H3ABioNet

Pan African Bioinformatics Network for H3Africa

Special Issue: BioRes Digest

Issue 20: March 2017

In This Issue

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Foreword

Welcome to issue 20 of the H3ABioNet newsletter. This issue reflects on the H3ABioNet AGM and SAB meeting which took place in November last year, and the directions the working groups have taken based on discussions and feedback from the meeting. Of interest at the SAB meeting was that many of the presentations were given by junior academics as they are increasingly playing leading roles in driving projects and tasks in the network.

On a sad note, at the end of 2016 we heard of the passing of our dear colleague Dr James Brandful, the PI of the NMIMR node in Ghana. The newsletter contains an article from his node in memory of Dr Brandful, but I would like to add my condolences to his family and the team at NMIMR. His kind, generous nature as well as his valuable advice and contributions will be missed.

Some new and exciting activities are underway, including the social media strategy to increase our visibility, and the revamp of our website. We need to move our focus to be a bit more outward looking, particularly as long term sustainability becomes an increasingly important issue.

Also new in this issue is the first BioRes Digest, a product of the RSWG, which aims to highlight interesting papers as a means to remain current with new tools and technologies in bioinformatics.

I hope you enjoy reading about our activities.

Issue 20: March 2017

H3ABioNet SAB/ AGM 2016

Continue the conversation:

#bioinformatics #Africa #H3ABioNet @H3ABionet

H3ABioNet Scientific Advisory Board and Annual General Meeting 2016

The 2016 H3ABioNet Annual General Meeting (AGM) and Scientific Advisory Board (SAB) meeting were held from the 1st to 4th November 2016 in Cape Town, South Africa. We were fortunate to have Prof. Jessica Kissinger, Prof. Francis Ouellette, Dr. Adebowale Adeyemo, and Prof. Gerard Tromp from the SAB attending in person, and Prof David Roos from the SAB attending remotely, as well as Dr. Jennifer Troyer as an invited H3Africa Programme Officer observer as advocated by the SAB in previous meetings. This meeting was a deviation from previous meets where usually the H3ABioNet AGM and working group meetings are held on the first two days and the report back to the SAB, as well as the one-on-one engagement sessions between the various Node PIs, the SAB and various teams from the Central Node, are held on the last two days.

The meeting kicked off with an annual progress report of the H3ABioNet project for year 4 by Prof. Nicola Mulder who provided an exciting overview of the numerous activities and progress made by the network to the SAB and the NIH Programme Officer. A summary of compiled year 4 Node highlights was presented to the SAB and the NIH Programme Officer by Sumir Panji. Shaun Aron, Amel Ghouila, Jonathan Kayondo, and Suresh Maslamoney provided working group updates to the SAB members for the Education and Training, Research, User Support, and Infrastructure working groups, respectively. After an opportunity for one-on-one engagement sessions, day 1 ended with research talks focused on projects that H3ABioNet has contributed to the H3Africa consortium. The research talks were on HTrainDB, the H3Africa Participant Recruitment Database, and the design of a genotyping array specific for African populations, which were provided by Zahra Mungloo-Dilmohamud, Mamana Mbiyavanga, and Ayton Meintjes, respectively.

Day 2 of the SAB meeting began with talks on specific H3ABioNet projects with Ziyaad Parker providing a presentation on the work he has accomplished in a short time on the H3Africa Data Archive dashboard, which aims to help H3Africa projects track their data submission progress and provide reports to the funders. Dr. Shakuntala Baichoo presented on the H3ABioNet Cloud Hackathon aimed at Dockerising the H3ABioNet workflows for heterogeneous and cloud computing environments. Dr Jean-Baka Domelevo Entfellner presented on the DREAM challenges hackathon held by the Research Working Group, IBM research Africa, and the University of Notre Dame, which was significant in that data generated by a American group was brought to Africa to be worked on by African scientists. Both Dr. Baichoo's and Dr. Domelevo Entefellner's talks signalled an important milestone for H3ABioNet as these are hackathon based activities that predominantly involved African scientists, and demonstrate that H3ABioNet is building the capacity and skills to work on a range of research problems in a collaborative manner. The session was well rounded off with the presentation by Kim Gurwitz on the highly successful Introduction to Bioinformatics course (IBT), organized and delivered by herself and the IBT core team which received lots of questions, comments, and enthusiastic reviews that carried on well into the tea-break. The afternoon session started with a talk on the grant application for H3ABioNet Phase II by Prof. Nicola Mulder, which outlined the structure of how the new network will operate and its priorities. The SAB had extra time to deliberate their initial feedback and responses, which they provided to the consortium members at the close of the day.

