Searching for a technology-driven acute rheumatic fever test: the START study protocol

Anna P Ralph,1,2 Rachel Webb,3,4,5 Nicole J Moreland,6 Reuben McGregor,6 Anthony Bosco,7 David Broadhurst,6 Timo Lassmann,7 Timothy C Barnett,7 Rym Benothman,7 Jennifer Yan,1,2 Bo Remenyi,1,2 Julie Bennett,9 Nigel Wilson,4 Mark Mayo,1 Glenn Pearson,7 Tobias Kollmann,7 Jonathan R Carapetis7,10,11

ABSTRACT
Introduction The absence of a diagnostic test for acute rheumatic fever (ARF) is a major impediment in managing this serious childhood condition. ARF is an autoimmune condition triggered by infection with group A Streptococcus. It is the precursor to rheumatic heart disease (RHD), a leading cause of health inequity and premature mortality for Indigenous peoples of Australia, New Zealand and internationally.

Methods and analysis Searching for a Technology-Driven Acute Rheumatic Fever Test (START) is a biomarker discovery study that aims to detect and test a biomarker signature that distinguishes ARF cases from non-ARF, and use systems biology and serology to better understand ARF pathogenesis. Eligible participants with ARF diagnosed by an expert clinical panel according to the 2015 Revised Jones Criteria, aged 5–30 years, will be recruited from three hospitals in Australia and New Zealand. Age, sex and ethnicity-matched individuals who are healthy or have non-ARF acute diagnoses or RHD, will be recruited as controls. In the discovery cohort, blood samples collected at baseline, and during convalescence in a subset, will be interrogated by comprehensive profiling to generate possible diagnostic biomarker signatures. A biomarker validation cohort will subsequently be used to test promising combinations of biomarkers. By defining the first biomarker signatures able to discriminate between ARF and other clinical conditions, the START study has the potential to transform the approach to ARF diagnosis and RHD prevention.

Ethics and dissemination The study has approval from the Northern Territory Department of Health and Menzies School of Health Research ethics committee and the New Zealand Health and Disability Ethics Committee. It will be conducted according to ethical standards for research involving Indigenous Australians and New Zealand Māori and Pacific Peoples. Indigenous investigators and governance groups will provide oversight of study processes and advise on cultural matters.

INTRODUCTION
The absence of a diagnostic test for acute rheumatic fever (ARF) is a major impediment to management and epidemiological understanding of this serious childhood condition. ARF is one of the best examples of human autoimmunity caused by an infection, with group A Streptococcus (GAS) the identified trigger. It is the precursor to rheumatic heart disease (RHD), a leading cause of health inequity and premature mortality for Indigenous peoples of Australia, New Zealand (NZ) and internationally. Prompt diagnosis of ARF allows timely commencement of secondary prophylaxis with long-acting penicillin to prevent repeated GAS infections that drive ARF recurrences. ARF recurrences are in turn the main mechanism responsible for...
development of RHD. Underdiagnosis of ARF appears to be a major contributor to the high rates of RHD in Aboriginal communities in Australia’s Northern Territory (NT), where 76% of patients with RHD lack a prior ARF diagnosis.6 Similarly a cohort study in NZ found the majority of RHD cases (65%) had never been previously hospitalised with ARF.3 Similar findings are reported in African countries.4

The Jones criteria (table 1) have been the diagnostic tool for ARF for nearly 80 years.5 This is a regularly reviewed clinical algorithm, now available in a user-friendly smart device application (‘app’),6 most recently modified to optimise sensitivity for use in high-risk settings and optimise specificity for low-risk populations.7 Despite these revisions, ARF diagnosis remains challenging and subjective. Diagnostic biomarkers for ARF that improve the performance of these criteria or provide a definitive diagnostic test would be a major advance in improving ARF detection, and thereby, RHD prevention.8 Accurate case ascertainment is also needed to monitor progress towards disease control targets.

