centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

Centre for Population Genomics

Annual Report 2022

Agenda

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

01	Achievements	• Grant Success • CPG Leaders • Impact: Science, Data & Community Engagement • National Influence • Our Partners
02	Our Science Strategy	• CPG Impact Mission • Path to Impact • Strategic Focus • OurDNA & Population Analysis • Rare Disease Analysis
03	Initiatives, Outcomes and Next Steps	• Project Management • Data Infrastructure • Inclusive Genomics
04	Areas of Ongoing Focus	• Operational Scalability • Participant-centric Science • Remote-first Environment • Professional Capability
05	Funding and Finance	• Funding Structure • Philanthropic Funding • Government & Industry Funding • Financial Report

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*

Achievements

Grant and Industry Funding

centre for population genomics

murdoch children's research institute | Garvan Institute of Medical Research

$18m

Government

Federal grant success rate of 65% for applications involving CPG

- Medical Research Future Fund (MRFF)
- National Health and Medical Research Council (NHMRC)
- Australian Genomics
- National Institutes of Health (NIH)

$1m

Industry

Partners in the United Kingdom and the United States

- Microsoft funding via collaboration with Broad Institute
- Open Targets with Sanger Institute
- >$2M in funding currently in negotiation over next 12 months

- CPG has attracted over $19M in external funding in the last 18 months
- The successful grants are highly aligned with our strategic goals (large-scale funding for rare disease, and the establishment of a reference database)
- This includes federal grants as well as industry agreements (Microsoft, Open Targets)
- Total Government funding confirmed to date: $18,260,064 (Aus Gov and NIH)
- Federal grants success rate: 65% for federal grants involving CPG as a partner; 2/2 CPG-led grants successful to date

Leadership: New Hires

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

Cas Simons

Rare Disease Analysis Lead

previously faculty member at MCRI

Katie de Lange

Population Analysis Lead

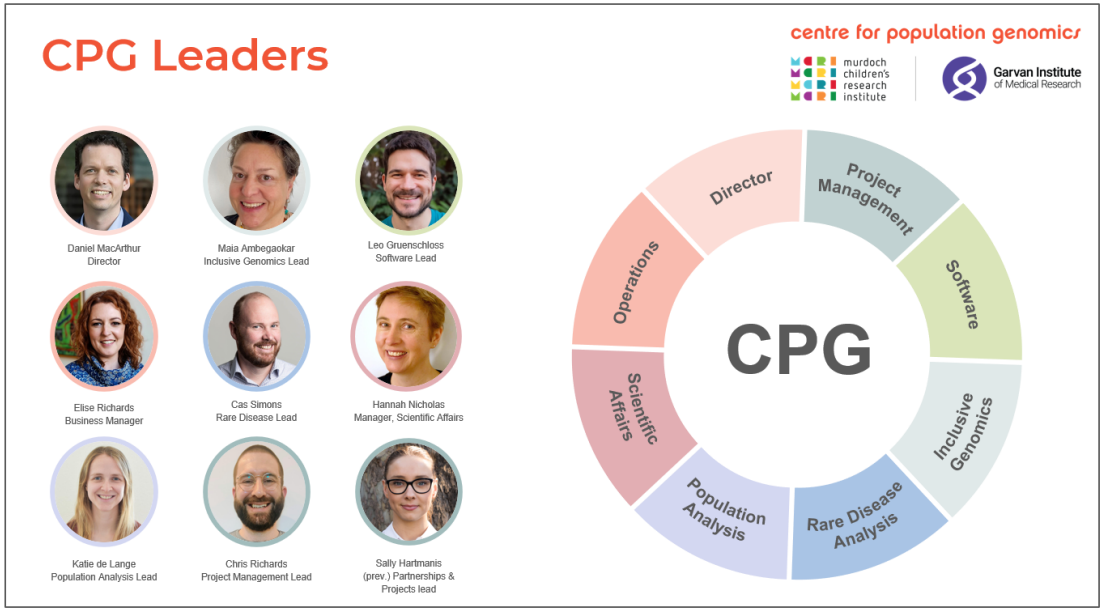
previously Data Scientist at Novartis

Elise Richards

Business Manager

previously Director of Operations at Wintermute Biomedical

Since our last annual report we've brought in three new team leads, completing our leadership team, with Cas and Katie building two new Analysis Teams focused on delivering scientific and clinical impact from our data generation projects, and Elise now overseeing operations across the Centre, liaising with teams at both institutes.



The core of CPG is eight teams comprising 35 staff based in Sydney, Melbourne and New Zealand.

Team leads come from a wide variety of backgrounds (Google, biotech, pharma, community engagement, academia)

Teams are entirely composed of professional (non-academic track) staff - CPG has built new career structures and KPIs to align with impact.

In 2022 we have bid farewell to Sally Hartmanis, who built the Centre's project management infrastructure but has accepted a PhD position in Oxford, and welcomed Chris Richards as the new Project Management Lead

CPG Research Faculty

centre for population genomics

murdoch children's research institute

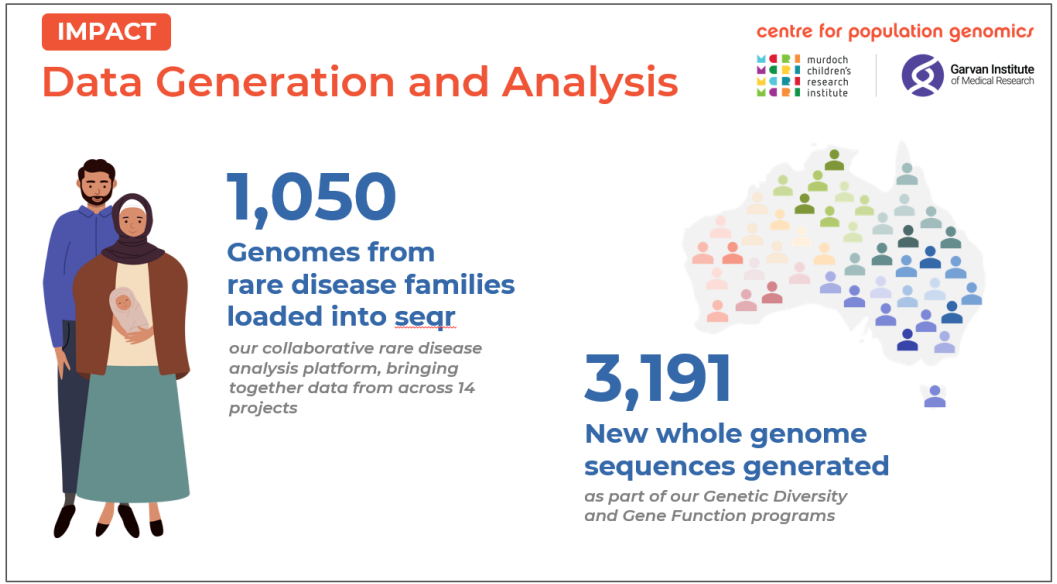
Garvan Institute of Medical Research

Jodie Ingles
Clinical Genomics

Owen Siggs
Genomic Medecine

Ira Deveson
Genomic Technologies

- **Three** CPG-affiliated faculty research groups (36 group members) at **Garvan**
- Faculty groups define their own research and fund-raising strategies, but collaborate closely with CPG to make use of central resources and data
- Two groups focused on **rare disease diagnosis** and a third focused on **genomic technology**
- Plans for further CPG-affiliated research groups at MCRI



- The last 12 months have seen rapid progress in data generation and aggregation
- In our Rare Disease programs we've either generated or reprocessed over 1,000 whole genomes from patients and unaffected relatives in families affected by rare disease, and all of these data sets are now loaded into the seqr platform that the CPG has established for collaborative analysis
- We have also overseen the sequencing of more than 3,000 new whole genomes, including the largest cohort of Australian Indigenous individuals sequenced to date, and two cohorts with accompanying large-scale cellular genomic data
- These data sets have provided critical pilot projects for our longer-term goals, as well as opportunities for direct clinical and scientific impact

IMPACT

Community engagement

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

Key partnerships

Established with two leading community groups

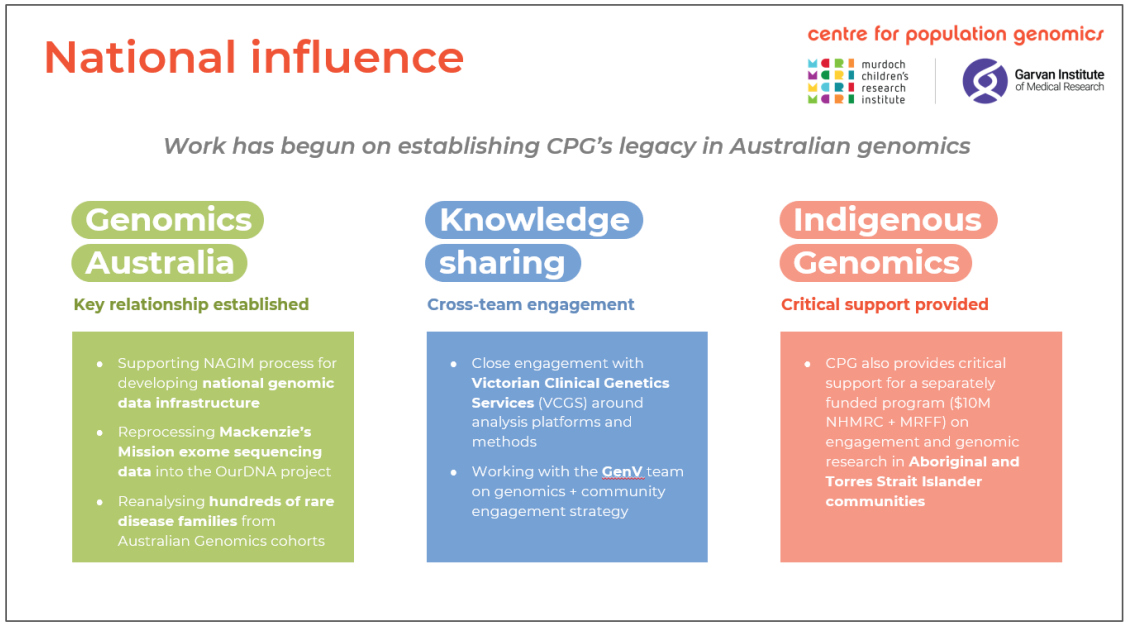
Australian Red Cross Lifeblood®

FECCA

CPG website

Release scheduled for early September

- In preparation for the intense community engagement work required for the OurDNA reference database project, we've established strong relationships with key community groups representing diverse communities, as well as a strategic partnership with the blood donation service Lifeblood
- The CPG website has been designed in consultation with both institutes and will be launched in mid-September



