

2016 Annual Report

Victorian Life Sciences Computation Initiative

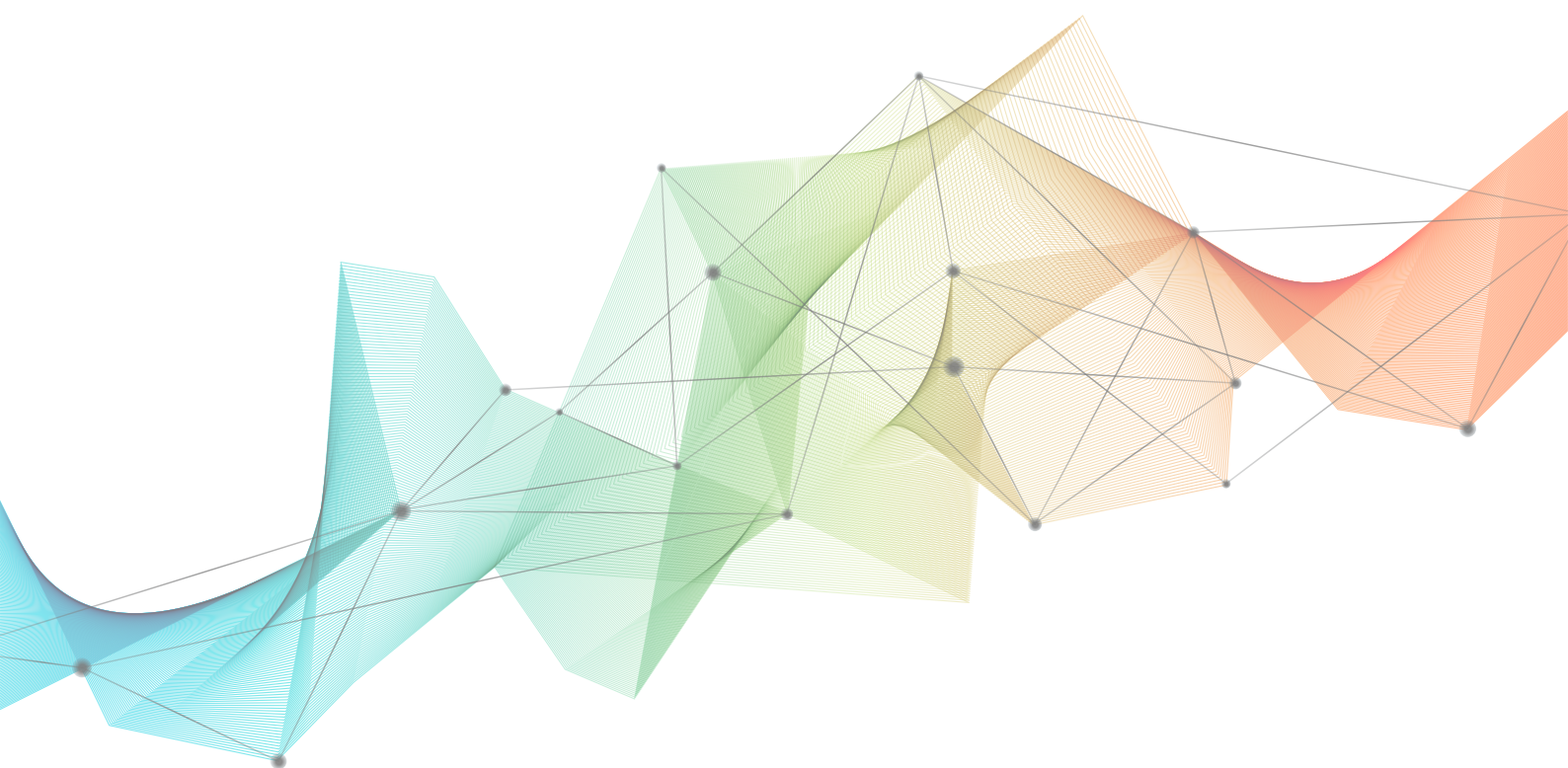


Table of Contents

Table of Contents	2
Director's Report	3
Governance	5
<i>University of Melbourne Life Sciences Computing Steering Group</i>	5
<i>VLSCI Advisory Council</i>	5
Our People	7
<i>VLSCI Staff</i>	7
High-end computing	8
<i>Support</i>	8
<i>Expertise</i>	8
<i>Software</i>	8
<i>Compute Systems</i>	9
<i>Storage infrastructure</i>	9
<i>Hardware Hosting</i>	9
Our Projects - Highlights	10
<i>Cancer Genomics Theme</i>	10
<i>Clinical Genomics Theme</i>	10
<i>EMBL Australia Bioinformatics Resource</i>	10
<i>Microbial Genomics Theme</i>	11
<i>Melbourne Bioinformatics Platform</i>	12
<i>Software Development</i>	13
Our Users	15
Our Collaborations	17
<i>Collaborators</i>	18
<i>Publications resulting from VLSCI staff collaborations</i>	20
Our Community	24
<i>Skills development</i>	24
<i>Hands-on Workshops</i>	26
<i>Contributions to external training</i>	27
<i>Support for Students</i>	29
<i>MSc (Bioinformatics) Students</i>	30
<i>Sponsorship of Meetings and Conferences</i>	31
Financials	32
Glossary	36

Director's Report

The Victorian Government funding of the VLSCI under its 2015-16 Grant Agreement concluded on 31 December 2016. The milestone marked the end of the Victorian Government's formal relationship with VLSCI, having funded the initiative since its inception in 2009. It is timely to acknowledge the foresight and support of the Victorian Government, (through the (now) Department of Economic Development, Jobs, Transport and Resources) from 2009 – 2016, the University of Melbourne (as host of the VLSCI), and its institutional partners: Deakin, La Trobe and RMIT Universities. Due acknowledgement must also go to the researchers and users of VLSCI's systems, IBM Global Financing Australia Limited and IBM Australia Limited.

This 2016 Annual Report is therefore the perfect opportunity to take stock of VLSCI's achievements and evolution into a hub for specialist bioinformatics expertise and services located in Melbourne's key biomedical and biosciences precinct.

In November, the University of Melbourne's Deputy Vice-Chancellor (Research) Prof Jim McCluskey reflected on the growth of the VLSCI and its important role in building momentum for life science computation, which was very much needed back in 2009:

This concentrated project helped to build a state-wide capacity for leading research which, for the University of Melbourne, has meant the development of a team of experts who are continuing to add value to research efforts and VLSCI's institutional, state and national capability.

The impacts of VLSCI activities and collaborations reach far beyond the advice or compute we provide to researchers. This report documents VLSCI's contributions to journal publications and successful grant applications, to professional training workshops and the development of the next wave of graduate bioinformaticians. While VLSCI is firmly embedded in the layers of the life sciences ecosystem, its success has come from its agility and commitment to meeting the ever-changing needs of researchers.

Striving to keeping abreast of the changing computational needs of researchers, our introduction of a fair-share resource allocation scheme in 2016 - with operational elements to deliver greater flexibility and less administrative burden to users - was highly successful. Anticipating user needs beyond local systems, VLSCI staff continued to be central to discussions with Nectar, RDS, Research Platforms and Amazon around optimising long-term options for compute and storage support. Much interest has been expressed internationally regarding Australia's expertise in use of the cloud for genomics analysis and training, and the international uptake of the VLSCI-led Genomics Virtual Laboratory by the UK and Canada for microbial genomics was facilitated by developers at VLSCI.

The agreements between VLSCI members and partners concluded on 31 December 2016 in line with the end of Victorian Government funding and consequent decommissioning of the Blue Gene/Q "Avoca" high performance computer. Given its age and rising maintenance costs, the decision was made by the VLSCI Advisory Council, partner institutions and with the endorsement of the Deputy Vice-Chancellor (Research), to decommission Avoca the Blue Gene/Q by year's end.

Meanwhile, VLSCI has been actively looking to the horizon post-2016. Two member institutions (RMIT and Deakin Universities) joined VLSCI's host institution, University of Melbourne, in a successful ARC LIEF bid for funds to purchase a new GPU cluster in 2017 – adding to the systems that provide an ongoing legacy of the Victorian Government funding of the Initiative.

VLSCI's focus in 2016 was aligned with the strategic development of the emerging national research infrastructure in Australia. Building relationships with international bioinformatics experts and establishing linkages with international institutes on training, data, tools and standards led to my keynote address to the ELIXIR "All-Hands" meeting in Barcelona, Spain, in March.

Through a formal relationship with Bioplatforms Australia, the EMBL-Australia Bioinformatics Resource (EMBL-ABR) was established, with roles for VLSCI as the EMBL-ABR Hub and also as the EMBL-ABR:VLSCI Node. Assoc Prof Vicky Schneider started in February as Deputy Director, with the remit to

drive the national and international engagement and bioinformatics infrastructure agenda of EMBL-ABR. Assoc Prof Schneider swiftly garnered enthusiastic support for a national bioinformatics resource network, built an operational model and devised a business plan from which to launch an ambitious set of activities throughout 2016.

A series of specialised workshops were presented on uncommon but crucial topics - many involving international experts who had not offered training in Australia before. An exceptional International Science Advisory Group (ISAG) was established, comprising representatives from European Bioinformatics Institute (EU), National Institutes of Health (USA), CyVerse (USA), ELIXIR (EU) and key Australian institutes. Their high calibre and expertise reflects the increasing transformation of biology into a rich data science, impacting health, society and industry.

The inaugural ISAG meeting was held at VLSCI in December, and the associated “All-Hands” meeting attracted sixty-two participants from the bioinformatics communities and Nodes across New South Wales, Queensland, Victoria, South Australia, Western Australia and Tasmania.

The goodwill of a network of key institutes and international experts was successfully harnessed to establish a truly national research infrastructure in Australia. The Chair, ISAG, Jason Williams, offered his reflection on the strengthening of this coalition of expertise and effort:

I think the ISAG was very impressed with the progress of EMBL-ABR. One of the reasons for having such a group is to ensure that the quality of the science EMBL-ABR will enable is world-class. We certainly saw this standard in the calibre of the science being presented, the professional organisation of the meetings, and the timely workshop topics. Australia’s unique datasets and research infrastructure are of tremendous value, and we look forward to the positive impact EMBL-ABR will have in realising greater opportunities for discovery.

The University of Melbourne has recognised the value of the ongoing relationships and activities and committed to supporting VLSCI and EMBL-ABR through 2017-18 to enable the broadening uptake of bioinformatics best practice expertise and research. (From 2017, VLSCI will be known as Melbourne Bioinformatics.)

VLSCI has built a team of highly skilled professionals who are dedicated to the future growth and consolidation of VLSCI’s landmark achievements in life sciences in Victoria over its first seven years. With thanks and due acknowledgment to all, I commend the 2016 VLSCI Annual Report.

Director
Assoc Prof Andrew Lonie



Governance

In accordance with key deliverable requirements of the current Grant Agreement (the 2015-16 Grant), two key milestone obligations were delivered by the VLSCI to the Victorian Innovation & Technology Research, Department of Economic Development Jobs, Transport and Resources in 2016.

In September 2016 VLSCI submitted a Milestone 3 and Sustainability Report that confirmed the delivery of activities and research benefits of significant value to Victoria; enabling research outcomes that raise the profile of and delivered advanced computational life sciences education, training and expertise to Victorian based life sciences researchers, students and employees.

The Sustainability Report focused on VLSCI's progress of the Project, future positioning and potential Commonwealth research infrastructure funding opportunities and an exploration of sustainability initiatives of the VLSCI post-2016.

In April 2016 VLSCI submitted its 2015 Annual Report (including audited accounts) detailed progress and achievements and outcomes for the Project, including the activities and research benefits of significant value to the State of Victoria and beyond.

University of Melbourne Life Sciences Computing Steering Group

In 2016 the University of Melbourne Life Sciences Computing Steering Group continued to provide strategic guidance and oversight of the VLSCI Advisory Council, by monitoring contractual obligations between the University, as host and institutional partner of the VLSCI, and the Victorian State Government as its major funder through the current Grant Agreement (the 2015-16 Grant) to its conclusion of the Group and the Victorian Government's funding of the VLSCI.