A further barrier to ARF management is the absence of an identified effective immunomodulatory treatment to alter long-term cardiac outcomes. Non-steroidal anti-inflammatory drugs alleviate joint symptoms but do not alter the long-term disease trajectory. Corticosteroids

Table 1: Revised Jones criteria

<table>
<thead>
<tr>
<th></th>
<th>High-risk groups*</th>
<th>Low-risk groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite initial episode of ARF</td>
<td>2 major manifestations + evidence of preceding GAS infection. 1 major + 2 minor manifestations + evidence of preceding GAS infection.†</td>
<td></td>
</tr>
<tr>
<td>Definite recurrent‡ episode of ARF in a patient with a documented history of ARF or RHD</td>
<td>2 major manifestations + evidence of preceding GAS infection. 1 major + 2 minor manifestations + evidence of preceding GAS infection. 3 minor manifestations + evidence of a preceding GAS infection.†</td>
<td></td>
</tr>
<tr>
<td>Probable or possible ARF (first episode or recurrence)‡</td>
<td>A clinical presentation in which ARF is considered a likely diagnosis but fails short in meeting the criteria by either:  ► one major or one minor manifestation.  ► no evidence of preceding GAS infection (streptococcal titres within normal limits or titres not measured). Such cases should be further categorised according to the level of confidence with which the diagnosis is made:  ► Probable ARF (previously termed ‘probable: highly suspected’).  ► Possible ARF (previously termed ‘probable: uncertain’).</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Major manifestations</th>
<th>Carditis (including subclinical evidence of rheumatic valvulitis on echocardiogram)</th>
<th>Carditis (including subclinical evidence of rheumatic valvulitis on echocardiogram)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Polyarthritis§ or aseptic monarthritis or polyarthralgia</td>
<td>Polyarthritis§</td>
</tr>
<tr>
<td></td>
<td>Sydenham chorea¶</td>
<td>Sydenham chorea¶</td>
</tr>
<tr>
<td></td>
<td>Erythema marginatum**</td>
<td>Erythema marginatum**</td>
</tr>
<tr>
<td></td>
<td>Subcutaneous nodules</td>
<td>Subcutaneous nodules</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Minor manifestations</th>
<th>Fever ††≥38°C</th>
<th>Fever ≥38.5°C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monarthralgia ‡‡</td>
<td>Polyarthralgia or aseptic monarthritis‡‡</td>
</tr>
<tr>
<td></td>
<td>ESR ≥30 mm/h or CRP ≥30 mg/L</td>
<td>ESR ≥60 mm/h or CRP ≥30 mg/L</td>
</tr>
<tr>
<td></td>
<td>Prolonged P-R interval on ECG‡§</td>
<td>Prolonged P-R interval on ECG‡§</td>
</tr>
</tbody>
</table>

Reproduced with permission from The 2020 Australian guideline for prevention, diagnosis and management of ARF and RHD (third edition).2 *High-risk groups are those living in communities with high rates of ARF (incidence >30/100 000 per year in 5–14 years olds) or RHD (all-age prevalence >2/1000). Aboriginal and Torres Strait Islander peoples living in rural or remote settings are known to be at high risk. Data are not available for other populations but Aboriginal and Torres Strait Islander peoples living in urban settings, Māori and Pacific Islanders, and potentially immigrants from developing countries, may also be at high risk.†Elevated or rising antistreptolysin O or other streptococcal antibody, or a positive throat culture or rapid antigen or nucleic acid test for GAS infection. †‡Recurrent definite, probable or possible ARF requires a time period of more than 90 days after the onset of symptoms from the previous episode of definite, probable or possible ARF. §A definite history of arthritis is sufficient to satisfy this manifestation. Note that if polyarthritis is present as a major manifestation, polyarthralgia or aseptic monarthritis cannot be considered an additional minor manifestation in the same person. ¶Chorea does not require other manifestations or evidence of preceding GAS infection, provided other causes of chorea are excluded. **Care should be taken not to label other rashes, particularly non-specific viral exanthems, as erythema marginatum. ††In high-risk groups, fever can be considered a minor manifestation based on a reliable history (in the absence of documented temperature) if anti-inflammatory medication has already been administered. ‡‡If polyarthritis is present as a major criterion, monarthritis or arthralgia cannot be considered an additional minor manifestation. §§If carditis is present as a major manifestation, a prolonged P-R interval cannot be considered an additional minor manifestation. ARF, acute rheumatic fever; CRP, C reactive protein; ESR, erythrocyte sedimentation rate; GAS, Group A Streptococcus; RHD, rheumatic heart disease.
appear ineffective in limiting development or severity of RHD, but in some settings are prescribed as a treatment of last resort for severe carditis. Hydroxychloroquine use has recently been reported in two cases of ARF with carditis, based on plausible in vitro data but efficacy is as yet unknown. Improved knowledge of ARF immune pathogenesis is needed to inform potential immune modulatory strategies targeted to the immune pathology. Compounding complexity, different pathological processes may occur, accounting for the heterogeneous clinical phenotypes of ARF (table 1, see major manifestations).

