

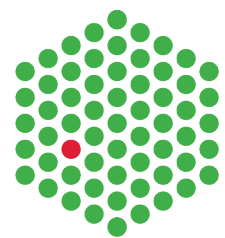
Annual Report 2016



EMBL

Australia

Bioinformatics Resource





EMBL Australia Bioinformatics Resource (EMBL-ABR) is a distributed national research infrastructure providing bioinformatics support to Australian life science researchers. The Resource was set up as a collaboration with the European Bioinformatics Institute (EMBL-EBI) to maximise Australia's bioinformatics capability. This close partnership is made possible in the context of Australia's associate membership of EMBL. EMBL-ABR is supported by Bioplatforms Australia via Australian Government NCRIS investment, and the University of Melbourne.



BIOPLATFORMS
AUSTRALIA

NCRIS

National Research
Infrastructure for Australia

An Australian Government Initiative

EMBL
Australia





Table of Contents

2016 Highlights	4
From the Directorate	6
About EMBL-ABR	7
Our Mission	9
From the Chair, International Science Advisory Group	9
The Hub	10
International Science Advisory Group	12
Heads of Nodes Group	12
The Nodes and their activities	12
Key Areas Coordinators Group	16
Community Leads Group	16
Outreach & Community Development	18
Training & Workshops	23
Media	24
EMBL-ABR People	25
Glossary	30
Appendix: Node Profiles	30
Publishing information	41

2016 Highlights



One Hub, Ten Nodes and a high-calibre International Scientific Advisory Group (ISAG)

By the end of the year, the Hub had been established, it had initiated a process for inviting applications for Nodes across Australia and ten Nodes were signed on. This new Head of Nodes group is consulted on all decision-making, with the ISAG providing strategic consultation and advice for the team at the Hub to ensure transparency, due diligence and good communication channels.



20 Interviews in new Series

20 international and national bioinformatics experts participated in our interview series published on the new EMBL-ABR website. The Series developed into a significant resource of independent views about the needs, trends and problems currently experienced by the life science community and it is freely available to inform future policy in the field.



One unique new workshop series

The October Data Life Cycle workshop series launched a new level of training for life scientists working on human, animal, plant and non-model organisms. International faculty joined local experts in Melbourne for an engaging, week-long program of workshops.

Other novel workshops were provided by international experts coming to Australia to contribute to EMBL-ABR planning, governance and lobbying activities.

600 followers, 5000 visitors, 270 subscribers



- o Following its March launch, by the end of the year the new website - www.embl-abr.org.au - had recorded over 5000 unique visitors.
- o Established early in the year, the @EMBL_ABR twitter handle attracted 600 followers.
- o Subscribers to EMBL-ABR News grew to 270, with 25% being international.

2016 Highlights



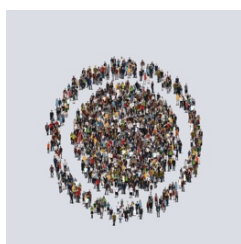
123 participants in Survey of Bioinformatics and Computational needs in Australia

Responses from this survey of Australian life science researchers are now providing essential data for priority planning in training, resources, data services and infrastructure. The survey outcomes have been made freely available on the website and shared with the community.



Five new International Resources

Through our international networks at EBI and the EU new resources for the Australian research community were implemented: ToolsAU, ISA Tools, BioSharing, iAnn (an event-sharing platform) and the Search for Training Materials (STM).



62 stakeholders attended first All Hands meeting

In December, coinciding with the first face-to-face International Scientific Advisory Group meeting in Melbourne, Node members and scientific leads participated in an inaugural All Hands meeting and an Australian Biosciences Cloud network meeting.

EMBL-ABR was represented at meetings of the Australian bioinformatics professional bodies such as ABACBS and COMBINE, as well as important communities at CSIRO and AGRF.



Many important International Networks established

- o Galaxy Australia (@galaxyaustralia) was launched with an announcement of the inaugural Galaxy Australasia conference to be held in Melbourne in February 2017 (GaME2017).
- o The Australian-made Genomics Virtual Laboratory (GVL) was well received at conferences and meetings in the USA, Europe and the UK.
- o Relations were developed with important international bioinformatics communities at CyVerse, BD2K, EMBL-EBI, ELIXIR, Biosharing, GOBLET and at the ISMB and ECCB conferences.



From the Directorate

It is our pleasure to present our community and stakeholders with the 2016 Annual Report. It is hard to isolate highlights for this introduction as it was a year of highlights. So much has been achieved in a short time and we are now looking forward to 2017 when we will consolidate these activities to demonstrate how essential it is for us to take time from our day to day research tasks to secure infrastructure support for Australian bioscience research. We see this work as vital if we are to keep up with international trends in biology.

Funding from Bioplatforms Australia and the University of Melbourne enabled the EMBL-ABR Hub to begin its work connecting communities across Australia and linking Australian researchers to the best practices of major international bioinformatics efforts at EMBL-European Bioinformatics Institute (EBI), ELIXIR, CyVerse and the NIH's Big Data to Knowledge (BD2K). Along with the arrival of Assoc Prof Vicky Schneider as Deputy Director in February, in March Dr Philippa Griffin was appointed in the half-time role as Open Data Coordinator. Together with the assistance of the professional staff at the Victorian Life Science Computation Initiative (from 2017, "Melbourne Bioinformatics") we pushed ahead with a busy schedule of communicating, training and strategic and governance planning. By the end of the year we had a comprehensive website, ten completed Node descriptions and important groups established: the International Science Advisory Group (ISAG); the Head of Nodes Group; the Key Areas Coordinators Group; and the Biosciences Community Leads Group.

The new website quickly expanded to contain all this diverse activity. A feature was the regular interviews with national and international bioinformatics experts sharing their insights into the needs for bioinformatics and making suggestions for how to develop EMBL-ABR. This built into an informative overview of the current needs and preoccupations of the bioinformatics community and quickly circulated around the world through our growing networks.

In October we welcomed international guests for our intensive Data Life Cycle workshop series. In December we were able to offer more workshops to coincide with experts visiting for the ISAG, when our Australian colleagues were also at our Melbourne Hub for the inaugural All Hands meeting. We were very pleased to welcome colleagues to the first face-to-face ISAG meeting. It presented an opportunity for them to hear about the exciting developments in major Australian life science projects and of efforts being made by EMBL-ABR and other groups, such as the Australian Bioinformatics and Computational Biology Society (ABACBS), to build support for bioinformatics here. We were all extremely grateful for the ISAG's time and efforts: their high calibre and expertise reflects the increasing transformation of biology into a rich data science impacting health, society and industry. We welcome their insights and advice as we work out the best way to develop this crucial network.

Having worked hard on phase one of this project - establishing a network of contacts and expertise to develop a coherent and comprehensive vision for bioinformatics support in Australia - in 2017 we lead on with the second phase, pursuing research infrastructure funding from the Federal Government to secure the required resources to continue to develop this important biology framework. In doing so, we will be accounting for our activities guided by some ISAG-recommended metrics to ensure we are being transparent, collaborative and inclusive whilst aiming to maximise our effectiveness and community outreach.

We thank all our stakeholders for their ongoing support, we hope you enjoy reading about our work in 2016.



Andrew Lonie, Director



Vicky Schneider, Deputy Director

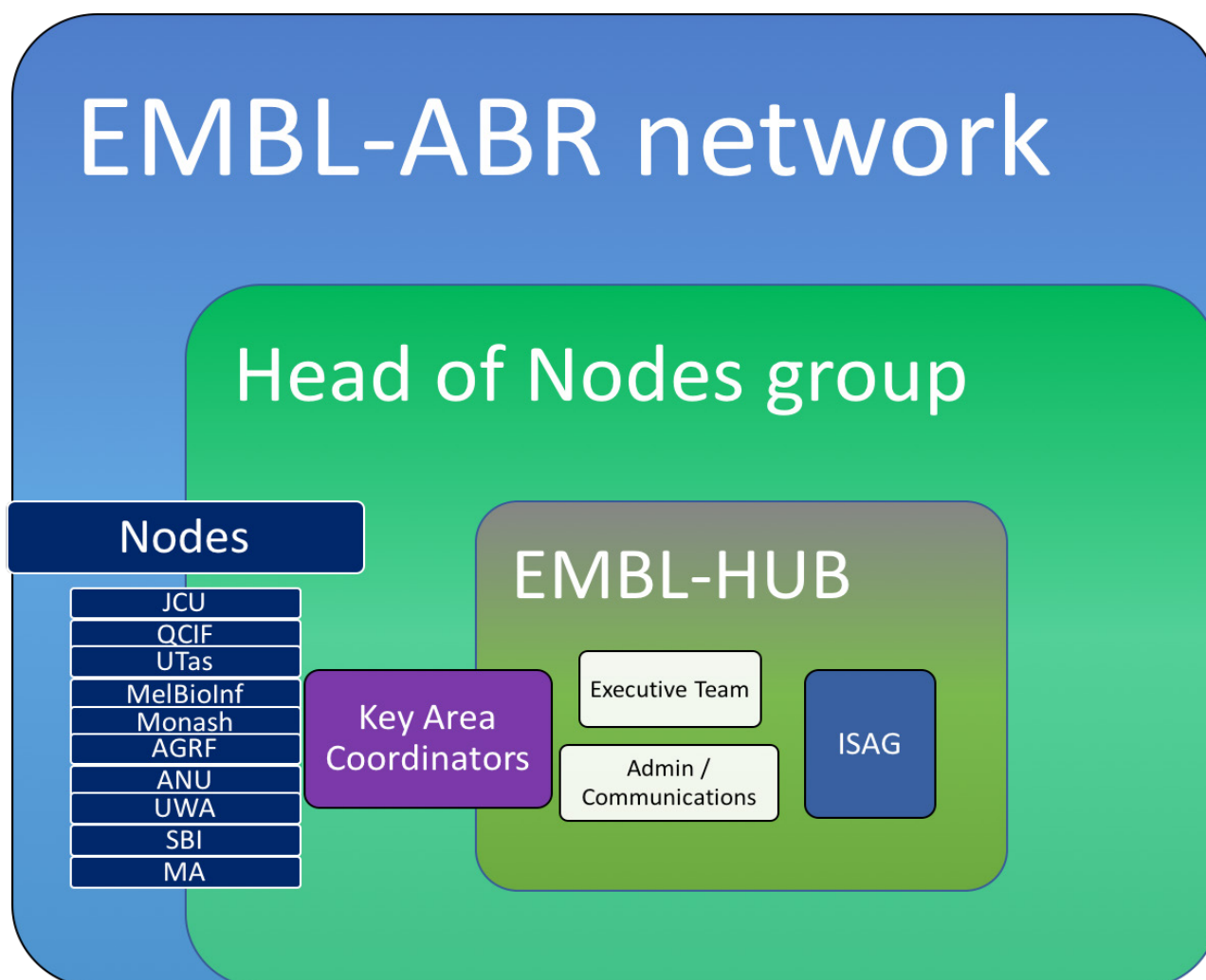
About EMBL-ABR

EMBL Australia Bioinformatics Resource is a distributed national research infrastructure providing bioinformatics support to Australian life science researchers. EMBL-ABR (the successor to BRAEMBL) was set up as a collaboration with the European Bioinformatics Institute (EMBL-EBI) to maximise Australia's bioinformatics capability. This close partnership is made possible in the context of Australia's associate membership of EMBL. EMBL-ABR is supported by Bioplatforms Australia (BPA) via the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) investment, and the University of Melbourne.

Structure and Governance

Aiming for a distributed, representative model for coordinating national bioinformatics activity, in early 2016 the Executive implemented a structure based upon a small Hub supporting key areas focussed in ten Nodes around the country. During the year, potential Nodes were invited to submit expressions of interest indicating what they would be able to offer the network in the key areas of Data, Tools, Compute, Standards, Training and Platforms. One of the first tasks of the ISAG was to review those Node proposals and approve the process, as a demonstration of how bioinformatics activities could be implemented across Australia through EMBL-ABR.

With Hub administrative and communications support provided by VLSCI, through host institution, the University of Melbourne, EMBL-ABR's Executive team convened the International Science Advisory Group (ISAG), the Head of Nodes Group (HoN), the Key Areas Coordinators' Group (KAGroup) and the Biosciences Community Leads Group. It is intended that with future funding success, an Executive Group would be convened from representatives across the network to oversee key decisions.