The LSCSG met in February, August and a final meeting held in November with the conclusion of its oversight obligations. Its membership comprised Prof Jim McCluskey, Deputy Vice-Chancellor (Research) (Chair), Prof Liz Sonenberg, Pro Vice-Chancellor (Research Collaboration and Infrastructure) (and DVCR alternate), nominees from the Deans, Faculty of Medicine, Dentistry and Health Sciences, Science and Engineering, Prof Justin Zobel, Head, Department of Computing and Information Systems, Melbourne School of Engineering and Dr Steven Manos, Associate Director Research Platforms, University Services and Assoc Prof Andrew Lonie as Director, VLSCI & EMBL-Australia Bioinformatics Resource.

VLSCI Advisory Council

As a cross-institutional member body, the primary role of the VLSCI Advisory Council was to assist and support the VLSCI in the strategic planning and use of resources for the benefit of the life sciences community within Victoria and beyond during the 2015-16 funding period.

The Advisory Council met quarterly in 2016, in February, May, August, and November.

2016 saw the VLSCI Advisory Council focus its discussions on exploring future funding opportunities for VLSCI within the scope of its emerging future direction in computational infrastructure.

The swift establishment and development of the EMBL-ABR research infrastructure network has harnessed capacity building of computational research infrastructure across projects, institutions and domains and this development will continue into 2017.

With the retirement of the IBM Blue Gene/Q supercomputer "Avoca" due at the end of 2016, the VLSCI Advisory Council and its partner institutions, Deakin, La Trobe, Melbourne and RMIT Universities agreed to decommission "Avoca" by 31 December 2016 whilst still maintaining its commitment to providing compute resources on alternate VLSCI systems.

With the completion of the 2015-16 Grant Agreement and its primary advisory obligations completed, the VLSCI Advisory Council concluded at its 8 November 2016 meeting with every confidence that whilst 2016 would see the end the BG/Q, it also marked the start of an exciting new era for VLSCI.

Members

Assoc Prof Jan Tennent (Chair), CEO, Biomedical Research Victoria

Prof Ricky Johnstone (LSCC Community Representative), Assistant Director, Cancer Research; Co-head, Cancer Therapeutics Program; Group Leader, Gene Regulation Laboratory, Peter MacCallum Cancer Centre

Prof Heinrich Schmidt (Institutional Subscriber), eResearch Director, Research & Innovation, RMIT University; Australia-India Research Centre for Automation Software Engineering, RMIT University; Adjunct Professor, Mälardalen University, Västerås, Sweden

Prof Brian Smith (Institutional Subscriber), Head, School of Molecular Sciences, La Trobe University

Assoc Prof Michael Kearney (University of Melbourne Representative), Conservation and Wildlife Biology, School of Biosciences, University of Melbourne

Prof Melissa Southey (University of Melbourne Representative), NHMRC Senior Research / Deputy Head of Department, Pathology, University of Melbourne

Prof Tiffany Walsh (Institutional Subscriber), Veski Innovation Fellow; Professor of Bionanotechnology, Institute for Frontier Materials, Deakin University

Observers

Assoc Prof Andrew Lonie, Director, VLSCI

Dr Michael Lynch, Senior Program/Policy Officer, Innovation & Technology Research, Department of Economic Development, Jobs, Transport and Resources

VLSCI Management Team Representative (a member of the VLSCI Management Team whose attendance is to be rotated as business requires, with the approval of the Chair and VLSCI Director).

Secretariat

Ms Fiona Kerr, Executive Officer, VLSCI



Left: The VLSCI Advisory Council in discussion at one of the 2016 meetings

Right: The VLSCI Advisory Council L-R: back row H. Schmidt, A. Lonie; middle row J. Tennent, M. Lynch; front step M. Southey, B. Smith; floor M. Kearney, T. Walsh (Note: R. Johnstone absent).



Our People

VLSCI Staff

Directorate

Assoc Prof Andrew Lonie [Director, VLSCI & EMBL-ABR](#)
Ms Karin Diamond [Business Manager, VLSCI & EMBL-ABR](#)
Ms Fiona Kerr [Executive Officer, VLSCI & EMBL-ABR](#)
Ms Claudia Curcio [Reception/Administrative Officer, VLSCI & EMBL-ABR](#)

Communications, Skills and Development

Ms Helen Gardiner [Communications Manager, VLSCI & EMBL-ABR](#)
Dr Christina Hall [Communications Officer, VLSCI & EMBL-ABR](#)

High-End Computing Experts

Dr Andrew Isaac [Platform Lead and Specialist Programmer](#)
Dr Jeff Tan [HPC Specialist IBM \(to October 2016\)](#)
Mr Chetan Babji [HPC Specialist IBM](#)
Mr Chris Samuel [Senior Systems Administrator](#)
Mr Carl Thomas [Storage and Infrastructure Administrator \(to April 2016\)](#)
Mr Ben Moran [Systems Administrator](#)
Ms Jin Zhang [Systems Administrator](#)

EMBL Australia Bioinformatics Resource

Assoc Prof Vicky Schneider [Deputy Director \(from January 2016\)](#)
Dr Philippa Griffin [Research Scientist \(from January 2016\)](#)

Bioinformaticians

Assoc Prof Torsten Seemann [Senior Research Scientist](#)
Assoc Prof Daniel Park [Head, Melbourne Bioinformatics Platform](#)
Dr Dieter Bulach [Senior Research Scientist](#)
Dr Bernard Pope [Lead Bioinformatician, Human Genomics](#)
Dr Mathew Wakefield [Senior Research Fellow](#)
Dr Michael Kuiper [Computational Molecular Scientist](#)
Dr Clare Sloggett [Research Scientist](#)
Mr Anthony Marty [Research Scientist](#)
Dr Gayle Philip [Research Scientist](#)
Dr Khalid Mahmood [Research Scientist](#)
Dr Juan Nunez-Iglesias [Research Scientist](#)
Dr Nuwan Goonasekera [Software Engineer](#)
Dr Chol-hee Jung [Research Scientist](#)
Mr Simon Gladman [Research Scientist](#)
Dr Yousef Kowsar [Scientific Software Developer](#)
Dr Anna Syme [Bioinformatician](#)
Ms Jessica Chung [Bioinformatician/Research Assistant](#)
Mr Peter Georgeson [Bioinformatics Platforms Developer](#)
Ms Madison Flannery [Bioinformatics Platforms Developer \(from Feb 2016 to 31 Dec 2016\)](#)
Ms Juny Kesumadewi [Software Developer \(from June 2016\)](#)
Mr Michael Milton [Software Developer \(From July 2016\)](#)

High-end computing

The computing facility at VLSCI is a world leader in the provision of high end computer processing infrastructure and computational expertise specialised for life sciences research. The petascale facility consists of three systems built on two architectures (BlueGene/Q and x86), that provide a variety of hardware configurations to support computational and data-processing workloads from genomics through to molecular dynamics.

Recognising our expertise in supporting Life Sciences research, and the growing range of resource needs, 2016 has seen many developments towards ensuring researchers have easy access to a range of resources suited to their research. This includes continued support of specialised high-performance systems, simplified access to resources and simplified resource management.

Avoca was retired 31 December 2016 after proving to be a remarkable machine that provided well over 1,500,000,000 CPU hours (the equivalent of 180,000 years' worth of single CPU computing) and running more than 5,000,000 jobs. A total of 15 national projects were supported through the National Computation Merit Allocation Scheme (NCMAS) where 15% of the BlueGene/Q system was available. Anticipating the retirement of the BlueGene/Q (Avoca) at the end of 2016, we successfully applied for a LIEF grant for a new GPU-based high-performance system (arriving mid-2017).

Avoca
5,000,000+ jobs
1,500,000,000 CPU hours
Equivalent of 180,000 years' worth of single CPU computing

A new resource application process and fair-share job scheduling were introduced to simplify access and management to high-end compute resources. The new processes removed the requirement for existing projects to reapply for resources, resulting in 231 active projects.

We look forward to continuing to broaden and simplify access to a variety of resources tailored to researchers' needs.

Support

We pride ourselves on the active support we provide and the in-depth experience we have gained through providing infrastructure and support for researchers. In 2016, we responded to over 1,500 help tickets and 250 software installation tickets (many tickets with long lists of requested software!).

Expertise

- HPC Systems Administration
- HPC programming
- Specialist HPC programming – parallelising of algorithms
- Molecular Modelling and Dynamics

Software

From 'ACG' to 'Zlib' over 220 licensed software applications specific to life sciences are installed on VLSCI's systems.

Compute Systems

IBM Blue Gene/Q - Avoca

- Peak performance of 838.86 teraFLOPS
- 65,536 PowerPC based 1.6GHz cores
- A total of 64TB RAM
- Interconnect between compute nodes forms a five-dimensional torus providing excellent nearest neighbour and bisection bandwidth
- Suitable for large-scale parallel processing
- Compute nodes run a custom lightweight operating system called Compute
- Node Kernel (CNK) that is similar to Linux and mostly POSIX compliant
- The head node runs the RHEL 6 operating system, a variety of Linux

IBM iDataplex x86 system – Barcoo

- Peak performance - compute nodes currently performing at 20 teraFLOPS - with Xeon Phi cards running nominally at 1 teraFLOPS each
- 1120 Intel Sandybridge compute cores running at 2.7GHz
- 67 nodes with 256GB RAM and 16 cores per node
- 3 nodes with 512GB RAM and 16 cores per node
- 20 Xeon Phi 5110P cards installed across 10 nodes

LENOVO NeXtScale – Snowy

- Peak performance of 28 teraFLOPS
- 31 nodes, each with 32 cores of the latest Haswell processor, running at 2.3GHz
- 29 nodes have 128GB of memory
- 2 FAT nodes have 512GB of memory

IBM iDataplex x86 system – Merri

- Re-purposed as specialised high-memory machine for the Melbourne Genomics Health Alliance, a \$25m Victorian clinical genomics initiative in which VLSCI is a funded participant and contributor.
- 3 nodes with 1024GB RAM and 16 cores per node.
- 1 node with 1.5 TB RAM and 32 cores per node.

All x86 compute clusters run the RHEL 6 operating system, a variety of Linux and are connected to a high speed, low latency Mellanox FDR14 InfiniBand switch for inter-process communications.

Storage infrastructure

The storage infrastructure is connected via a high speed, low latency Mellanox FDR14 InfiniBand network for fast data-handling.

- 700TB GPFS Parallel Data Store (shared by Barcoo, Snowy, Merri and Avoca)
- 1PB HSM tape system, made available through GPFS (shared by Barcoo, Snowy, Merri and Avoca).

Hardware Hosting

VLSCI hosts hardware for Life Sciences research groups as requested. This includes;

- The University of Melbourne's Centre for Systems Genomics' Helix cluster (a 384 core, 3.1Tb RAM, 200Tb storage) machine,
- A specialised large memory compute node for Assoc Prof Andrew Turpin from the University of Melbourne.

Our Projects - Highlights

Cancer Genomics Theme

The Cancer Genomics theme applies state-of-the-art bioinformatics techniques to understand the biology underpinning cancer and related conditions. In 2016 we continued with our long-running collaborations with the Genetic Epidemiology Laboratory at the University of Melbourne and the Cancer Council of Victoria around inherited risk for breast (and other) cancers. Through that work we have developed computational pipelines for detecting rare disease-related DNA mutations to identify cancer risk factors. We have also developed filtering and prioritisation tools which compare mutation frequencies in a disease group with background population frequencies, to highlight genes which might carry an unexpectedly high burden of mutation. In other work we initiated a new partnership with the Colorectal Oncogenomics group at the University of Melbourne, beginning with an analysis of complex structural mutations related to Lynch syndrome (an inherited genetic condition that is closely linked to colon cancer and other cancers). The results of this work fed into a successful NHMRC grant application, which will fund the continuation of this project for the next three years. 2016 also saw the foundation of the Pan Prostate Cancer Consortium, consisting of several high-profile international partner institutions who have joined together to apply standardised analysis pipelines to the largest and highest-quality collection of prostate cancer samples ever assembled. We lead the Australian bioinformatics component of the project, and have begun processing hundreds of whole genomes collected from Australian patients. These results will be pooled with analyses performed by our international partners, yielding a highly significant resource for cancer researchers worldwide.