Most investigations into the immune basis of ARF have used biased (directed) approaches, with limited utility. For example, elevated tumour necrosis factor-α, interleukin 6 (IL-6) and IL-8 have been reported in patients with active ARF and RHD, but are unreliable disease markers. In difficult-to-diagnose diseases, the use of a multiplicity of biomarkers (‘multivariate’) to identify the presence or progression of disease has proven useful. Using microarrays, characteristic gene expression signatures have been reported in white blood cells from patients with Kawasaki disease (another inflammatory cardiac condition), lupus and tuberculosis demonstrating that multivariate expression profiling has potential to identify biomarker composite signatures as diagnostic tools. Multivariate approaches to analysing biomarkers in sera or plasma have also shown promise. In a bead-based multiplex assay, a number of Mycobacterium tuberculosis-reactive antibodies were successfully used to diagnose tuberculosis in non-human primates, and antibody glycosylation patterns can discriminate latent from active TB in humans. Most importantly, our recent studies show proof-of-principle for immunopathogenesis studies to understand ARF, identifying a dysregulated cytokine axis using multivariate approaches including RNAseq and flow cytometry as well as a linked IgG3-C4 response in early ARF with multiplex bead-based assays. Serumborne and plasmaborne biomarkers in addition to transcriptional profiling and circulating cellular components thus constitute a strong and accessible means of diagnosis once useful biomarkers have been identified.

Our hypothesis is that unbiased, multivariate analyses hold the greatest potential for identifying meaningful, translatable information on the immunopathogenesis of ARF. The aims of the ‘Searching for a Technology-Driven Acute Rheumatic Fever Test’ (START) study are to determine whether this approach can identify a biomarker signature that accurately classifies ARF diagnoses according to a defined ‘gold-standard’ diagnosis (the 2015 revised Jones criteria applied by a panel of clinical experts), and to better understand ARF immune pathogenesis (table 2). Specifically, we aim to develop, in a discovery cohort and validate in a second cohort, a profile of metabolic and immunological biomarkers that distinguishes ARF cases from a range of non-ARF conditions and healthy controls.

Table 2

<table>
<thead>
<tr>
<th>Control groups</th>
<th>Non-ARF streptococcal infections or toxin mediated condition or poststreptococcal condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A (n=30):</td>
<td>Examples: Active infection with a proven or likely causative streptococcal organism (eg, β-haemolytic streptococcus, α-haemolytic streptococcus, S. pneumoniae etc) from sterile or non-sterile site. Scarlet fever. Acute poststreptococcal glomerulonephritis. No evidence of definite RHD on echocardiogram.</td>
</tr>
<tr>
<td>Group B (n=30):</td>
<td>Other acute condition Examples: Inflammatory conditions, including autoimmune inflammatory conditions (eg, lupus with arthritis), pancreatitis, etc. Infective condition: confirmed arboviral infection (eg, Ross River virus), septic arthritis, bacterial endocarditis, etc. Trauma requiring hospitalisation or operative procedure. No evidence of definite RHD on echocardiogram.</td>
</tr>
<tr>
<td>Group C (n=30):</td>
<td>RHD Echocardiographically confirmed RHD with no active inflammation. History of definite ARF but no current disease activity.</td>
</tr>
<tr>
<td>Group D (n=30):</td>
<td>Healthy No intercurrent medical condition as determined by clinical questionnaire. Minor trauma allowed. No evidence of definite RHD on echocardiogram.</td>
</tr>
</tbody>
</table>