- Australia is well-placed to be an international leader in genomic medicine, because of our population size, national clinical genomics networks, strong research community, and public health system
- To enable CPG to help shape the future of genomics in the country we have built relationships with multiple key stakeholders, including Australian Genomics (soon to be Genomics Australia), multiple clinical labs, large biobanks and cohorts, and a new national consortium focused on Indigenous genomics
- We continue to explore other critical areas including:
    - Opportunities for regional impact through collaboration with groups in Aotearoa and the broader Pacific region
    - Groups in scientific areas critical for our future impact, such as cellular genomics and stem cells

Our Partners

- We have established a number of **high-impact partnerships** with key local and international stakeholders and organisations
- Several of these organisations have committed **direct research support** or **significant in-kind contributions**
- These partnerships span **academia**, **not-for-profits**, **philanthropic** organisations and **industry**

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

BROAD INSTITUTE

Google

Sano

illumina

Microsoft

SONIC HEALTHCARE

Australian Red Cross Lifeblood

Australian Genomics

GENV

SaxInstitute | 45 and Up Study

INVITAE

FECCA

VCGS leaders in genetic health

We've also established formal partnerships with other academic, for-profit, and non-profit organisations who are aligned with our delivery goals. In several cases these partnerships include substantial direct research support (e.g. Microsoft, Australian Genomics) or significant in-kind contributions (e.g. reagent discounts from Illumina; discounted data storage and compute from Google).

Scientific Strategy

Following the development of our two new Analysis teams we've formalised the Centre's scientific strategy, which allows us to organise our ongoing programs into clear domains, and to make streamlined decisions about future opportunities.

CPG Impact Mission

The Centre aims to have impact in all three of the areas where genomics is most clearly set to transform healthcare: in the diagnosis, prevention, and treatment of disease. While the Centre will continue to explicitly steer clear of direct clinical practice, all our programs and projects are selected on the basis of having a clear path towards clinical impact.

DIAGNOSIS

Enable the diagnosis of disease through improving genomic reference databases

PREVENTION

Identify high-risk patient populations prior to disease onset

TREATMENT

Leverage genetic insights to accelerate drug discovery and personalise medicine

Genetic variation and health

centre for population genomics

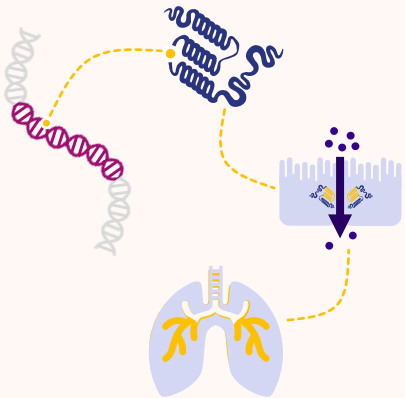
murdoch children's research institute

Garvan Institute of Medical Research

Most of our **genetic code** is identical from person to person

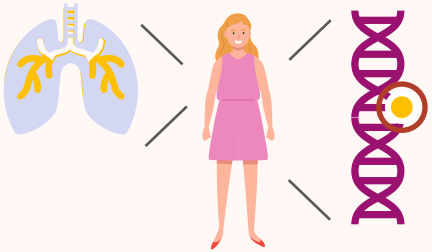
But everyone has small differences in their DNA that make them unique: these differences are called **genetic variation**

This genetic variation alters our **physical traits** by affecting the products of **genes**, such as **proteins**



The effect of each **genetic variant** could include changes to the function of the protein itself, or changes in the amount of protein produced

Most genetic variation has a relatively small impact, but sometimes a variant can have a large effect



Rare **genetic diseases** are (often severe) disorders caused by rare genetic variation that has a very large negative effect

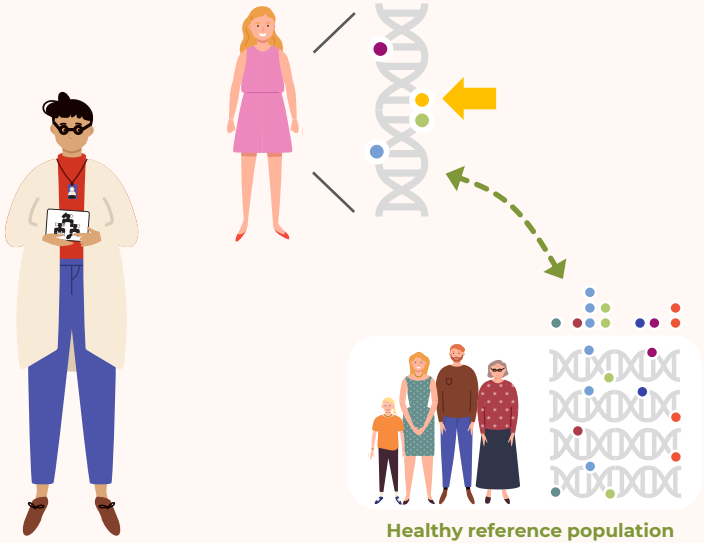
Diagnosing genetic diseases

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

To diagnose a **genetic disease**, we look for genetic variants in the patient that are rare or absent in **healthy people**



Healthy reference population

- At present, maps of healthy variation are based on only some groups of people, mainly **European** or **American**, and some **East Asian** and **South Asian**.
- This means when someone of an **underrepresented group** has a serious genetic illness, it is often **much harder** to pinpoint which change causes their disease.
- Lack of a diagnosis has **implications for treatment**, access to **clinical trials**, and many other aspects of healthcare
- It also makes it less likely that future treatments for these diseases will benefit patients from under-represented communities

A problem for Australia

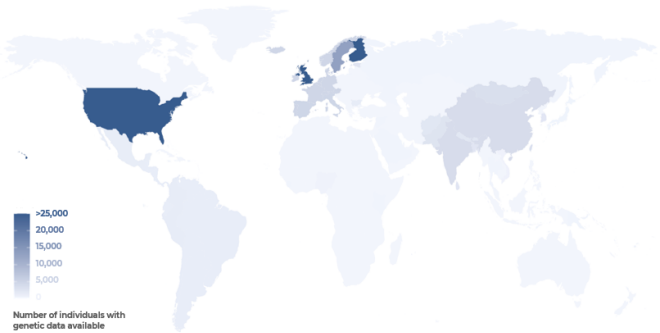
centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

- Existing global resources of genetic variation, mainly built in the US and UK, **don't include many large Australian communities**, spanning several million Australians
- This results in **inequity in access** to accurate diagnosis and future developments in genomic medicine
- If we fail to address this, the impact of genomic medicine will **exacerbate existing health inequities** - leaving many Australians behind

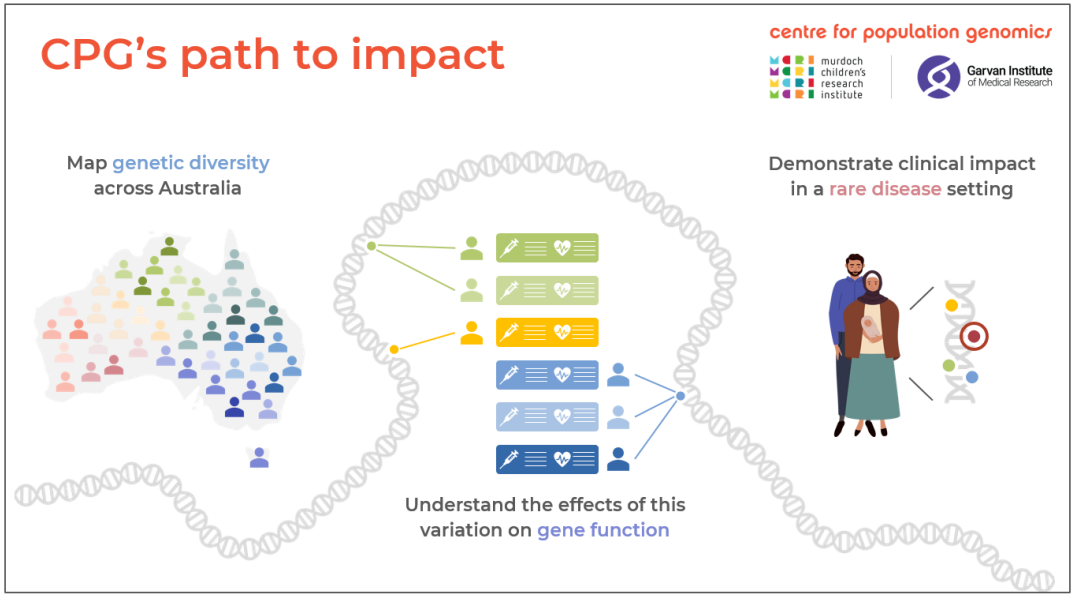
The global availability of genetic data by ancestry group origin



We've reviewed available international genomic resources to identify key gaps in the representation of Australian communities.

