Scientific Summary (Scientific Report)

A. What were the main scientific objectives of the grant?

Because of a lack of breakthrough new medicines, patients suffering from systemic lupus erythematosus (SLE, lupus) continued to be treated with non-specific immunosuppressants, which have limited effectiveness and result in significant toxicity, largely unchanged since the 1960s, with only one new drug (belimumab) being licensed in 2011 [1]; it is not on the Australian PBS. As a result, patients suffer accrual of irreversible organ damage [2], low quality of life, and dramatically reduced life expectancy [3].

Failed high-profile clinical trials of targeted therapies in SLE are now understood to reflect the application of these drugs to untargeted patients, i.e. a failure to take into account the biological diversity of SLE [4]. To improve patient outcomes in SLE, there is therefore an urgent need for tools that have utility for assigning therapies based on biologically-determined stratification [4, 5]. Due to the striking biological and clinical heterogeneity of SLE, an unbiased and large-scale multi-omics approach is required for understanding different patient clusters and predicting which patients may benefit from targeted therapies.

Among Australian SLE patients, Indigenous Australians are substantially overrepresented both in terms of overall prevalence and the subgroup of SLE patients with severe organ- or life-threating disease. The reason for this is unclear, but is likely to reflect a combination of genetic and environmental risk factors [6]. These risk factors have not been characterised, and how they are reflected in terms of the 'expressed genome' (i.e. transcriptome, proteome, metabolome etc) is unknown. My hypothesis is that *unique protein and metabolomic signatures will identify biological subsets of patients with SLE* enabling assignment to targeted therapies. This knowledge will identify which biological pathways are dominant in Indigenous Australian lupus in order to prescribe them the most effective targeted therapy, such as drugs targeting the cytokine B cell-activating factor (BAFF) (belimumab) or the Type I interferon (IFN) pathway (anifrolumab, successful Phase III clinical trial in 2020). To address this, I sought to establish the first national registry and biobank of Indigenous Australians with SLE. Over the course of this project, and beyond, biobanking (serum and mRNA) will allow me to analyse proteome, metabolome and transcriptome data from a population with corresponding demographic and longitudinal clinical data. Using this new registry, I will test the hypothesis that *the specific pathways are dominant in Indigenous Australians with SLE*, and will create new capacity to test the utility of other biological drugs in Indigenous Australians as they are developed.

In this estimated 5-year research program, I aim to:

- **Aim 1:** Align protein and metabolome profiling on transcriptome signatures to identify biological subsets of patients with SLE.
- **Aim 2:** Generate a definitive biological profile of Indigenous Australians with SLE.

In the first year of this research program, supported by the AFA/ARA Health Fellowship the main sub-Aims were to:

- **Sub-Aim 1.1**: re-analyse our pilot serum cytokine proteomics SLE data in collaboration with Dr James Peters' team at the Imperial College London, in order to develop an optimised proteomics analysis pipeline.
- **Sub-Aim 2.1**: advance the development of the registry and biobank of Indigenous Australians SLE.
- **Sub-Aim 2.2**: secure more funding and develop key collaborations for -omics layers data generation and analytics.

These goals have been achieved.

B. What were the main scientific achievements of the grant? Your answer should be at least 200 words.

Sub-Aim 1.1:
Because SLE is a highly heterogeneous and dynamically unpredictable autoimmune disease, a better understanding of the molecular differences between patients is needed in order to direct emerging targeted therapies. The aim of this pilot study was to evaluate whether analysis of serum proteomic profiles of SLE patients using unbiased approaches would yield clinically meaningful results. Here, we used a large multiplex approach to identify differentially expressed serum proteins in SLE, compared to healthy controls (HC). 211 proteins were quantified, and clinical data, including disease activity and organ damage, were collected prospectively. A machine learning approach was used to cluster study participants based on proteomic data only.

198 SLE patients and 38 healthy subjects were recruited. We found 37 serum proteins significantly differentially expressed in SLE including 33 upregulated and 4 downregulated, and amongst which 13 are reported here for the first time. Two clusters of proteins were identified from a correlation matrix of the 37 proteins, one particularly comprised of interferon-regulated chemokines. Unsupervised feature selection methods identified reduced set of proteins, comprising between 24 and 53 relevant analytes. Patients were clustered, according to the analytes that remained after feature selection, using k-means algorithm, and these biologically determined clusters significantly differed cross-sectionally in clinical characteristics, including overall and renal disease activity, as well as organ damage. In conclusion, unsupervised analytics of large serum protein profiles in SLE yielded a tractable reduced analyte set, which revealed biologically distinct clusters of patients who in turn had significant differences in clinical profile. These findings indicate the potential for identifying clinically meaningful biological subsets of SLE based on serum protein profiling.