ARF, acute rheumatic fever; RHD, rheumatic heart disease.
METHODS

Design
START is a cross-national, prospective, observational study with a quasi case-control design. We will prospectively recruit children and young adults with ARF, and people matched on age, sex and ethnicity who have eligible control diagnoses, in Australia and NZ. The study will comprise a discovery cohort (up to 65 definite cases and 65 controls) and a validation cohort (up to 25 definite cases and 65 controls). Probable and possible ARF diagnoses will then be tested independently (figure 1). Eligible control conditions are shown in table 2. Participants will have blood collected for comprehensive immune profiling.

Ethics and dissemination
The study has approval from the NT Department of Health and Menzies School of Health Research ethics committee (18/3126) and the NZ Health and Disability Ethics Committee (18/CEN/197). It will be conducted according to Good Clinical Practice, Good Laboratory Practice and the National Health and Medical Research Council and ethical standards for research involving Indigenous Australians and NZ Māori and Pacific Peoples.

Knowledge dissemination will be through academic channels and community discussion forums. The study team is well placed to foster translation of findings into practical tools, namely, an improved diagnostic and immunomodulatory therapy.

Study sites
Participating hospitals across the two international sites: Royal Darwin Hospital, Darwin, NT, Australia, and Starship and Middlemore Hospitals, Auckland, NZ. Clinical guidelines at all three study sites require that all people with suspected ARF are hospitalised for diagnostic workup and initiation of management.10 24 Royal Darwin Hospital is the major tertiary referral hospital in Australia’s NT. Thirty per cent of the NT population is Indigenous (Aboriginal or Torres Strait Islander).25 ARF occurs at rates >300/100 000 total Aboriginal population in highest-burden communities in the NT.26 Starship and Middlemore Hospitals serve a combined urban population of approximately 1.1 million people in the Auckland region. Māori and Pacific peoples make up approximately 27% of the population of greater Auckland region, where the incidence of ARF among Māori and Pacific children is approximately 33/100 000 and 93/100 000, respectively, for those 5–15 years old.24 27

Study participants
Inclusion criteria for ARF cases are: individuals with suspected or confirmed ARF (the vast majority of whom are expected to identify as being of Aboriginal, Torres
People with suspected ARF will be enrolled as early as possible during admission, prior to a final diagnosis being assigned, to ensure that an acute sample is collected. A final diagnosis will be assigned once all required diagnostic information has become available. Participants in group C (RHD) may additionally be recruited through outpatient clinics. Group D (healthy) will be sought from among family members and friends of other enrolled participants, or healthy members of the community in Darwin.

Clinical data collection

After assigning a sequential study code, clinical details will be recorded on a paper case report form (CRF) then entered into an electronic database (Medrio Electronic Data Capture System), including: date of illness onset, date of blood sample collection, demographics, clinical presentation, Jones criteria, clinical laboratory results, medications received and diagnosis assigned by the clinical treating team.

Assignment of diagnosis

Patients presenting with suspected ARF may have a final diagnosis of definite ARF, probable ARF, possible ARF or a non-ARF condition. Probable ARF is defined in Australian guidelines as an acute presentation not fulfilling criteria, missing a major or minor criterion or lacking evidence of preceding streptococcal infection, but ARF is still considered the most likely diagnosis. Possible ARF applies to the same presentation type, but where ARF is considered uncertain but cannot be ruled out.10 For this study, participants ultimately diagnosed with a non-ARF condition will be allocated by the clinical panel as group A or B controls.