The SAB are impressed with the progress and cohesion that has been fostered in the network, which bodes well for the future. In terms of training, the SAB stressed that there needs to be more coordination between the education and training in bioinformatics and getting training materials curated so they are discoverable. The SAB urged the network to investigate the use of social media, place some of the training materials and seminars on YouTube and to investigate the use of subtitles for the recorded training material to help make them more accessible to non-English speakers. Another point emphasised by the SAB is that H3ABioNet should prepare for the deluge of data that will occur when the H3Africa projects rush to submit their data and ensure that rigorous quality control measures are put in place. The SAB stressed that in the coming years, H3ABioNet will need to take on a more user support role to assist the current and next round of H3Africa projects funded if the grant application is successful. The SAB also encouraged all Nodes within H3ABioNet to collaborate a lot more to build joint capacity.

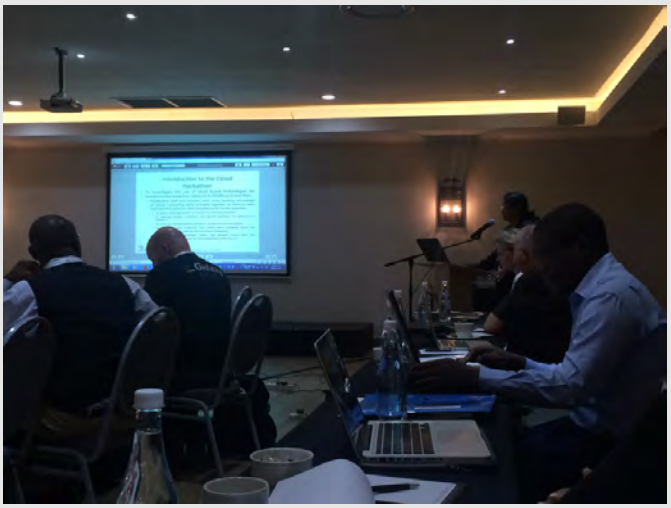
Issue 20: March 2017

H3ABioNet SAB/ AGM, 2016

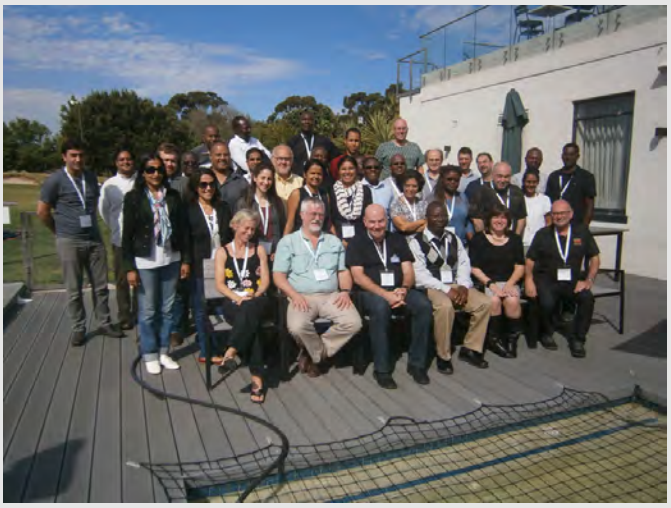
Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

The last two days of the 2016 H3ABioNet SAB/ AGM meeting comprised of working group meetings where the milestones and goals for each group were presented and discussed in detail, as outlined in the working group updates below in the current edition of the newsletter. An exciting development has been the refocus of the USWG to work on more outreach and social media activities in order to promote H3ABioNet and a strategy will be implemented soon.



Dr. Shakuntala Baichoo presenting the H3ABioNet Cloud Hackathon.



Group picture of attendees to the H3ABioNet SAB/ AGM 2016.

The next generation of African Bioinformaticians - courtesy of H3ABioNet.