Financial agreements, contractual obligations

Acting on recommendations from BPA's 2014 review of operations of BRAEMBL, an agreement was struck between the University of Melbourne and BPA to provide funding of \$500,000 over 2014/15 to cover Queensland Cyber Infrastructure Foundation (QCIF) staffing and administrative work to transfer management of EMBL-ABR to the University of Melbourne, and begin to implement the review's other recommendations.

In 2016, a further \$400,000 was allocated to this project to:

- employ the Deputy Director
- employ the Open Data Coordinator (half time role)
- cover outreach activities driven from the Hub
- continue to contribute to costs involved with maintaining the data chaperoning as carried out at QCIF.

These funds have been fully expended in 2016 and any other activities carried out by this project in 2016 were made possible through both the in-kind and financial contributions of the University of Melbourne through the VLSCI and other collaborative efforts from the Nodes.

On 22 February 2016, University of Melbourne Associate Professor Maria Victoria Schneider joined EMBL-ABR as Deputy Director and together with VLSCI/Melbourne Bioinformatics Director Associate Professor Andrew Lonie formed the Executive team of the EMBL-ABR Hub and planned to:

- develop the organisational structure to support EMBL-ABR's role
- deliver bioinformatics advice and support to a range of researchers and projects
- develop data and method services and implement those on local and Commonwealth infrastructure
- build and deploy infrastructure to democratise access to bioinformatics tools and platforms
- continue to lead in training and education; and liaise closely with EBI across these activities
- employ an Open Data Coordinator for the Hub (Dr Philippa Griffin was appointed to that role in March),



Proposed 2016 activities:

- develop a rationale and protocol for expanding the Hub to Nodes, offering differential capability including but not limited to expanded domain expertise, additional infrastructural capacity and capability and geographic representation
- optimise exploitation of the tools and data of bioinformatics by Australian scientists by provision of eResearch infrastructure and expertise to the Australian life science research community including aspects such as High Performance Computing, the Nectar Research Cloud (Nectar), the Genomics Virtual Laboratory (GVL), International reference biomolecular data, researcher acquired biomolecular data, Research Data Storage (RDS), and, data management
- contribute to the global biomolecular information infrastructure in a way which showcases Australian science via support of several agreed key research projects with expertise as above
- provide broad access to infrastructure as above
- facilitate Australian researchers exchanging data with EBI by mirroring data where appropriate and chaperoning Australian data sets into relevant international data reference sets
- interact with EU counterparts at EBI and ELIXIR
- strengthen training and education, delivering local activities as appropriate
- coordinate national activities relating to BPA and network partners
- extend adoption of the GVL by the biology researcher community through the network.

Key performance measures for the agreement included: quantitative and qualitative metrics including adoption and usage of services, training evaluations, feedback from community representatives and successful implementation of key research projects.

This report acquits the revised 2015/16 contract. The Hub and ten Nodes have been established and EMBL-ABR is now a mature, established bioinformatics resource in the Australian research landscape.

Our Mission

EMBL-ABR aims to:

- increase Australia's capacity to collect, integrate, analyse, exploit, share and archive the large heterogeneous data sets now part of modern life sciences research
- contribute to the development and provision of training in data, tools and platforms to enable Australia's life science researchers to undertake research in the age of big data.
- showcase Australian research and datasets at an international level
- 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.

From the Chair, International Science Advisory Group


The International Scientific Advisory Group (ISAG) was formed to offer independent advice and feedback from national and international experts in bioinformatics, genomics, biomedicine, and life science education to EMBL-ABR. In the context of this annual report, the ISAG expresses its satisfaction with the excellent progress in just one short year. We feel strongly that the Executive team and the teams at EMBL-ABR Nodes are now well-positioned to capitalise on the emerging international opportunities for Australian life science researchers and students.

In 2016, the ISAG met three times; in June to establish the goals and purposes of the group, in October to familiarise ISAG members with Hub and Node activities, and finally in-person at VLSCI in December. From the outset, we agreed that our goal was to be pro-active in pursuit of opportunities to represent and promote EMBL-ABR and its Nodes. Given the current funding situation for research infrastructure in Australia, the ISAG will offer advice and insight from experiences of established infrastructure models in Europe and the USA. With members from three continents, ISAG also serves the need to help EMBL-ABR 'speak from experience' to the Federal Government and other decision-makers about the need for EMBL-ABR, and how investments in international science can be leveraged and replicated in Australia through EMBL-ABR.



ISAG members, L-R: Tony Papenfuss, Rebecca Johnson, Vivien Bonazzi, Jason Williams, Paul Flicek, Sean Grimmond, Jaap Heringa, Delphine Fleury, Mark Walker, (absent, Jenny Martin).

The ISAG was impressed with the progress made to draw in networks across Australia, having reviewed ten successful applications for establishment of Nodes of the EMBL-ABR Hub. Modelled somewhat on ELIXIR in Europe, this phase is designed to generate goodwill, stock-take capabilities and share resources. With no attached funding, it was good to see participants willing to share, commit and participate; to demonstrate how such a structure can usefully service the needs of Australia's diverse



bioscience research community as well as provide a mechanism through which local activity can be supported, on the proviso that it is then shared with the rest of the community. As bioscience matures as a data science, there are now many more reasons for researchers around the world to curate their data to international standards, share technologies, federate data sets and focus on reproducibility. Raising awareness of Australia's world-class science contributions will ensure its participation in large, internationally-funded research projects. Australia's high-value datasets and research capabilities can be greatly advanced through EMBL-ABR's work to unify biological and bioinformatics researchers and keep them in touch with international efforts.

During the December meeting, the ISAG was impressed with the variety of presentations highlighting the unique data being generated in Australia. With each Node and Key Area Coordinator's presentations, it also became clear that there is still a lot of work to do to ensure that Australian data is treated according to FAIR principles; that it is, Findable, Accessible, Interoperable and Re-usable – a potentially daunting task in an environment where researchers are not funded to do so. One of the recommendations of the ISAG was for EMBL-ABR to continue to work on some profile-raising, demonstration projects which will highlight the value of putting resources into such work. This is a task for 2017.

A full report from the ISAG was presented to the Executive following the meeting and formed the basis informing planning for 2017. Some members of the ISAG stayed on for the All Hands meeting held adjacent to the ISAG meeting, and we were grateful to those who also participated in a showcase to Federal Government representatives in Melbourne. Once again, our congratulations to the team on making great progress in 2016. The ISAG looks forward to supporting EMBL-ABR to grow and mature in 2017.

Jason Williams, Chair, ISAG

The Hub

Envisaging how a federated resource would be structured under a future funded model, the Hub was established to create that structure. Its current role is to curate the centralised guidelines for communication and marketing as well as all project, activities and reports, including network activity and impact measurements.

This Hub's future role is vital for actively building relationships with end-users, industry partners and national and international offices relevant to and collaborating with EMBL-ABR. The Hub also supports and monitors the EMBL-ABR bioinformatics training and scientific activities, and works on best practice guidelines for academic and industry users of EMBL-ABR services.

2016 activities:

- focussed the scientific strategy on recent scientific advances and front-end technologies relevant to the Australian research community
- implemented procedures to achieve objectives, monitored and captured Node activities to ensure better coordination of domain activity
- began exploring and trialling ways to capture end-users' needs and getting feedback for service improvement
- started defining quality control and quality assurance monitoring processes for the Nodes and network as a whole
- explored concepts for key national and international strategies and activities and investigated relevant funding opportunities for these
- reviewed mature, international bioinformatics infrastructure models and strategies for ensuring the sustainability of the network
- convened the inaugural All-Hands meeting and handled all related logistics
- launched, administered and led the Head of Nodes and the Key Area Coordinators' Group meetings
- convened, launched, administered and led meetings for a high-profile ISAG
- built and launched a new website
- launched a series of communication and marketing actions, including the EMBL-ABR Interview Series, On focus News and a Monthly Newsletter
- developed a Communications strategy, drafted a Gender and Balance Policy, drafted an Impact capturing model and related key documents.

EMBL-ABR

A distributed national research infrastructure

EMBL



Australia

Bioinformatics Resource

EMBL Australia Bioinformatics Resource (EMBL-ABR) is a distributed national research infrastructure providing bioinformatics support to life science researchers in Australia.



Data

- showcasing Australian research and datasets at an international level
- supporting best practice experimental design and analysis in bioinformatics, to encourage the cost-effective collection of high-quality, informative and reusable data
- cataloguing and delivering the data Australia needs



Tools

- assessing needs for bioinformatics software hosting, maintenance and support among Australian researchers and bioinformaticians
- implementing Australian Bioinformatics Tools Registry, ToolsAU
- contributing to the development and dissemination of bioinformatics tools



Training

- providing workshops on crucial topics not covered elsewhere in national bioinformatics training
- bringing expertise to Australia through visits by international experts and improved access to training resources
- coordination and dissemination of node end-user training activities



Platforms

- enabling access to bioinformatics platforms that link multiple tools, facilitate data sharing and analysis, and trace and record analysis pipelines
- leveraging the existing expertise and knowledge in platform development through the Genomics Virtual Lab (GVL) supported by the Nectar cloud



Compute

- assisting in the design, architecture and delivery of infrastructure
- supporting network activities and ensuring interoperability across the Australia and with international efforts
- identifying pathways to respond to future computation, analysis and storage infrastructure requirements



Standards

- fostering adoption of standardised file formats, metadata, vocabularies and identifiers by the Australia bioscience community
- bringing Australian needs in standards development, awareness and implementation to the international existing and emerging efforts in this area
- ensuring Australia's input into the development and, where appropriate, coordination of activities in this area



"Bioinformatics is an increasingly international endeavor and building a truly global infrastructure for data requires tight collaboration between partners across continents. EMBL-ABR is well placed to serve the local needs of its users and at the same time act as the reference point for global collaborations in data transfer, exchange, storage, compute and training."

Niklas Blomberg, ELIXIR Director



CONTACT **Assoc Prof Andrew Lonie**
Director, EMBL-ABR and
Node Head, EMBL-ABR: VLSCI Node

ADDRESS Lab-14, 700 Swanston Street
University of Melbourne 3010 Victoria AUSTRALIA
PHONE +61 3 8344 1395
EMAIL alonie@unimelb.edu.au

CONTACT **Assoc Prof Vicky Schneider**
Deputy Director, EMBL-ABR

ADDRESS Lab-14, 700 Swanston Street
University of Melbourne 3010 Victoria AUSTRALIA
PHONE +61 3 8344 9265
EMAIL mv.schneider@unimelb.edu.au



International Science Advisory Group

The ISAG brings together several highly-regarded scientists from around the world, to discuss and advise on strategy, activities and priorities. The ISAG is convened by the Deputy Director and aims to meet face to face once a year at the EMBL-ABR Hub in Melbourne. Following teleconferences in June and October, the Group met in Melbourne over 5-6 December, to learn about the unique and common issues faced by the Australian life science community. At the close of the meeting the Director expressed his gratitude for the dedicated input and advice from such a significant group of national and international experts. Their subsequent report to the Executive is now being incorporated into EMBL-ABR's strategic planning process.

Membership

- EMBL-EBI: Dr Paul Flicek (Senior Scientist and Team Leader, Vertebrate Genomics, European Bioinformatics Institute)
- NIH BD2K/Commons: Dr Vivien Bonazzi (Senior Advisor for Data Science Technologies & Innovation, Office of Associate Director for Data Science, National Institutes of Health)
- CyVerse: Mr Jason Williams (CyVerse Education, Training and Outreach Lead; Assistant Director for External Collaboration, Cold Spring Harbor Laboratories DNA Learning Center, New York)
- ELIXIR: Prof Jaap Heringa (Scientific Director, Netherlands Bioinformatics Centre, and Deputy Head, ELIXIR-Netherlands)
- ABACBS: A/Prof Tony Papenfuss (Head, Computational Biology and Lab Head, Bioinformatics, The Walter and Eliza Hall Institute of Medical Research)
- Australian Expert in Plant Research: Dr Delphine Fleury (Program Leader, Australian Centre for Plant Functional Genomics)
- Australian Expert in Animal Research: Dr Rebecca Johnson (Director, Australian Museum Research Institute, Science and Learning Division)
- Australian Expert in Microbial Research: Prof Mark Walker (Director, Australian Infectious Disease Research Centre, University of Queensland)
- Australian Expert in Medical Research: Prof Sean Grimmond (Director of Cancer Research, University of Melbourne)
- Diversity and Industry Expert: Prof Jenny Martin (Director, Eskitis Institute for Drug Discovery, Griffith University, Queensland).