While large studies such as gnomAD have done a good job of creating resources reflecting the variation present in individuals of European ancestry, as well as some other communities (especially those present in large numbers in the US and UK), there is still a near-complete absence of information about communities in many parts of the world, especially in the global south.

At least 3.5 million Australians come from communities that are absent or highly under-represented in global databases.



We see our core path to impact as follows: by discovering the full range of genetic variation present in Australian communities; bring genetic data together with clinical and genomic data sets to understand the impact of variation on gene function and human health; and finally, to focus on rare genetic disorders as our major near-term area for clinical impact.

We've divided our work broadly across those three functional areas, while ensuring that there is plenty of cross-talk between domains to allow more efficient use of resources and maximal impact. We've also carefully selected our early projects to ensure we have flagship work ongoing in each domain to drive capability development while also providing opportunities for both near-term and long-term impact

In this section we'll provide overviews of our strategy in each of these domains and outline several of the key flagship projects currently underway.

Strategic Focus Areas

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

	Genetic diversity	Gene function	Rare disease
	Increasing inclusion of under-represented communities in genomic research	*Using naturally occurring genetic variation to explore the function of human genes*	*Leveraging novel genomic and analysis methods at scale to improve diagnosis*
Flagship project	**OurDNA** Recruitment and WGS of >7,000 CALD participants	**Cell + Genome cohorts** Paired WGS and single cell RNA-seq of ~2,000 people	**Automated pipelines** Reanalysis and interpretation for rare disease diagnosis
Immediate impact	Population-specific allele frequency data to enable rare disease diagnosis in high-need communities	Identification of novel gene effects of potential therapeutic relevance	Increasing the speed, sensitivity, and accuracy of genomic diagnosis
Long term goal	Development of a larger, representative biobank and database	Industry collaborations on follow-up of variants of high therapeutic interest	Improved treatment of rare disease using genetics

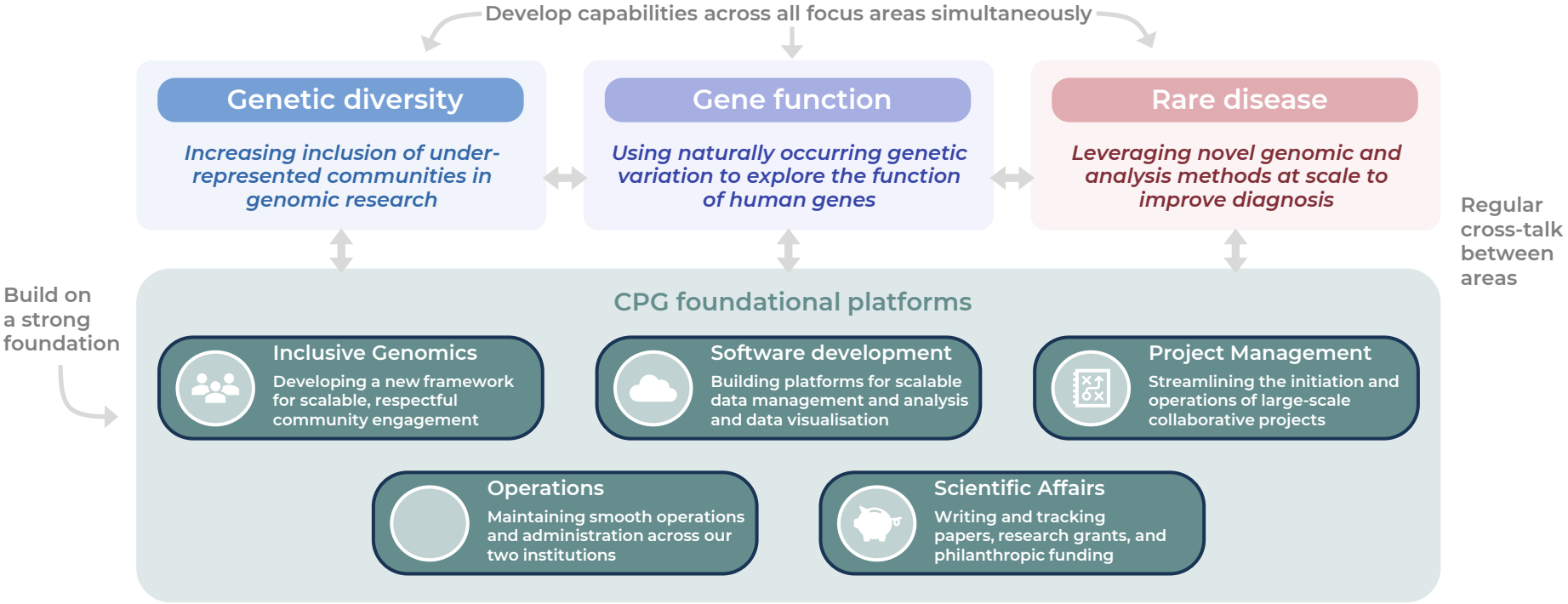
Overview of the three domains of the Centre's scientific strategy – in each we have described some of the key flagship projects, and outlined the areas we see as the major opportunities for both near-term and long-term impact.

Execution Plan

centre for population genomics

murdoch children's research institute

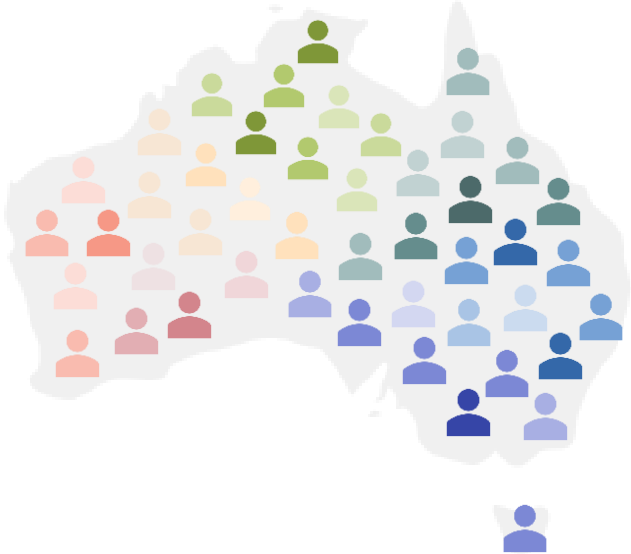
Garvan Institute of Medical Research

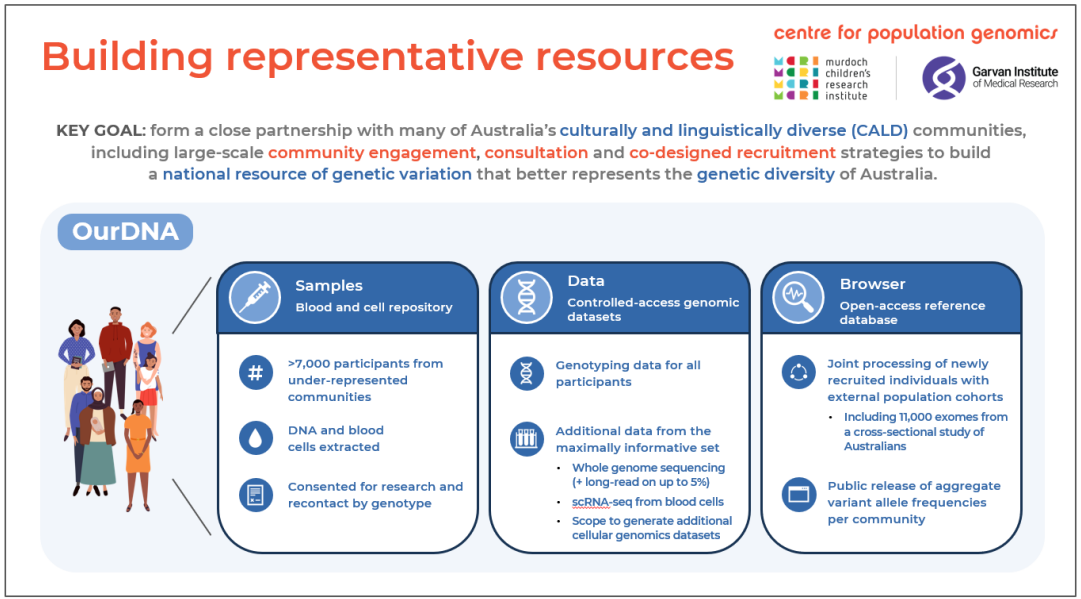


These three domains all rest on a set of foundational platforms that empower our work. These platforms are substantial bodies of work in their own right, and have often required developing entirely novel approaches to solve complex technical and/or logistical problems.

Genetic Diversity

Increasing inclusion of under-represented communities in genomic resources





OurDNA is the flagship project in our Genetic Diversity domain, and will be the largest program ever undertaken specifically focused on increasing the representation of diverse Australian communities in genomics.

The project will involve extensive community engagement, followed by recruitment and blood collection from at least 7,000 individuals from under-represented communities to create a new biobank for future research, and the generation and sharing of whole-genome sequencing and other data sets to empower research as well as improving the diagnosis of families from these communities affected by rare genetic disorders.

In addition to generating our own WGS data, the program will also work with existing national programs such as Mackenzie's Mission to incorporate genomic data already collected from these communities.