Management will be directed by treating clinicians. Clinical diagnoses will be assigned according to local Australian and NZ Diagnostic Criteria by a panel of ≥2 study clinicians expert in ARF. The decision on diagnostic category requires clinical judgement, after evaluating all aspects of the case. The Jones criteria app which provides a result of definite, probable or possible ARF or not ARF depending on information provided to the algorithm, will be used by the panel at their discretion. The diagnosis of carditis will be made if rheumatic valvulitis is evident on echocardiogram as per Jones criteria. The diagnosis of RHD will be made in accordance with features described by the World Heart Federation.29 The 2015 revised Jones Criteria28 will be applied as the gold standard diagnostic for definite ARF. This represents a departure from normal practice in NZ where local diagnostic criteria differ: polyarthritis is not considered a major criterion in NZ.21 In addition to the Jones criteria, comprehensive clinical and laboratory information, such as results of synovial fluid testing, serology for alternative infectious aetiologies, autoantibodies, radiological findings, diagnosis assigned by the treating team, family history and local epidemiology, will be considered in adjudicating the diagnosis. ARF type will be further specified...
as initial episode or recurrence. Where the adjudication panel’s diagnosis differs from that of the treating team with implications for management, this will be communicated to the treating specialist. The clinical panel will also assign final diagnoses for controls.

Blood collection
Blood will be collected at baseline, and for patients with ARF, on follow-up occasions during the convalescent period as able, for example, during prolonged hospitalisation or at a later outpatient appointment (table 3). A safe maximum volume of 28.5 mL depending on age will be obtained in Australia,31 and 14.5 mL in NZ. Baseline samples will be collected as soon as possible during the acute presentation, timed to coincide with routine blood testing wherever possible. Convalescent samples will be used to determine persistence and decay of immunopathological signatures post ARF. After collection, samples will be transported immediately to research laboratories at Menzies School of Health Research (Australia) or the University of Auckland (NZ) for centrifugation, serum/plasma separation and viable peripheral blood mononuclear cell (PBMC) preparation. Timely freezing and storage of serum/plasma aliquots, PAXgene tubes at −80°C and PBMCs in gas-phase liquid nitrogen. Samples will be shipped periodically to relevant laboratories.

Sample size
It is not possible to predict a priori the combined discriminatory ability of independently measured factors. This study has elected to enrol up to 120 ARF cases (comprising approximately 90 definite ARF, 15 probable, 15 possible) and 130 controls (30 each of control groups A, B and C, and 40 control group D; figure 1 and table 2). These numbers are feasible and should enable characterisation of different ARF clinical phenotypes (carditis, arthritis, chorea, skin/soft-tissue manifestations or a combination of these) and phases (first or recurrent episodes), acknowledging heterogeneity among participants (ethnicity, age, sex; on corticosteroids or not; at different stages of illness).

In the multivariate approach, it is expected that most, if not all, of the influential factors in the optimal model will be significant (p<0.05) in univariate testing. Assuming α of 5% and 95% power (β=0.05) the minimum sample size is 26 cases. These calculations are only a guideline for removing clear false positives, as the standard procedure for producing robust screening models is to have a separate blinded analysis of a test subset of samples. We will use discovery and validation cohorts, together with stratified bootstrap cross-validation to internally optimise the structural parameters in each model. We further used the standard inferential approach to sample size estimation in diagnostic test studies of biomedical informatics.32 An effective multivariate predictor of ARF with positive predictive value of 0.8 (clinically useful), and a 95% CI of ±0.1, would require 65 ARF cases in the discovery cohort.

The discovery cohort will, therefore, comprise a target of 65 definite ARF cases and 65 controls. The validation cohort will comprise a target of 25 definite ARF cases and 65 controls. Probable and possible diagnoses (~15 in each group) will be tested independently (figure 1). Cases will be allocated to the discovery or validation cohorts using a computer-generated random selection of unique study identification numbers.

Laboratory methods
The overarching aim is to develop and validate a profile of host related biomarkers that distinguishes ARF cases from non-ARF conditions and healthy controls. This will be achieved through the following analyses on peripheral blood samples.

