Pajusalu S, Donald KA, Bruwer Z, Ravenscroft G, Laing NG, MacArthur DG, Rehm HL, Talkowski ME, Brand H, O'Donnell-Luria A. **Exome copy number variant detection, analysis, and classification in a large cohort of families with undiagnosed rare genetic disease**. Am J Hum Genet. 2024 May 2;111(5):863-876. doi: 10.1016/j.ajhg.2024.03.008. PMID: 38565148.

Lowther C, Valkanas E, Giordano JL, Wang HZ, Currall BB, O'Keefe K, Pierce-Hoffman E, Kurtas NE, Whelan CW, Hao SP, Weisburd B, Jalili V, Fu J, Wong I, Collins RL, Zhao X, Austin-Tse CA, Evangelista E, Lemire G, Aggarwal VS, Lucente D, Gauthier LD, Tolonen C, Sahakian N, Stevens C, An JY, Dong S, Norton ME, MacKenzie TC, Devlin B, Gilmore K, Powell BC, Brandt A, Vetrini F, DiVito M, Sanders SJ, MacArthur DG, Hodge JC, O'Donnell-Luria A, Rehm HL, Vora NL, Levy B, Brand H, Wapner RJ, Talkowski ME. **Systematic evaluation of genome sequencing for the diagnostic assessment of autism spectrum disorder and fetal structural anomalies**. Am J Hum Genet. 2023 Sep 7;110(9):1454-1469. doi: 10.1016/j.ajhg.2023.07.010. PMID: 37595579.

Marchant RG, Bryen SJ, Bahlo M, Cairns A, Chao KR, Corbett A, Davis MR, Ganesh VS, Ghaoui R, Jones KJ, Kornberg AJ, Lek M, Liang C, MacArthur DG, Oates EC, O'Donnell-Luria A, O'Grady GL, Osei-Owusu IA, Rafehi H, Reddel SW, Roxburgh RH, Ryan MM, Sandaradura SA, Scott LW, Valkanas E, Weisburd B, Young H, Evesson FJ, Waddell LB, Cooper ST. **Genome and RNA sequencing boost neuromuscular diagnoses to 62% from 34% with exome sequencing alone**. Ann Clin Transl Neurol. 2024 May;11(5):1250-1266. doi: 10.1002/acn3.52041. PMID: 38544359.

McGurk KA, Zhang X, Theotokis P, Thomson K, Harper A, Buchan RJ, Mazaika E, Ormondroyd E, Wright WT, Macaya D, Pua CJ, Funke B, MacArthur DG, Prasad SK, Cook SA, Allouba M, Aguib Y, Yacoub MH, O'Regan DP, Barton PJR, Watkins H, Bottolo L, Ware JS. **The penetrance of rare variants in cardiomyopathy-associated genes: A cross-sectional approach to estimating penetrance for secondary findings**. Am J Hum Genet. 2023 Sep 7;110(9):1482-1495. doi: 10.1016/j.ajhg.2023.08.003. PMID: 37652022.

Passchier EMJ, Bisseling Q, Helman G, van Spaendonk RML, Simons C, Olsthoorn RCL, van der Veen H, Abbink TEM, van der Knaap MS, Min R. **Megalencephalic leukoencephalopathy with subcortical cysts: a variant update and review of the literature**. Front Genet. 2024 Feb 29;15:1352947. doi: 10.3389/fgene.2024.1352947. PMID: 38487253.

Singer-Berk M, Gudmundsson S, Baxter S, Seaby EG, England E, Wood JC, Son RG, Watts NA, Karczewski KJ, Harrison SM, MacArthur DG, Rehm HL, O'Donnell-Luria A. **Advanced variant classification framework reduces the false positive rate of predicted loss-of-function variants in population sequencing data**. Am J Hum Genet. 2023 Sep 7;110(9):1496-1508. doi: 10.1016/j.ajhg.2023.08.005. PMID: 37633279.

Tanudisastro HA, Deveson IW, Dashnow H, MacArthur DG. **Sequencing and characterizing short tandem repeats in the human genome**. Nat Rev Genet. 2024 Jul;25(7):460-475. doi: 10.1038/s41576-024-00692-3. PMID: 38366034.