Clinical Genomics Theme

The Clinical Genomics theme applies best-practices in software engineering and bioinformatics to bring the benefits of DNA sequencing and related technologies into clinical practice, simultaneously improving patient outcomes and reducing hospital costs. In 2016 we continued our ongoing collaboration with the Melbourne Genomics Health Alliance, producing a clinical-grade variant calling pipeline and variant curation database. These tools are deployed within accredited clinical laboratories within Victoria and have been successfully used to complete the analysis of the first phase of project flagships. We also began analysing the next phase of project flagships, with the inclusion of a new cancer analysis pipeline. As the project moves from the pilot phase to production, we participated in requirements analysis for the future state production platform, and the review of candidate solutions. We also engaged with the clinical genomics community at the national level by participating in the newly established Australian Genomics Health Alliance. Through that project we have begun a curated registry of clinical pipelines in use throughout Australia with the aim of developing a standard vocabulary for pipeline comparison, and a structured representation for improved pipeline elucidation and visualisation.

EMBL Australia Bioinformatics Resource

Deputy Director, EMBL-ABR, Vicky Schneider, arrived in February and promptly got to work delivering on the vision for this Bioplatforms Australia-funded project which aims to:

1. increase Australia's capacity to collect, integrate, analyse, exploit, share and archive the large heterogeneous data sets now part of modern life sciences research
2. contribute to the development of and provide training in data, tools and platforms to enable Australia's life science researchers to undertake research in the age of big data
3. showcase Australian research and datasets at an international level
4. enable engagement in international programs that create, deploy and develop best practice approaches to data management, software tools and methods, computational platforms and bioinformatics services.

At the end of 2016, EMBL-ABR is now well established as a national resource. Ten Nodes have been organised around six Key Areas: Data, Tools, Compute, Standards, Training and Platforms, mapped

to their respective expertise in terms of biosciences domains. An International Scientific Advisory Group (ISAG) has been assembled with expertise from around the world connecting Australian researchers directly to the best practices of major international bioinformatics efforts at EMBL-EBI, ELIXIR, CyVerse and the NIH's Big Data to Knowledge (BD2K). There is now a comprehensive website, an active Head of Nodes Group, a Key Areas Coordinators' Group and a Biosciences Community Leads Group.

Highlights included a series of Data Life Cycle workshops held in October, an interview series with national and international bioinformatics experts sharing their insights into the needs for bioinformatics and making suggestions for how to develop EMBL-ABR, and workshops and an All Hands meeting held to coincide with the ISAG's third meeting for the year held in person in Melbourne in December. EMBL-ABR was represented by information booths at ISMB in Orlando, USA, and ECCB in The Hague, Netherlands to increase international awareness of our activities, establish important relationships and gather information about the current knowledge and future needs of Australian life sciences and bioinformatics.

In its role as advocate and counsel, the ISAG is helping to plan the next phase of the project, pursuing research infrastructure funding from the Federal Government to secure the required resources to continue to develop this important biology backbone. This will be carried out in consultation with the wider research community and facilitated by the Nodes in each State.

Microbial Genomics Theme

The Microbial Genomics theme focuses on using bioinformatics and genomics to better understand and monitor the spread of pathogens and antimicrobial resistance at the local, national and international level. In 2016 we contributed specialist expertise to several national collaborative projects, and worked on developing platforms for, and providing leadership to, a wide range of training activities.

The NCRIS-funded RDS Omics project will provide the first Australian platform to allow genomics (DNA), transcriptomics (RNA), proteomics (proteins) and metabolomics (small molecules) data to be efficiently co-analysed, stored, shared and published from one common interface. VLSCI staff are contributing to the development of a structured data environment, as well as tools and workflows for the 'omics and cross-omics systems biology. These were incorporated into the platform and deployed on RDS storage and the Nectar Research Cloud in 2016.

Anna Syme has generated a suite of new training materials that were used at multiple teaching events in Canada, USA, Sydney and Perth. These are all publicly available and are currently being used by the research community.

The basis for these materials stemmed from the protocols and methods developed by Anna for her work on the Antibiotic Resistant Pathogen Initiative, funded by Bioplatforms Australia. Anna has reconstructed 26 complete genomes from Pacbio and Illumina data that will act as fundamental references for corresponding RNA-Seq and proteomics analyses.

In his role at Doherty Applied Microbial Genomics, Dieter Bulach and Torsten Seemann convened the *Introductory Course in Public Health Microbial Genomics for the Asia Pacific Region*. Attendees came from an array of countries including Korea, Hong Kong, Vietnam, Fiji and New Zealand and resulted in several invitations to run training sessions in other countries.

Work at the Microbiological Diagnostic Unit Public Health Laboratory to develop the Nullarbor pipeline for analysing national outbreak data resulted in a milestone in Australian public health and clinical microbiology. As part of its management of this serious outbreak, it produced the first national "case definition" employing analyses and information from whole genome sequencing (WGS) directly. Managed by OzFoodNet and provided to the federal Health Department, it clearly demonstrated of the ability of WGS and bioinformatics to transform pathogen surveillance and respond to public health threats.

The Microbial Genomics group authored more than 20 papers in 2016. A couple of Torsten Seemann's publications had a particularly notable year, with a paper on his rapid prokaryotic genome annotation tool

“Prokka” reaching 800 citations and his Gigascience “Ten Recommendations” paper becoming the most downloaded paper at the journal (with over 15,000 downloads).

Genomics Virtual Laboratory

A major activity of the Microbial Genomics Theme, the Genomics Virtual Laboratory (GVL) continued to deliver many returns to the Australian research community and beyond. This computational workbench was developed by VLSCI and the University of Queensland in 2012 and is now administered by VLSCI. It provides a highly accessible cloud-based genomics analysis environment preconfigured with best-practice genomics tools and supplemented with comprehensive tutorial materials and protocols, managed services and user support.

Based on Galaxy, the open, web-based platform for accessible, reproducible, and transparent computational biomedical research, the GVL has brought recognition for Australian bioinformatics practice through opportunities to showcase the GVL in several international forums. Several new ‘flavours’ of the GVL are now well-developed: base, tutorial and microbial and throughout 2016 it has been used for delivery of training to life scientists across Australia and around the world.

In late June we sent a team to Bloomington, Indiana, USA to attend and teach at the Galaxy Community Conference 2016 where the GVL training materials were very well received, both at conference sessions and across the globe. Training sessions run by Philippa Griffin (EMBL-ABR), Annette McGrath (CSIRO), Simon Gladman and Torsten Seemann (VLSCI) – with local support from VLSCI experts Anna Syme and Madison Flannery – were rated as the best organised and presented. Later, when Galaxy PIs gave their annual talk they singled out the GVL community training materials as demonstrating best practice and Galaxy PI Anton Nekrutenko, Professor of Biochemistry and Molecular Biology, Pennsylvania State University, USA, requested permission to use these materials as the basis for Galaxy group training materials in the future.

Our strong presence at this conference certainly gave more impetus and further opportunity to connect with key Galaxy people and progress our plans for a more structured Galaxy Australia community network. By the end of 2016 the community group Galaxy Australia had been launched and arrangements were firmly in place for Melbourne to host the inaugural Galaxy Australasia Conference in February 2017.

Melbourne Bioinformatics Platform

Melbourne Bioinformatics has worked with leading collaborators across a wide range of life science disciplines.

Working closely with Professor David Ritchie’s group, we are identifying markers that predict patient outcomes following bone marrow transplants to inform future treatment practices. With Professor Peter Rogers, we are working to better understand endometriosis via a large competitive grant that was recently awarded. We are named investigators on another recent award, led by Professor Peter Cowan, that aims to reduce problems of rejection in organ transplant medicine. With Professor Judy Savage, we have identified the probable underlying causes in patients with familial kidney diseases. We have developed automated imaging systems for improved screening of cell responses to pharmaceutical agents and, working with Prof Leanne Tilley, methods to visualise blood cell structural organisation during malarial infection. We are working with government agencies to develop technology for rapid, phenotypic, microbial detection and are in the process of establishing collaboration with Professor Stephen Swearer to apply microbial meta-genomics for improved ecological monitoring in the marine setting. We have led the development of an exciting new technology, Hi-Plex Origin, which allows unprecedented access to high-accuracy, low-cost targeted DNA sequence screening – this technology is being adopted rapidly by large international consortia. Through molecular simulation, we helped Professor Michael Parker’s group to achieve structural insights into how certain bacteria target the cellular receptor CD59, normally used to protect cells from the body’s own immune system, with their pore-forming toxins. Understanding these molecular interactions paves the way for more effective treatment options for these serious, often life threatening, infections.

Software Development

The following open source projects have had significant input into their design and implementation from our experts (nb. attribution will sometimes refer to VLSCI's new name of Melbourne Bioinformatics):

ABRicate Mass screening of contigs for antimicrobial and virulence genes.
github.com/tseemann/abricate

Annokey Gene-based search for key-terms in the NCBI gene database and associated PubMed abstracts. bjpop.github.io/annokey/. Published in Source Code for Biology and Medicine, 2014. www.scfbm.org/content/9/1/15. Collaboration with the GEL lab, Department of Pathology, UoM.

Bpipe A bioinformatics pipeline. github.com/ssadedin/bpipe. Published in Bioinformatics 2012. bioinformatics.oxfordjournals.org/content/early/2012/04/11/bioinformatics.bts167.abstract. Collaboration with MCRI.

Cpipe a shared variant detection pipeline designed for diagnostic settings, published in Genome Medicine: genomemedicine.biomedcentral.com/articles/10.1186/s13073-015-0191-x

Degust A web tool to help analyse, visualise and fully appreciate Differential Gene Expression data from RNA-seq experiments. www.vicbioinformatics.com/software/degust.shtml. Collaboration with VBC, Monash.

FAVR Filtering and Annotation of Variants that are Rare. github.com/bjpop/favr. Published in BMC Bioinformatics, 2013. www.biomedcentral.com/1471-2105/14/65/abstract. Collaboration with the GEL lab, Department of Pathology, UoM.

Genomics Virtual Laboratory A Practical Bioinformatics Workbench for the Cloud. Published in PLOS ONE: journals.plos.org/plosone/article?id=10.1371/journal.pone.0140829

Hi-Plex A streamlined highly-multiplexed PCR-based target-enrichment system for massively parallel sequencing suitable for an extensive range of clinical and research applications. Hi-Plex is complemented by software for primer design and variant calling. The Hi-Plex tool and workflow (April 2015) Paper "Abridged adapter primers increase the target scope of Hi-Plex" is listed in Top Ten Biotechniques Peer reviewed papers for 2015.

HiTIME High-resolution Twin-Ion Metabolic Extraction. github.com/bjpop/HiTIME. Collaboration with Department of Chemistry, UoM. Published in Analytical Chemistry: High-Resolution Twin-Ion Metabolite Extraction (HiTIME) Mass Spectrometry: Nontargeted Detection of Unknown Drug Metabolites by Isotope Labeling, Liquid Chromatography Mass Spectrometry, and Automated High-Performance Computing pubs.acs.org/doi/abs/10.1021/ac504767d

Methpat A program for summarising CpG methylation patterns. github.com/bjpop/methpat. Collaboration with the Ludwig Institute for Cancer Research. Published in Gigascience: Exemplary multiplex bisulfite amplicon data used to demonstrate the utility of Methpat. gigascience.biomedcentral.com/articles/10.1186/s13742-015-0098-x Methpat in the news, 26/6/2016

MLST For scanning contig files against PubMLST typing schemes. github.com/tseemann/mlst

Nullarbor Pipeline to generate complete public health microbiology reports from sequenced isolates. github.com/tseemann/nullarbor

Prokka A software tool for the rapid annotation of prokaryotic genomes. www.vicbioinformatics.com/software/prokka.shtml. Published in Bioinformatics, 2014. Collaboration with VBC, Monash. This high-profile software has reached 374 citations in less than two years.

Rover Read-pair overlap considerate variant-calling software for PCR-based massively parallel sequencing datasets. github.com/bjpop/rover. Published in Source Code for Biology and Medicine, 2014. www.scfbm.org/content/9/1/3. Collaboration with the GEL lab, Pathology, UoM. Part of the Hi-Plex project: www.hiplex.org/ Simple, low-cost, modular targeted DNA sequencing technology. Supported by NHMRC project grant APP 1025879.

Rubra A bioinformatics pipeline. github.com/bjpop/rubra. Presented at the Bioinformatics Open Source Conference, 2013 (BOSC2013). Used by many bioinformatics projects at VLSCI.