Heads of Nodes Group

Once established, the HoN plus the Executive team conducted monthly catch-ups to report on activity, issues and to set the agenda and direction of EMBL-ABR. The Executive team regularly consults with the Heads of Nodes and provides updates via a mailing list. All documentation for this Group is accessible via the website.

The Nodes and their activities

Early in 2016 the Deputy Director identified relevant groups with bioinformatics expertise and invited them to submit expressions to become a Node. Ten applications were received, and ten Nodes were approved. They have each undertaken to contribute their expertise and resources to the network, as documented in the following flyers produced in time for the December meetings.

AGRF

AGRF Bioinformatician Sonika Tyagi volunteered to work for EMBL-ABR on Training Coordination and in that role attended the EMBL-ABR booth at the ECCB2016 with Hub personnel, submitted a poster to the November ABACBS conference, and attended the represented the GOBLET AGM. Sonika has now joined the ABACBS Education Sub-Committee and throughout October to December, reviewed and delivered training for AGRF clients incorporating EMBL-ABR best practice standards. AGRF also invited Vicky to present as part of the regular AGRF Seminar Series in November.



Australian National University (ANU)

This Node was established late in 2016 with the generous cooperation of Sylvain Forêt, first Head of Node. Late in 2016 we were all saddened to learn of Sylvain's untimely death. We would like to take this opportunity to thank him for his generous contribution to the work of EMBL-ABR. We also acknowledge ANU's hosting of the Canberra stage of the visit to Australia in late 2015 of Niklas Blomberg, Director, ELIXIR.

James Cook University (JCU)

Work at this Node initially focussed on training and community building through discussion groups (journal club) and seminars (now with local funding to attract interstate visitors) with particular focus on the needs of marine biology, climate and tropical researchers: non-model organisms, genomics in ecology and tropical medicine.

They are now working on strengthening the local support network through implementation of the GVL for training and research and have been granted funds from the Australian Institute of Tropical Health and Medicine to tailor the GVL setup and tool development for their use, in conjunction with the QCIF Node.

Mid-year they ran the first JCU Software Carpentry workshop which was attended by 20 biologists.

Metabolomics Australia (MA)

This Node has already built a range of resources for use by the research community and its role here is to work with EMBL-ABR to extend their reach. Regular workshops and engagement with international data standards groups enables two-way exchanges to ensure MA activity is aligning with international community efforts in metabolomics and Australian researchers are operating at world standards.

This work includes: building a mass spectral database for all detected metabolites in ~50 strains of bacteria that are a major cause of sepsis in hospitals (as part of the Biplatforms Australia Antibiotic Resistant Pathogens Initiative); implementation of the MASTR-MS as an innovative Laboratory Information Management System (LIMS) which has been adopted by other MA Nodes as well as a number of other internationally renowned Metabolomics labs; ongoing testing to apply MASTR-MS as a LIMS for proteomics workflows; further development of MAMBO-MS, a mass spectral library framework that can be used by metabolomics labs to build and populate their in house mass spectral library; publication of PyMS, a tool for rapid alignment and identification of GC/MS peaks and MAR, an R-based statistical analysis tool key for analysing metabolomics datasets.

Monash University

This Node has extended its Instrument Integration work in 2016, connecting gene sequencers to cloud data management (MyTardis); it has established a prototype, user-driven RNAseq pipeline accessing Amazon Web Services' cloud for cost efficiency and speed; created a new one-day 'R Best Practices' course with some of the Shiny interactive web data visualisations delivered at ABACBS / GOBLET meeting in November.

The Node bid successfully for NCRIS RDS 2016/17 Transition Plan funding to integrate instruments and informatics (mostly for imaging technology) to boost its imaging informatics pipeline work whereby nearly 50 instruments are integrated with cloud-based data management software to ensure that data generated by these instruments is automatically captured, managed and can be delivered to the cloud or downloaded to a desktop for analysis. The project has developed a working list of 100 national and institutional instruments across Australia.



Queensland Cyber Infrastructure Foundation (QCIF)

The QCIF Node added BioCyc, a tool requested by researchers to analyse multi-omics datasets from a metabolism perspective, to the GVL; co-developed an online tool to help make decisions about storing human data on research data stores; conducted security and privacy audit in preparation for QRIScloud medical and health data service; and, participated in the investigation of a national collaborative environment for the storage of health and medical research data. Galaxy and RStudio research services were also deployed on the QRIScloud, along with the Galaxy-tut service for training researchers on QRIScloud.

The Node also commenced deployment of metagenomics capabilities for the microbial GVL; carried out data chaperoning on two toxins for two unique Australian spiders; finalised Bioplatforms Australia's BASE and Marine Microbes; and, continued to operate ArachnoServer/ToxNote.

Over 40 training events were held using the GVL, Software Carpentry, Biostatistics with R bootcamp, with successful implementation of online training in Introduction to R and Data Processing with R on GitHub. Materials are in development for proteomics training for Australia's 'Omics platform.

Systems Biology Initiative, University of New South Wales

In collaboration with the Koala Genome Consortium, this Node has assembled a high-quality Koala (marsupial) genome and in collaboration with the Cane Toad Genome Consortium, commenced assembly of the Cane toad hybrid genome. Work follows with a transcriptomic study of behavioural plasticity in the Australian black cricket.

Work also commenced on a project to see if deisotoping of tandem mass spectra, from Q-exactive instruments, can improve protein identification rates and the team produced a PTMOracle plugin for Cytoscape, integrating protein modification data with protein interaction networks (now available on the Cytoscape Apps store).

University of Tasmania (UTAS)

Community building has been a focus for this Node with the local bioinformatics network now connected through a new website/blog and the establishment of a bi-weekly interdisciplinary seminar series. Software Carpentry workshops, Nectar training sessions and new graduate certificates in Bioinformatics and R have added to the training opportunities for Tasmanian researchers and students.

UTAS eResearch has been collaborating with the Tasmanian Nectar Node to provide data storage for human genomes, medical data, and rare datasets generated by local research teams.

University of Western Australia (UWA)

This Node participated in review of Wheat Information System conducted by the International Rice Informatics Consortium; had a paper outlining *B. oleracea* pangenome based on 9 lines accepted in Nature Communications; submitted publication on SUBA4 and another on the Australian bread wheat pangenome.

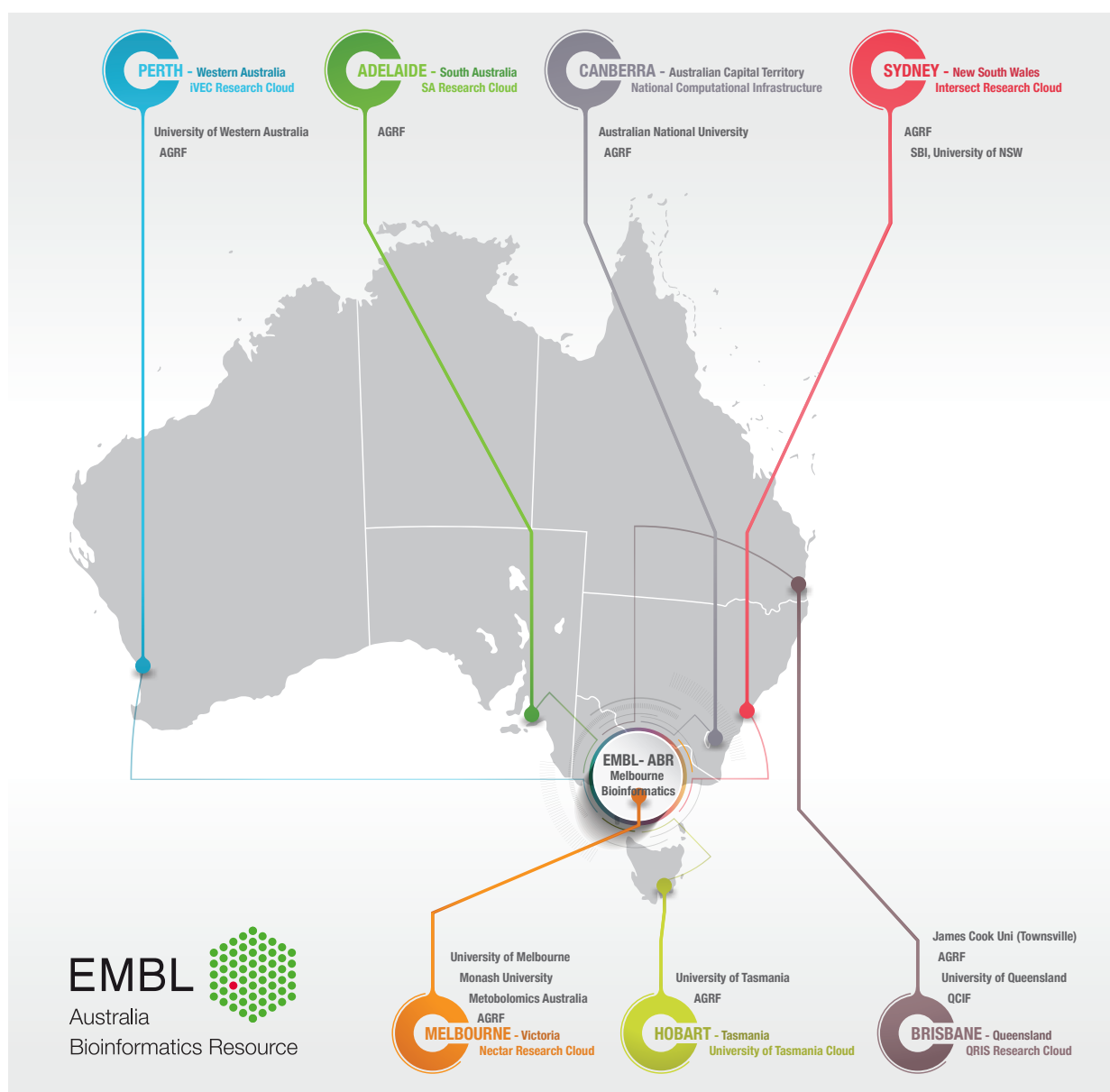
Training was supported through hosting of the first UWA Data Carpentry workshop with UWA Library and COMBINE in October, regular Software Carpentry workshops in conjunction with Curtin Library, and the establishment of a Bioinformatics course for Biotechnology/Biomedicine MSc students.

VLSCI (now Melbourne Bioinformatics)

In addition, work continued managing high-throughput clusters and brokering researcher access to national cloud and data storage / archiving resources; supporting researchers to expose their high-value datasets; leading the development and ongoing roll-out of the Genomics Virtual Laboratory (GVL) as a training and research platform, tailoring it to specific user groups such as microbiologists, plant scientists etc. The GVL was also used to showcase Australian expertise in bioinformatics tools and tool development in overseas forums such as the ELIXIR All Hands meeting.

A range of workshops designed for the bio-medical research community concentrated in the local precinct were delivered and planning commenced for the inaugural Galaxy Australasia Meeting 2017, following the establishment of the Galaxy Australia community (@GalaxyAustralia).

The team also supported the establishment of EMBL-ABR Nodes around Australia; built capacity in significant research programs including the Melbourne Genomics Health Alliance and the Bioplatforms Australia-funded Antibiotic Resistant Pathogens Project; and, contributed expertise to NCRIS-funded infrastructure projects led by Research Data Services (RDS), Australian National Data Services (ANDS), Bioplatforms Australia and EMBL Australia.





Key Areas Coordinators Group

With good will and enthusiasm from the Nodes, champions for the nominated 'Key Areas' volunteered to form the KAGroup to focus and prioritise efforts for impacting Australian biosciences. Established initially by the Deputy Director, the KAGroup then convened monthly meetings to progress efforts to federate activity within Australia, aligned to international efforts in Standards (Saravanan Dayalan), Training (Sonika Tyagi), Data (Philippa Griffin), Compute (Simon Gladman) and International (Vicky Schneider).

On recommendation of the ISAG, the KAGroup will now seek to draw in more community engagement and action through the formation of Special Interest Groups. All the work of this Group is documented and accessible on the website.

Community Leads Group

Mindful that any resource such as EMBL-ABR needs to have a constant dialog with its stakeholders, Vicky Schneider invited several research community leads in key biological areas to help determine where support for biological infrastructure could impact Australia's international profile and engagement. At the end of the year, four projects had been identified for further investigation in 2017:

An adaptable bioinformatics platform for Prokaryotes led by Dieter Bulach, VLSCI

Rapid identification of pathogens by using next generation sequencing techniques is the new gold standard in public health and food safety. This can only be achieved when the right processes and data access arrangements to support the bioinformatics required for annotating each suspect genome are in place.