Issue 20: March 2017

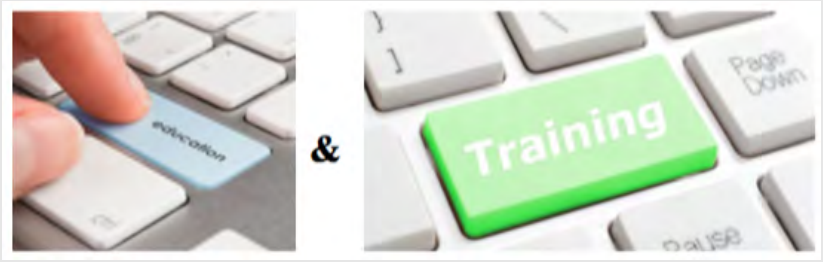
H3ABioNet

Education and Training Working Group

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Education and Training Working Group



With 2017 already in full swing, the E&T working group has been continuing work on established projects, as well as reflecting on the training requirements for the upcoming final period of the current grant cycle. The AGM/ SAB meeting held in November 2016 presented the working group with the valuable opportunity to discuss the various projects with the rest of the consortium and gain some insight into the focus areas for the upcoming year. The working group is looking forward to an exciting, albeit busy year ahead. Below we recap some of the highlights discussed during the E&T working group session at the AGM/ SAB meeting together with the projects that the working group will focus on for 2017.

- The 2016 iteration of the Introduction to Bioinformatics course (IBT_2016) was reviewed and plans and suggestions for improving the course were discussed. In addition to the course review surveys administered to staff and participants, there was a suggestion for each module of the course to be reviewed by an expert to ensure that the modules are up to date and relevant. There were also further suggestions on additional modules to include and how to encourage more interaction during the contact sessions, possibly by using a variation of the flipped classroom model. Finally, it was suggested that should a node want to apply for accreditation for the IBT course, they should initiate this through their own institutes. Taking into consideration these suggestions for improving the course, planning is already underway for IBT_2017, with the course confirmed to begin at the end of April 2017.

- Prior to the meeting, two new task forces were established to carry out the curation and updating of the MSc curriculum website and H3ABioNet training material. Both are valuable resources that can be beneficial to other bioinformatics groups if presented in a standardised and well-organised format. Efficient strategies for completing these two tasks were discussed and the task forces are hard at work implementing these suggestions to complete the tasks.

- Leading on from the discussions pertaining to improving the visibility and dissemination of H3ABioNet developed tools, courses, and resources, there was a suggestion to create online tutorials and short YouTube videos showcasing the tools and resources that the Network has developed. The working group is currently assessing which resources would benefit most from either a short YouTube demo video, online tutorial, or improved documentation, as well as the associated time and resource investment required to generate these.

- Finally, the training requirements for both within H3ABioNet and H3Africa were extensively discussed to prioritise which training courses the working group will focus on developing and running in 2017. Below is a tentative list of upcoming training events in the pipeline for 2017.
  - IBT_2017
  - H3Africa chip data analysis workshop
  - Flu data analysis jamboree
  - Pharmacogenomics workshop
  - Career development workshop

#H3ABioNetEducationAndTraining **Nicky Mulder and Shaun Aron** Back to Contents

Issue 20: March 2017

H3ABioNet

Infrastructure Working Group

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Infrastructure Working Group

The ISWG had a good meeting at the AGM in November last year. We reviewed progress and we noted that although most milestones had been met, we needed to have broader membership attending. There are a number of projects that we will be involved with, but the key issues that were discussed, and which will be priorities for 2017, are:

- Data Management Task Force - in 2017, research groups will be submitting their data to the EGA through H3ABioNet. This is a mission critical activity, and we need to ensure that the process works well and is reliable. Good progress has been made and we've been able to test processes with pilot submissions.
- A second project for the Data Management Task Force is building a query interface to allow researchers to query across all H3Africa projects. For this, we need to work closely with the biorepositories.
- Our Globus Online and Netmap projects must continue rollout. We have a good number of nodes participating. We will start reporting regularly to the MC about statistics from the various nodes. Data transfer is an important responsibility for us.
- Work on our Cloud Projects, including building pipelines for H3Africa projects and evaluating cloud architectures, will be very important work. This will help the Network as a whole to show capacity and spread skills in the Network.
- Other work of the System Admin Taskforce - such as developing and maintaining documentation, work on the e-Biokits - must be sustained.
- Planning for the second round of the H3Africa project must start soon in case of renewal. A particular concern is the H3Africa Participants Recruitment Database, which was developed recently for the funders under very tight deadlines and works very well. However, there is much greater functionality that could be built for both funders and research groups for the second round so early development of this will be a priority.