Snippy Rapid haploid variant calling and core SNP phylogeny. github.com/tseemann/snippy

SRST2 Short Read Sequence Typing for Bacterial Pathogen. github.com/katholt/srst2. Accepted for publication in Genome Medicine, 2014. Collaboration with UoM researchers Dr Kat Holt and Dr Mike Inouye. BioMedCentral Announcement re SRST2, 18/12/2014 Kat Holt's blog announcement re SRST2, 26/11/2014

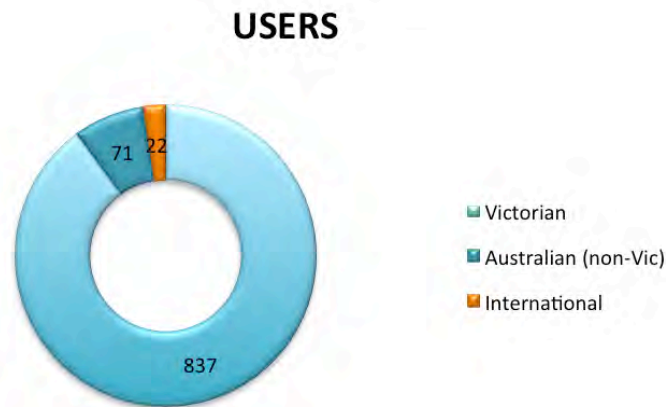
UNDR ROVER A fast and accurate DNA mutation detector for targeted DNA sequencing. github.com/bjpop/undr_rover. Published in BMC Bioinformatics. Created by VLSCI experts Dr Daniel J. Park, Mr Roger Li, Mr Edmund Lau, Mr Peter Georgeson, Dr Tú Nguyen-Dumont and Dr Bernard J. Pope.

Over 50 'recipes' for Homebrew Science to make it easy for people to install bioinformatics software have been published by our experts.



Our Users

There were 930 researchers from 73 different research institutions or organisations using VLSCI systems in 2016. Of these, 90% were from Victoria, with 71 individual users from other states, and 22 logging on from international institutions.



All projects using VLSCI systems in 2016 were asked to complete a report with a range of questions and reporting criteria. The figures presented here represent data from the 104 responses received from the 123 Chief Investigators (CIs) contacted.

VLSCI's role as an accessible, quality source of HPC resources for the academic community is important to the progress of many aspects of the research we carry out.

*It's the combination of **exceptional support from dedicated VLSCI research staff** as well as the actual CPU/GPU processing power that is the key for us - our in-house expertise is sufficient to know what we want to do and what is likely to be possible, but the actual hands-on HPC use and data analysis relies on the presence of VLSCI experts.*

Craig Morton

Group Leader, ACRF Rational Drug Discovery Centre
St Vincent's Institute

*VLSCI has not only enabled our project to produce some high quality results without the associated investment in prohibitively expensive computing resources, but also provided **expert IT support**. We appreciate this service and would like to see it continue in the future.*

Douglas Hilton

Director, Walter and Eliza Hall Institute
Head, Molecular Medicine Division, WEHI
Head, Department of Medical Biology, University of Melbourne

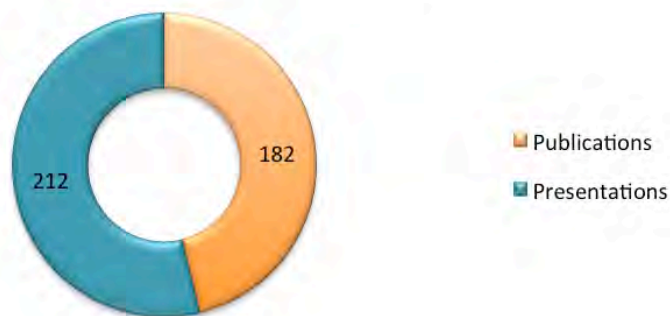
*VLSCI has provided a consistent resource which has been **reliable and accessible**. This has been **critical for our project's success**. It has allowed the development of an intervention study for prevention of post-stroke dementia using aerobic exercise. We look forward to a long collaboration. Our project has been extended so we hope to continue to utilise this resource until 2022.*

Amy Brodtmann

Division Head, Behavioural Neuroscience
Florey Institute of Neuroscience and Mental Health

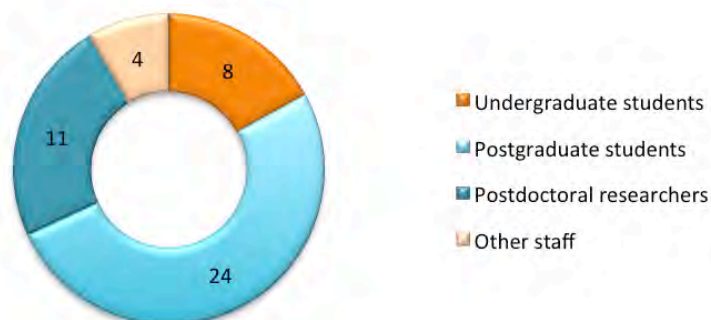
There were a reported 183 papers published involving research on VLSCI systems, and 212 presentations made. Asked to report the details of their five best publications or presentations, CIs reported on 90 publications with an average Journal Impact Factor of 5.99. More than 74% of the journals in which these publications from VLSCI projects were in the first quartile of their discipline category, as ranked by Thomson Reuters.

COMMUNICATING OUTCOMES



23 CIs allocated resources on VLSCI systems reported increased capacity to employ additional staff or students.

EMPLOYMENT



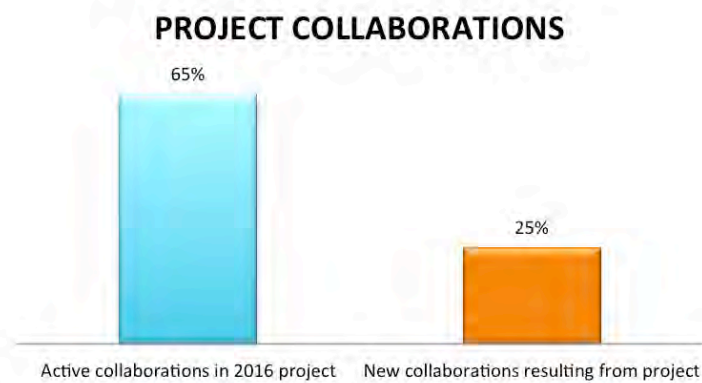
Project CIs were asked what grant income supported their work on VLSCI machines. From a total of \$47,395,606 in grants were received in 2016 by the projects that responded.

GRANT INCOME

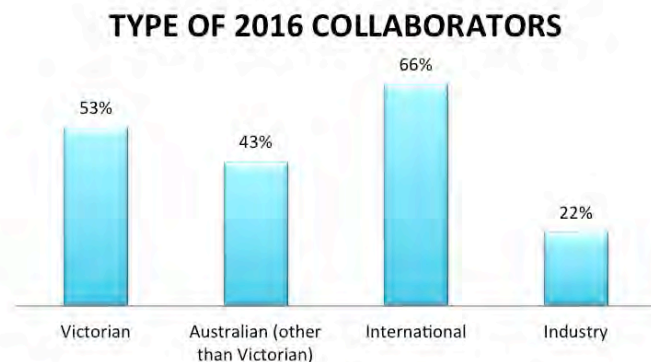


Our Collaborations

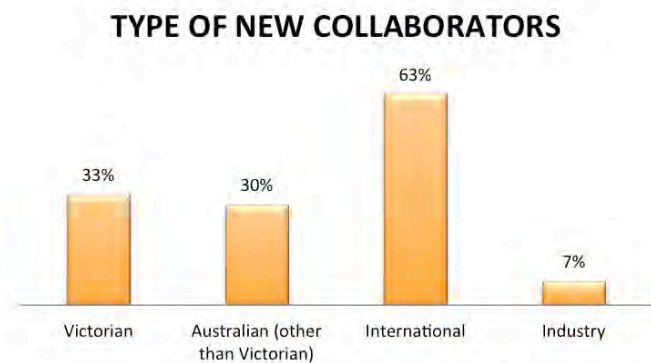
Researchers with projects running on VLSCI machines were actively involved in a variety of collaborations. Collaborators were involved in 65% of projects, and 25% reported that the 2016 project resulted in the formation of new collaborations.



Many projects had a variety of types of collaborations, with 53% reporting having Victorian collaborators, 43% having other Australian collaborators, 66% International, and 22% nominating Industry collaborators. Of the 68 projects who had external collaborators (65% of total), 22% of CIs reported their projects had industry collaborators (representing 14% of all projects).



Most of new collaborations that resulted from 2016 projects were with International collaborators.



Project Collaborators

Victorian Institutions

AgriBio
Baker IDI Heart and Diabetes Institute
Bio21 Institute
Bionics Institute
Cancer Council Victoria
Deakin University
Florey Institute of Neuroscience and Mental Health
La Trobe University
Monash University
Olivia Newton-John Cancer Research Institute
Peter Doherty Institute
Peter MacCallum Cancer Centre
RMIT University
Royal Botanic Gardens Victoria
Royal Melbourne Hospital
St Vincent's Hospital
St Vincent's Research Institute
Swinburne University
Walter and Eliza Hall Institute

Australian Institutions

ANDS
ARC Centre of Excellence in Plant Cell Walls
ARC Research Hub for Australian Steel Manufacturing
Australian National University
Bioplatforms Australia
Centre for Cancer Biology, SA Pathology
Charles Darwin University
Charles Perkins Centre
CSIRO
EmPathy Breast Cancer Network
Griffith University
Menzies School of Health Research
Murdoch University
National Vision Research Institute
Nectar
South Australian Research and Development Institute
Transport Accident Commission
University of Adelaide
University of New South Wales
University of Queensland
University of Sydney
University of Western Australia
University of Wollongong
Victor Chang Cardiac Research Institute

International Institutions

Baylor College of Medicine, USA

California Institute of Technology, USA
Central European Institute of Technology, Czech Republic
Central Institute of Medicinal and Aromatic Plants, India
CERMAV, France
Erasmus University Medical Centre, The Netherlands
European Commission
Florida Atlantic University, USA
George Washington University, USA
Imperial College, UK
Indian Institute of Technology, India
Institute of Genomics and Integrative Biology, India
Institute of Infectious Disease and Molecular Medicine, South Africa
Institute Pasteur, France
International Crops Research Institute for the Semi-Arid Tropics, India
Jawaharlal Nehru University, India
King's College London, UK
Kobe University, Japan
Laboratory of Epidemiology, Neuroimaging & Telemedicine, Italy
Max Planck Institute for Medical Research, Germany
Michigan State University, USA
Nanyang Technological University, Singapore
National Institute of Nuclear and Particle Physics, France
New York University, USA
Ontario Cancer Institute, Canada
René Rachou Research Center, Brazil
Scuola Normale Superiore, Italy
Swiss Federal Institute of Technology, Switzerland
Swiss National Science Foundation, Switzerland
Tongji University, China
University College Dublin, Ireland
University College London, UK
University of Alabama
University of Birmingham, UK
University of Cambridge, UK
University of Copenhagen, Denmark
University of Economics & Technology, Turkey
University of Exeter, UK
University of Illinois, USA
University of Konstanz, Germany
University of Leuven, Belgium
University of Lugano, Switzerland
University of Michigan, USA
University of Missouri, USA
University of Oklahoma, USA
University of Utrecht, The Netherlands
University of Washington, USA

Vanderbilt University, USA
Vrije University, The Netherlands
Weierstrass Institute, Germany
York University, Canada
Zhejiang University, China

Department of Environment and Primary
Industries, Victoria
Department of Health and Human Services,
Victoria
General Electric
Haag-Streit, Switzerland
Heidelberg Engineering, Germany
IBM Research Australia
Melbourne Genomics Health Alliance
Neurosolutions Ltd.

Industry Collaborators

BlueScope Steel
Capsugel, France
CSL Ltd.
Department of Economic Development, Jobs,
Transport and Resources, Victoria

Industry Engagement

Our location within Lab-14 in the Carlton Connect Initiative allows VLSCI to interact closely with the established stakeholders within the Parkville Precinct as well as introduce ourselves to a wide range of visitors. Broad awareness of VLSCI is increased by regularly providing tours of our facilities and briefings on our research and expertise. For example, on request from the Research and Economic Group, Strategic Policy Projects, Research and Higher Education Infrastructure Branch of the Australian Government Department of Education and Training, VLSCI hosted part of a tour of BPA-funded facilities by foreign government officials in February 2016. Participants from South Africa, Italy, Thailand, Malaysia and Australia were briefed by VLSCI's director before visiting our machine room and being escorted by staff through the Precinct.