Such a process involves looking at the genomic sequence, working out where the genes themselves are and what they might do. This is done by comparison with genes of known function (for example, from other bacteria). Similarly, for disease control, bioinformatics is being used to adapt a miniaturised sequencing device to conduct live environmental surveillance, enabling researchers to deliver real-time experimental genetic data for immediate analysis. Such bioinformatics applications can also be used to develop a sensing system by sequencing environmental samples, containing DNA from hundreds or thousands of different organisms.

The ability to set up a tool box and adequate resources for accessibility, storage and exploration of prokaryotes' information relies on a robust bioinformatics infrastructure. This needs to be portable, interoperable and, when it comes to data, accessible through a sustainable and scalable environment for the exploration of 'omics data for particular organisms of interest as well as across different types of data and species.

Australia is already a leader in this field and so it makes sense to highlight and support this area of research and help to build the capability and generate further international interest.

Bioinformaticians at the EMBL-ABR: VLSCI Node have played a significant role in establishing the bioinformatics methods to enable the efficient collection, storage, analysis and publication of these genomic data. This is part of an NCRIS-funded Australia-wide project: the Antibiotic Resistant Pathogens Initiative.

EMBL-ABR has a significant role in the future of increasingly big data driven fields such as microbiology. Through existing projects, we have already brokered access to national computing infrastructure, provided access and training on the GVL (EMBL-ABR: VLSCI Node is part of the team developing this 'life-scientist friendly' analysis platform) and enabled access to a range of large, publicly-available, data sets.



Wheat Annotation Platform led by Rudi Appels, Murdoch University & Gabriel Keeble-Gagnere, Victorian Department of Economic Development, Jobs, Transport and Resources

New genomes are being sequenced at an increasing rate, including genomes of unique and valuable Australian species. However, much of the value of a genome assembly comes from identifying where the important features are: the genes, splice variants, repetitive regions and small RNAs. This process, called ‘annotation’, can be very time-consuming, requiring careful attention. Automated annotation approaches need manual curation and annotation to reach a high standard. The most successful genome annotation projects have typically involved numerous researchers and multiple lines of evidence, including RNA and protein data obtained from the target species, data from related species, and computational gene model predictions based on the DNA sequence.

There is a need for collaborative bioinformatics infrastructure to incorporate multiple sources of information, and support multiple users with the ability to check others’ annotations. The current best-practice approach to this problem is the Apollo software, however, it still requires hosting with sufficient memory and compute resources, which may be inaccessible locally to some Australian researchers. It may also be challenging to set up, especially for biologists without a computing background. This infrastructure-level problem is being solved by providing a ‘pre-packaged’ Apollo setup in the Genome Annotation Platform, accessible to everyone with an Australian research affiliation via the national Nectar Research Cloud.

Based upon Prof Appels’ work on wheat chromosome 7A, as part of the International Wheat Genome Sequencing Consortium, EMBL-ABR is working on the Genome Annotation Platform setup for the Australian research community. This is a collaborative cloud-based platform for Australian researchers (and international collaborators) to annotate a reference genome using one or more lines of evidence. It is set up as an Apollo server running in a custom GVL instance on the Nectar cloud. The first step is a Genome Annotation Platform for the wheat chromosome 7A. This will serve as a pilot project towards other domain and species-specific genome annotation platforms.

Invertebrate ‘Omics led by Kevin Dudley, QUT

Invertebrates - from insects to protozoans - are key to revealing insights in a variety of biological discoveries. Over the last decade, advances in next-generation sequencing, proteomics, metabolomics and other ‘omics, have been applied to ecological, evolutionary and developmental questions. However, for non-model organisms such as invertebrates, there is still a gap in the necessary bioinformatics resources and tools.

For this community, EMBL-ABR is exploring what key activities would be most beneficial in invertebrate ‘omics. Initial efforts will look at best practice in the data life cycle across different biological domains, which we will also extend to invertebrate ‘omics. This will encompass experimental design principles, sample collection, processing procedures, sequencing, assembly and annotation approaches. It will also highlight any gaps and the need to create an annotation and curation environment for invertebrate ‘omics. Future activities will be defined that can establish methodologies to reproducibly deliver high quality reference genomes, proteomes, metabolomes, supplemented with biologically relevant annotation datasets, to researchers working in areas of national significance. Within this area, visualisation of datasets and patterns also plays an important role, and will be included into the focus of this project.

Oz mammals bioinformatics led by the EMBL-ABR Hub

Over 2016, we have been looking at how to maximise the genomics data production of unique Australian mammals, for both the Australian mammals research community and for widening its exposure to international researchers. The Deputy Director investigated the requisite specifications and drafted a proposal with a group of experts in the technical and research aspects towards an Open Bioinformatics platform for Oz mammals. This plan progressed at the ISAG and All Hands meeting and is gearing up in 2017.



Outreach & Community Development

Effective communication is a key enabler for the network and needed at the local, national and international levels. A strategic plan was developed which focused on building relationships across the Nodes and engaging with other stakeholders.

International networks

In addition to Andrew Lonie being invited by Niklas Blomberg, Director, ELIXIR, to deliver a keynote address to the ELIXIR All Hands meeting held in Barcelona in March, he and the Communications Manager, Helen Gardiner, visited EMBL headquarters in Heidelberg, Germany to meet with international counterparts (Silke Schumacher, Director, International Relations and Dan Noyes, Joint Head, Strategy and Communications) and to introduce the new EMBL-ABR concept. Vlcky Schneider also met with Silke Schumacher and Premysl Velek (now Joint Head of Government & EU Relations, EMBL) in the UK in August.

Further, key relationships were pursued with four major international entities and EMBL-ABR was represented at two international conferences.

Galaxy Australia @galaxyaustralia

Capturing the already significant use of Galaxy across Australia, this year EMBL-ABR launched Galaxy Australia and announced its first activity, convening the inaugural Galaxy Australasia Conference (GaME20) for February 2017 to comprise a day of life science researcher training, a two-day conference, and a four-day workshop for Galaxy administrators. Galaxy Australia aims to bring Australian Galaxy users and administrators closer together; identify and address the needs of the community; and, encourage interaction and collaboration.

The Australian-made Genomics Virtual Laboratory

In late June, our community was represented at the Galaxy Community Conference 2016 held in Bloomington, Indiana, USA, where the GVL training materials were reported to have been 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 Principal Investigators (PI) gave their annual talk they singled out the GVL community training materials as demonstrating best practice, with 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.

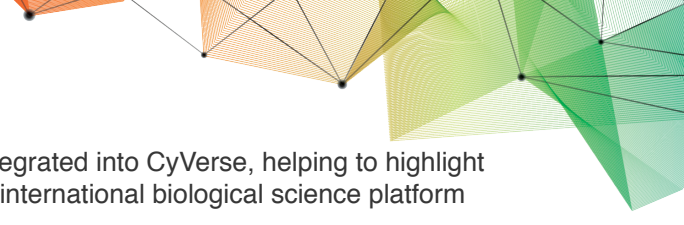
Our strong presence at this conference certainly gave more impetus and further opportunity to connect with key Galaxy people and progress our plans for a more structured Galaxy Australia community network.

CyVerse

Preliminary discussions were held with Jason Williams (Chair, ISAG and Assistant Director for External Collaboration, CyVerse, at Cold Spring Harbor Laboratories) on the viability of establishing CyVerse in Australia. Over 2016 CyVerse and EMBL-ABR discussed ways to co-operate including:

- identifying the data and tool requirements in the Australian biological research community
- understanding how Australian researchers are currently using structured, integrated analysis applications and workflows
- making a roadmap towards democratising Australian science so that research groups lacking capacity to build close working relationships with experts from other disciplines can work on large datasets using publicly available workflows in a single online location with consistent rules, formatting and required skills.

Both groups are now working together to identify how to federate the parts of CyVerse most useful to and demanded by the community in Australia.



In turn, we are investigating which Australian tools can be integrated into CyVerse, helping to highlight Australian research at the global scale and build a common, international biological science platform while preventing duplication of effort and funding.

ELIXIR

EMBL-ABR has pro-actively approached and Vicky regularly engaged with the ELIXIR Hub and some of its Nodes. Throughout August the Deputy Director was hosted by the Directorate of the ELIXIR Hub, working daily with the ELIXIR Chief Technical Officer, Rafael Jimenez on Open Source Policy and working with Frederik Coppens at ELIXIR-Belgium and Steven Newhouse (Head of Technical Services – Technology and Science Integration) on the potential adoption of the Genomics Virtual Lab and its potential for providing workflows as well as optimising download/upload of large datasets for those working in Australia.

Another resources being pursued from this visit is OMICS DI, a dataset repository that can handle multiple datatypes across different 'omics levels. Training opportunities were also pursued with John Hancock (Node Coordinator of ELIXIR-UK) and Pedro Fernandes (ELIXIR-PT).

This was a very productive time spent at the heart of bioinformatics in Europe and the UK and our thanks to the ELIXIR Hub team for making this a very welcoming and productive extended visit.

EMBL-EBI

In the UK, the Deputy Director met with Paul Flicek (ISAG member and Senior Scientist and Team Leader, Vertebrate Genomics, EBI) regarding visualisation of genomes and how Australian researchers can make the best use of EBI's Ensembl. Ensembl's goal is to enable more effective genomics and biological research for all species. Australia's engagement will centre around our unique collections and interesting species as well as more 'traditional' and very active areas of human and mouse genomics.

Further discussions were held with key EMBL-EBI people including Sandra Orchard (Team Leader – Orchard team: Molecular Interactions), Dan Bolser (Ensembl Plants Project Leader – Kersey team: Non-vertebrate genomics), Paul Kersey (Team Leader – Kersey team: Non-vertebrate genomics), tackling issues of metadata and looking at what annotation pipelines would be useful to have federated in Australia. This is regarded as a high-priority as access to these resources would equip Australian life scientists with the tools to annotate and share their data most effectively by adopting best practices already established at EMBL-EBI.

Monthly follow-up calls between Vicky and Paul explored Australia's priorities for EBI services and how best to meet them. This resulted in a collaborative proposal prioritising awareness-raising about the vast resources and services already present at EMBL-EBI and collecting the domain-specific requirements and needs of Australian bioscientists, so that a joint discussion on how to promote these can be most productive.

BIOSHARING

While in the UK, Vicky met with Susanna Sansone, Oxford e-Research Centre Associate Director, and together they launched a collaboration between EMBL-ABR and Biosharing. This will see the EMBL-ABR Hub working with Dr Peter McQuilton, Knowledge Engineer/Senior Research Associate and content lead for the BioSharing project, on Biosharing collections, and with Dr Philippe Rocca-Serra Technical Project Leader on ISA-tools.

GOBLET

A collaboration agreement was made with the Global Organisation for Bioinformatics Learning, Education and Training (GOBLET), to identify common interests and find opportunities to enhance training opportunities which complement existing efforts. In further efforts at community and network building, EMBL-ABR also financially supported the successful bid to have the GOBLET Annual General Meeting held in November in Brisbane as part of the 2016 ABACBS national conference the B3 Symposium, making a festival of bioinformatics: AB3ACBS-2016.



ISMB and ECCB 2016

EMBL-ABR attended two major international bioinformatics conferences during 2016: ISMB in Orlando, Florida, USA; and ECCB in The Hague, The Netherlands. Our booth gave the EMBL-ABR team direct access to conference participants who were surveyed with a set of standard questions. Through this survey, we actively engaged with 250 individual attendees (119 at ECCB and 139 at ISMB).

The survey indicated a general lack of international awareness of Australian bioinformatics research and Australian bioinformatics resources. This is at odds with the representation of Australian science in the published/refereed literature in biosciences, where Australia produces a higher proportion of publications than expected from its population size. Clearly there is more work to do for a concerted effort to highlight Australian bioinformatics internationally and this will be a priority in 2017.

National

Community building around the country was a key driver in 2016, with activities being both formal and informal, and as presentations, meetings and workshops.

In his capacity as Director, Melbourne Bioinformatics, Andrew submitted a Node Expression of Interest to the ISAG for consideration and this was reported to the VLSCI/Melbourne Bioinformatics Advisory Council in May.

Most states were visited by Vicky over April/May, with potential Nodes contacted in Sydney, Hobart, Brisbane, Adelaide and Melbourne. She also met with the University of Western Australia representative at the Australasian Genomic Technologies Association conference in New Zealand.