Immunophenotyping
PBMC will be stained with labelled antibody panels to identify specific cell populations. Flow cytometry raw
data will be analysed manually using Flowjo software and via an automated gating platform as described.\textsuperscript{33 34} Such high throughput, automated analysis of big flow cytometry data offers several advantages over manual gating, including increased throughput while increasing quality control (such as preprocessing removal of anomalous events via the flowCut algorithm), and identification of specific cell populations with up to 50-dimensional datasets.\textsuperscript{34–36} Plasma samples will be analysed using multiplex cytokine assays to quantify cytokine and chemokine plasma concentrations, detailed elsewhere.\textsuperscript{33 37–39}

**Metabolome analyses**

Untargeted metabolomic profiling (>1000 metabolites) will be performed on plasma samples using liquid chromatography coupled to high-resolution mass-spectrometry (LC-MS). Data will be acquired using three modes of operation: reverse-phase/ultraperformance liquid chromatography (UPLC)-MS/MS with positive ion mode electrospray ionisation (ESI), reverse-phase/UPLC-MS/MS with negative ion mode ESI, and Hydrophilic interaction (HILIC)/UPLC-MS/MS with negative ion mode ESI. All identified metabolites will be annotated using appropriate orthogonal analytical techniques applied to the metabolite of interest against a chemical reference standard.

**Blood transcriptomics**

RNA will be extracted from stabilised whole blood samples (PAXgene tubes) as described,\textsuperscript{40–41} libraries prepared (TruSeq Stranded Total RNA with Ribo-Zero Globin reduction, Illumina) and NextGen sequencing undertaken (Illumina HiSeq2500, 50bp single-end reads). Read alignment and gene-level quantification (counts) will be performed using Hisat.\textsuperscript{42} Negative binomial models will be employed for differential expression analysis, with false discovery rate control for multiple testing.\textsuperscript{43} The analyses will be adjusted for batch effects and variations in cellular composition, which will be estimated employing Remove Unwanted Variation from RNA-Seq Data (RUVseq) and CIBERSORT, respectively.\textsuperscript{44 45} Genes will be mapped to blood transcriptional modules to provide a systems-level view of the responses and reduce the dimensionality of the data.\textsuperscript{46}

**CD4 T cell transcriptomic responses to GAS**

A two-phased approach will be used to examine CD4 T cell responses to GAS. First, PBMCs from a subset of definite ARF cases and healthy controls will be cultured under a variety of conditions to identify optimal conditions for the second phase. Cells will be harvested at multiple time points poststimulation (6 hours, 24 hours and 48 hours) with heat-killed and antibiotic-killed ARF-associated GAS strains\textsuperscript{12 47–48} and a selection of candidate ARF antigens. Total RNA from PBMCs cultured under each condition will be profiled by RNA-Seq (Illumina 100bp paired-end reads, 20M reads), and culture supernatants will be examined by Luminex (Bio-Plex 48-plex Pro Human Cytokine Screening Panel) to identify optimal conditions for maximal discrimination of responses between cases and controls. In the second phase, an expanded set of ARF cases and controls for which PBMC samples are available will be examined following culture under the identified, optimal conditions. Gene expression patterns and cytokine production will be profiled by RNA-Seq and Luminex, as described above.

**Antibody analysis**

For unbiased investigation of autoantibodies, selected sera will be screened against planar protein microarrays comprised of over 42000 protein fragments representing some 19000 human proteins.\textsuperscript{49} Protein fragments that are more significantly bound by autoantibodies in ARF compared with controls will be identified using a >0.05 and fold change of 2.0 as cut-off. A suspension bead array will then be designed\textsuperscript{50} composed of up to 380 potential autoantigens to assess autoantigen reactivity in all sera. The bead-antigens will be selected based on the planar array results, previously completed screens using high-content protein arrays,\textsuperscript{51} and targets from the literature. Finally, candidate ARF autoantigens identified in the suspension bead-array will be orthogonally validated as individual antigens in ELISA or Luminex bead-based assays to determine sensitivity and specificity.