OurDNA workflow

centre for population genomics

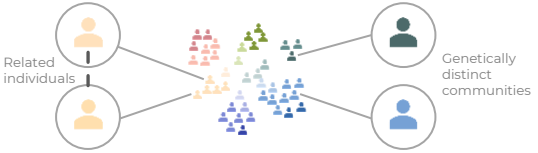
murdoch children's research institute

Garvan Institute of Medical Research

1 Community engagement and recruitment of individuals from under-represented communities

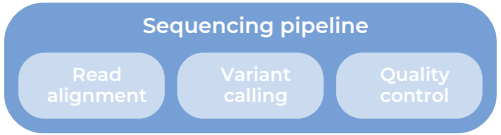
2 Genotype everyone to assess sample quality, identify related individuals and infer genetic ancestry

3 Growing map of genetic diversity highlights communities that need further targeted recruitment or prioritisation for sequencing



4 Maximally informative subset of individuals selected for sequencing

5 Whole genome sequencing, plus single cell RNA-sequencing where possible



6 Generation and release of aggregate allele frequencies and data resources

HIGHLIGHTS

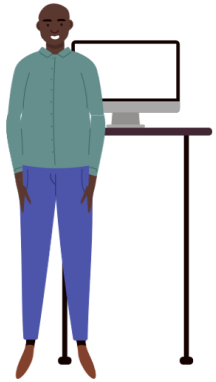
Recruitment prioritisation

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

- Led by the **Inclusive Genomics** team, a careful assessment was made to prioritise Australian communities for the initial OurDNA recruitment
- Prioritisation inputs:
  - Under-representation in existing global databases
  - Size of each community from the Australian census
  - Clinical need based on a survey of Australian clinicians
  - Inclusion of groups from all under-represented regions
  - Partnership with Lifeblood
- Top priority regions: **Oceanian, South-East Asian, Middle Eastern,** and **East African**
- Beginning with Australians from these ancestries: **Samoan, Fijian, Tongan, Vietnamese, Filipino, Lebanese,** and **Sudanese**



There are dozens of ancestry communities in Australia who are under represented in genomic resources. We needed to prioritise in order to decide where to start our engagement work. Research on the ancestries represented in open-access databases such as gnomAD indicates that the peoples of Europe, the Americas and some parts of East Asia and South Asia are largely covered. Missing are many groups from Oceania, South-East Asia, the Middle-East and Africa.

To decide which groups from those large regions to prioritise initially, we looked at 1) census data to determine the size of the group in Australia, 2) the results of our survey of clinical geneticists and genetic counsellors working in Australia to understand which ancestry groups are being seen in genetic services but diagnoses are difficult because of a lack of data, 3) overlap with the priority groups of our collaborators at Lifeblood, and 4) including at least one community from each of the large regions lacking sufficient representation in global databases.

This resulted in our 7 priority groups for the first phases of our work: Australians of Samoan, Fijian, Tongan, Vietnamese, Filipino, Lebanese, and Sudanese backgrounds.

HIGHLIGHTS

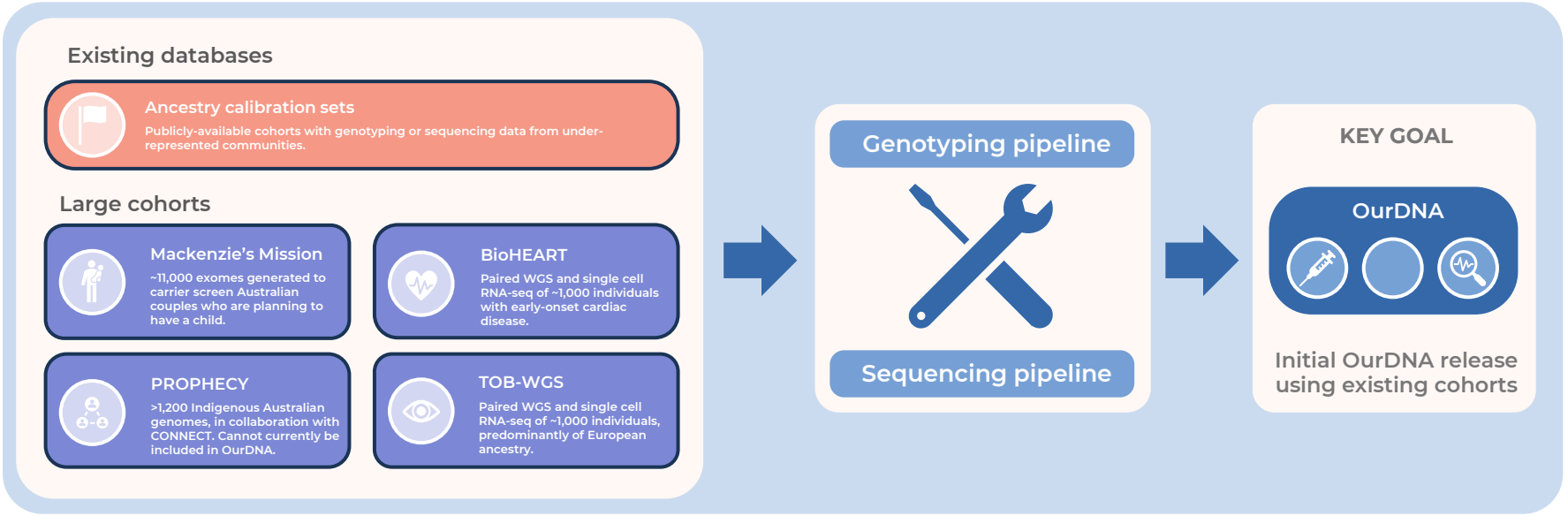
Establishing core pipelines

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

While recruitment is underway, core pipelines are already being established using existing cohorts



Project Management | Software Development | Population Analysis

LOOKING FORWARD

Genetic diversity roadmap

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

The OurDNA project provides a **strong foundation** for downstream **projects and impacts**

OurDNA outcomes

- Established models for respectful engagement and scalable recruitment from diverse communities, including culturally-aligned tools, resources and communication materials
- Large-scale resources to enable effective population-scale genetic studies and targeted research into the effects of population-specific variation

Enabling population-scale studies

- Future community research projects with hundreds of thousands of individuals
- Design of population-specific genotyping arrays, with matched reference panels for imputations
- Population-scale cellular genomics to explore the effects of genetic variation at a cellular level

Helping to understand genetic disease

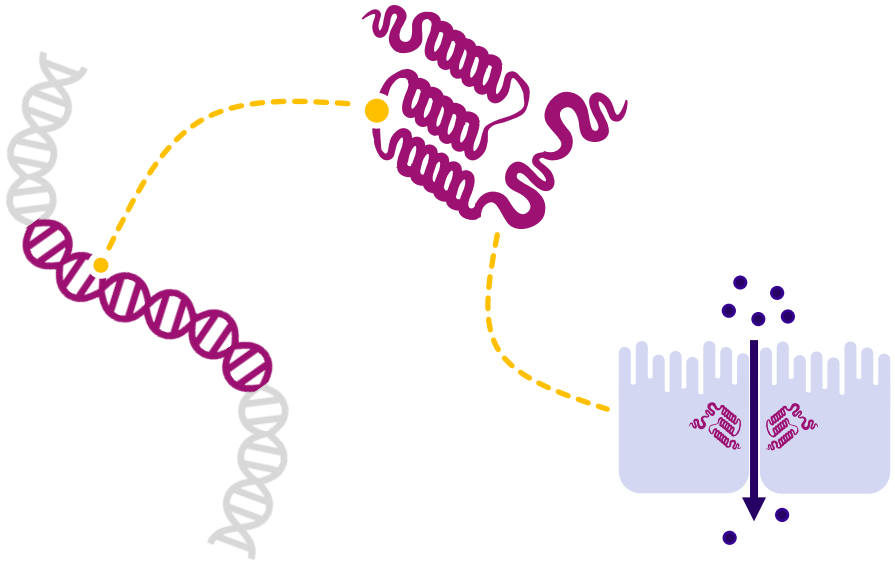
- Enabling studies into the causes of common disease
- Increasing accuracy of polygenic risk predictions in non-European communities
- Genotype-based recontact studies that deeply characterise the effects of specific genetic variants

Informing the development of treatments

- Proposing novel targets and mechanisms of action for therapeutic development
- Using polygenic risk scores to prioritise high-risk individuals for early interventions or clinical trials
- Understanding safety and efficacy implications of genetic variation on genome-targeted therapies

Gene Function

Using naturally occurring genetic variation to explore the function of human genes and inform therapeutic development



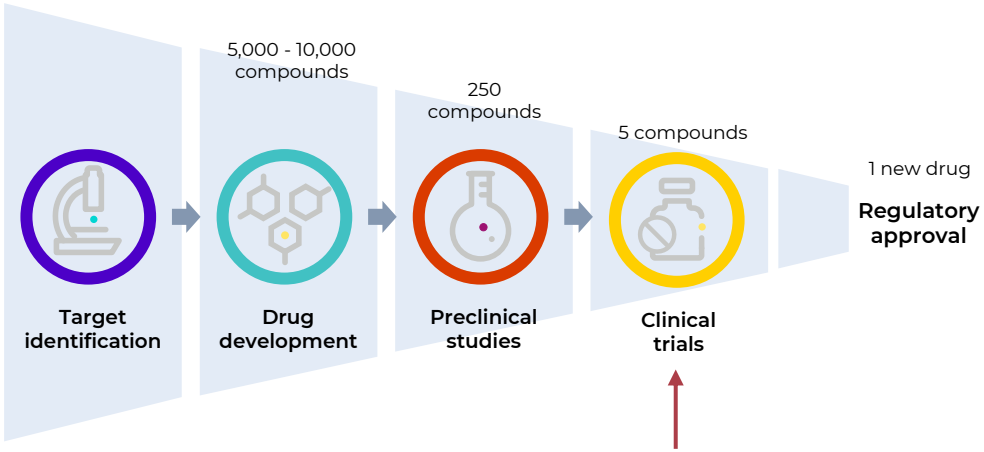
Genomics for drug discovery

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

The classic drug discovery pipeline is an expensive, failure-prone process



One of the biggest and most expensive failure points is a lack of efficacy in clinical trials

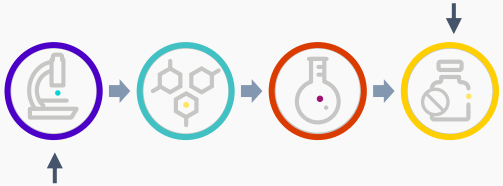
A safe drug that successfully hits its target was developed, but it was the wrong target for the disease!