Sub-Aims 2.1/2.2:
Ethics/Governance clearance in Northern Territory
Through the network of collaborators that I have initiated, developed and maintained, there is a unique opportunity to obtain clinical biospecimens from Indigenous Australians with SLE. One of the main research program objectives was to create and implement the first ever registry and biobank of Indigenous Australians SLE nested within the Australian Lupus Registry & Biobank (ALRB). We are now ready to enrol Indigenous Australian SLE patients particularly via collaborations with Dr Stephen Brady, Dr Vipin Tayal and Dr Sachin Khetan (NT), and

A/Prof. Maureen Rischmueller (SA and NT). We recently reached a key milestone in this project, being granted Ethics and Governance approvals for the two pivotal participating sites located in Northern Territory, namely Alice Springs Hospital (HREC Ref: CA-19-3570; approved 03/2020; Governance approval: 05/2021) and Royal Darwin Hospital (HREC Ref: 2019-3329; approved 10/2019; Governance approval: 03/2021), to be part of the ALRB. In addition, we were successful in obtaining Governance approvals for Monash Health (10/2019) and The Queen Elizabeth Hospital (09/2020) participating sites to start this project, the latter including a key approval from the Aboriginal Health Council of South Australia (AHCSA; 08/2020) as a pre-requisite. Overall, this means that SLE patient recruitment is starting at active participating sites in NT, SA and Vic.

Funding Support to reach ~$1M AUD

Another key element for this research program to be successful is the funding support not only for studying the stacked multi-omics layers, but also to enable implementation of the project logistics in participating sites, particularly in NT. This includes, but is not limited to, implementing biosample collection, processing and storage on site, biosample shipment, as well as demographics/clinical data collection, use of Aboriginal interpreters, etc. I have been successful in securing ~$1M AUD in Grants and Fellowships (all CIA) from seven different funding bodies to allow this project to start. This includes the initiation and development of a key collaboration with CSL, who will cover the cost for and perform transcriptomics studies (genome-wide RNA-seq studies).

Key International Collaboration Development

Finally, I have initiated and developed an international collaboration with Dr James Peters (Imperial College London), one of the few rheumatologist-bioinformaticians in the world, with dual training in clinical medicine and computational biology/genomics. Dr James Peters will provide guidance and advice on analytic strategy for -omics/multi-omics data. We have since started our collaborative work by re-analysing our data from our pilot serum cytokine proteomics project in SLE (*See Sub-Aim 1.1*).

C. What problems, if any, did you encounter in achieving the project's objectives, and how did you address them?

The progress of our research program suffered from delay in site-Governance approvals for the activation of the two Northern Territory sites, initiated in March 2020 and only recently completed. The current COVID-19 situation was a contributing factor.

D. Have you disseminated, or plan to disseminate, the results of this research?

Please tell us about:

- ***References for peer-reviewed papers that have been published (please provide pdf copies of papers if possible)***
- ***Papers that have been submitted and/or accepted for publication***
- ***Meetings/conferences at which you have presented this research, or are due to present it (please provide abstracts if possible)***
- ***Any other ways in which you may have disseminated the research, including to the public and the media (please provide urls to relevant press releases or media articles)***

International Meeting & Journal Publications

We have presented some results of our serum cytokine proteomics pilot project at the 2020 American College of Rheumatology (ACR) Convergence annual meeting, as detailed below. After improvement of our research analytics pipeline in collaboration with Dr Peters, we are now aiming to publish outcomes of this project in a top-tier peer-reviewed journal, targeting submission to *Nature Communications*, *Med*, *Lancet Rheumatology* or *Cell Reports Medicine*.

> **Vincent FB**, Ong J, Nim HT, Boyd SE, Hoi AY, and Morand EF. Serum Cytokine Profiling in Systemic Lupus Erythematosus, Analysed using Unsupervised Machine Learning, Reveals Clusters that are Clinically Distinct. Poster session presented [online] at: American College of Rheumatology (ACR) Convergence 2020 [Abstract ID 909010]; 2020 Nov 5-9; Washington, USA.

Supported by my Fellowship I have published six articles in peer-reviewed journals and one book chapter, and two book chapters have been submitted. Please find below the list of the above-mentioned journal publications and book chapters.