**Bioinformatics and statistical analyses**

The data collected for each patient across the immune phenotyping, transcriptomic, proteomic and metabolomic technologies will be integrated computationally to identify patterns of covariance within and between designated clinical outcome groups. Convergence of signatures across diverse systems biology domains will provide independent functional validation. The use of ‘multiomic’ data integration techniques will also enable the possibility of deriving novel biological information that will not be revealed in a single dataset alone. Data integration methods applying different computational strategies will be used to explore the complex covariance structure within and between the multiomic data blocks. Initially, analysis will focus on each data block in isolation employing a combination of classical generalised linear modelling, multivariate projection regression models (Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA)) and unsupervised cluster analysis. The multiple blocks will then be integrated into a single computational model to enable multiomic functional mapping, which in turn will allow us to uncover the underlying biochemical mechanisms. Several methods for performing multiblock data integration will be investigated, from which a consensus model will be derived. These will include: protein–protein interaction networks using NetworkAnalyst,\textsuperscript{52} Multiblock component analysis,\textsuperscript{53} regularised canonical correlation analysis,\textsuperscript{54} the Data Integration Analysis for Biomarker discovery using Latent Component framework\textsuperscript{55} and similarity network fusion.\textsuperscript{56}
This systems biology analysis should result in both a domain-specific and domain-integrative summary of biological phenomena that are associated with clinical outcome. Likely outcomes are a signature biological pathway associated with ARF as well as individual candidate markers that may provide the basis for the development of novel diagnostic tests. Furthermore, the data generated through this effort will allow crisply defined power calculations for future narrowly targeted assessment of potential biomarkers in clinical trials.

**DISCUSSION**

We anticipate that study findings will provide the most comprehensive knowledge of the immunopathogenesis of ARF to date, be used as the basis for development of a diagnostic test and provide a pathway towards development of targeted immunomodulatory treatment(s). START commenced recruitment in Australia in November 2018 and in NZ in May 2019. Anticipated completion of recruitment is end-2021. Laboratory analyses will be batched and run together at completion of recruitment.

Both underdiagnosed and overdiagnosis of ARF pose major challenges to individuals and health systems. Whether the use of the ‘probable’ and ‘possible’ ARF diagnostic categories is reducing underdiagnosis or contributing to overdiagnosis is currently unknown. Should findings from this study successfully identify a discriminatory biomarker profile differentiating definite ARF from non-ARF, this will provide a mechanism for accurate diagnosis to guide appropriate management. This will be a critically important advance in the diagnosis and management of ARF globally.

A minority of any population are at risk of ARF after exposure to group A streptococci. The estimated lifetime cumulative incidence of ARF was previously calculated at 5.7% in Australia’s NT. This was considered likely to be an underestimate, and indeed more recent data highlight the burden of unrecognised ARF in Aboriginal communities in the NT where repeated infection with GAS is ubiquitous from early childhood, suggesting that the lifetime risk of ARF in these populations is higher than 6%. In US military camps, 2%–3% of recruits developed ARF from non-ARF, this will provide a mechanism for accurate diagnosis to guide appropriate management. This will be a critically important advance in the diagnosis and management of ARF globally.

Previous efforts to develop ARF diagnostic tests and elucidate the immunopathogenesis have laid foundations for the START study design and analysis. Kim et al analysed responses to GAS challenge of PBMC from ARF patients at Royal Darwin Hospital using multiplex cytokine array, flow cytometric analysis and global gene expression analysis by RNA sequencing. They identified a dysregulated IL-1β-granulocyte-macrophage colony-stimulating factor cytokine axis in PBMCs from ARF patients, and the potential to suppress this response by hydroxychloroquine. The authors proposed that hydroxychloroquine, an immunomodulatory agent, could be repurposed to reduce the risk or severity of RHD after ARF. Clinical use of hydroxychloroquine, now reported as safe in two cases, requires further investigation. A study using sera from patients with ARF in NZ combined multiplex bead-based assays and systems immunology data analysis to identify a linked IgG3-C4 response that may have utility as a clinical biomarker in early ARF. However, small sample sizes have been a limiting factor. Key strengths of the current study include the larger sample size, multi-centre enrolment across different countries to maximise relevance and use of the latest laboratory approaches.