Estimated 12-15 years and over $1 billion from idea to launch of a finished product

Hughes et al (2011)
Matthews et al (2016)

However, human genomics has the potential to revolutionise this process

The classical model doesn't test **safety and efficacy** in ***humans*** until the clinical trial stage



But human genetics can help us to assess it upfront: **targets with human genetic support are ~ 2x more likely to succeed in the clinic.**

Nelson et al (2015)
King et al (2019)

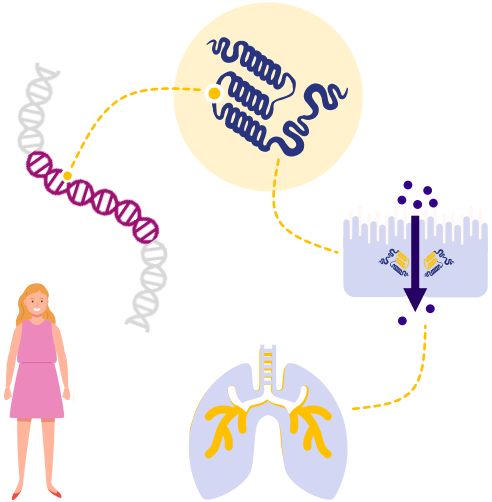
Finding safe, effective drug targets

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

Identifying genetic variants that lead to disease, and understanding how they have this effect, can highlight potential drug targets



Not all targets can be safely modulated: human genomic evidence can indicate when a target is likely to be unsafe

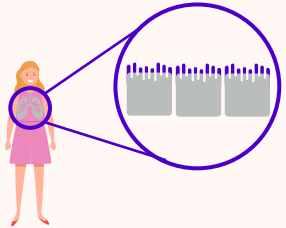
Genetic variants can increase or decrease the function of a gene

Disease-protective or neutral variants that have the same effect as a proposed drug target are evidence for safety

Some genes are constrained: they have very little genetic variation, as modulating them is generally incompatible with life

These are less likely to be safe drug targets!

Some genetic variants only have an effect in very specific cell types and conditions



While designing a drug for these is tricky, highly specific effects can be safer targets

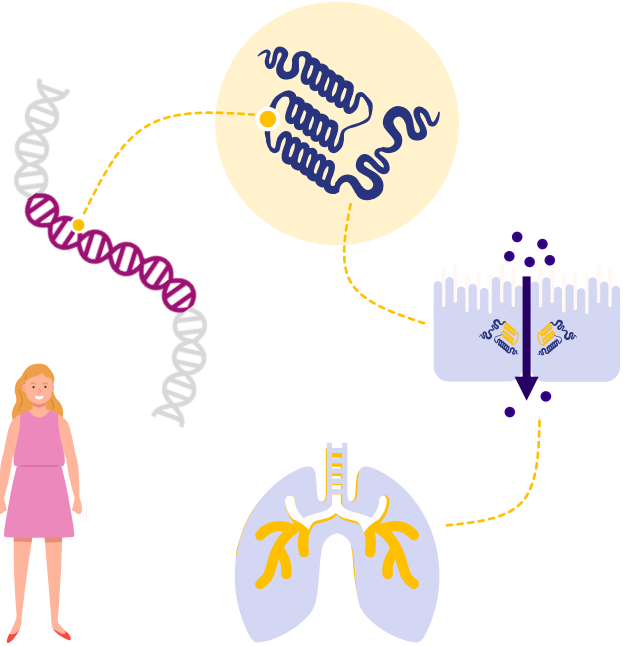
Understanding gene function

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

- The promise of human genomics for drug discovery means that the pharmaceutical industry is investing **billions of dollars** into **large-scale human genetic efforts**
- However, these efforts are highly-skewed towards individuals of **European ancestry**: this restricted view of the genetic landscape limits the variation that can be explored and means valuable **safety and efficacy insights will be missed**
- Diverse resources like OurDNA provide an opportunity to empower therapeutic discovery, by moving from a catalogue of genetic variation to insights into *how* genetic variation may impact health and disease
- In this domain we will build new genomic resources that provide critical insight into the **biological effects of genetic variation** and enable the **testing of therapeutic hypotheses**



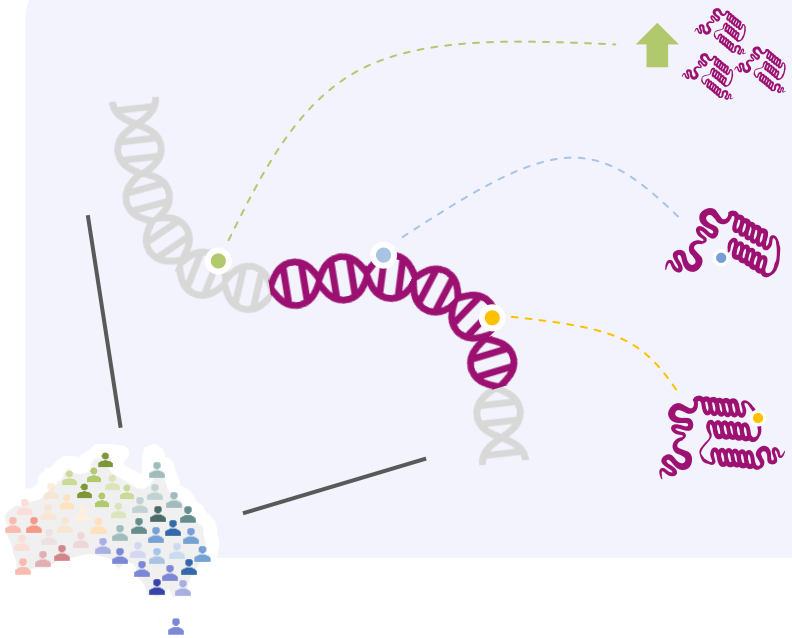
Gene Function workflow

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

Develop large cohorts with paired genetic and cellular genomics data → **Leverage this to gain biological insights into gene function** → **Empower improved disease diagnosis and treatment**



Regulatory variation

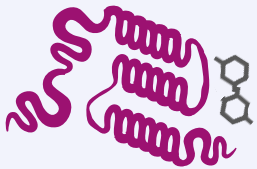
Link genetic variation to cell-type-specific changes in gene expression.

Loss-of-function variation

Validate that predicted loss-of-function variation actually disrupts the expression and/or function of genes.

Missense variation

Model genetic constraint, identifying genes and proteins, and their sub-regions, that are highly tolerant or intolerant to genetic changes.



Current focus areas

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

Regulatory variation

Link genetic variation to cell-type-specific changes in gene expression

- Partnership with Joseph Powell's team, the Tasmanian Ophthalmic Biobank (TOB) and BioHEART initiative
- Paired whole genomes and single-cell RNA-seq of **~2,000** people in the TOB and BioHEART cohorts
- Identify variants associated with gene expression at a cellular level

Loss-of-function variation

Validate that predicted loss-of-function variation disrupts the function of genes

- Industry/academic collaboration with Open Targets and Genes & Health
- Curate **~10,000** predicted loss of function variants in gnomAD and Genes & Health datasets
- Results publicly accessible via integration into the Open Targets portal

Missense variation

Examine missense constraint within protein structures to improve variant interpretation

- Close collaboration with the gnomAD team at the Broad Institute
- Data on an unprecedented scale: **~20,000** Alphafold2 structures and **~800,000** gnomAD v4 sequences
- Develop, assess and release a new structurally-aware metric of genetic constraint

HIGHLIGHT **1,952 whole genomes** sequenced from cohorts with cells available

LOOKING FORWARD

Gene function roadmap

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

Near-term

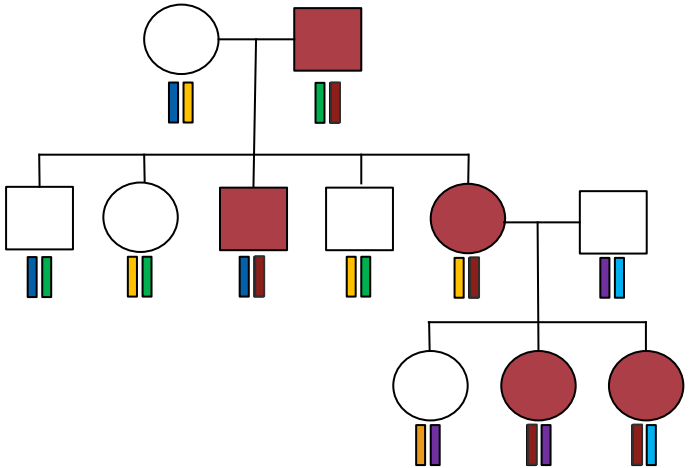
- Generate the largest cohorts in the world with paired **whole genome sequencing** and **single cell RNA-sequencing data**, leading to the **discovery of novel biology** and high-impact publications
- Empower **variant interpretation in rare disease** using **structural constraint metrics**, including the release of public visualisation tools through seqr and the OurDNA browser
- Establish capabilities in cellular genomics and loss-of-function variant interpretation

Long-term

- Generate expression quantitative trait loci (eQTLs) at a cellular level, providing crucial insight into the precise context (**which gene**, in **which cell**, under **what conditions**) in which a genetic variant acts and allowing **more effective translation of disease-associated variation into biological mechanisms and novel therapeutics**
- Rare GEMs: large-scale recruitment of individuals from Australian consanguineous communities to identify "**human knockouts**" with loss-of-function variation affecting both copies of a gene, providing invaluable **safety information** regarding the likely effects of therapeutics targeting that gene