During 2016, VLSCI continued to maintain close working relationships with the following groups:

Australian National Data Service
Biomedical Research Victoria
BioMelbourne Network
Bioplatforms Australia
COMBINE
Convergence Science Network
Melbourne Genomics Health Alliance
Melbourne Python Users Group
National Computational Merit Allocation Scheme
Nectar
Research Data Storage Infrastructure
Victorian Government's Technology Voucher Program

Contributing Organisations



THE UNIVERSITY OF
MELBOURNE



Publications resulting from VLSCI staff collaborations

The dedicated team of bioinformaticians, computational biologists and high performance computing specialists at VLSCI work with external collaborators in research, training and platform development capacities. As a result of these activities our staff members authored the many high quality publications in 2016.

Afgan, E., Lonie, A., Taylor, J., Skala, K. and Goonasekera, N., 2016, May. Architectural models for deploying and running virtual laboratories in the cloud. In Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2016 39th International Convention on (pp. 282-286). IEEE.

Baglietto, L., Ponzi, E., Haycock, P., Hodge, A., Bianca Assumma, M., Jung, C.H., Chung, J., Fasanelli, F., Guida, F., Campanella, G. and Chadeau - Hyam, M., 2017. DNA methylation changes measured in pre - diagnostic peripheral blood samples are associated with smoking and lung cancer risk. International Journal of Cancer, 140(1), pp.50-61.

Baines, S.L., Howden, B.P., Heffernan, H., Stinear, T.P., Carter, G.P., Seemann, T., Kwong, J.C., Ritchie, S.R. and Williamson, D.A., 2016. Rapid emergence and evolution of *Staphylococcus aureus* clones harboring *fusC*-containing staphylococcal cassette chromosome elements. Antimicrobial agents and chemotherapy, 60(4), pp.2359-2365.

Boritsch, E.C., Khanna, V., Pawlik, A., Honoré, N., Navas, V.H., Ma, L., Bouchier, C., Seemann, T., Supply, P., Stinear, T.P. and Brosch, R., 2016. Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. Proceedings of the National Academy of Sciences, 113(35), pp.9876-9881.

Buultjens, A.H., Lam, M.M., Ballard, S., Monk, I.R., Mahony, A.A., Grabsch, E.A., Grayson, M.L., Pang, S., Coombs, G.W., Robinson, J.O. and Seemann, T., 2016. Evolutionary origins of the emergent ST796 clone of vancomycin resistant *Enterococcus faecium* (No. e2562v1). PeerJ Preprints.

Carter, G.P., Buultjens, A.H., Ballard, S.A., Baines, S.L., Tomita, T., Strachan, J., Johnson, P.D., Ferguson, J.K., Seemann, T., Stinear, T.P. and Howden, B.P., 2016. Emergence of endemic MLST non-typeable vancomycin-resistant *Enterococcus faecium*. Journal of Antimicrobial Chemotherapy, p.dkw314.

Caruana, N.J., Cooke, I.R., Faou, P., Finn, J., Hall, N.E., Norman, M., Pineda, S.S. and Strugnell, J.M., 2016. A combined proteomic and transcriptomic analysis of slime secreted by the southern bottletail squid, *Sepiadarium austrinum* (Cephalopoda). Journal of Proteomics, 148, pp.170-182.

Doo, N.W., Makalic, E., Schmidt, D., Joo, J.E., Jung, C.H. and Giles, G.G., 2016. Aberrant DNA Methylation Patterns in Peripheral Blood Are Detectable Years before the Diagnosis of Mature B Cell Neoplasms.

Dugué, P.A., Brinkman, M.T., Milne, R.L., Wong, E.M., FitzGerald, L.M., Bassett, J.K., Joo, J.E., Jung, C.H., Makalic, E., Schmidt, D.F. and Park, D.J., 2016. Genome-wide measures of DNA methylation in peripheral blood and the risk of urothelial cell carcinoma: a prospective nested case-control study. British Journal of Cancer.

Dugué, P.A., English, D.R., MacInnis, R.J., Jung, C.H., Bassett, J.K., FitzGerald, L.M., Wong, E.M., Joo, J.E., Hopper, J.L., Southey, M.C. and Giles, G.G., 2016. Reliability of DNA methylation measures from dried blood spots and mononuclear cells using the HumanMethylation450k BeadArray. Scientific Reports, 6.

Goonasekera, N., Lonie, A., Taylor, J. and Afgan, E., 2016, July. CloudBridge: a Simple Cross-Cloud Python Library. In Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale (p. 37). ACM.

Gorrell, R.J., Zwickel, N., Reynolds, J., Bulach, D. and Kwok, T., 2016. *Helicobacter pylori* CagL hypervariable motif: a global analysis of geographical diversity and association with gastric cancer. Journal of Infectious Diseases, p.jiw060.

Gouillart, E., Nunez-Iglesias, J. and van der Walt, S., 2017. Analyzing microtomography

data with Python and the scikit-image library. *Advanced Structural and Chemical Imaging*, 2(1), p.18.

Griffin, P.C., Hangartner, S.B., Fournier-Level, A. and Hoffmann, A.A., 2016. Genomic Trajectories to Desiccation Resistance: Convergence and Divergence Among Replicate Selected *Drosophila* Lines. *Genetics*, pp.genetics-116.

Hall, N.E., Hanzak, J., Allcock, A.L., Cooke, I.R., Ogura, A. and Strugnelli, J.M., 2016. The complete mitochondrial genome of the pygmy squid, *Idiosepius* (Cephalopoda: Decapodiformes): the first representative from the family Idiosepiidae. *Mitochondrial DNA Part A*, 27(1), pp.5-6.

Hasmad, H.N., Lai, K.N., Wen, W.X., Park, D.J., Nguyen-Dumont, T., Kang, P.C.E., Thirthagiri, E., Ma'som, M., Lim, B.K., Southey, M. and Woo, Y.L., 2016. Evaluation of germline BRCA1 and BRCA2 mutations in a multi-ethnic Asian cohort of ovarian cancer patients. *Gynecologic oncology*, 141(2), pp.318-322.

Jung, C.H., Kim, K.J., Kim, B.Y., Kim, C.H., Kang, S.K. and Mok, J.O., 2016. Relationship between vitamin D status and vascular complications in patients with type 2 diabetes mellitus. *Nutrition Research*, 36(2), pp.117-124.

Kpeli, G., Otchere, I.D., Lamelas, A., Buultjens, A., Bulach, D., Baines, S.L., Seemann, T., Giulieri, S., Nakobu, Z., Aboagye, S.Y. and Owusu-Mireku, E., 2016. Possible healthcare-associated transmission as a cause of secondary infection and population structure of *Staphylococcus aureus* isolates from two wound treatment centres in Ghana. *New microbes and new infections*, 13, pp.92-101.

Kwong, J.C., da Silva, A.G., Dyet, K., Williamson, D.A., Stinear, T.P., Howden, B.P. and Seemann, T., 2016. NGMASTER: *in silico* multi-antigen sequence typing for *Neisseria gonorrhoeae*. *Microbial Genomics*, 2(8).

Kwong, J.C., Mercoullia, K., Tomita, T., Easton, M., Li, H.Y., Bulach, D.M., Stinear, T.P., Seemann, T. and Howden, B.P., 2016. Prospective whole-genome sequencing enhances national surveillance of *Listeria*

monocytogenes. *Journal of clinical microbiology*, 54(2), pp.333-342.

Kwong, J.C., Stafford, R., Strain, E., Stinear, T.P., Seemann, T. and Howden, B.P., 2016. Sharing is caring: international sharing of data enhances genomic surveillance of *Listeria monocytogenes*. *Clinical Infectious Diseases*, p.ciw359.

Lawrence, S.L., Gorman, M.A., Feil, S.C., Mulhern, T.D., Kuiper, M.J., Ratner, A.J., Tweten, R.K., Morton, C.J. and Parker, M.W., 2016. Structural Basis for Receptor Recognition by the Human CD59-Responsive Cholesterol-Dependent Cytolysins. *Structure*, 24(9), pp.1488-1498.

Lee, J.Y., Monk, I.R., Pidot, S.J., Singh, S., Chua, K.Y., Seemann, T., Stinear, T.P. and Howden, B.P., 2016. Functional analysis of the first complete genome sequence of a multidrug resistant sequence type 2 *Staphylococcus epidermidis*. *Microbial Genomics*, 2(9).

Lonsdale, A., Penington, J.S., Rice, T., Walker, M. and Dashnow, H., 2016. Ten simple rules for a bioinformatics journal club. *PLoS computational biology*, 12(1), p.e1004526.

MacInnis, R.J., Schmidt, D.F., Makalic, E., Severi, G., FitzGerald, L.M., Reumann, M., Kapuscinski, M.K., Kowalczyk, A., Zhou, Z., Goudey, B. and Qian, G., 2016. Use of a Novel Nonparametric Version of DEPTH to Identify Genomic Regions Associated with Prostate Cancer Risk. *Cancer Epidemiology and Prevention Biomarkers*.

Miles, M.A., Shekhar, T.M., Hall, N.E. and Hawkins, C.J., 2016. TRAIL causes deletions at the HPRT and TK1 loci of clonogenically competent cells. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 787, pp.15-31.

Moffatt, C.R.M., Greig, A., Valcanis, M., Gao, W., Seemann, T., Howden, B.P. and Kirk, M.D., 2016. A large outbreak of *Campylobacter jejuni* infection in a university college caused by chicken liver pâté, Australia, 2013. *Epidemiology and infection*, pp.1-8.

Mofiz, E., Holt, D.C., Seemann, T., Currie, B.J., Fischer, K. and Papenfuss, A.T., 2016.

Genomic resources and draft assemblies of the human and porcine varieties of scabies mites, *Sarcoptes scabiei* var. *hominis* and var. *suis*. *GigaScience*, 5(1), p.23.

Mofiz, E., Seemann, T., Bahlo, M., Holt, D., Currie, B.J., Fischer, K. and Papenfuss, A.T., 2016. Mitochondrial genome sequence of the scabies mite provides insight into the genetic diversity of individual scabies infections. *PLoS Negl Trop Dis*, 10(2), p.e0004384.

Monk, I.R., Seemann, T., Howden, B.P. and Stinear, T.P., 2016. Refutation: Structure and mechanism of the essential two-component signal-transduction system WalkR in *Staphylococcus aureus*. *bioRxiv*, p.058842.

Ng, A.C., Poudel, G., Stout, J.C., Churchyard, A., Chua, P., Egan, G.F. and Georgiou-Karistianis, N., 2016. Iron accumulation in the basal ganglia in Huntington's disease: cross-sectional data from the IMAGE-HD study. *Journal of Neurology, Neurosurgery & Psychiatry*, 87(5), pp.545-549.

Nor, B., Young, N.D., Korhonen, P.K., Hall, R.S., Tan, P., Lonie, A. and Gasser, R.B., 2016. Pipeline for the identification and classification of ion channels in parasitic flatworms. *Parasites & vectors*, 9(1), p.155.

Page, A.J., Steinbiss, S., Taylor, B., Seemann, T. and Keane, J.A., 2016. {GFF} 3toEMBL: Preparing annotated assemblies for submission to {EMBL}. *The Journal of Open Source Software*, 1(6).

Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A. and Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*, 2(4).

Park, D.J., Li, R., Lau, E., Georgeson, P., Nguyen-Dumont, T. and Pope, B.J., 2016. UNDR ROVER-a fast and accurate variant caller for targeted DNA sequencing. *BMC bioinformatics*, 17(1), p.165.

Pope, B.J., Mahmood, K., Jung, C.H. and Park, D.J., 2016. Fine resolution mapping of double-strand break sites for human ribosomal DNA units. *Genomics Data*, 10, pp.19-21.

Pope, B.J., Mahmood, K., Jung, C.H., Georgeson, P. and Park, D.J., 2017. Single nucleotide-level mapping of DNA double-strand breaks in human HEK293T cells. *Genomics Data*, 11, pp.43-45.

Quek, C., Bellingham, S.A., Jung, C.H., Scicluna, B., Shambrook, M.C., Sharples, R., Cheng, L. and Hill, A.F., 2016. Defining the purity of exosomes required for diagnostic profiling of small RNA suitable for biomarker discovery. *RNA biology*, (just-accepted), pp.00-00.

Raina, J.B., Tapiolas, D., Motti, C.A., Foret, S., Seemann, T., Tebben, J., Willis, B.L. and Bourne, D.G., 2016. Isolation of an antimicrobial compound produced by bacteria associated with reef-building corals. *PeerJ*, 4, p.e2275.