Limitations include that capture of eligible participants early in disease is challenging, due to the variable time taken for participants to be hospitalised after disease onset, to identify and gain consent from participants, and obtain blood samples. At the Australian site, the majority of eligible participants will be Aboriginal children living in remote communities, often hundreds of kilometres distant from the enrolment site. Although Auckland is a large urban city inequitable healthcare access is experienced by many Māori and Pacific people living in socioeconomic deprivation and this can lead to delayed hospitalisation for many children presenting with ARF. Given these inherent delays, a proportion of baseline blood samples will be collected after the peak inflammatory phase has passed. Another limitation is heterogeneity among people with ARF with regards to ARF type (first or recurrent), diagnostic category (possible, probable, definite), clinical phenotype (carditis, chorea, arthritis, etc) and demographic variations; however, we limited enrolment to people below 30 years and only of Aboriginal, Torres Strait Islander, Māori or Pacific ethnicity to minimise heterogeneity. Although a distinguishing biomarker signature may be identified, the ability to readily translate study findings into a feasible diagnostic able to be used in health service laboratories outside a research environment, is uncertain.

Successful completion of this study will considerably improve knowledge of ARF immunology. Innovative strategies to improve the clinical management of ARF are core components of the overall suite of activities required to achieve the goal of RHD elimination.

**Author affiliations**

1. Global and Tropical Health, Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia
2. Royal Darwin Hospital, Darwin, Northern Territory, Australia
3. KidzFirst Hospital, Counties Manukau District Health Board, Auckland, New Zealand
4. Starship Children’s Hospital, Auckland, New Zealand
5. Department of Paediatrics; Child and Youth Health, University of Auckland, Auckland, New Zealand
6. School of Medical Sciences and Maurice Wilkins Centre, The University of Auckland, Auckland, New Zealand
7. Westfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, Perth, Western Australia, Australia
8. Centre for Integrative Metabolomics and Computational Biology, Edith Cowan University, Perth, Western Australia, Australia
9. Department of Paediatrics, University of Otago, Wellington, New Zealand
10. Department of Infectious Diseases, Perth Children’s Hospital, Perth, Western Australia, Australia
11. School of Medicine, University of Western Australia, Perth, Western Australia, Australia
Acknowledgements We are very grateful to Caroline Motteram for project management, Yolanda Hernandez, Laura Francis and Vincent Mitton for recruiting participants in Darwin and Shirley Lawrene, Rhonda Holloway, Renee Clark, Alicia Stanley and Roxanne Buchanan for recruiting participants in New Zealand. We thank Alex Kaethner and NT Cardiac for facilitating echocardiography. Vanessa Riga, Erin Gargan, Gabriella Minigo and Michael Serralha are gratefully acknowledged for laboratory support; Kryan Kuzio, Tasha Cole, Alana Whitcombe, Lauren Carlton and Timothy Ho for laboratory assistance; Amy Baker and Rebecca Trowman for program and financial management.

Contributors JC, AR, RW, NJM, RM, AB, DB, TCB, MM, GP and TK conceived and designed the study. NJM, RM, AB, DB, TL, TCB, RB, JB and TK will perform the assays and analyse the data. GP and MM provide project governance. AR, RW, JY, BR, NW and JC provided clinical care, patient recruitment, assessment and clinical review panel participation. AR created the first draft and managed the edits to the manuscript. All authors reviewed and approved the study protocol.

Funding The study is funded by the Australian National Health and Medical Research Council (NHMRC) grant #1147531. AR is supported by NHMRC fellowship #1142011. TCB is supported by the Western Australia Department of Health and a Research Council (NHMRC) grant #1131932. AR is supported by NHMRC fellowship R10. BMJ Open: e053720. doi:10.1136/bmjopen-2021-053720. 2021; 11: e053720.

Patient consent for publication The Methods section for further details.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; peer reviewed for ethical and funding approval prior to submission.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD Anna P Ralph http://orcid.org/0000-0002-2253-5749

REFERENCES