Rare Disease

Leveraging novel genomic and analysis methods at scale to improve diagnosis



Rare disease:
A focus area for direct clinical impact

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

- Rare genetic disorders are **collectively common**, **under-diagnosed** and **extremely under-treated**
- These disorders represent the clearest path to direct clinical impact from CPG's work, with our near-term focus being on the **improvement of diagnosis**:
  - OurDNA will empower diagnosis in under-represented communities
  - The application of novel genomic and analysis methods to rare disease families will directly provide new genetic diagnoses and identify new disease genes
  - Close collaboration with clinical labs (*e.g.,* VCGS) will allow rapid translation of novel approaches into routine clinical practice
- Long-term, our focus is on the **development of new therapeutics**:
  - New therapeutic modalities (*e.g.,* RNA, CRISPR) will drive a transformation of the therapeutic landscape for rare disorders over the next 5-10 years
  - CPG will position itself for partnerships leveraging both rare disease and population data to identify and validate therapeutic targets and strategies for these disorders

- Despite recent advances in clinical genomics, rare diseases still result in a substantial health burden
- This burden is even greater for patients of non-European ancestry, where the diagnostic yield from genomic testing can be reduced by up to 50%
- CPG is uniquely placed in the Australian setting to deliver a large, near-term impact on the rare disease field
- We are developing a scalable platform that can bring together our RD genomics knowledge, the outputs of the OurDNA program with the cohorts and knowledge of many of Australia's best rare disease clinical researchers
- This will deliver diagnoses for patients and empower the development of novel therapeutics

A national rare disease platform

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

We have built foundational capabilities in rare disease diagnosis and begun large-scale analysis

Collaborations

Partnerships with rare disease cohorts of national significance

- UDN-Aus (National)
- CIRCA (Garvan)
- KIDGEN (National)
- RDNow (MCRI)
- Mito-MDT (National)

Analysis pipelines

Cloud-based rare disease genomic analysis platform

Production pipeline

WGS and exome best practices analysis pipeline

Automated interpretation pipeline

Enabling continuous reanalysis of existing cohorts

seqr

Cloud-native variant curation and prioritisation portal

Empowering researchers

Seqr provides the tools for researchers and clinicians to drive discovery

Variant analysis team

CPG's analysts partner with cohort leaders, ensuring expert analysis

Improved patient diagnosis

HIGHLIGHT **1,050 genomes** from rare disease families spanning 14 projects now in analysis

A national rare disease platform

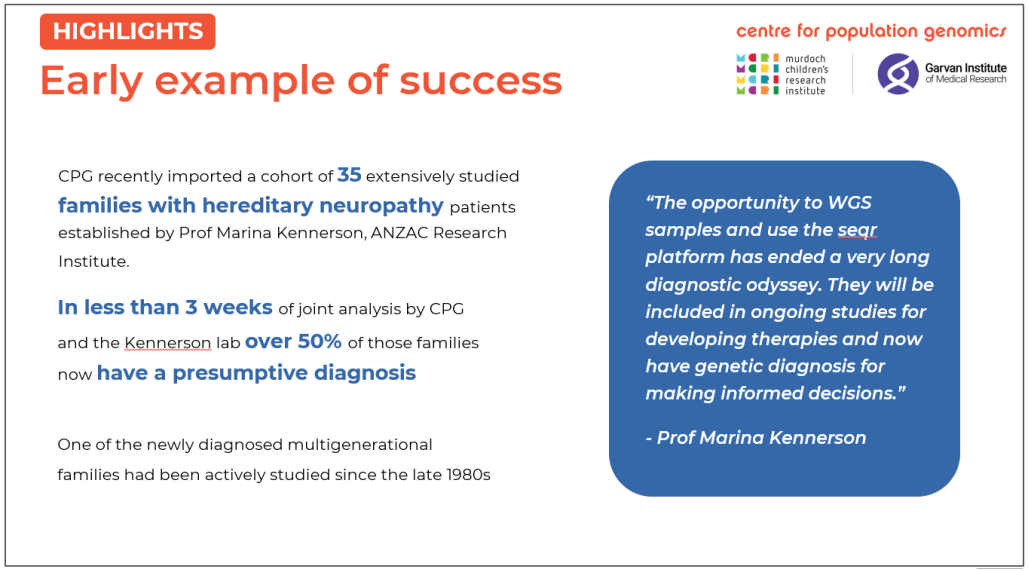
centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

- To deliver a rapid impact on improving diagnosis rates, CPG has partnered with several rare disease research programs of national significance. Our RD platform is empowering these programs by providing an integrated end-to-end genomic analysis environment.
- The genomic data from each cohort is processed through our harmonised production pipeline allowing direct comparison of results across all RD cohorts
- Patient genomes are loaded into seqr, our web-based collaborative variant analysis tool where our expert collaborators together with CPG's own variant analysis team can intuitively mine genomic variants to uncover diagnoses.
- We have recently complemented this process through the development of the "Automated Interpretation Pipeline" (AIP), a tool designed to continuously reanalyze genomes using the latest clinical genomic databases and notify researchers when new potentially causative variants can be identified.

- In this domain we have focused on identifying areas where genomic approaches are likely to have the most rapid impact on improving diagnosis rates, and on working with collaborators across Australia to bring together large cohorts of undiagnosed families who could benefit from such approaches
- We have partnered with multiple national cohorts, including several with MRFF funding, focused either or national-level multi-disease initiatives or on building large cohorts in specific disease areas
- All of these cohorts go through a harmonised production pipeline, and an automated interpretation pipeline developed in collaboration with partners at MCRI, VCGS, Broad, and Microsoft
- We also load all of the data into our analysis platform *seqr*, which allows it to be analysed by our expert collaborators, as well as by the CPG's own variant analysis team



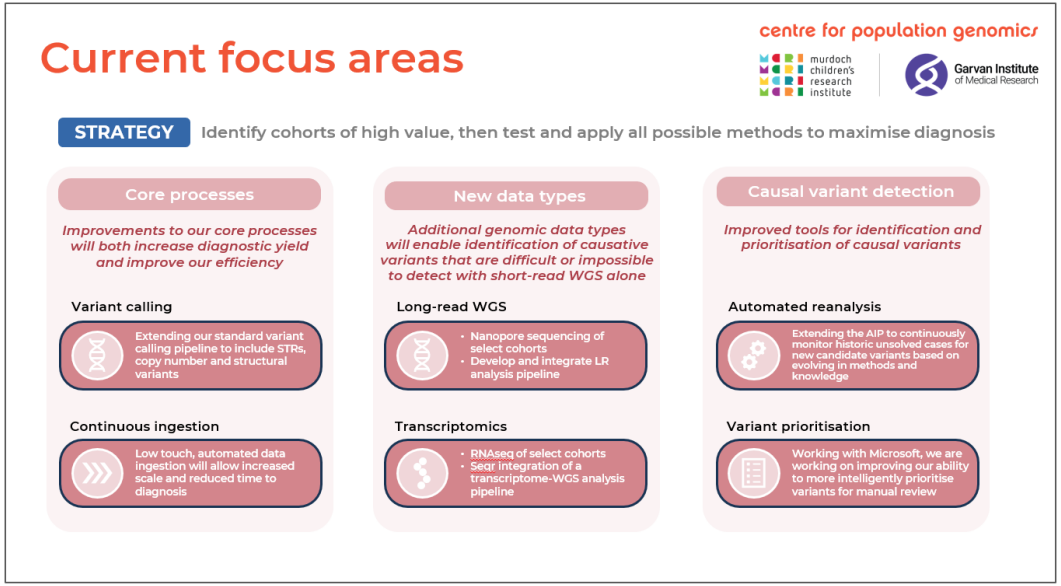
- An example of our ability to enable and empower Rare Disease research in Australia is our collaboration with Prof Marina Kennerson as part of the GHFM funded "Closing the gap in diagnosis of neurological disorders including ataxias and neuropathies" project

- The collaboration was initiated with 35 families from a historic cohort that had been extensively studied for many years.

- GPG processed the WGS data from these 35 families making it available via seqr

- Within 3 weeks of joint analysis by the Kennerson lab and a CPG variant analyst presumptive diagnoses have been made for half of the families including one multi-generational family that had been part of the research study since the 1980s.

- This vignette is a demonstration of the impact that can be realised by bringing together CPGs large-scale technical and genomics expertise with the disease-specific domain expertise and cohorts of Australia's clinical researchers.



- Our strategic focus for rare diseases is on identifying cohorts of high value and then test and applying all possible methods to maximise diagnosis.
- To enable this, we are extending the foundation of our Rare Disease Platform to enable integrated analysis of multiple variant types and genomic data types.
- This includes strategic application of both long read and RNA sequencing, supported by the development of analysis workflows that can integrate directly with the seqr platform
- We are also working to extend the capability of our AIP - an automated framework that can be applied regularly across all research cohorts and also deployed in clinical diagnostics.
- This tool should remove most of the burden associated with manual reanalysis of historic unsolved cases. As the fields knowledge of genes and variants associated with disease evolves, this knowledge can be applied to historic cohorts in near real-time to identify disease-causing variants
- Building on the foundation of the AIP, we are also working with Microsoft and the Broad to develop new approaches to improve our ability for automated prioritization of candidate causative variants.