Töpf A, Cox D, Zaharieva IT, Di Leo V, Sarparanta J, Jonson PH, Sealy IM, Smolnikov A, White RJ, Vihola A, Savarese M, Merteroglu M, Wali N, Laricchia KM, Venturini C, Vroling B, Stenton SL, Cummings BB, Harris E, Marini-Bettolo C, Diaz-Manera J, Henderson M, Barresi R, Duff J, England EM, Patrick J, Al-Husayni S, Biancalana V, Beggs AH, Bodi I, Bommireddipalli S, Bönnemann CG, Cairns A, Chiew MT, Claeys KG, Cooper ST, Davis MR, Donkervoort S, Erasmus CE, Fassad MR, Genetti CA, Grosmann C, Jungbluth H, Kamsteeg EJ, Lornage X, Löscher WN, Malfatti E, Manzur A, Martí P, Mongini TE, Muelas N, Nishikawa A, O'Donnell-Luria A, Ogonuki N, O'Grady GL, O'Heir E, Paquay S, Phadke R, Pletcher BA, Romero NB, Schouten M, Shah S, Smuts I, Sznajer Y, Tasca G, Taylor RW, Tuite A, Van den Bergh P, VanNoy G, Voermans NC, Wanschitz JV, Wraige E, Yoshimura K, Oates EC, Nakagawa O, Nishino I, Laporte J, Vilchez JJ, MacArthur DG, Sarkozy A, Cordell HJ, Udd B, Busch-Nentwich EM, Muntoni F, Straub V. **Digenic inheritance involving a muscle-specific protein kinase and the giant titin protein causes a skeletal muscle myopathy**. Nat Genet. 2024 Mar;56(3):395-407. doi: 10.1038/s41588-023-01651-0. PMID: 38429495.

Wieder N, D'Souza EN, Martin-Geary AC, Lassen FH, Talbot-Martin J, Fernandes M, Chothani SP, Rackham OJL, Schafer S, Aspden JL, MacArthur DG, Davies RW, Whiffin N. **Differences in 5'untranslated regions highlight the importance of translational regulation of dosage sensitive genes**. Genome Biol. 2024 Apr 29;25(1):111. doi: 10.1186/s13059-024-03248-0. PMID: 38685090.

Wojcik MH, Lemire G, Berger E, Zaki MS, Wissmann M, Win W, White SM, Weisburd B, Wieczorek D, Waddell LB, Verboon JM, VanNoy GE, Töpf A, Tan TY, Syrbe S, Strehlow V, Straub V, Stenton SL, Snow H, Singer-Berk M, Silver J, Shril S, Seaby EG, Schneider R, Sankaran VG, Sanchis-Juan A, Russell KA, Reinson K, Ravenscroft G, Radtke M, Popp D, Polster T, Platzer K, Pierce EA, Place EM, Pajusalu S, Pais L, Õunap K, Osei-Owusu I, Opperman H, Okur V, Oja KT, O'Leary M, O'Heir E, Morel CF, Merkenschlager A, Marchant RG, Mangilog BE, Madden JA, MacArthur D, Lovgren A, Lerner-Ellis JP, Lin J, Laing N, Hildebrandt F, Hentschel J, Groopman E, Goodrich J, Gleeson JG, Ghaoui R, Genetti CA, Gburek-Augustat J, Gazda HT, Ganesh VS, Ganapathi M, Gallacher L, Fu JM, Evangelista E, England E, Donkervoort S, DiTroia S, Cooper ST, Chung WK, Christodoulou J, Chao KR, Cato LD, Bujakowska KM, Bryen SJ, Brand H, Bönnemann CG, Beggs AH, Baxter SM, Bartolomaeus T, Agrawal PB, Talkowski M, Austin-Tse C, Abou Jamra R, Rehm HL, O'Donnell-Luria A. **Genome Sequencing for Diagnosing Rare Diseases**. N Engl J Med. 2024 Jun 6;390(21):1985-1997. doi: 10.1056/NEJMoa2314761. PMID: 38838312.

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

The Centre for Population Genomics is a joint national initiative of Garvan Institute of Medical Research and the Murdoch Children's Research Institute. The Institutes are national leaders in genomic research, with complementary strengths spanning large-scale genomics, data science, population health, and clinical impact.

populationgenomics.org.au