After a short break over the new year, the group has been very busy in 2017.

The Cloud Computing task force is working on finishing the development of the pipelines with some groups farther in the process then others. The microbiome Docker pipeline has undergone some testing and will be tested by an external group while the NGS variant calling pipeline will need to undergo testing. The current QC genotyping and imputation pipelines are being worked on and will be tested in due course. The final steps remaining will be to deposit the pipelines to Quay.IO and wrap up draft publications by the various team leaders and groups for the different pipelines. Mali ICER has launched its Globus Online end point - we'll be bugging other sites to get their end points up and running.

The System Administration task force has met this year and has significant new blood, with over 20 members. A new project, working with the South African National Research Network evaluating Aspera against Globus Online for large data transfers, will start soon. The Data Management Task force has been busy, mainly the members from CBIO. They are working with several H3Africa groups on the submission of data. Ziyaad Parker has been working hard, the dashboard is up and running, and we're getting feedback on the GUI. Documentation is the next task.

#H3ABioNetInfrastructure **Scott Hazelhurst and Suresh Maslamoney**

Back to Contents

Issue 20: March 2017

Research Working Group

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Research Working Group

Research working group briefing during the SAB/ AGM 2016
The RSWG Co-chair gave a brief overview of all activities of the working group during the fourth year of the grant as well as the milestones and activities of the upcoming fifth year in front of the SAB and members of the H3ABioNet consortium on the 1st day of the AGM 2016. The SAB and H3ABioNet members were also briefed on the Mozilla study group, which aims to encourage peer learning amongst communities, including the existing and the newly established study groups within the H3ABioNet nodes as well as the idea of establishing regional study groups. The idea of linking the H3ABioNet SIGs to the ISCB COSI groups was also highlighted. The RSWG Chair presented the idea of the seminars evaluation to the SAB and H3ABioNet members who attended the 3rd day of the AGM 2016. The RSWG Chair summarized the suggested metrics for evaluating the seminars ((Click-through rate (CTR), Attendee ratio (AR), Online polls, Post seminar survey, Audience retention, and On-demand viewing)) and the idea of designation of the best presentation. The best presentation will be selected and the presenter will be offered the possibility to attend the next H3ABioNet annual meeting. The students' seminar evaluation will be based on the abstract quality, the presentation skills, and the manner of handling the audience questions. The possibility of assessing the number of attendees could be used based on the Mconf metrics. In Addition, a brief review of the seminar planning for 2017 including the proposed speakers and topics suggestions was also discussed.

RSWG focus for the 5th Year (2017):
The RSWG is trying to identify a mechanism by which to register H3ABioNet as an organization within ISCB in order to have access and the possibility for the SIGs members to participate. This point has to be discussed further with Prof. Nicky Mulder. The idea of the H3ABioNet Bioinformatics Research Digest (BioRes Digest) and having an issue each 3 months has been discussed and approved. Have a look at the first edition of the BioRes Digest in this month's issue of the newsletter. The digest focuses on the importance of a Data Management Plan (DMP) and how to develop a DMP. The content will be similar to what was presented during the H3Africa meeting, which took place in Dakar, Senegal last May. The importance of linking the BioRes Digest to the new USWG journal club has been highlighted. The idea is to encourage journal club participants to submit their summaries together with their biographies to the BioRes Digest.

H3ABioNet is working on redesigning the website to better represent H3ABioNet activities. Each working group has been tasked to come up with suggestions to re-structure their working group page to better reflect the accomplished activities being done within each group. Some of the RSWG members were volunteered for assisting the RSWG Chairs in identifying modifications and re-structuring ideas for the RSWG page. The different teams of the DREAM Malaria Hackathon submitted their reports together with supporting materials (e.g. scripts and models) that they developed in January 2017. The finalization and submission of the concept paper and preliminary data analysis paper will be in the coming months. Launching the DREAM challenge is expected to be in the Summer of 2017. The integration of H3ABioNet pipelines and tools into JMS and some of the technical issues faced was discussed during the AGM. The RSWG will follow up with Prof. Ozlem Tastan-Bishop and her team on what has been integrated so far to the JMS, the issues faced and when the integration of the submitted pipelines will be completed.