LOOKING FORWARD

Rare disease roadmap

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

Near-term

- Complete establishment of **rare disease platform capabilities**
- Impact the lives of rare disease families through **diagnosis**, **open tools** and **improved knowledge of disease mechanisms**
- Cement CPG's position as the **national partner of choice** for rare disease research genomics and build a strong, productive collaborative network
- Transfer knowledge/capabilities to **clinical lab partners**

Long-term

- Integrate novel **machine learning approaches** to **improve diagnosis and reanalysis**
- **Empower therapeutic development**
  - Track development in **targeted therapeutics** to identify industry partnership opportunities
  - Integrate data from **rare disease families** (gene/mechanism discoveries) and **population analysis** (target validation/safety signals)

Initiatives, outcomes and next steps for delivery

- Project Management
- Software Development
- Inclusive Genomics

PROJECT MANAGEMENT

Project Management Framework

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

- Designed a robust and scalable project management framework
  - Developed and implemented foundational data logistics workflows
  - Delivering core components on collaborative projects
  - Established key partnerships with commercial and academic organisations
- What this looks like in practice
  - 24 project or dataset related legal agreements executed
  - 7 HREC approved protocols
  - 3 data logistics workflows established
  - 19 unique collaborations encompassing 33 datasets

PROJECT MANAGEMENT

Project Management Framework

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

CPG has undergone a rapid expansion phase as we work towards delivering our core goals. Under the leadership of **Sally Hartmanis**, the project management team have established a project management framework that has delivered on project outcomes whilst scaling to meet the expanding needs of the CPG.

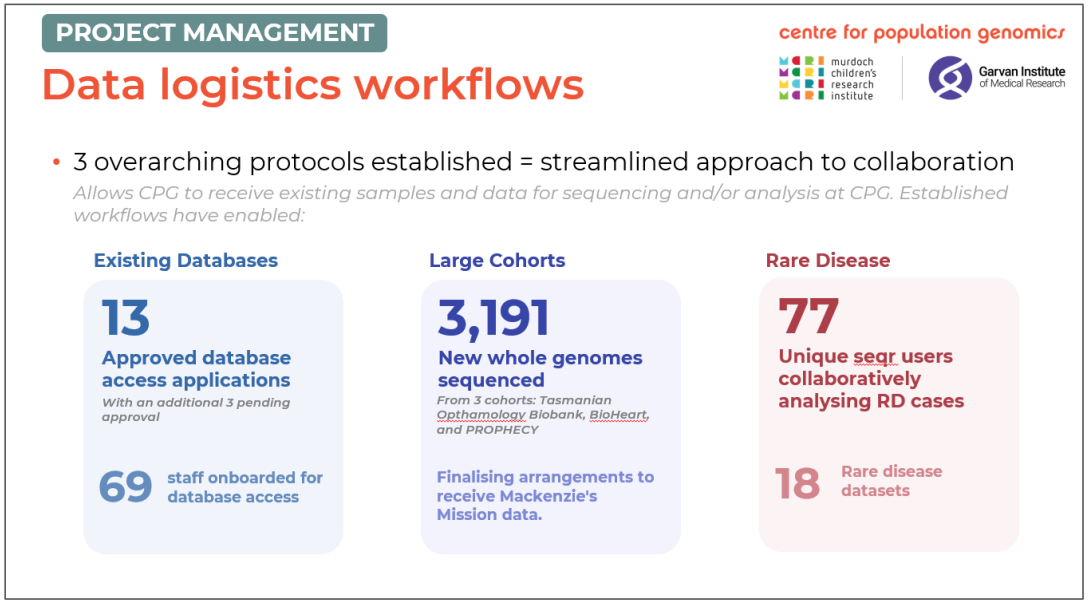
Essential to the framework has been the establishment of foundational data logistics workflows, that allow us to ingest collaborator data and enable key project outcomes to be delivered. Our workflows were rapidly established and designed to be able to adapt and scale to meet the needs of future initiatives. Supported by these workflows, CPG is delivering core components of multiple collaborative projects from across Australia, and our framework ensures we deliver on project deliverables and coordinate with our collaborators effectively.

Significant time and effort has also been dedicated to the rigorous development of, and planning for, the end-to-end sample workflow for the our flagship OurDNA program. Including identifying and forming key partnerships with commercial and academic organizations specializing in each step of the workflow to drive efficiency gains and maximize the total sample size. With this solid foundation, we are now well positioned to enter hte delivery phase of OurDNA.

The figures on this slide provide a snapshot into the work that has been undertaken by the project management team and forms a crucial part of delivering the project management framework.

CPG has a project management framework, supported by foundational data logistics workflows, that is able to adapt, scale and evolve over time and meet the unique needs of delivering large scale genomics research projects and initiatives. CPG is

bringing together the best parts of structured project management that drive efficiency and at the same time support innovation in research. With a full complement of CPG teams, we are working on refining our framework and finding the right balance between the formal project management and academic approaches.



We have established data logistics workflows that allows CPG to efficiently engage with collaborators and receive existing genomic samples or data to facilitate sequencing and/or analysis at CPG.

These workflows encompass collaborator engagement, database applications, legal and ethical submissions and responsibilities, project staff onboarding and management of access permissions. A core component of our workflows is the establishment of three overarching HREC approved protocols for (i) existing databases, (ii) large cohort and (iii) rare diseases.

These protocols allow CPG to adopt a streamlined collaboration model which means we can manage many collaborations through a single protocol - this significantly reduces the administrative and governance burden on ourselves and our collaborators.

The figures on this slide highlight activity that has been enabled through our overarching protocols and data logistics workflows. These outputs are the result of a combined effort from all CPG teams.

SOFTWARE DEVELOPMENT

Data infrastructure highlights

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

- Developed **robust**, **reusable**, and **scalable** open-source production systems, and are ready to hand over the running of these production pipelines to the analysis teams

Population datasets

~21,000

existing and novel genomes processed

through joint-calling and quality control pipelines

Rare Disease

1,050

genomes loaded into seqr *from 18 datasets*

Platforms

Metamist

Scalable metadata management, single source of truth for automation

Automated billing reports

Automated, aggregated billing reports that can be used by collaborators, such as those using seqr

- Built custom **front-end resources** and **data visualisations**
  - Single-cell expression quantitative trait loci (eQTL) browser, displaying data from the Tasmanian Ophthalmic Biobank project, was developed in collaboration with GWCCG and will be publicly released very soon

With both the rare disease and population analysis teams now in place, we've started the process of handing over the running of the production pipelines (and associated downstream QC) from the software team to the analysis teams. This lets the software team focus on the generalizable aspects of the infrastructure, while the analysis teams can apply their domain expertise to concrete analyses for each dataset.

*** ~21,000 genomes includes ~18,000 genomes processed through NAGIM.**

SOFTWARE DEVELOPMENT

Data infrastructure partners

centre for population genomics

murdoch children's research institute

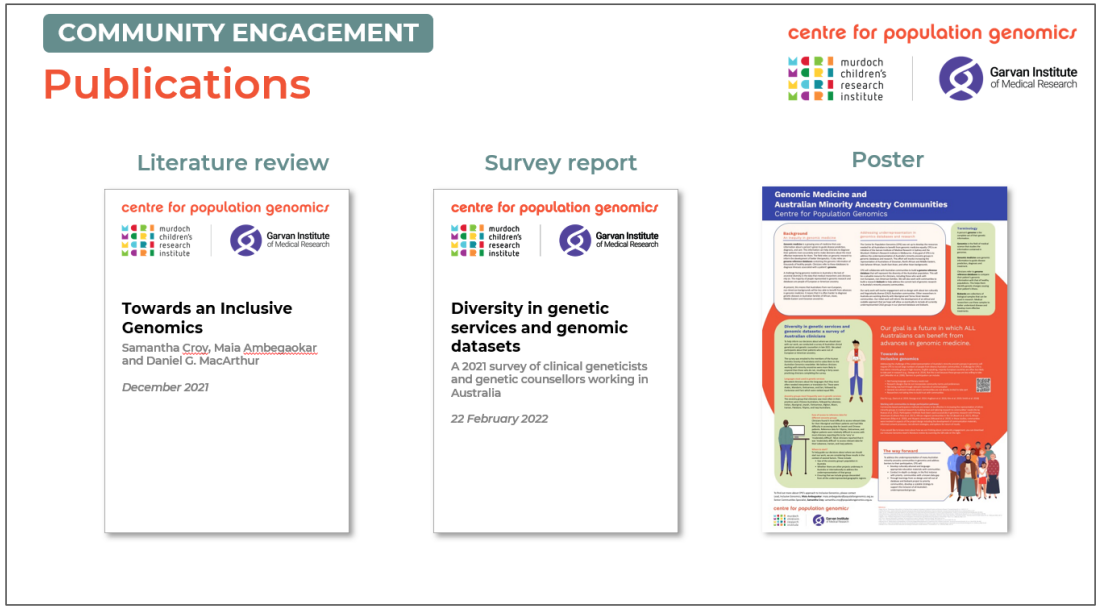
Garvan Institute of Medical Research

- **NAGIM** (National Approach to Genomic Information Management)
  - Resulted in recommendations to government resembling CPG's cloud-based infrastructure
  - Joint work with **KCCG**, **Google**, and **Australian Genomics**
- Multi-cloud deployment: **Google Cloud** + **Microsoft Azure**
  - Better discount negotiation position, avoids lock-in
  - More portable (and reusable) pipelines
- Continued joint development with the **Broad Institute**
  - Largest deployments of Hail Batch and seqr outside the Broad
  - Working together on gnomAD v4 production / release

Australian Genomics

BROAD INSTITUTE

A core goal of these collaborations is to build scalable and cost-sustainable genomic infrastructure for CPG (and beyond). While we're still far from gnomAD scale in terms of the number of samples we process, all of our pipelines and methods are designed in a way to allow for similar scale.



In December, we shared online our white paper Towards an Inclusive Genomics, a literature review of evidence regarding engaging ethnic minority groups in genomics and medical research. We have been pleased by the response from fellow researchers in Australia who are also seeking to increase the involvement of communities who are underrepresented in genomic research. No other summary of this literature from an Australian perspective existed before.