Peer-Reviewed Journal Publications 2020-21

1. Santarelli DM, **Vincent FB**, Rudloff I, Nold-Petry CA, Nold MF, Russo MA. Circulating interleukin-37 levels in healthy adult humans – establishing a reference range. *Front Immunol* 2021; 12:708425
2. Gottschalk TA, **Vincent FB**, Hoi AY, Hibbs ML. Granulocyte colony-stimulating factor is not pathogenic in lupus nephritis. *Immun Inflamm Dis* 2021; doi: 10.1002/iid3.430
3. Nocturne G, Ly B, Paoletti A, Pascaud J, Seror R, Nicco C, Mackay F, **Vincent FB**, Lazure T, Ferlicot S, Stimmer L, Pascal Q, Roulland S, Krzysiek R, Hacein-Bey S, Batteux F, Mariette X. Long-term exposure to monoclonal anti-TNF is associated with an increased risk of lymphoma in BAFF-transgenic mice. *Clin. Exp. Immunol* 2021; 205:169-181
4. Nataraja C, Dankers W, Flynn J, Lee JPW, Zhu W, **Vincent FB**, Gearing J, Ooi J, Pervin M, Sherlock R, Hasnat Md, Harris J, Morand EF and Jones SA. GILZ regulates the expression of proinflammatory cytokines and protects against end-organ damage in a model of lupus. *Front Immunol* 2021; 12:652800
5. Apostolopoulos D, **Vincent FB**, Hoi AY, Morand EF. Associations of Metabolic Syndrome in SLE. *Lupus Sci Med* 2020; 7:e000436.
6. **Vincent FB**, Kandane-Rathnayake R, Koelmeyer R, Hoi AY, Harris J, Mackay F and Morand EF. Associations of serum soluble Fas and Fas ligand (FasL) with outcomes in systemic lupus erythematosus. *Lupus Sci Med* 2020; 7:e000375

Book Chapters 2020-21

1. **Vincent FB**, Figgett WA, Hibbs ML. (2021) Hallmark of systemic lupus erythematous: role of B cell hyperactivity. In: Hoi AY (ed) Pathogenesis of systemic lupus erythematous - insights from translational research. (*Submitted*)
2. **Vincent FB**, Figgett WA, Hibbs ML. (2021) B cell-specific targeted therapies in systemic lupus erythematous. In: Hoi AY (ed) Pathogenesis of systemic lupus erythematous - insights from translational research. (*Submitted*)
3. **Vincent FB**, Lang T. (2020) Measuring MIF in Biological Fluids. In: Harris J., Morand E. (eds) Macrophage Migration Inhibitory Factor. Methods in Molecular Biology, vol 2080. Humana, New York, NY

Research Funding

During the time of this fellowship, and in support of my growing independence, I have been awarded a highly competitive 5-year NHMRC Investigator Grant (EL1) to support this work identifying molecular signatures in Indigenous Australian SLE. I have also initiated and developed a key collaboration with CSL, who will cover the cost for and perform transcriptomics studies (genome-wide RNA-seq studies) in biosamples from Indigenous Australians with SLE (N = 100), and those from all control groups [healthy Indigenous Australians (N = 30); healthy non-Indigenous Australians (N = 30); non-Indigenous Australians SLE (N = 60)] (*estimated value: $110K AUD*). In 2021, I have further developed

our collaboration with CSL, who will provide the same support for an extra number of 120 biosamples, worth ~$60K AUD. In addition, I have initiated a collaboration with SomaLogic to test another wide-angled proteomics platform (SomaScan Discovery) in a pilot project. The overall funding support for this research program now totals ~$1M in Grants and Fellowships (all CIA) from seven different funding bodies. Please find below the list of the above-mentioned Grant/Fellowship obtained in 2020-21.

Research Funding 2020-21
Total 3 awards; worth ~ $830K AUD
2021-2025 NHMRC Investigator Grant – Emerging Leader 1 ($645,205 AUD) (CIA)
2020-2022 CSL INNOVATION PTY LTD Research Contract (estimated value: $110,000 AUD + $60,000 AUD) (CIA)
2020-2021 Janssen-Cilag Pty Ltd Medical Grant ($15,000 AUD) (CIA)

2020 ACR Abstract
--

Title *(character limit: 250 characters, excluding spaces)* ***(Chrs count: 127)***
Serum Cytokine Profiling in Systemic Lupus Erythematosus, Analysed using Unsupervised Machine Learning, Reveals Clinically Relevant Clusters

Fabien B Vincent[1], Jannina Ong[1], Alberta Y Hoi[1], Sarah E Boyd[1], Hieu T Nim[1], and Eric F Morand[1]
[1] Monash University, Clayton, Victoria 3168, Australia.

Keywords: Biomarker; Cluster; Cytokine; Feature selection; Machine learning; Proteomics; Systemic lupus erythematosus (SLE).

Body character limit: 2,750 characters, which EXCLUDES the title, names of authors/co-authors, authors' affiliations, spacing, and disclosures.