Richardson, K.M., Schiffer, M., Griffin, P.C., Lee, S.F. and Hoffmann, A.A., 2016. Tropical *Drosophila pandora* carry *Wolbachia* infections causing cytoplasmic incompatibility or male killing. *Evolution*, 70(8), pp.1791-1802.

Robinson, N.A., Hall, N.E., Ross, E.M., Cooke, I.R., Shiel, B.P., Robinson, A.J. and Strugnelli, J.M., 2016. The complete mitochondrial genome of *Haliotis laevigata* (Gastropoda: Haliotidae) using MiSeq and HiSeq sequencing. *Mitochondrial DNA Part A*, 27(1), pp.437-438.

Shafee, T.M., Robinson, A.J., van der Weerden, N. and Anderson, M.A., 2016. Structural homology guided alignment of cysteine rich proteins. *SpringerPlus*, 5(1), p.27.

Smibert, O., Aung, A.K., Woolnough, E., Seemann, T., Carter, G., Spelman, D. and Peleg, A., 2016, December. Healthcare Worker Hand-Held Devices as Vectors for Multidrug-Resistant Organisms in Intensive Care Patients. In *Open Forum Infectious Diseases* (Vol. 3, No. suppl 1, p. 254). Oxford University Press.

Stark, Z., Tan, T.Y., Chong, B., Brett, G.R., Yap, P., Walsh, M., Yeung, A., Peters, H., Mordaunt, D., Cowie, S. and Amor, D.J., 2016. A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genetics in Medicine*.

Unterweger, B., Bulach, D.M., Scoble, J., Midgley, D.J., Greenfield, P., Lyras, D., Johanesen, P. and Dumsday, G.J., 2016. CYP101J2, CYP101J3, and CYP101J4, 1, 8-Cineole-Hydroxylating Cytochrome P450 Monooxygenases from *Sphingobium yanoikuyae* Strain B2. Applied and Environmental Microbiology, 82(22), pp.6507-6517.

Villacorta-Rath, C., Ilyushkina, I., Strugnell, J.M., Green, B.S., Murphy, N.P., Doyle, S.R., Hall, N.E., Robinson, A.J. and Bell, J.J., 2016. Outlier SNPs enable food traceability of the southern rock lobster, *Jasus edwardsii*. Marine Biology, 163(11), p.223.

Wakefield, M.J., 2016. Survival Volume: interactive volume threshold survival graphs. Journal of Open Source Software, 1(8).

Wakefield, M.J., 2016. Xenomapper: Mapping reads in a mixed species context. Journal of Open Source Software, 1(1).

Whitelaw, B.L., Strugnell, J.M., Faou, P., da Fonseca, R.R., Hall, N.E., Norman, M., Finn, J. and Cooke, I.R., 2016. Combined transcriptomic and proteomic analysis of the posterior salivary gland from the southern blue-ringed octopus and the southern sand octopus. Journal of proteome research, 15(9), pp.3284-3297.

Wong, N.C., Pope, B.J., Candiloro, I.L., Korbie, D., Trau, M., Wong, S.Q., Mikeska, T., Zhang, X., Pitman, M., Eggers, S. and Doyle, S.R., 2016. MethPat: a tool for the analysis and visualisation of complex methylation patterns obtained by massively parallel sequencing. BMC bioinformatics, 17(1), p.98.

*The VLSCI resources were important to many of our projects. The computations that that we have been doing required more CPU time and memory than was otherwise available to us. Without access to the VLSCI, a number of the publications that are currently under review and in preparation would not have happened. For some of our projects the **VLSCI was a catalyst** for undertaking the project, since without the HPC resources being reliably and conveniently available, it is likely that these lines of enquiry could not have been pursued.*

Anthony Burkitt

Research Director, Bionic Vision Australia
Professor of Engineering
University of Melbourne

*VLSCI has been a valuable contributor to grant applications and publications. With regard to the latter, MD studies may have helped publish in **higher impact journals**.*

Michael Parker

Director, Bio21 Institute
Head of Structural Biology, ACRF Rational Drug Discovery Centre, St. Vincent's Institute of Medical Research
NHMRC Senior Principal Research Fellow Professor, Department of Biochemistry and Molecular Biology
University of Melbourne

*This resource remains **crucial to our analysis** of genomic and transcriptomic data - we cite it in grant applications, and it is central to our production of publications on this theme.*

Stuart Ralph

Associate Professor and NHMRC RD Wright Biomedical Research Fellow
Department of Biochemistry and Molecular Biology
University of Melbourne

Our Community

Skills development

Now firmly embedded in the Melbourne bioinformatics landscape, VLSCI assists hundreds of aspiring researchers each year through activities that reach different levels of the research community. A range of activities can help build the careers of computational life scientists, from work experience for high school students, to lecturing undergraduates, and professional placements for postgrads. We invest in skills development and the sharing of expertise because we know they are essential to the sustained growth of a healthy research ecosystem.

In 2016 it was particularly rewarding to nurture the development of these passionate and skilled young women for a brief time in their promising careers.

Chelsea

Encouraging young people to investigate careers in computational biology is essential to our research community's sustainability, so VLSCI regularly participates in events such as coding competitions and Open Days. From time to time our staff also host 'work experience' for particularly enthusiastic high school students who have shown an aptitude for science and computing. We aim to inspire their curiosity in scientific problem solving, inform their notions about what it is like to work as a scientist or in a professional environment, and also encourage them to strive for making the most of their education – at high school, and then potentially at university.

Following her interest in sciences and aptitude for maths and computing, Chelsea Giofrelle had to spend a week away from her family and hometown of Robinvale to work with us at VLSCI. Taking it all in her stride, the intelligent and social Year 11 student worked on a variety of projects including researching the mutations and 3D printing a protein associated with cystic fibrosis with Dr Michael Kuiper. Talented Chelsea worked hard and told her supervisors:

I have had an amazing time and have learned so much about bioinformatics and what it means to be in this field of science. I know I still have a lot to learn about it but you all have inspired me to work hard and focus on my studies.

Careers Transition Pathway Officer from Robinvale College, Julie Kelly followed up with a message of thanks:

Chelsea had a great week and worked alongside an array of people and found the whole experience so very rewarding. Staff at the VLSCI were so very supportive and Chelsea has come back from Melbourne even more determined to follow her pathway into science/computing or bioinformatics.

Priyanka

With a background in software engineering, and an undergraduate degree in bioinformatics, Priyanka Pillai was sure the MSc (Bioinformatics) course at the University of Melbourne was the right one for her. Ultimately aspiring to use her computer science and bioinformatics proficiencies to create effective and economical computational biology tools, Priyanka hopes to "work towards holistic solutions in health care and eventually work with a philanthropic association".

As part of her course, coordinated by VLSCI Director Assoc Prof Andrew Lonie, Priyanka is undertaking research that could eventually help reduce the prevalence of a condition affecting the mental, emotional and sexual health of women. The project, 'DNA Sequencing approach to classify sub-species causing Bacterial Vaginosis' is supervised by Melbourne Sexual Health Centre's Assoc Prof Catriona Bradshaw and VLSCI's Dr Dieter Bulach. Priyanka's exceptional communication skills came recommended to us when we needed someone in the VLSCI office to assist with organising

and evaluating our bioinformatics training activities. During her time working at VLSCI, this keen networker made many new connections and eventually moved on to take up an internship with Dr David J Lynn, EMBL Australia group leader in the Infection and Immunity Theme at the South Australian Health and Medical Research Institute.

Madison

Immediately after graduating from the MSc (Bioinformatics) program, Madison Flannery successfully applied for a job as a Bioinformatics Platform Developer at VLSCI in 2016. With a background in computer science, Maddy was seeking a role that would position her at the intersection of biology and computing science.

Maddy's work was primarily focussed on building and supporting the cloud-based bioinformatics analysis platform, the Genomics Virtual Laboratory (GVL) - adding new tools to ensure all the gold standard tools are available, as well as undertaking general maintenance and improvement to ensure system reliability and ease of use. Her work allowed more Australian scientists to conduct efficient, collaborative research, and get faster outcomes.

My job allows me to support all kinds of bioinformatics and life sciences research around Australia. Through my contributions to the GVL, I am enabling this potentially life-changing research, by ensuring researchers have access to the user-friendly tools they need. This is the most exciting thing about my job: knowing that my work - which sometimes can be writing a few lines of code - is always contributing to something much larger and something that can have an impact directly on patients in hospitals.

Pursuing her goal to make a tangible difference to people's lives, Maddy took up a Software Engineer position with artificial intelligence company, DeepMind Technologies, in the UK in late 2016.



Clockwise from left: Bioinformatics Platform Developer, Madison: work experience student Chelsea from Robinvale College; casual employee and MSc(Bioinf) student Priyanka Pillai presents to VLSCI staff.

*VLSCI's efforts in community building have been **key in connecting groups** around the greater Melbourne area that have common research interests, and would rarely meet otherwise.*

Patrick Sexton
Theme Leader, Drug Discovery Biology, Professor of Pharmacology
NHMRC Principal Research Fellow
Monash Institute of Pharmaceutical Sciences
Monash University

Hands-on Workshops

VLSCI offers a regular curriculum of workshops providing introductions for biologists to a range of tools and techniques. In 2016 there were 21 local workshops offered across three locations, reaching a total of 359 Victorian participants from a wide variety of research institutes or organisations.

21 WORKSHOPS



*Figures include workshops organised by VLSCI in its capacity as an EMBL-ABR Node, and exclude La Trobe University's workshops

WORKSHOP	LOCATION	DATE
Introduction to Unix	La Trobe	4/04/16
Introduction to High Performance Computing	La Trobe	6/04/16
EMBL-ABR CyVerse Training Day	VLSCI	11/04/16
HPC Archiving	La Trobe	7/05/16
Introduction to Unix	VLSCI	02/06/16 3/11/16
Introduction to HPC	VLSCI	2/06/16 3/11/16
Introduction to the GVL and Galaxy	VLSCI	7/06/16 7/11/16
Partek Genomics Suite	La Trobe	6/07/16
RNA-Seq Differential Gene Expression Analysis Using the GVL & Galaxy	VLSCI	14/06/16
Variant Calling with Galaxy and the GVL for Beginners	VLSCI	8/06/16 13/5/16
CLC Genomics	La Trobe	16/06/16
CLC Genomics Workbench/Server	La Trobe	16/06/16
HPC Archiving	La Trobe	21/06/16
Introduction to EBI Bioinformatics Resources	VLSCI	23/06/16
Introduction to Molecular Modelling and Visualisation for Life Sciences	VLSCI	4/07/16 29/11/16
Intermediate Molecular Modelling and Visualisation for Life Sciences	VLSCI	6/07/16
Open Source Science with Git and GitHub	VLSCI	2/08/16
EMBL-ABR Data Life Cycle Best Practice Workshop Series	VLSCI	24/10/16 -28/10/16
EMBL-ABR/GOBLET Hands on RNA-Seq: Data Analysis	VLSCI	14/11/16
EMBL-ABR All-Hands Meeting	VLSCI	7/12/16
EMBL-ABR/ELIXIR Workshop: Bioinformatics Software	VLSCI	8/12/16
EMBL-ABR/ELIXIR Workshop: Registries in Bioinformatics: Tools, Datasets, Standards, Events & Training	VLSCI	9/12/16
EMBL-ABR Open and Scalable Training Workshop	VLSCI	9/12/16