EMBL-ABR organised and funded a June network meeting for prospective Heads of Nodes. Following the formal Node approval process, the Heads of Nodes group was formed and from October held monthly video calls with the EMBL-ABR Executive team. Similarly, once established, the KAGroup met via monthly video calls.

In July, Andrew represented EMBL-ABR at the EMBL Australia Showcase held at the Australian National University as well as at the EMBL Australia Council meeting.

The data life cycle workshops in October were aimed at all Nodes. Due to a lack of travel funds, participation was less than desired as most wanted to be present at the All Hands meeting in December.

In Melbourne Vicky presented at the March CSIRO/DATA61 Open FOAM meeting and at the November AGRF Special Interest Group Seminar, sharing her insights about EBI and ELIXIR and how EMBL-ABR is working to get access to more resources for Australian researchers and practising bioinformaticians through these connections.

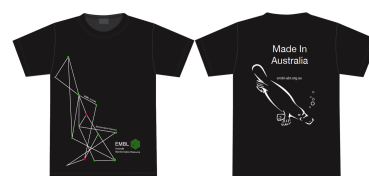
In November Andrew spoke about EMBL-ABR at the Bioinformatics meeting held at Walter & Eliza Hall Institute.

December was a busy month with presentations to and coordination of the face to face ISAG meeting, the All Hands meeting and a briefing for representatives from the Australian Government's Department of Education and Training.

In time for the December meetings, Assoc Prof Tony Papenfuss, President, Australian Bioinformatics and Computational Biology Society (ABACBS), accepted our invitation to join the ISAG. His presence will ensure good communication across our activities and allow us to maximise our collective efforts to raise the profile of the profession and increase resourcing for bioinformatics in Australia.

Conference representation

Hub staff represented EMBL-ABR at three major bioinformatics and genomics conferences in Australasia: ABACBS, BioInfoSummer and AGTA.



Survey of Bioinformatics and Computational needs in Australia 2016

A survey of the bioinformatics and computational biology needs among Australian life scientists and medical researchers was conducted over July to September. In total 123 responses were received from around Australia, distributed across all states. Most respondents were from Victoria (37%), New South Wales (21%), Queensland (18%) and South Australia (15%). Postdoctoral fellows constituted 31% of the respondents, primary investigators 27% and PhD students 26% of the overall respondents. More than 35 different institutions, universities and research institutes were named. This survey was done in collaboration with CyVerse and Cold Spring Harbor Laboratory where the survey was originally developed. The data has been made available to all on the website and will inform further planning in 2017.

All Hands Meeting

The inaugural All-Hands meeting took place on 7 December 2016 and hosted 62 representatives from EMBL-ABR Nodes and many others interested in bioinformatics for life sciences and medical research. The keynotes and Head of Nodes presentations as well as the suggested 2017 priority areas arising from group discussions are available on the website.



EBI's Paul Flicek addressing All Hands Meeting

Showcase

EMBL-ABR hosted a showcase about its vision, mission and current evolution, with participation from national and international members of the ISAG as well as key representatives of Australia sciences and from the University of Melbourne and invited guests from the Australian Government Department of Education and Training.

Hub Resource Development

These activities were driven by the Hub for the network, with support from the EMBL-ABR: Melbourne Bioinformatics Node and the University of Melbourne.



ToolsAU: Australian Bioinformatics Tools

An Australian instance of the ELIXIR Tools Registry as led by ELIXIR-Denmark, has been implemented online to help Australian tool developers make their tools easier to find by researchers around the world. Anyone can now upload their link to the tool with accompanying information, applying some minimum standards to documentation, versioning, availability etc.

Data – ISA Tools and BioSharing

In the process of understanding how aware and prepared bioscientists are for data sharing and adhering to FAIR data principles, the team has liaised directly with ANDS and a series of joint efforts are currently being discussed.

Following interactions and establishment of a collaboration with BioSharing.org, work is now underway to see how EMBL-ABR could adopt and contribute to isatools. Built around the ‘Investigation’ (the project context), ‘Study’ (a unit of research) and ‘Assay’ (analytical measurement) data model and serialisations (tabular, JSON and RDF), the ISA framework helps researchers provide rich descriptions of the experimental metadata so that the resulting data and discoveries are reproducible and reusable. To this end, an EMBL-ABR BioSharing collection has been established and effort is now being placed on how to coordinate the development of ISA standards in experimental design.

Data Indexes such as NIH BD2K’s DataMed and OmicsDI are being explored for possible implementation and adoption within Australia. The Hub team will continue to promote best practice through such activities as the poster presented on *The Data Life Cycle and FAIR Principles* at the ABACBS Conference.

Search for Training Materials: EMBL-ABR STM

An existing customised Google search that creates a registry for training materials was adapted for the EMBL-ABR site by Madison Flannery and it is now collecting and listing existing ‘open’ training materials which have Australian URLs, to gradually provide an overview of Australian training materials. Longer term, these could further inform discussions leading to the collaborative creation of new materials. This, however, will rely on community adoption of standards and FAIR principles.

Standards

EMBL-ABR has initiated discussions around standards through engagement with Biosharing, **isatools** and the Bioschemas Tool Group, aiming to:

- develop a common schema to facilitate description and dissemination of a “tool” as an extension of schema.org
- create a specification including a metadata schema, minimum information guidelines, recommended vocabularies as well as documentation and examples on how to use the tool schema
- engage the community in describing a file science tool and to participate in shaping the specification.

These initiatives are all part of existing international efforts to make data more accessible and transparent, and to enable more transdisciplinary research.

Advertising

Hub staff implemented iAnn: an event-sharing platform for the life sciences. Known Australian bioinformatics training courses and workshops were uploaded onto iAnn and a plug-in located on the EMBL-ABR site. Stage two will see it promoted to Nodes and ABACBS as a plug-in which is able to be implemented on any website, and Hub staff are now arranging for local representatives to have access to input their own training. A list of all courses offered in 2016 was collected on iAnn also collated for the website.

Training & Workshops

Training is a key task for EMBL-ABR, as for the entire life science community. Research was undertaken to establish what training already exists in Australia, who is delivering it, and how might EMBL-ABR complement, support and extend existing training offered by BPA, CSIRO, ABACBS and other institutions. Focus was placed on:

- providing Hub workshops on crucial topics not covered elsewhere in national bioinformatics training (e.g. data life cycle)
- bringing expertise to Australia through visits by international experts and improved access to training resources
- coordinating and disseminating Node end-user training activities
- seeking out distance learning and flexible e-learning resources for bioinformatics collaborative training.



Day 1, Data Life Cycle Best Practice Workshop


Through connections with EMBL-EBI, GOBLET, ELIXIR and CyVerse, international presenters were invited to join with local experts to deliver workshops on subjects not currently offered in Australia. These workshops attracted a total of 90 attendees.

Due to funding constraints, they were all offered in Melbourne. The December events followed the All-Hands meeting, with many interstate visitors taking that opportunity to stay on to attend the workshops.

WORKSHOP	DATE
CyVerse Training Day with Jason Williams	11 April
Data Life Cycle Best Practice Workshop Series (details below)	24-28 October
Hands-on training: RNA-seq data analysis with Dr Eija Korpelainen, CSC-IT Center for Science Ltd, Finland & GOBLET	14-15 November
Bioinformatics Software with Rafael Jimenez, Chief Technical Officer, ELIXIR Hub, UK, and Jason Williams, CyVerse	8 December
Open and Scalable Training Workshop with Rafael Jimenez, ELIXIR Hub, UK	9 December
Workshop on Bioinformatics Registries with Rafael Jimenez, ELIXIR Hub, UK	9 December

Data Life Cycle Best Practice Workshop series

From 24-28 October 2016, the EMBL-ABR Hub ran a week-long Best Practice workshop series focussing on the Data Life Cycle for biological and bioinformatics data. International and local faculty members joined highly engaged participants from around Australia: students, postdocs, lecturers, other scientists, research assistants and research data managers. Five workshops covered the best practice data life cycle for researchers working with plants, animals, non-model, microbial and medically relevant genomes.



During the interactive workshops, participants could take their own data through the best-practice data life cycle following open data and FAIR principles. Specific Australian needs were also identified in the research domains of plants, animals, microbes and health data that will help EMBL-ABR target future efforts and advice. Participants came away with a greatly improved understanding of where to find public research data, the importance of rich metadata in making their own data useful and discoverable, ontologies and controlled vocabularies, and where and how to deposit biological datasets in public repositories. We also explored some of EBI tools that work with data from the primary archives, including the Omics Discovery Index (a user-friendly interface that searches multiple EBI archives, especially useful for multi-omics experiments) and Reactome tools (visualising and analysing biological pathways).

We were particularly grateful to have such high-calibre local and international experts on the Faculty for this series: Sandra Orchard (EMBL-EBI, UK), Jyoti Khadake (Uni Camb, UK), Dan Bolser (Ensembl Plants, UK), Suzanna Lewis (Berkeley Bioinformatics, USA), Andrew Pask (University of Melbourne), Torsten Seemann (VLSCI, University of Melbourne), Ute Roessner (University of Melbourne) and Bernard Pope (VLSCI). Following on from this workshop and on behalf of all the presenters, EMBL-ABR Open Data Coordinator Philippa Griffin coordinated the production of a data life cycle poster for presentation at the AB3ACBS 2016 Conference in Brisbane and is now helping to develop this into a paper for publication in 2017.

Media

The new website, monthly newsletter, social media activities and an engaging interview series drew attention to the aims and objectives of EMBL-ABR nationally and internationally.

Website

A new website was launched in March 2016, built by the Deputy Director and maintained by the Communications team. This could not have been achieved without the in-house expertise of Ben Moran. By the end of the year there had been over 5000 unique visits to the site, with an average of 90% of visits coming from Australia however the site's reach extended to all continents thanks to our international networks and the interest generated by the Interview Series (see below). It is worth noting that most local traffic on the website is dominated by interest in training and workshop opportunities, directly reflecting what we know to be the growing need across the biology community.

Newsletter

From June, a monthly newsletter was launched, aimed at documenting the many activities, initiatives and training opportunities being developed. With 270 subscribers by year-end, the audience extended to key stakeholders and members of the bioinformatics community, including 25% of subscribers from outside Australia. Reaching out to more people through workshops, institutional mailing lists, social media and greater engagement from the Nodes, subscription numbers are expected to continue to grow. The newsletter maintains a 50% open rate, which is a very positive result and reflects the relevance of the material to the subscribers.

Social media

The @EMBL_ABR twitter channel was launched in March to develop awareness of Australian bioinformatics across our extensive, global bioinformatics and biosciences infrastructure networks. Operated by both the Communications team and the Deputy Director, it had over 600 followers at the end of 2016.

Interview Series

During the year, we contacted Australian and international expert bioinformaticians and asked them six questions about bioinformatics today, as part of an extensive interview series for the website. The variety of their responses reflects the diversity of our community and the challenges we face. It forms a helpful resource for the community.

Director profiles

In April 2014 a feature article on *Why big data is a big deal*, was published in the University of Melbourne's online magazine *Pursuit*. Featuring both Andrew Lonie and Vicky Schneider, the article was written by science communicator, Dr Andi Horvath.




Vicky was also interviewed on the growing demand for skills in bioinformatics, as part of a *Spotlight on Bioinformatics* feature for *NatureJobs*:

"More dialogue between users and developers results in better tools," she says. For instance, a computer science-trained "developer might create a powerful tool. But if someone with a biology background doesn't know how to use it, that tool is useless. The two sides need to work together to develop user interfaces."