(Chrs count: 2,558)

Background/Purpose:
SLE is a heterogeneous disease, where a better understanding of molecular differences between patients is needed in order to direct therapy. Existing approaches generally examine mRNA expression, whereas therapeutic targets are mostly soluble or cellular proteins. We aimed to evaluate whether a serum cytokine proteomic profile of SLE patients analysed using an unbiased approach would yield clinically meaningful results.

Methods:
Demographic and clinical data of SLE patients (ACR criteria) including disease activity (SLEDAI2K) and organ damage (SLICC Damage Index (SDI)), and matching serum samples, were collected prospectively. A wide-angled serum cytokine proteomics analysis was conducted using Quantibody, Luminex, and ELISA platforms. To reduce heterogeneous data complexity and remove redundant variables, unsupervised feature selection methods were applied. For clustering, machine learning approaches were used, and optimal cluster number confirmed by consensus clustering.

Results:
198 SLE patients (median [IQR] age 46.7 [36.7, 56.3] years, 88.4% female, median SLEDAI2K 4 [2, 6], 52.5% with organ damage) and 37 sex/ethnicity-matched healthy subjects were recruited. 211 serum analytes were measured. Using unsupervised feature selection methods, we identified a reduced set of 10/211 analytes (MCP-1, GH, CTACK, HCC-1, CD14, ErbB3, E-Selectin, Trappin-2, Cathepsin S and IL-18). The dataset was then clustered, according to the analytes that remained after feature selection, using unsupervised machine learning approaches. k-means analysis produced a distinct result in terms of separability of two clusters that associated only in SLE (**Figure 1)**, and was chosen as the final clustering algorithm. We analysed clinical parameters in patients categorised by these biomarker clusters, and found the clusters differed in clinical characteristics including organ damage (P = 0.04), proportion of patients with high ESR (P = 0.03), and SLEDAI2K (P = 0.08). Using multivariable linear regression models, cross-sectional associations with patient clinical characteristics were found for 8/10 analytes. We next assessed associations of baseline analytes with longitudinal disease outcomes, using multivariable logistic regression models. Trappin-2 and IL-18 were significantly associated with damage accrual. IL-18 and CTACK had positive and negative associations respectively with lupus low disease activity state (LLDAS) attainment, and converse associations with indices of active disease over time.

Conclusion:
Unsupervised analytics of wide-angle serum cytokine profiles in SLE yielded a tractable reduced analyte set, which in turn revealed two biologically distinct clusters of patients who had significant differences in clinical profile, individual analytes from which were associated with longitudinal outcomes. These findings indicate the potential for biological subsets of SLE to be based on serum cytokine profiling.

Disclosures: FV Vincent, None; **J Ong**, None; **AY Hoi**, None; **SE Boyd**, None; **HT Nim**, None; **EF Morand**, None.

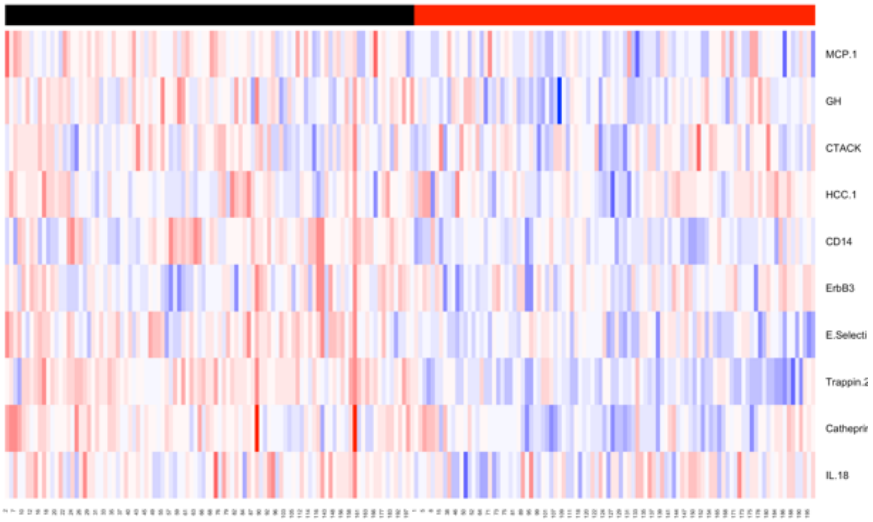

Figure 1. Heat map of 10 analytes, selected using unsupervised feature selection on 211 analytes measured in healthy subjects and SLE patients and clustered using k-means techniques, in SLE patients (n=198). Two clusters (indicated by red or black on top row) were identified in SLE.