#H3ABioNetResearch **Faisal Fadlelmola and Amel Ghouila**

Back to Contents

Issue 20: March 2017

H3ABioNet

User Support Working Group

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

User Support Working Group

This period finds the work group fresh from the fifth consortium SAB/ AGM, held in November 2016, and so has turned it into a hive of activities. The strategy for Year 5 activities, discussed and refined at the general meeting, is quickly being implemented.

Deliverables for Yr5 2016/2017:

The user support working group (USWG), previously operating in areas with substantial overlap with other technical working groups, has now been reconstituted and its activity focus evolved to lay more emphasis on outreach and network promotion through various avenues among which are:

- Node ambassadors
- Network Promotional materials
- Newsletter
- Website review and overhaul
- Social Media
- Journal Club
- Help Desk article

The above activities have been adopted. In addition, round table discussions at the AGM came up with/ endorsed the following plans, ideas and outreach tools to realize key work group deliverables:

1. Coordinated media strategy

A proposal for a communication, and network activity promotion strategy encompassing social/professional networks (Twitter, Facebook, LinkedIn, blogs), newsletter, and youtube was presented and discussed as a coordinated outreach plan. This will involve continuing the revived newsletter, and utilizing the social media taskforce for content coordination and design of matrices for impact assessment. Also another taskforce, separate from the social media one, will be set up to lead creation of a consortium youtube channel under the H3ABioNet google account where short videos of our tools, tutorials, and topical interviews on node activities among others will be captured. Blogs comprising personal description/experience/or opinion of consortium activities/products will be added to the website (see below).

Twitter will be our primary tool for interactive engagement with social media savvy audiences, and consortium ground rules will be developed and led by the social media taskforce to guide the process. As a result, upcoming events such as training, seminars, or launch of consortium products such as new publications, blogs and newsletters among others will be tweeted. Twitter and Facebook feeds could be strategically combined.

Issue 20: March 2017

User Support Working Group

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

2. New consortium website

The H3ABioNet consortium website will be redesigned to a better structured and integrated platform. All website content, including help desk SOPs, will be reviewed and missing links updated. Sections of the website shall be assigned to different working groups for in-depth scrutiny and internal evaluation under various user case scenarios to help guide 're-think' for the new improved user interface design. Lots of suggestions for how to improve the website outlook and helpdesk interface were discussed during the meeting and will be factored into the upcoming version. Also various assessment matrixes (e.g. click through rate, attendee ratio, online polls, exit surveys etc) to help improve user website experiences and opportunities will be collected.

3. Journal club

This is a forum envisaged as an avenue for increased bioinformatics tool awareness. The plan is for monthly sessions with rotations among nodes. Where possible it could be combined with seminar proceedings to cut down on numbers of separate meetings being set up. A taskforce, already in place, will be spearheading this initiative.

4. Helpdesk article

This article is intended to promote the Helpdesk, a flag-ship deliverable for the USWG. A concept has already been developed and will be circulated within the workgroup for wider group input and authorship.

Progress so far:

The USWG has embarked on a fortnightly meet frequency to jump-start these activities. During this period the USWG has continued to refine its strategy in reaching out and offering support to H3ABioNet service users. We have developed and implemented integrated and coordinated social media platforms to reach out to our audience. We have begun to exploit social media tools including Twitter, Facebook, LinkedIn, and a monthly newsletter to announce events and also to engage with the scientific community and the public at large. The social media tools are managed by a dedicated team to ensure timely, consistent and coordinated communication of H3ABioNet services and tools across various media platforms thereby maximizing our visibility, impact and to reach a wider global audience.

The USWG page on the H3ABioNet website is also being redesigned, incorporating features that allows for easy and quick access to H3ABioNet services and tools that are available online. The H3ABioNet social media sites are linked to the websites and supported with RSS feed to collate updates from the social media platforms and blogs of interest.

The USWG has also developed and implemented strategy to identify and gather H3ABioNet events, announcements, products, success stories, and activities being undertaken in various H3ABioNet nodes for publication across the media platforms. Please have a look at the Announcements and Upcoming Events sections of the current newsletter for links to forms that may be used to tell us about what is happening at your Node. The USWG is also offering support to Nodes to reach out to a global audience on their activities via the H3ABioNet ever-presence on social media platforms.