EMBL-ABR People

Our thanks to the following people from across Australia, and including the Nodes, who have generously contributed their time and resources to help build the network as it stood at the end of 2016. Included here also are any international colleagues who contributed to training and/or community meetings:






<i>Name</i>	<i>Role and/or expertise</i>	<i>Institution</i>	<i>EMBL-ABR Role</i>
Delphine Fleury	Plant Genomics	ACPFG	Presenter, All Hands meeting Wheat Genome Annotation Platform
Nathan Watson-Haigh	Bioinformatics	ACPFG	Key Area coordinator, Tools Wheat Genome Annotation Platform Interview subject All Hands meeting
Radoslaw Suchecki	Bioinformatics	ACPFG	All Hands meeting
Ute Baumann	Transcriptomics and Bioinformatics	ACPFG	Wheat Genome Annotation Platform
Lavinia Gordon	Bioinformatics	AGRF	Interview subject Federal Government briefing All Hands meeting
Sonika Tyagi	Bioinformatics	AGRF	Head of Node Key Area coordinator, Training June network meeting ECCB booth All Hands meeting December workshops
Andrew Treloar	Director, Technology	ANDS	All Hands meeting
Angeletta Leggio	Partnerships	ANDS	All Hands meeting
Kate LeMay	Research Data Specialist	ANDS	Interview subject All Hands meeting
Ross Wilkinson	Executive Director	ANDS	All Hands meeting
Dan Andrews	Bioinformatician	ANU	Acting Head of Node
Sylvain Forêt	Bioinformatician	ANU	Head of Node All Hands meeting
Rebecca Johnson	Australian animal genomics	Australian Museum	Presenter, All Hands meeting
Saravanan Dayalan	Bioinformatics, Biostatistics	Metabolomics Australia	Head of Node Key Area coordinator, Standards All Hands meeting
Andrew Gilbert	Director	Bioplatforms Australia	Funding body Federal Government briefing
Jason Williams	Assistant Director, DNA Learning Centre	Cold Spring Harbor Laboratory	December workshops All Hands meeting Federal Government briefing Dec network meeting ISAG meetings
Angus Macoustra	Deputy CIO, Scientific Computing	CSIRO	All Hands meeting
Annette McGrath	Data 61	CSIRO	June network meeting All Hands meeting
Jaap Heringa	Bioinformatics	Dutch Centre for Life Sciences	All Hands meeting
Maree Agiazis	Chief Operating Officer	EMBL Australia	All Hands meeting
Rafael Jimenez	Chief Technical Officer	ELIXIR Hub, UK	December workshops



<i>Name</i>	<i>Role and/or expertise</i>	<i>Institution</i>	<i>EMBL-ABR Role</i>
Paul Flicek	Senior Scientist & Team Leader	EMBL-EBI	All Hands meeting
Geraldine Kong	Biology	Florey	All Hands meeting
Vinh Nguyen	Biology student	La Trobe	All Hands meeting
Ira Cooke	Comparative Genomics Centre	JCU	Head of Node All Hands meeting
David Powell	Bioinformatics	Monash	Acting Head of Node All Hands meeting
Steven Androulakis	Monash Bioinformatics Platform	Monash	Head of Node June network meeting All Hands meeting
Xiaochuan Wang	Bioinformatics	Monash	All Hands meeting
Rudi Appels	School of Veterinary and Life Sciences Research	Murdoch University	Community Lead, Wheat Genome Annotation Platform Interview subject
Glenn Moloney	Director	Nectar	All Hands meeting
Michelle Barker	Deputy Director	Nectar	All Hands meeting
Vivien Bonazzi	Senior Advisor, BD2K	NIH	All Hands meeting December network meeting Federal Government briefing
Maria Doyle	Bioinformatics	PeterMac	All Hands meeting
Kevin Dudley	Bioscience	QUT	Community Lead, Invertebrate 'Omics June network meeting
David Lynn	EMBL Australia Group Leader in Biomedical Informatics	SAHMRI	June network meeting
Marc Wilkins	Director	SBI	Head of Node All Hands meeting
David Adelson	Center for Bioinformatics and Computational Biology	UAdelaide	June network meeting
Aditya Kala	Biology	UniMelb	All Hands meeting
Andrew Isaac	Platforms Lead, VLSCI	UniMelb	All Hands meeting
Andrew Lonie	VLSCI (Melbourne Bioinformatics)	UniMelb	Director, EMBL-ABR Head of Node All meetings
Andrew Pask	Pask Lab	UniMelb	Federal Government briefing Presenter, Data Life Cycle workshop
Anna Syme	Bioinformatics	UniMelb	All Hands meeting
Ben Moran	Systems Engineer, VLSCI	UniMelb	Web development
Bernard Pope	Bioinformatics, VLSCI	UniMelb	All Hands meeting
Christina Hall	EMBL-ABR, VLSCI	UniMelb	Communications and Training
Daniel Park	Bioinformatics	UniMelb	All Hands meeting



<i>Name</i>	<i>Role and/or expertise</i>	<i>Institution</i>	<i>EMBL-ABR Role</i>
Dieter Bulach	Bioinformatics	UniMelb	Community Lead Interview subject
Felicia Burtcher	Bioinformatics	UniMelb	All Hands meeting
Fiona Kerr	EMBL-ABR, VLSCI	UniMelb	Executive Officer
Harriet Dashnow	COMBINE Melbourne	UniMelb	All Hands meeting
Helen Gardiner	EMBL-ABR, VLSCI	UniMelb	Media and Communications
James McCluskey	DVC Research	UniMelb	Funding Body Federal Government briefing
Khalid Mahmood	Bioinformatics	UniMelb	All Hands meeting
Michael Milton	Software developer, VLSCI	UniMelb	All Hands meeting
Gayle Philip	Bioinformatics, VLSCI	UniMelb	All Hands meeting
Liz Sonenberg	Department of Computing and Information Systems	UniMelb	Funding Body Federal Government briefing
Madison Flannery	Computer Programmer, VLSCI	UniMelb	Search for Training Materials initiative Wheat Genome Annotation Platform
Malcolm McConville	Director, Bio21 Institute	UniMelb	All Hands meeting
Matthew Wakefield	Bioninformatics, VLSCI & WEHI	UniMelb	All hands meeting Staff photographer
Michael Kuiper	Molecular Modelling & Dynamics, VLSCI	UniMelb	All Hands meeting
Neil Young	Bioinformatics	UniMelb	All Hands meeting
Pasi Korhonen	Bioinformatics	UniMelb	All Hands meeting
Peter Georgeson	Bioinformatician	UniMelb	All Hands meeting
Philippa Griffin	EMBL-ABR	UniMelb	Open Data Coordinator Co-convenor and presenter, Data Life Cycle workshop All Hands meeting
Saravanan Dayalan	Metabolomics Australia	UniMelb	Head of Node
Sean Grimmond	Cancer Medicine	UniMelb	Dec ISAG meeting All Hands meeting
Simon Gladman	Specialist Programmer, VLSCI	UniMelb	GVL, MicroGVL Wheat Genome Annotation Platform Key Area coordinator, Compute All Hands meeting
Steven Manos	Director, Research Platform Services	UniMelb	All Hands meeting
Tony Bacic	Metabolomics Australia	UniMelb	Federal Government briefing All Hands meeting
Torsten Seemann	Melbourne Bioinformatics & Peter Doherty Institute	UniMelb	MicroGVL All Hands meeting Presenter, Data Life Cycle workshop



<i>Name</i>	<i>Role and/or expertise</i>	<i>Institution</i>	<i>EMBL-ABR Role</i>
Ute Roessner	Metabolomics Australia	UniMelb	Presenter, Data Life Cycle workshop
Vicky Schneider	EMBL-ABR	UniMelb	Deputy Director All meetings, workshops Co-convenor and presenter, Data Life Cycle workshop
William Ho	Bioinformatics	UniMelb	All Hands meeting
Ye Tian	Student	UniMelb	All Hands meeting
Marc Wilkins	Systems Biology Initiative	UNSW	Head of Node June network meeting Interview subject All Hands meeting
Richard Edwards	Edwards Lab Research	UNSW	Dec network meeting Interview subject
Dominique Gorse	QFAB/QCIF	UQ	Head of Node Dec network meeting All Hands meeting
Mark Ragan	Genomics of Development and Disease Division	UQ	Interview subject
Mark Walker	Microbial Genomics	UQ	Presenter, All Hands meeting
Rob Cook	Director, QCIF	UQ	Acting Head of Node All Hands meeting
Cali Willet	Sydney Informatics Core Research Facility	USyd	All Hands meeting
Jonathan Arthur	Bioinformatics, Children's Medical Research Institute	USyd	Node discussions
Jac Charlesworth	Bioinformatics, Menzies	UTas	Acting Head of Node Interview subject
Michael Charleston	Bioinformatics	UTas	Head of Node June network meeting All Hands meeting
David Edwards	Plant Biology	UWA	Federal Government briefing All Hands meeting Head of Node
Philipp Bayer	Plant Biology	UWA	June network meeting All Hands meeting Acting Head of Node
Gabriel Keeble-Gagnere	AgriBio	Victorian Government DEDJTR	Community Lead, Wheat Genome Annotation Platform
Matthew Ritchie	Bioinformatics	WEHI	All Hands meeting
Tony Papenfuss	ABACBS and Computational Biology	WEHI	ISAG



Diversity and Gender

EMBL-ABR pro-actively manages gender balance when planning activities and this is reflected in Committee membership and role allocations. A Diversity and Gender draft proposed action plan is in place for when EMBL-ABR's governance requires it. Until then, each Node's hosting institution is bound by its own policies in this area.

Glossary

ABACBS	Australian Bioinformatics and Computational Biology Society
ACPGF	Australian Centre for Plant Functional Genomics
AGTA	Australasian Genomic Technologies Association
AGRF	Australian Genome Research Facility Ltd
AMSI	Australian Mathematical Sciences Institute
ANDS	Australian National Data Services
ANU	Australian National University
BD2K	Big Data to Knowledge
BCLGroup	Biosciences Community Leads Group
Bio21	Bio21 Institute
BPA	Bioplatforms Australia
BRAEMBL	Bioinformatics Resource Australia, EMBL
DEDTJR	Victorian Government Department of Economic Development, Transport, Jobs & Resources
ECCB	European Conference on Computational Biology
EMBL	European Molecular Biology Laboratory
EMBL-ABR	EMBL Australia Bioinformatics Resource
EMBL-EBI	EMBL European Bioinformatics Institute
FLOREY	Florey Institute of Neuroscience and Mental Health
GOBLET	Global Organisation for Bioinformatics Learning, Education & Training
GVL	Genomics Virtual Laboratory
HoN	Heads of Nodes Group
ISAG	International Scientific Advisory Group
JCU	James Cook University
KAGroup	Key Areas Coordinators Group
La Trobe	La Trobe University
MA	Metabolomics Australia
MDU PHL	Microbiological Diagnostic Unit Public Health Laboratory
MelBioInf	Melbourne Bioinformatics
Monash	Monash University
NCRIS	National Collaborative Research Infrastructure Strategy
NIH	National Institutes of Health
QCIF	Queensland Cyber Infrastructure Foundation
QFAB	Queensland Facility for Advanced Bioinformatics
QUT	Queensland University of Technology
RDS	Research Data Services
SBI, UNSW	Systems Biology Initiative, The University of New South Wales
SAHMRI	South Australian Health and Medical Research Institute
SIGs	Special Interest Groups
UAdelaide	University of Adelaide
UCSD	University of California, San Diego
UTas	University of Tasmania
UQ	University of Queensland
UWA	University of Western Australia
VLSCI	Victorian Life Sciences Computation Initiative
WEHI	Walter & Eliza Hall Institute of Medical Research
UniMelb	University of Melbourne

Appendix: Node profiles

EMBL-ABR: **AGRF NODE**

Enabler of world-class genomic science

EMBL



Australia

Bioinformatics Resource

The Australian Genome Research Facility is involved in biological data generation and research programs involving Australia animals, crop-plants, bacterial, fungi and corals, soil microbiomes and melanoma sequencing in collaboration with Bioplatforms Australia.

As an enabler of world-class genomic science through the provision of genomics technology and services, new data integration platform development initiatives are already in progress. AGRF has an existing national reach in Australia, allowing rapid dissemination and uptake of new resources and information, and strengthening the EMBL-ABR network.

Training



AGRF will develop training activities targeted to life scientists working on non model organisms. These interactive materials will be available online and will also form part of EMBL-ABR's 'Training & Learning' program.



Tools



AGRF will share databases, tools and pipelines developed via active collaborations. As well as exchanging tailored pipelines and specific tools AGRF will work with EMBL-ABR to nurture a sustainable environment for a variety of tools and pipelines produced in Australia.

"Life science is data science. I think we have to be open to a broader definition of data science. Biology is very diverse - a doctor, a field ecologist, or a geneticist may have a different need for data science. Their work may require working in the data science domain on a problem of searchability, metadata, or visualisation. Since no one can be an expert in everything, a major task is to make challenges clear. Life Scientists need clear standards and uniform processes so they have 'guide books' that demarcate what types of analyses are appropriate, how to evaluate the quality of data and analyses, and where pitfalls are located."

Mr Jason Williams: Assistant Director, External Collaborations, Cold Spring Harbor Laboratory's DNA Learning Center and Education, Outreach, and Training Lead, CyVerse – the U.S. National Cyber infrastructure for Life Science.