In February, we shared a report of our survey of clinical geneticists and genetic counsellors regarding the the ethnic and language groups seen in genetic services. Conducted via the Human Genetics Society of Australia and Australian Genomics, the survey helped to inform our prioritisation, described below.

In June, we presented a poster on CPG's planned work with Australian minority ancestry communities at the FECCA 2022 conference. The Federation of Ethnic Communities Councils is a key collaborator and supporter of our work. The conference was an important networking event helping us to make new contacts, as described below.

COMMUNITY ENGAGEMENT

Outreach

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

Networking

Ethnic Communities' Council of NSW
Representing All Communities

eccv
Ethnic Communities' Council of Victoria

Partnerships

Australian Red Cross Lifeblood®

FECCA

Membership

Australian Multicultural Health Collaborative

COMMUNITY ENGAGEMENT

Participant information

- **Australian Genomics** grant-funded work
- Plain English versions of genomic research materials
  - Glossary
  - Genes, genomes & health
  - Genomic research and testing
  - Possible results
  - Risk & benefits
  - Standardised Patient Information and Consent form
- Translation and focus groups
  - Arabic, Dari, Farsi, Fijian, Hazaragi, Samoan, Tagalog, Tongan, Urdu and Vietnamese

Australian Genomics

CultureVerse

a think/hq company

In Australia, there has so far been essentially no development of genomics information materials in languages other than English, despite the fact that more than 20% of Australians speak a language other than English at home.

In June we were delighted to receive funding from the Australian Genomics Genomic Implementation Projects Grant Program for our project Development of culturally-aligned and language-appropriate participant information and education resources for Australian ethnic minority ancestry groups under-represented in clinical and population genomics research. The two stage project, underway now, will first develop plain English versions of genomics information materials and test and refine these via review by experts from across Australian genomics research. The materials will be appropriate for use in both clinical and population research projects. These will then be translated into ten first-priority languages for genomic information: Arabic, Dari, Farsi, Fijian, Hazaragi, Samoan, Tagalog, Tongan, Urdu, and Vietnamese.

CultureVerse will lead the translation process and our Inclusive Genomics team will lead focus groups with bi-lingual speakers of the 10 languages to refine the terms and phrases used to ensure that the materials are culturally appropriate and understood. We will make the materials openly available for use by any genomics research program in Australia, and will share our methodology to allow further languages to be added. Within CPG, the materials will serve as the foundational content for our own recruitment and consent portal.

Areas of ongoing focus

Scalability

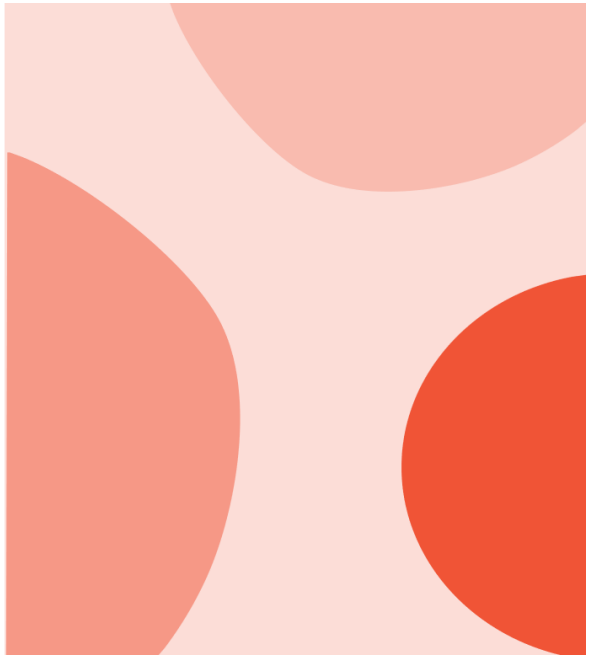
WHY CPG's project goals require extremely rapid scaling of capabilities across multiple domains

Progress to date:

- Cloud-based computing platform allows secure data storage and analysis at nationally unprecedented scale: NAGIM prototyping exercise (with Garvan DSP) in 2021 involved 14,000 WGS, largest Australian dataset ever assembled
- Rare disease analysis platform *seqr* now includes more than 1,000 WGS, engineered for further rapid growth
- Project management framework allows efficient coordination of many complex collaborations

Areas of ongoing focus:

- Community engagement and participant recruitment workflow for OurDNA program still in development
- Need to reduce complexity of legal and administrative workflows across Garvan/MCRI to allow more rapid addition of new partners and collaborators



The Centre has now moved from the building phase to consolidation and scaling.

We have built many of the foundational capabilities required for success, allowing us to focus on scaling and improving efficiency

- Key capabilities in place: ethics approvals, sample handling and data generation workflows, massively scalable cloud-based data platforms
- Data generation: now have data generated and iDn analysis for flagship projects in all three domains
- Key collaborations: network of active academic collaborators now spans 30 institutions in Australia, US, and UK
- National impact: active contributions to design of national genomic data infrastructure

Participant-centric

WHY Engagement of participants from diverse communities is critical to our goals

Progress to date:

- Developed and received ethics approval for co-design & community engagement process for OurDNA
- Now deeply embedded within the CONNECT consortium for Indigenous genomics, providing leadership around genomics and data management

Areas of ongoing focus:

- Approaches to data governance and benefit sharing, refined in consultation with stakeholders representing CALD and Indigenous communities
- Developing and testing a scalable framework for recruitment and ongoing engagement of participants from diverse communities

Remote-first

WHY The CPG's structure requires us to build and sustain a geographically distributed team

Progress to date:

- Team members in Melbourne, Sydney and New Zealand
- Robust culture of asynchronous communication using Slack and other platforms
- Opportunities for face-to-face contact created through quarterly meetings, custom space in Sydney/Melbourne

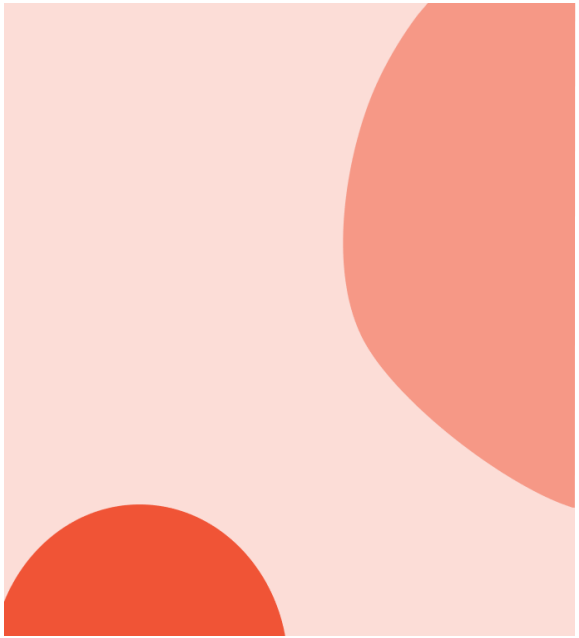
Areas of ongoing focus:

- Ensure that remote communications model is sustainable beyond the COVID pandemic
- Building and sustaining a strong culture through a combination of in-person events and deliberate creation of new norms around remote working

CPG was founded on the pretext that the organisation would be remote-first, permanently. Whilst the Centre certainly hasn't been immune to the challenges impacting team culture that have presented by the pandemic, our remote-first working environment has provided a measure of stability and allowed us to pivot as required. Further, creating a strong, staff-led culture for CPG will be critical to the success of the organisation, particularly in the remote-first environment.

The majority of team interactions occur via the well-known communication platform Slack, and the group are actively engaged in trying new methods of socialising and collaborating in a digital environment, such as Gather.Town. There are a number of initiatives scheduled to occur prior to the end of the calendar year. Several in-person events are scheduled for the remainder of 2022, and the implementation of an Agile-style retrospective initiative will encourage meaningful cross-team communication and form the basis for a new collaboration framework for the organisation. Direct staff member engagement will be collated via CPG's initial engagement activities with CultureAmp.

The leadership team recognises that the success of the organisation is dependent on our ability to create culture in a distributed environment. We look to others in our sector and beyond for the best ideas, we try them on to see how they fit and consult at every level to ensure we have developed the best solution. This is an iterative process that will gain momentum throughout the next year and beyond.



Professional capability

WHY CPG's focus on scale requires building teams unusual in traditional academia

Progress to date:

- Over 35 professional staff in seven CPG core teams
- Developed new career development pathways for project managers, software engineers, and computational biologists
- Staff have KPIs focused on outcomes rather than traditional academic metrics of success; quarterly career meetings

Areas of ongoing focus:

- Remuneration and salary increase frameworks better-suited to competition for talent with industry

CPG team leads come from a wide variety of backgrounds (Google, biotech, pharma, community engagement, academia) with diverse professional experience and the teams are entirely composed of professional (non-academic track) staff.

CPG has taken a thoughtful approach to professional career development across all areas. We have developed career frameworks that provide a defined career pathway for professionals in research from entry level positions through to management roles. By doing so, CPG can attract highly skilled professionals to the research sector and retain them by aligning KPI's with impact.

Other initiatives

centre for population genomics

murdoch children's research institute

Garvan Institute of Medical Research

The establishment of a strong operations team led by the CPG's Business Manager will allow us to improve operational efficiency and manage risk, through a number of key initiatives including:

- Process optimisation across Garvan & MCRI operations: financial approvals, inter-org invoicing, standardised salary brackets, joint legal review, event budget sign off, faculty affiliations
- Introduction of communication and collaboration framework for projects crossing team boundaries
- Support for leadership skills in junior managers
- Career progression framework for administrative and operations staff
- Instituting new formal frameworks for risk management, and for responsibility and accountability
- 2023 – 2024 budget and financial forecast
- Developing and refining models for long term organisational sustainability
- Instituting OKR-style performance management tools

Questions