#H3ABioNetUserSupport

Jonathan Kayondo and Pandam Salifu

Back to Contents

Issue 20: March 2017

In memory of Dr. James A.M. Brandful, NMIMR

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

In memory of Dr. James A.M. Brandful

1st February 1956 to 8th December 2016

Tribute by Members of the Noguchi Memorial Institute for Medical Research (NMIMR) H3ABioNet Node:

On the 8th of December 2016, Dr. James Brandful was called to glory, after a long battle with sicke cell disease. Dr. Brandful was the Principal Investigator and Leader of the Noguchi Memorial Institute for Medical Research (NMIMR) Node of the H3ABioNet in Ghana. Dr. Brandful played significant roles in the successful establishment of the Node and he is credited as being one of the pioneers of bioinformatics capacity building in West Africa.

As members of the Noguchi node of the H3ABioNet, we had the opportunity to directly learn from Jim, as we popularly called him. He played the role of an astute leader, communicator, organizer and above all, a father figure. He cared tremendously for the welfare and well-being of both the students and researchers within the Node. As an excellent mentor, Jim taught some of us the art and science of meticulous research pertaining to infectious diseases, including endemic viral infections. There were instances that we saw no road ahead of us, but with his scientific wisdom, he steered us to the destination of plausible research paradigms that will forever shape our thinking!

At the Noguchi Memorial Institute, Dr. Brandful was a Senior Research Fellow, Principal Investigator and Leader of several research studies during his time at the Noguchi Memorial Institute for Medical Research. Some of the major projects he undertook include, characterizing HIV strains in Ghana in the early days of the HIV epidemic in Ghana; development of anti HIV compounds from Ghanaian medicinal plants and surveillance on viral pathogens in animal handlers. Dr. Brandful received several awards for his work including visiting fellowships at Cambridge University UK, and AIDS Vaccine Institute and the Medical Research Council both in South Africa. In 2012 he received the Vice Chancellor's award for the outstanding doctoral thesis in the Sciences at University of Ghana. He has authored several scientific publications in reputable international journals on viral pathogens such as HIV, Coxsackie, Hepatitis and Human Papilloma Viruses among others. He also authored a book on HIV diagnosis. Dr. Brandful served the Institute and the University of Ghana in several capacities including Head of Department of Virology from October 2006 to July 2011. He was advisor to the Institute of Continuing and Distance Education and supported the establishment of MPhil and MA programmes in HIV/AIDS and AIDS Management in the Workplace. He was also an examiner in microbiology for the School of Graduate Studies University of Ghana and Kwame Nkrumah University of Science and Technology, Kumasi. At the Institute, he was the head of our Biosafety team, ensuring safe practices in the BSL3 laboratory. Dr. Brandful was very keen on passing on knowledge and skills to the younger generation and thus supervised and mentored several students that passed through his laboratory.

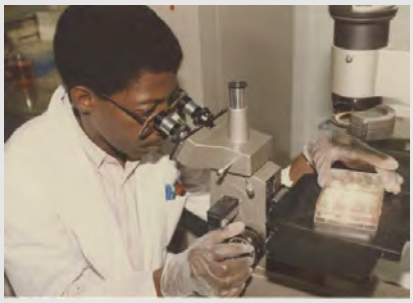
Our Jim is no more with us physically, but cognitively he lives within us forever and the tree of scientists that will evolve from the bioinformatics Node will remain deeply rooted in him and will treasure his contribution to computational biomedicine in Ghana. Even though we are perturbed by the demise of Jim, we will reorder the system dynamically to sustain his legacy! We will keep his good work running! We will immortalize his legacy as a way of appreciating his contribution to this emerging field. We will miss his quiet demeanor and analytical mind that he brought to dissect thorny issues in our deliberations.

Issue 20: March 2017

In memory of Dr. James A.M. Brandful, NMIMR

Continue the conversation:

#bioinformatics #Africa #H3ABioNet @H3ABionet



LHS: Members of the Noguchi Memorial Institute for Medical Research (NMIMR) H3ABioNet Node - (from left to right) Sector Amuzu, Dr. Samuel Kwofie, Dr. Anita Ghansah, the late Dr. James Brandful, Wisdom A. Akurugu. RHS: the late Dr. James Brandful.