Compute



AGRF will improve data transfer for researchers by pushing the data directly to a high performance computing environment. The workflows developed to reduce data transaction time and human error by accessing high performance compute for whole genome level studies can be adopted by EMBL-ABR nodes around Australia.

CONTACT

Dr Sonika Tyagi

Bioinformatics Supervisor
Australian Genome Research Facility Ltd

ADDRESS

Suite 219, 55 Flemington Road
North Melbourne VIC 3051

PHONE

+61 3 9345 2664

EMAIL

Sonika.tyagi@agrif.org.au



EMBL-ABR: ANU NODE

A high-end research computing service



The National Computational Infrastructure, located at the Australian National University (ANU), is a high-end research computing service, supported by a highly-integrated and expert e-infrastructure environment. The EMBL-ABR: ANU Node has four main domains of expertise: evolutionary genomics, functional genomics, high performance computing and medical genomics. Future activity will also focus on providing bioinformatics computational capability for the analysis of non-model organisms, particularly for the unique fauna and flora of Australia. The bioinformatics and molecular tools developed here and the unique datasets collected and analysed at the ANU Node will benefit the EMBL-ABR community and raise the profile of Australian biological research on the international stage. Training in genetics, genomics, bioinformatics and programming will be consolidated by and made more accessible through participation in the EMBL-ABR network.

Data & Tools & Platforms



The ANU Node has produced large genomic and transcriptomic datasets with some accompanying functional genomics datasets in:

- corals
- insects
- Eucalyptus and other native plants
- reptiles and mammals
- medical genomics



An extensive range of bioinformatics tools have been developed:

- pipelines for whole genome assembly and annotations
- tools for population genomics and landscape genomics
- molecular and bioinformatics methods for handling the genome and transcriptomes of symbiotic species
- tools for the analysis of epigenomic data.
- a tool suite for the analysis of large amounts of human genomic data deployed on a high performance computing platform.



As part of the EMBL-ABR network, these valuable resources will be extended to include the current in-house methods for the analysis of non-model organisms. Further platform and tool development will make unique Australian datasets more easily accessible and visible internationally.



"The future of biological research will be built, in part, on the foundation of strong bioinformatics. A coordinated, national approach will connect to international efforts while understanding and serving the unique needs of Australian science."

Paul Flicek Senior
Scientist/Team
Leader, Vertebrate
Genomics, EMBL-EBI

Training



The ANU Node will provide training in basic programming, and on best practices in the assembly and annotation of genomes and transcriptomes of non-model organisms. This will empower researchers working on many of Australia's endemic plant and animal species with the tools that they need to start exploring their data, and to have more fluent interactions with bioinformaticians. These activities could be integrated workshops, discussion and work groups, and would produce shareable training resources and may result in the publication of review papers on best practices.

CONTACT **Dr Thomas Daniel (Dan) Andrews**
Bioinformatics Fellow
Australian National University
ADDRESS 46 Sullivans Creek Road
Acton ACT 2601
PHONE +61 2 6125 2740
EMAIL dan.andrews@anu.edu.au



**Australian
National
University**

EMBL-ABR: JCU NODE

Unique bioinformatics strengths



James Cook University is a leader in Tropical Health and Medicine, Marine Science, Aquaculture and Tropical Ecosystems. Unique bioinformatics strengths have built around diverse research interests such as non-model organism research including corals and tropical aquaculture species, genomics and proteomics of tropical parasites and pathogens, and diseases that disproportionately affect Aboriginal and Torres Strait Islanders.

The JCU Node will provide training in big data analysis that cuts across these research areas. The node will provide better mechanisms for sharing this knowledge, both in North Queensland through regular local events, and nationally by providing specialist input to national projects, and by developing software tools and training materials with a particular focus on these specialties. The JCU Node is ideally positioned to engage with the bioinformatics communities of countries in South East Asia and Oceania. As one of the world's leading tropical universities, JCU will contribute unique tropical, marine and non-model organism expertise and resources to the EMBL-ABR network.

Training



Regular bioinformatics workshops at the JCU Node will cover a basic skills such as those covered by Data and Software Carpentry workshops, as well as specialist workshops on a range of topics including Epigenetics, Phylogenomics, Population genomics, Transcriptomics and Metagenomics. The JCU Node will work with the rest of the network to federate training materials, and training activities will be open to all Australian researchers.

Platforms



Genomics Virtual Lab instances will be built with tools required for coral genomics and metagenomics, proteomics, non-model organism transcriptomics that will specifically suit local training and research requirements. This will involve developing new tools for Galaxy and involve providing access to indexed genomic resources. The initial focus will be on developing and documenting GVL flavours for undergraduate bioinformatics training and metagenomics, which will be available to researchers throughout Australia.



JCU is working on the genome of the soft-coral *Lobophytum pauciflorum*. Image provided by recent JCU PhD graduate, Dr Wiebke Wessels.

"Data driven science has expanded sharply in the last ten years. Established life scientists often discover that they are missing an important set of skills and competencies in this area. It is a fantastic opportunity for training providers, worldwide. Barriers that we see in the use of data and computational resources need to be lowered. And we need to orient infrastructures globally towards greater standardisation of datasets and their annotation, to be made discoverable by placing them in coherent public repositories."

Pedro L. Fernandes Coordinator, Gulbenkian Training Programme in Bioinformatics, Portugal

CONTACT **Dr Ira Cooke**
Senior Lecturer
James Cook University
ADDRESS 1 James Cook Drive
Townsville QLD 4811
PHONE +61 7 4781 4354
EMAIL ira.cooke@jcu.edu.au



EMBL-ABR: MA NODE

Enabler of world-class genomic science



Metabolomics Australia (MA) is a research and service delivery national platform providing a range of advanced metabolomics services to researchers. Experts distributed nationally across five nodes have developed scientific tools, scientific databases and data management software platforms that have been adopted by the MA network and by international metabolomics facilities. The MA Node will complement the EMBL-ABR network by making these resources easily accessible, while continuing to provide specialist training workshops for researchers and students and contributing to metabolomics data standardisation through its collaborations with international partners.

Data



The MA Node generates scientific databases, including a current project to generate mass spectral database for metabolites of sepsis-causing bacteria. Utilising applied software developed in-house, the curated mass spectra database will be linked to genomic, transcriptomic and proteomic data collected on the same bacterial pathogens, providing a unique resource for researchers. This publicly available data will lead to better understanding of these pathogens by biological researchers and clinicians.



Tools



The MA Node has developed a number of software tools, including PyMS, for rapid alignment and identification of GC/MS peaks, and MAR, an R-based statistical analysis tool key for analyzing metabolomics datasets. These tools are open source and the MA Node will work on providing them for broader use via workflow solutions like Galaxy and Knime. In addition to assisting with their dissemination, EMBL-ABR will support the long term hosting of these tools.

Platforms



The MA Node will work with collaborators to incorporate international standards into tailored platforms for the metabolomic and proteomic research communities. The metabolomics-specific laboratory information system, MASTR-MS, was developed by MA and MASTR-MS is currently being tested for use with proteomics workflows. MAMBO-MS is a mass spectral library building and populate mass spectral libraries. Interaction with the EMBL-ABR network to disseminate and in the long term host these resources will support their development and accessibility.

Training

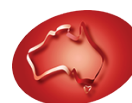


The MA Node will provide training workshops on metabolomics data processing and statistical analysis in conjunction with the EMBL-ABR network.

The metabolome is "equally, if not more, important that the human genome" for capturing "the fingerprint of human disease."
(Harvard Magazine)

Robert Gerszten,
Chief of Cardiology at
Beth Israel Deaconess
Medical Center,
Professor of Medicine
at Harvard Medical
School, Senior
Associate Member of
the Broad Institute.

CONTACT **Prof. Malcolm McConville**
Head of Bioinformatics
Metabolomics Australia
ADDRESS The University of Melbourne
Melbourne VIC 3010
PHONE +61 3 8344 2201
EMAIL malcolmm@unimelb.edu.au



METABOLOMICS
AUSTRALIA

EMBL-ABR: **MONASH NODE**

A leading centre of activity in bioinformatics

EMBL

Australia

Bioinformatics Resource



Monash University is a leading centre of activity in bioinformatics. Embedded in the Clayton Innovation Precinct, Monash holds key expertise in genomics, proteomics and imaging informatics, develops bioinformatics tools and platforms for Australian researchers, and is a leading participant in national and international bioinformatics initiatives. The Monash bioinformatics community is led, coordinated and underpinned by the Monash Bioinformatics Platform and the Monash Node will contribute high level training activities, sophisticated imaging and instrument platforms and high performance compute to the EMBL-ABR network.

Training



The Monash Node will improve bioinformatics practice through establishing best practice around data analysis lifecycles, and tool and training material development. The development of guidelines relating to community best-practices, technical leadership, implementation of infrastructure, development and delivery of training will benefit bioinformaticians and biologists facing the challenges of data intensive science.

Platforms



The ongoing coordinated deployment of the Characterisation Virtual Laboratory (CVL) will be facilitated by membership of the EMBL-ABR network. The CVL integrates imaging instruments with cloud-based computing and data storage, provides online environments for data analysis and visualisation, and makes a wide variety of data processing tools and techniques more easily accessible to a wide cohort of imaging users.

Compute



A share of the MASSIVE facility will be available to the bioinformatics community through the Monash Node. MASSIVE is a unique Australian facility with a focus on fast data processing, including processing data in-experiment, large-scale visualisation, and analysis of large-cohort and longitudinal research studies.



"Growing training and infrastructure together is vital. The measure of success will be when your work is completely transparent – the tools, training and infrastructure you have built is enabling users to do high quality life science research, exchanging data, interacting seamlessly across continents, knowing what agreed standards people require of data so that it is accessible to all. Cloud computing will obviously be instrumental in this too. All the effort needs to come from those of us working behind-the-scenes to build working collaborations and master new technologies. And high quality training for users is essential to bring this all together."

Dr John Hancock ELIXIR-UK Node Coordinator,
The Genome Analysis Centre, Norwich, UK

CONTACT **Steve Androulakis**
Manager, Monash Bioinformatics Platform
Monash University

ADDRESS 23 Innovation Walk, Monash University
Wellington Road, Clayton VIC 3800

PHONE +61 3 9902 9318

EMAIL steve.androulakis@monash.edu



MONASH
University

EMBL-ABR: QCIF NODE

Expert consulting and services in bioinformatics



QCIF is a consortium of seven leading Queensland universities and includes QFAB, the Queensland Facility for Advanced Bioinformatics. QCIF operates significant HPC, cloud computing and data storage resources, is part of the national eResearch infrastructure, offers a wide range of services across a broad spectrum of research domains and undertakes projects of local and national significance. QFAB is an established and mature centre for computational life sciences expertise, providing expert consulting and services in bioinformatics to life sciences and health researchers throughout Queensland and beyond. The QCIF Node will collaborate nationally through its EMBL-ABR data chaperoning services and continued partnering in the development and operation of the Genomics Virtual Lab.

Compute



The QCIF Node provides extensive computational facilities using high performance clusters, large memory nodes and Cloud resources to support bioinformatics data, tools and platforms. All researchers across Australia can access resources via QRIScompute. QRIScompute facilitates access to Nectar Cloud compute infrastructure including Euramoo - a cloud-based computer cluster built on the Nectar Cloud. FlashLite is a research computer that has been designed and optimised for data intensive computing, in particular for life science computing.

Data



The QCIF Node facilitates a range of services including tailored data chaperoning and solutions for processing, managing, sharing and archiving of high value data. Hosting and mirroring of national and international reference data collections such as the 1000 Genomes and the Bioplatforms Australia collections is provided.

Platforms

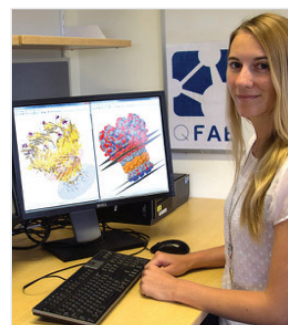


The QCIF Node is part of a national collaborative effort to develop a platform for multi-omics analysis and interpretation that will initially support Bioplatforms Australia's Antibiotic Resistance Pathogens Initiative. The QCIF Node is involved in the nationally funded eResearch infrastructure, med.data, which facilitate the discovery, storage and sharing of health datasets. The QCIF Node partners in the development of the Genomics Virtual lab (GVL) and provides user access and support throughout Queensland.

Training



Empowering life science and health researchers to get the most out of their research data, the QCIF Node provides a wide range of training options to up-skill students and researchers in digital infrastructure usage, and in best practice data analysis in bioinformatics, biostatistics and health informatics.