Contributions to external training

COURSE/ UNIVERSITY SUBJECT	TITLE	PRESENTER	LOCATION	DATE
Bio-Nano Computer Simulation & Modelling Workshop	Computer simulations of antifreeze proteins	Michael Kuiper	Melbourne	18/01/16
Bioplatforms Australia training course	Genomics Analysis Tool Kit Best Practice Workshop	Gayle Philip Jessica Chung Peter Georgeson	UoM	3/02/16 -4/02/16
Bioplatforms Australia training course	Hands-On Cancer Genomics	Gayle Philip	UoM	8/02/16 -10/02/16
CSIRO 3D Printing Workshop	3D printing in the life sciences	Michael Kuiper	Melbourne	29/02/16
Joint Academic Microbiology Seminars:	The Genomics Virtual Laboratory & An Introduction to Microbial Genomics	Simon Gladman	AM	14/03/16
Universidad de La Salle, Bogota	Introduction to Bacterial Genomics	Dieter Bulach	Colombia	21/04/16 -23/04/16
Galaxy Community Conference 2016	Galaxy Community Conference 2016 workshops: Genome Assembly, Genome Annotation, RNA-Seq, Human Variant Calling	Simon Gladman Torsten Seemann Philippa Griffin Jessica Chung Anna Syme	Bloomington, USA	1/06/16
McGill Summer Institute	Introduction to genomic epidemiology of infectious diseases	Torsten Seemann	Montreal, Canada	20/06/16 -24/06/16
Winter School in Mathematical and Computational Biology	De novo genome assembly & Comparing the variome and pan-genome of bacterial isolates	Torsten Seemann	Brisbane	4/07/16
ImageXD 2016	Image processing in Python	Juan Nunez-Iglesias	University of California, USA	6/07/16
Winter School in Mathematical and Computational Biology	Comparing the variome and pan-genome of bacterial isolates	Torsten Seemann	Brisbane	8/07/16
Bioplatforms Australia training course	Introduction to NGS	Gayle Philip	Curtin	12/07/16 -14/07/16
Bioplatforms Australia training course	Clinical Genomics	Gayle Philip	Garvin	18/07/16 -22/07/16
AINSE Winter School on Applications of Nuclear Techniques	Computer simulations of antifreeze proteins	Michael Kuiper	Lucas Heights, NSW	18/07/16
NGS Summer 2016 - Analyzing Next-Generation Sequencing Data	Various seminars & practicals	Torsten Seemann	Michigan, USA	8/08/16 -19/08/16
Microbial Genomics Course, University of Western Australia	Microbial Genomics Workshop	Anna Syme Simon Gladman	UWA	22/11/16 -23/11/16
Software Carpentry	Software Carpentry in R	Juan Nunez-Iglesias	ANU	4/12/16

Doherty Asia Pacific Genomics Course	Five day intensive course: various seminars & practicals	Torsten Seemann Dieter Bulach Simon Gladman	PDI	5/12/16 -9/12/16
SciPy 2016	Image processing in Python	Juan Nunez-Iglesias	University of Texas, USA	7/12/16
BINF90002 Elements of Bioinformatics	Full subject coordination	Bernie Pope	UoM	Semester 1
BINF90002 Elements of Bioinformatics	Molecular Modelling I & II	Michael Kuiper	UoM	Semester 1
GENE30001 Evolutionary Genetics and Genomics	Genome Assembly	Philippa Griffin	UoM	Semester 1
GENE30001 Evolutionary Genetics and Genomics	Genetic Markers I & II	Philippa Griffin	UoM	Semester 1
COMP90014 Algorithms for Functional Genomics	De novo Genome Assembly I & II	Simon Gladman	UoM	Semester 2



I really enjoyed the workshop. It gave me a much better level of understanding of the analysis of RNA-seq data. I will now be able to apply this to my own dataset. Before this, I had trouble communicating with the person doing the analysis. Now I think I could do it myself and get advice from the person originally doing the analysis.

PhD Student
University of Melbourne

Support for Students

STUDENT	HOST INSTITUTION	PRIMARY SUPERVISOR/S	PROJECT
RHD STUDENTS WITH VLSCI SUPERVISION			
Brett Shiel	PhD, La Trobe	Nathan Hall Ira Cooke Jan Strugnell	Understanding of abalone heat stress through transcriptomics
Anbarasu Karthikaichamy	PhD, Monash	Dieter Bulach Ross Coppel	Generating "live" gene expression tools for use in algal molecular manipulation
Tom Stent	PhD, Monash	Dieter Bulach Julian Rood	Clostridial necrotic enteritis
Andrew Buultjens	PhD, UoM	Torsten Seemann Tim Stinear	Tracking the source of mycobacterium ulcerans in Victoria
Sehrish Kanwal	PhD, UoM	Andrew Lonie Richard Sinnott	Informatics platforms for clinical genomics
Jason Kwong	PhD, UoM	Torsten Seemann Ben Howden	Genomics for public health microbiology
Camelia Quek	PhD, UoM	Chol-Hee Jung Andrew Hill	Implication of small transcriptome in exosomes during Prion disease
Nikeisha Caruana	PhD, La Trobe	Ira Cooke Jan Strugnell	Composition and function of Cephalopod slime
Muhammd Dilmir	MSc(Bioinf), UoM	Khalid Mahmood	A framework for prioritizing variants using protein structure
Lachlan Dryburgh	MSc(Bioinf), UoM	Juan Nunez-Iglesias	Improving segmentation tools for connectomics
Hongzi Luo	MSc(Bioinf), UoM	Gayle Philip Judith Savage	Analysis of Renal Disease associated gene variants in Miniature Bull Terriers
Priyanka Pillai	MSc(Bionf), UoM	Dieter Bulach Catriona Bradshaw	Whole Metagenome Sequencing to study the onset of Bacterial Vaginosis
Jonathan Reich	MSc(Bioinf), UoM	Juan Nunez-Iglesias	Who watches the watchers? Ranking error metrics in connectomics, and a vision for error visualisation
Dong (Don) Hao Teng	MSc(Bioinf), UoM	Juan Nunez-Iglesias Kaylene Simpson	Features and metrics to compare high content screen images
WORK EXPERIENCE STUDENTS			
Chelsea Giofrelle	Robinvale College	Michael Kuiper	
BEST PRESENTATION IN LIFE SCIENCES COMPUTATION – CIS DOCTORAL COLLOQUIUM			
Sehrish Kanwal	UoM	Andrew Lonie Richard Sinnott	Reproducibility of Genomic Workflows
MSC (BIOINFORMATICS) BURSARY RECIPIENTS			
David Elkanan (Systems Genomics bursary)	UoM	David Balding Christine Wells	Compilation of a blood development expression time course
Michal Rousset (VLSCI bursary)	UoM	Leann Tilley James McCaw	Mathematical modelling of Artemisinin treatment in <i>Plasmodium falciparum</i>

MSc (Bioinformatics) Students

NAME	HOST INSTITUTION	PRIMARY SUPERVISOR/S	PROJECT
Keely Aranyos	UoM	Alex Adrianopolous Karin Verspoor	Functional characterisation of Transcription Factors in Fungi
Kevin Cheng	UoM	Aaron Jex	Analysis tools for the MinION sequencer
Daniel Esposito	UoM	Melissa Davis	Analysis of functional modules in tissue- and cell-specific molecular interaction networks.
Kahli Flekac	UoM	Melissa Davis	Mapping the molecular networks of the early endosome
Peter Glover	SVI	Mark Chong	Genomic and molecular consequence of autoimmunity-associated genetic variation
Karen Gu	SVI	Mark Chong	Sequence and structure analysis of genes under mRNA-cleavage regulation in stem cells
Nisitha Jayatilleke	WEHI	Melanie Bahlo	Finding and analysing copy number variation in malaria high throughput sequencing data
Alexis Lucattini	AGRF/WEHI	Lavinia Gordon Matt Ritchie	Comparison of long reads _ MinION vs PacBio
Hong Zhi Luo	UoM/RMH	Judy Savige	Identification of further genes that are mutated in inherited renal disease
Cristian Pena	Peter Mac	Tony Pappenfuss	De novo genome assembly of the scabies mite
Tony Pham	UoM	Paul Waring Graham Taylor	Molecular Evolution of Meningiomas
Priyanka Pillai	RWH	Sepehr Tabrizi Dieter Bulach	A Whole Metagenome Sequencing approach for the investigation of the onset of bacterial vaginosis
Marion Shadbolt	UoM	Gordana Rasic	Genome-wide recombination dynamics in <i>Aedes aegypti</i> destined for field releases
Michael Silk	MGHA	Natalie Thorne	Clinical Bioinformatics

*VLSCI has provided a vital resource (both facilities and personnel) for the establishment and continuation of this project. It provides **computational and intellectual support** to many research students involved with computational structural biology, vital to their education and enabling successful completion of their projects.*

*From a purely scientific standpoint, as a result of work carried out using VLSCI facilities, we have made significant progress in understanding the interactions between dietary-source compounds and proteins involved in inflammation, which will lead to **high impact publications**, as well as potentially inform clinical and nutritional practice in the context of treatment of chronic wounds and cancer.*

*From a collaborative perspective, VLSCI has also enabled us to strengthen our cross-institute collaborations nationally and internationally. These collaborations will continue to yield research outputs, ultimately **leading to better health and disease treatment**. VLSCI has also helped us establish further links with the existing computational biology community, and these will also ultimately benefit our research and educational outcomes.*

*I would like to see VLSCI continue to serve its **vital role in the Australian computational biology community**, and that it will receive the support necessary to acquire and maintain world-class computational facilities.*

Andrew Hung

Applied Chemistry and Environmental Science, RMIT University

Sponsorship of Meetings and Conferences

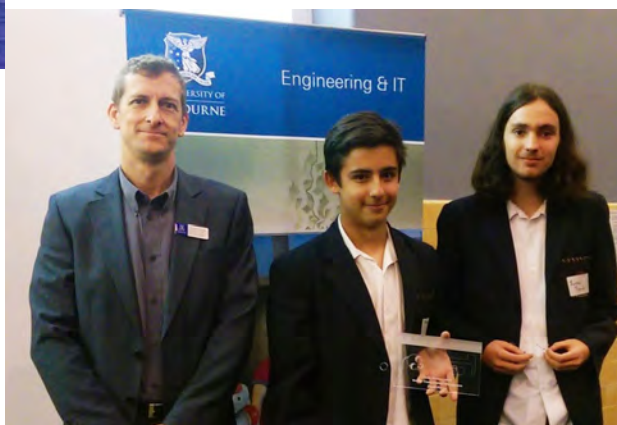
VLSCI contributed funds and in kind support to a range of organisations and events. In 2016 ongoing sponsorship of the bioinformatics/computational biology student group, COMBINE, and the Convergence Science Network strengthened existing training and networking activities. Providing financial support to a range of events - from student events such as UoM's Codemasters Programming competition for high school students and the CIS Doctoral Colloquium, to professional events such as Linux Conference Australia's Bioinformatics mini-conference - reach a diverse range of students and researchers.

DATE	ACTIVITY (LOCATION)	TYPE OF SPONSORSHIP
Annual sponsorship	COMBINE (student-run organisation for researchers in computational biology & bioinformatics)	Major sponsorship, plus in kind support for local workshops and WA workshop
Annual sponsorship	Convergence Science	Major sponsorship and provision of speakers
February	Code Masters Challenge (Computing & Information Systems, UoM)	Sponsored event
March	Graeme Clark Oration & activities (Melbourne Convention and Exhibition Centre)	Sponsored event, hosted a group of students at Oration and Oration Dinner.
July	Postdoctoral Colloquium (CIS, UoM)	Sponsored prize



Left: Sehrish Kanwal was presented the VLSCI Prize for Best Life Sciences Computation Presentation at the CIS Doctoral Colloquium by Assoc Prof Andrew Lonie

Right: Winners of the Seniors Division of the Code Masters Competition from Viewbank College with organizer Assoc Prof Andrew Turpin



This concentrated project helped to build a state-wide capacity for leading research

Prof Jim McCluskey
Deputy Vice-Chancellor (Research)
University of Melbourne

Financials

9 February, 2017

Karen Gadd,
Associate Auditor
Audit Assurance Services – Legal & Risk
The University of Melbourne
Level 4, Alan Gilbert Building
161 Barry Street
Carlton Victoria 3010

Dear Karen,

In relation to the VLSCI Accounts for the period 1 January 2016 to 31 December 2016, I hereby certify that the funds for VLSCI have been expended in accordance with the agreement between the Department of Economic Development, Jobs, Transport & Resources and The University of Melbourne.

Kind regards,



Karin Diamond
Business Manager
Victorian Life Sciences Computation Initiative
The University of Melbourne

Department of Economic Development, Jobs, Transport & Resources

ANNEXURE A - Statutory Declaration

Statutory Declaration


I, Allan Tait, do hereby solemnly and sincerely declare that:

1. I am the Chief Financial Officer of the University of Melbourne (the **Recipient**).
2. The Recipient has complied with all of its obligations under the Grant Agreement dated 29 September, 2014 (the **Agreement**) between the Recipient and the State of Victoria.
3. The Recipient has incurred \$6,650,000 of Project Expenditure as at 31 December 2016 in accordance with the terms of the Agreement.
4. At 31 December, 2016 all grant funds have been in accordance with the terms of the Agreement.
5. Previously submitted reports have confirmed that the Recipient secured \$6.67 million (Annual Report 2015) in corporate sponsorship/stakeholder contributions for the 2015 calendar year. To 30 June 2016, there was a further contribution from stakeholders of \$1.81 million (Milestone 3 Report) giving a total of \$8.48 million. For the 6 months from 1 July 2016 to 31 December 2016 there was a further stakeholder contribution of \$2.25 million. For the project (1 January 2015 to 31 December 2016) the total stakeholder contributions were \$10.73 million.
6. The accounts relating to the Project as attached to this Statutory Declaration are true and correct.*

AND I make this solemn declaration, conscientiously believing the same to be true, and by virtue of the provisions of an Act of the Parliament of Victoria rendering persons making a false declaration punishable for wilful and corrupt perjury.