Memorial piece for the late Dr. James Brandful, by Wisdom A. Akurugu:

I met Dr. James Brandful for the first time in July 2013, when I picked up my appointment letter to work with him as the bioinformatician at the NMIMR H3ABioNet Node. I have since worked with him for the past three (3) years. I introduced myself to him and he said he was glad to have me to assist him run the activities of the Node. We had deep and sincere discussions concerning the project. He was passionate about the work ahead and acknowledged his limitation in bioinformatics and thus most often relied on me for the work of the Node. He did his best even on his sick bed to run the Node.

Though he often fell sick and got admitted, he had hope in life. He always kept a smile - a sign of encouragement even during difficult times. His characteristic smile and dimple can't be forgotten. He cracked jokes and teased often and that made most people of the institute like to speak to him, particularly in times of distress. He was simply unique in his conversation, sharing his experiences in life and biomedical research.

Dr. Brandful recognized that everyone has his or her strengths and weaknesses. He would always thank me for whatever I did for him in relation to the work. He had confidence and trust in me though I have my challenges in the discharge of the assigned responsibilities. He wanted to develop and nurture anyone who worked with him. Thus he never hesitated in recommending me or anyone for any workshop or meeting for the sake of the work. He was fair and open. At the least opportunity, he would advocate for training and more training of Node members.

At a very personal level, he spoke to me like a father on the first day we met and has since interacted with me like my biological father. He has assisted me, as a family person, in many uncountable ways. Dr. Brandful is one of such people who believed that there is great potential in me and encouraged me to work hard. He gave me the opportunity to work and learn more in the field of bioinformatics. I owe him a lot of gratitude.

At his passing now, I can only pray for his gentle soul to rest peacefully with his maker and the comfort of God to be upon his immediate family. I pray for the energy and enthusiasm to keep the flame he has lit burning.

RIP, Dr. Brandful!!

Back to Contents

Issue 20: March 2017

Announcements

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Announcements

- Congratulations to the CBSB Node for successfully undertaking and accredited for the H3ABioNet Variant calling Node accreditation.

- Congratulations to Shakuntala Baichoo from the SANBio Node for her recent publication in Bioinformatics entitled *BioPAXViz: a cytoscape application for the visual exploration of metabolic pathway evolution* (*https://doi.org/10.1093/bioinformatics/btw813*)

- Congratulations to Gerrit Botha from the CBIO Node who is an award recipient of the *International Data Corporation (IDC) HPC Innovation Excellence Award 2016* and most probably the first African Scientist to win this award.

- The Wellcome Trust H3Africa call for proposals is now open. Preliminary application deadline: **24 March 2017**

Do you have an ANNOUNCEMENT for upcoming editions of the H3ABioNet newsletter or for the H3ABioNet social media pages?

Tell us about your anouncements *here*

Back to Contents

Issue 20: March 2017

Upcoming Events

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

Upcoming Events

- **April 2nd to April 12th 2017**: *1st Big Data Africa Summer School*. The school will focus on practical Data Science Applications to Astronomy and Bioinformatics and will take place in Cape Town, South Africa.

- **June 4th to June 16th 2017**: *EMBO Practical Course in Bioinformatics and Genomics*, will take place in Thermi, Greece.

- **July 19th to July 21st 2017**: *NextComp 2017: Next Generation Computing Application Conference*, will take place in Republic of Mauritius.

- **July 21st to July 25th**: *International Society for Computational Biology conference 2017 (ISCB ECCB 2017)*, will take place in Prague, Czech Republic.

- **August 13th to August 16th 2017**: *17th Biennial Congress of the Southern African Society for Human Genetics*, will take place in Durban, South Africa.
  Closing date for abstract submission: **1 April 2017**.

- **August 20th to August 23rd 2017**: *The 8th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*

  This year, the *Workshop on Algorithms in Bioinformatics (WABI)* will be co-located with ACM-BCB.

- **Every third week of every month**: *CPGR Foundation in Genomics Course*, from standard molecular technologies to advanced 'omics' application in 3 days, aimed at scientists who are new to 'omics' as well as researchers interested in an overview of a dynamically evolving field.

- For a comprehensive list of bioinformatics and genomics conferences, please consult: *Conference service - Bioinformatics*

Do you have an EVENT for upcoming editions of the H3ABioNet newsletter or for the H3ABioNet social media pages?