"Our biology colleagues want to be able to access the most important, common and critical tools without having to worry about what computer they are logged into. The challenge is putting all these required techniques together for individual biologists, or forming teams of biologists to do things using informatics, computational and data skills. Bioscience is a data science now and the biggest issue is training and skills."

Mark Ragan Professor and Co-head, Genomics of Disease and Development, Institute for Molecular Bioscience, and Professor (Adjunct), School of ITEE, The University of Queensland

CONTACT **Rob Cook**
CEO, Queensland Cyber Infrastructure Foundation Ltd

ADDRESS Axon Building 47, The University of Queensland
St Lucia QLD 4072

PHONE +61 7 3365 7274

EMAIL rob.cook@qcif.edu.au



EMBL-ABR: SBI NODE

A centre for molecular systems biology

EMBL

Australia

Bioinformatics Resource



The Systems Biology Initiative is a centre for molecular systems biology located at the University of New South Wales. It applies its expertise in 'omics technology and affiliated bioinformatics to address biological questions at a systems level. It has particular expertise in the use of next-generation sequencing for genomics and transcriptomics, and in the use of mass spectrometry for proteomic investigations. It has extensive experience in the generation and analysis of intracellular networks and their use in the contextual analysis of 'omics data. As part of the EMBL-ABR network, the SBI Node would support researchers using next-generation sequencing (NGS), microarray and/or mass spectrometry technology to answer systems-level questions in molecular biology.

Data



The SBI Node has expertise in the analysis of transcriptomics data, and in the assembly, QC and annotation of small to large genomes. This includes the submission of data to repositories such as GEO, SRA and GenBank. In proteomics, the SBI has a particular focus on analysis of post-translational modifications, including submission of data to ProteomeXchange, and PRIDE. This expertise will be accessible to the EMBL-ABR network, collaborators and an increasingly international audience.

Training



Staff at the SBI Node develop and teach hands-on workshops in NGS data analysis, with a focus on emerging technologies including long read sequencing. Being part of the EMBL-ABR will share skills and expertise and foster international best practice.

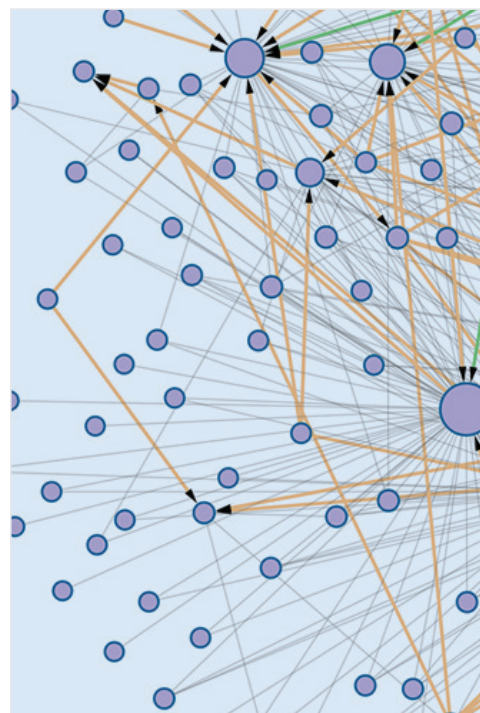
Tools



The SBI Node has a suite of resources available to independent users, including recently developed tools such as:

- PG Nexus for proteogenomics
- PTMOracle Cytoscape app for functional analysis of protein modification in protein interaction networks
- SLiMSuite for the identification of short linear motifs in protein interaction networks.

Participation in EMBL-ABR will assist in the dissemination of these tools and inform further development.



"There is value in the SBI being part of a larger, formal structure. Being part of the EMBL-ABR will facilitate bidirectional flow of skills and expertise, help establish and spread cutting edge approaches and help establish and maintain international best practice."

Marc Wilkins: Professor and Director of the New South Wales Systems Biology Initiative and Chair of Systems Biology in the School of Biotechnology and Biomolecular Sciences, University of New South Wales

CONTACT

Prof Marc Wilkins

Director Systems Biology Initiative,
Director Ramaciotti Centre for Genomics

ADDRESS

University of New South Wales
Kensington NSW 2052

PHONE

+61 2 9385 3633

EMAIL

m.wilkins@unsw.edu.au



SYSTEMS BIOLOGY
INITIATIVE

EMBL-ABR: **UTAS NODE**

Representing genomics diversity



The University of Tasmania is host to richly diverse and truly interdisciplinary research. UTAS bioinformaticians represent a broad range of research interests based on data of national and global importance - from the impact of climate change on rare species through to vaccine discovery for Devil Facial Tumour Disease. The UTAS bioinformatics research community frequently works on uncommon informatics problems, including a high number of local plant and animal species, often with no reference genome. UTAS has made significant investments into establishing and developing computational infrastructure and services for bioinformatics training and research that will, along with its unique expertise and data, strengthen the EMBL-ABR network.

Training



The UTAS Node will facilitate new bioinformatics workshops covering a range of topics. Requested content is likely to range from basic to advanced tutorials on subjects including RNAseq and NGS, working without a reference genome, building a reference genome and setting up your own GVL environment in NeCTAR.

Data



The UTAS Node will develop protocols for best practice in research data management. Decision trees for suitable data storage, privacy and management based on data type, funding sources and other considerations will be generated through collaborations with the EMBL-ABR network and new international interactions.

Platforms



The UTAS Node will configure a targeted GVL instance for unique training and research needs. Local research encompassing a wide range of organisms and data types requires tailoring cloud-based platforms to meet a range of needs. Through the documentation of the process of set up and tailoring of the GVL instance to the needs of plant science, Antarctic science, maths, physics, zoology and health, UTAS will contribute to the GVL project.



UTAS researchers are investigating Tasmanian devil facial tumour disease. Image: Greg Woods

"We need to build a bioinformatics community in Australia. A network that gives bioinformaticians a voice – both to talk to each other and build collaborative ties, but also a voice to funding bodies to let them know that bioinformatics is a field of scientific research in its own right, not a support science."

Jac Charlesworth: Senior Research Fellow, Computational Genomics, Menzies Institute for Medical Research, The University of Tasmania

CONTACT **Dr Jac Charlesworth**
Senior Research Fellow – Computational Genomics
Menzies Institute for Medical Research, University of Tasmania

ADDRESS Private Bag 23
Hobart TAS 7001

PHONE +61 3 6226 4607

EMAIL Jac.Charlesworth@utas.edu.au



EMBL-ABR: UWA NODE

An international reputation for excellence

EMBL



Australia

Bioinformatics Resource

The University of Western Australia (UWA) is one of Australia's leading universities and has an international reputation for excellence in teaching, learning and research. It is home to a number of industry-based research centres and institutes, and is affiliated with many others across the country. The EMBL-ABR: UWA Node includes research groups with valuable experience and knowledge of complex plant genomes not found anywhere else in Australia. This Node is vital to several international plant genome assembly efforts involving model species such as *Arabidopsis* and crop species such as wheat. Knowledge, tools, platforms and data from groups within the UWA Node have contributed significantly to collaborative international scientific efforts and will broaden and strengthen the reach and relevance of the EMBL-ABR network.

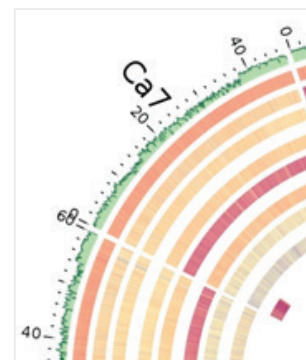
Data & Tools



The UWA Node is engaged in a diverse range of projects that collect, curate and host data while providing powerful tools specifically developed to enable interaction with the publically available data:



- Wheat, Brassica and chickpea pangenomes
- SUBA3, the SUBcellular protein localisation database for proteins
- cropPAL, the compendium of Crop Proteins with Annotated Locations
- Whiteflybase hosts curated barcode sequence data and nomenclatural information for *Bemisia tabaci* species
- plant stress methylome browsers presenting *A. thaliana* and *O. sativa* data, with more species to come.
- herbicide discovery tool to visualise and interpret the physicochemical properties of herbicides
- QTLNetMiner, a graph-database for biological information with *B. oleracea* and *C. arietinum* instances hosted at UWA, with more to come.



Platforms



UWA hosts genome databases displaying predicted genes and other annotations for reference genome assemblies of several species including *C. arietinum*, *B. oleracea*, *T. aestivum*, *B. napus* and seagrass. The WheatIs portal collates biological resources and bioinformatics tools for wheat research community and is hosted at UWA.

Training



A non-model *de novo* transcriptomics workshop has been developed and will be rolled out across Australia. New bioinformatics training materials under development will be made public in a central repository in liaison with EMBL-ABR network and ABACBS.

"As datasets become larger, there is a requirement to extend international collaborations, developing consortia to share approaches for the maintenance, integration and analysis of data. By working together and sharing our knowledge, experience and ideas, we can start to address some of the greatest challenges in biology."

Prof David Edwards:

Head of Applied
Bioinformatics,
University of
Western Australia

CONTACT **Prof David Edwards**
School of Plant Biology
University of Western Australia
ADDRESS 35 Stirling Highway
Crawley, Perth WA 6009
PHONE +61 8 6488 2415
EMAIL dave.edwards@uwa.edu.au



THE UNIVERSITY OF
**WESTERN
AUSTRALIA**

EMBL-ABR: VLSCI NODE

A centre for computational life science expertise



The VLSCI is a centre for computational life science expertise, hosted at the University of Melbourne, with significant computing resources and supporting a wide range of projects and services of local and national significance. The VLSCI Node has a proven record in platform development, bioinformatics training, compute and data resources, and tools development and will fortify the EMBL-ABR network with its breadth of experience and expertise.

Compute



The VLSCI Node will manage significant high throughput clusters and broker access to national cloud and data storage/archiving resources.

Data



The VLSCI Node will support researchers in exposing their high value datasets through technical expertise and provision of appropriate compute/data resources. This will include GUI portals for specific biological communities supported by EMBL-ABR. This activity will generate a series of best practice guidelines that will inform the whole EMBL-ABR network.

Platforms



The Genomics Virtual Lab, a cloud-based suite of genomics tools and workflow/analysis engines, will be provided through the national research cloud, on a per-user basis and also institutionally as managed services.

Training

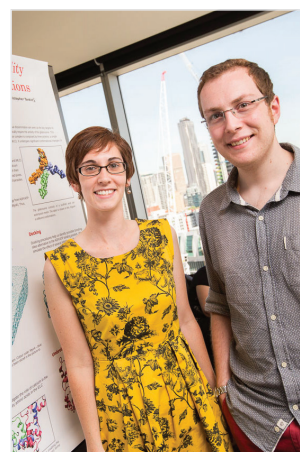


The VLSCI Node will deliver training across a broad spectrum of bioinformatics-focused topics to researchers, postdocs, clinicians and students. Topics include Data Carpentry, Software Carpentry, Using HPC for Life Sciences, Galaxy, and a range of specific topics in human and microbial genomics. Training will be delivered through face-to-face workshops using comprehensive training resources, which will be available online. Training will be conducted on VLSCI compute resources and especially on the Genomics Virtual Laboratory platform.

Tools



The VLSCI Node will support Australian-developed bioinformatics tools through local expertise and compute resources. National funding to support tool developers would be required and VLSCI would be well placed to deliver this given appropriate funding.



"Modern Biology is data-driven, and largely species agnostic when it comes to managing Big Data, solving visualization challenges, and modeling interactions. A national framework makes this work efficient, and accelerates progress by allowing solutions and resources to follow freely across different communities and areas of expertise."

Jason Williams:

Education, Outreach and Training Lead, CyVerse.

CONTACT **Assoc Prof Andrew Lonie**
Director, VLSCI
ADDRESS Lab-14, 700 Swanston Street
University of Melbourne 3010 Victoria AUSTRALIA
PHONE +61 3 8344 1395
EMAIL alonie@unimelb.edu.au





Publishing information

EMBL-ABR Hub
Melbourne Bioinformatics
Lab-14, 700 Swanston Street
University of Melbourne
Victoria 3010
Australia

Telephone +61 3 9035 5822
Email enquiries@embl-abr.org.au
Web www.embl-abr.org.au

Coordination & Design
Published by

EMBL-ABR & Melbourne Bioinformatics Communications
EMBL-ABR Executive



EMBL

Australia

Bioinformatics Resource