Declared at Melbourne

This 6th day of April 2017


Signature of person making this declaration
(to be signed in front of an Authorised
Witness)

Before me,


Signature of Authorised Witness

LEANNE FISHER JP
C/O HONORARY JUSTICE OFFICE
18/121 EXHIBITION STREET
MELBOURNE 3001
JUSTICE OF THE PEACE FOR VICTORIA
REG NO. 12694

An Australian lawyer (within the meaning of the *Legal Profession Act 2004*) Justice of the Peace;
A senior officer of a council (as defined in the *Local Government Act 1989*); or
Any other person empowered under Section 107A of the *Evidence (Miscellaneous Provisions)
Act 1958*

**Note: The Project accounts should, at a minimum, itemise the Project Expenditure and, in relation to each item, describe the nature of expenditure, the amount and the connection to the Project.*

ANNEXURE B – Audit Opinion – Project Expenditure

Date: 20 March 2017

The Chief Executive
Department of State Development, Business and Innovation
121 Exhibition Street
Melbourne VIC 3000

University of Melbourne

This audit opinion is prepared for the purposes of the Grant Agreement dated 29 September 2014 (**Agreement**) between the State of Victoria and The University of Melbourne (**Recipient**) for the transitional operations funding for the Victorian Life Sciences Computation Initiative (**VLSCI**) as further described in the Grant Details of the Agreement (**the Project**).

Scope

We have conducted an independent audit in accordance with Australian Auditing Standards of the attached Financial Statement of Income and Expenditure for the six month period from 1 July 2016 to 31 December 2016 provided by the recipient.

Our audit involved an examination, on a test basis, of evidence supporting the amount of the Project Expenditure incurred and the value of equipment installed and used at the Project Site. This included an examination of the Recipient's financial records and receipts, and an evaluation of the policies and procedures used to calculate the Project Expenditure. These procedures have been undertaken to form an opinion as to whether the methodology used to calculate the Project Expenditure, is in accordance with the Agreement, and that the figure is stated true and fair.

This Audit Opinion expressed in this report has been formed on the above basis.

Audit Opinion

We confirm that in our opinion, the Financial Statements of Income and Expenditure (2015 & 2016) present fairly the expenditure of grant funds for the VLSCI for the period 1 January 2015 to 31 December 2016. As at 31 December 2016 the Recipient had incurred:

- \$6.65 million cumulative project expenditure in accordance with the terms of the **VLSCI** Agreement with \$0 of carried forward expenditure as at 31 December 2016
- As at 31 December 2016 grant funding of \$3million received for the **VLSCI** had been fully expensed in accordance with the funding agreement



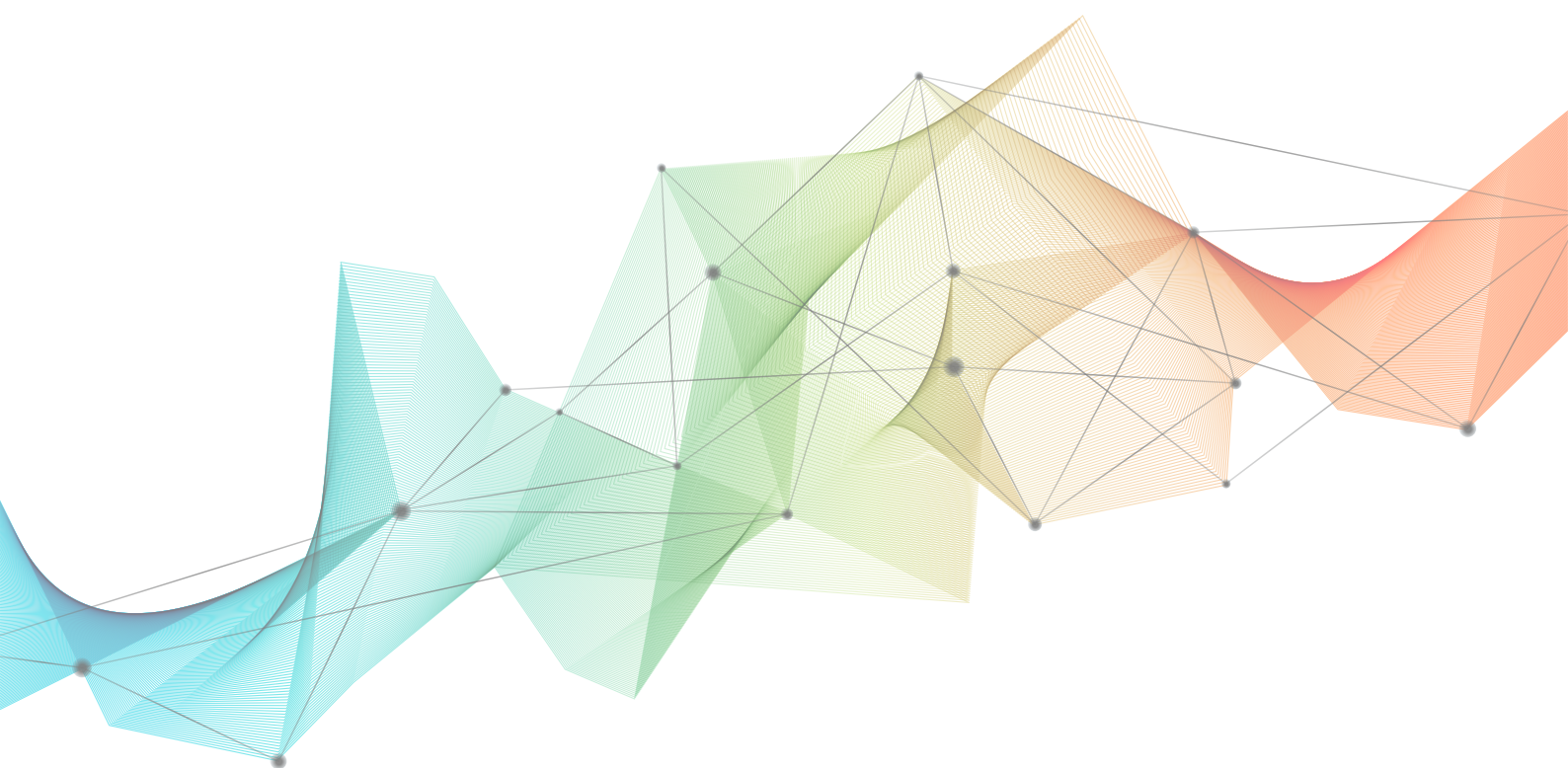
Lisa Tripodi

Date: 20 March 2017
Partner, Accounting and Assurance

Financial Statement 31 December 2016	YTD Actual to 31/12/2016 (\$)	BREAKDOWN (\$)	
CARRY FORWARD		Grant	University
Carry Forward (University of Melbourne)	3,285,496.46		3,285,496.46
Carry Forward (ECODEV) - 016477 Trust	750,000.00	750,000.00	
TOTAL CARRY FORWARDS	4,035,496.46	750,000.00	3,285,496.46
INCOME			
Other Grant Income	1,546,903.11		1,546,903.11
Fee for Service Income	157,000.00		157,000.00
Investment Income	4,309.88		4,309.88
Consulting and Contracted Services	778,896.11		778,896.11
Internal Allocation	286,356.52		286,356.52
University Contributions & Levy Income	1,290,000.00		1,290,000.00
Grant Income ECODEV (Milestone 2 Drawdown)	2,250,000.00	2,250,000.00	
TOTAL INCOME + GRANT Draw Down	6,313,465.62	2,250,000.00	4,063,465.62
EXPENDITURE			
Academic Salaries	2,496,169.97	1,116,078.90	1,380,091.07
Professional Salaries	1,723,094.75	728,389.66	994,705.09
TOTAL SALARY EXPENDITURE	4,219,264.72	1,844,468.56	2,374,796.16
Non Salary Expenditure			
Finance Related Costs	13,102.73		13,102.73
Student Support	12,711.59		12,711.59
Consumable Goods and Services	66,299.04		66,299.04
Expert Services	277,937.59		277,937.59
LaTrobe Salaries	165,307.54		165,307.54
Lab 14 Rent	91,615.73		91,615.73
Travel, Conf & Entertainment	229,860.19		229,860.19
Computer Services, Software	367,274.44		367,274.44
Hardware Maintenance (IBM Blue Gene)	1,221,017.14	1,155,531.44	65,485.70
Asset Expense < \$10,000*	18,773.74		18,773.74
Infrastructure Related Assets	32,705.74		32,705.74
Electricity Costs	268,409.01		268,409.01
TOTAL NON SALARY EXPENDITURE	2,765,014.48	1,155,531.44	1,609,483.04
TOTAL EXPENDITURE	6,984,279.20	3,000,000.00	3,984,279.20
Balance 31 December, 2016	3,364,682.88	0.00	3,364,682.88
31 December 2016 Balance (Grant Final)		\$0.00	

Glossary

AGRF	Australian Genome Research Facility Ltd
ANDS	Australian National Data Service
ARC	Australian Research Council
Baker IDI	Baker IDI Heart and Diabetes Institute
Bio21	Bio21 Institute
Burnet	Burnet Institute
CCS	Monash University's Central Clinical School
CCV	Cancer Council Victoria
CEC	Cancer Council Victoria's Cancer Epidemiology Centre
CERA	Centre for Eye Research Australia
Curtin	Curtin University
Deakin	Deakin University
DEPI	Department of Environment & Primary Industries Victoria
EMBL	European Molecular Biology Laboratory
Eng UoM	Melbourne School of Engineering, University of Melbourne
FedUni	Federation University
Florey	Florey Institute of Neuroscience and Mental Health
La Trobe	La Trobe University
LSCC	Life Sciences Computation Centre – VLSCI
Ludwig	Ludwig Institute for Cancer Research
MBG	Melbourne Brain Centre
MCRI	Murdoch Children's Research Institute
MDS	Melbourne Dental School, University of Melbourne
MHTP	Monash Health Translation Precinct Medical Genomics Facility
MIMR	Monash Institute of Medical Research
MIPS	Monash Institute of Pharmaceutical Sciences
Museum Vic	Museum Victoria
NCI	National Computational Infrastructure
NCMAS	National Computational Merit Allocation Scheme
Nectar	National eResearch Collaboration Tools and Resources
NHMRC	National Health and Medical Research Council
NICTA	National Information and Communications Technology Research Centre of Excellence
PCF	Peak Computing Facility - VLSCI
PDI	Peter Doherty Institute for Infection and Immunity
Peter Mac	Peter MacCallum Cancer Centre
RAS	Resource Allocation Scheme
RBG	Royal Botanic Gardens
RCH	The Royal Children's Hospital
RMH	Royal Melbourne Hospital
RMIT	RMIT University
RWH	Royal Women's Hospital
SVH	St Vincent's Hospital
SVI	St Vincent's Institute
Swinburne	Swinburne University of Technology
SydneyUni	The University of Sydney
UC	University of Canterbury
UNSW	University of New South Wales
UoA	The University of Adelaide
UoM	The University of Melbourne
UoN	University of Newcastle
UoSC	University of the Sunshine Coast
UoW	University of Wollongong
UQ	The University of Queensland
UWA	The University of Western Australia
VCB	Victorian Cancer Biobank
VIDRL	Victorian Infectious Diseases Reference Laboratory
WEHI	Walter & Eliza Hall Institute of Medical Research



VLSCI & EMBL Australia Bioinformatics Resource

Lab-14, 700 Swanston Street, Carlton 3053, Victoria, Australia

PHONE +61 (3) 9035 5822 WEB www.vlsci.org.au WEB www.embl-abr.org.au

VLSCI is funded by the Victorian Government and contributing institutions and is hosted by the University of Melbourne.
The EMBL Australia Bioinformatics Resource is hosted at VLSCI through a funding agreement
between the University of Melbourne and Bioplatforms Australia.