Tell us about your events *here*

Back to Contents

Issue 20: March 2017

Upcoming H3ABioNet working group meeting schedule

Upcoming H3ABioNet working group meeting schedule*

*Schedule until end April 2017

Summary of H3ABioNet upcoming working group meetings

Month	Date	Day	Working Group (WG)	Time (UTC)
April	7th	Friday	User Support WG	9:00
April	14th	Friday	Research WG	13:00
April	18th	Tuesday	Education and Training WG	11:00
April	21st	Friday	User Support WG	9:00
May	5th	Friday	User Support WG	9:00
May	12th	Friday	Research WG	13:00

Timezone conversions to UTC for all H3ABioNet working group meetings

UTC Time Offset	Time Zone Name	Region/ Country in the Time Zone offset
-6 hours	CDT	Chicago, USA
0 hours	GMT	Burkina Faso, Ghana, Mali, Morocco, Senegal
+1 hour	WAT	Cameroon, Chad, Gabon, Namibia, Nigeria, Niger, Tunisia
+2 hours	CAT	Botswana, Egypt, Malawi, South Africa, Sudan, Zambia
+3 hours	WAT	Ethiopia, Kenya,Tanzania, Uganda

This edition of the newsletter was compiled and edited by Kim Gurwitz. For any corrections, please contact Kim at kim.gurwitz@uct.ac.za

Back to Contents

Please scroll down for the latest issue of BioRes Digest.

Continue the conversation:

#bioinformatics #Africa
#H3ABioNet @H3ABionet

BioRes Digest

H3ABioNet Bioinformatics Research Digest | January - March 2017 | Issue: 01

In This Issue

Follow us:

#H3ABioNetResearch
#H3ABioNet @H3ABionet

Foreword

The H3ABioNet Research Working Group is launching a quarterly Bioinformatics Research Digest on the latest bioinformatics topics relevant to H3Africa projects. The main focus is on briefing the H3ABioNet and H3Africa community on the latest articles and topics in bioinformatics and computational biology.

Main Article Summary: Ten Simple Rules for Creating a Data Management Plan

by Sumir Panji and Faisal Fadlelmola

A recent published article entitled *Ten Simple Rules for creating a DMP* (*doi:10.1371/journal.pcbi.1004525*) was selected to be the topic of one of the sessions of the **H3ABioNet Data Support Workshop** during the 8th H3Africa Consortium Meeting that was held at the Radisson Blu Hotel, Dakar, Senegal on 12th May 2016.

Data Management

The planned collection, quality control, storage, retrieval, analysis, and dissemination of results, helps to control information generated during a research project. Managing data helps to control information generated in a structured manner. Data management is an increasingly integral part of biomedical research, especially as datasets are getting much larger. Biomedical sciences following a similar trend as physics where there are large datasets generated and teams of analysts around the world work on them.

Data Management Plan (DMP)

A DMP is a document that describes how various aspects of research data generated for a project are handled during the project's life cycle as well as (especially in genomics) after the project is over. The *common elements in a DMP* usually include: 1. Description of the data to be collected/ created, 2. Standards/ methodologies for data collection and management, 3. Ethics and Intellectual Property concerns or **restrictions**, 4. Plans for data sharing and access, 5. Strategy for long-term preservation.

BioRes Digest

H3ABioNet

Main Article Summary

Importance of DMP

Funding agencies are recognizing the importance of research data as community resources that accelerate the pace of discovery. In some cases, the data generated is funded by tax payers (e.g. NIH funded projects) and hence should be readily available. Other funding agencies such as *Wellcome Trust* recognize the Fort Lauderdale and Toronto statement on datasets that form community resources.

Most (if not all) genomic projects generate large datasets and when applying for grants, genomics projects will always most certainly need a DMP. These DMPs do undergo peer review to evaluate the merit of a submitted grant proposal. Although different agencies place varying levels of emphasis on the DMP, they will all definitely scrutinize your DMP and can revisit the DMP during performance reviews. A DMP is an integral part for the success of a research project as the outcome is dependent on how well the data is managed.

Good data management can be challenging; especially when studies involve multiple sites, multiple PIs, site specific protocols, naming conventions, measurement scales, and ethics. The NASA Mars Climate Orbiter, costing USD 125 million, was lost in 1999 because two sets of engineers working on different systems failed to convert metric units to imperial units.

A good DMP helps project affiliated researchers to:

- Define the various data attributes (metric or imperial)
- Easily find files for analyses
- Share and store their analyses with their multi-site partners
- Support the results of the published work.

In the subsequent paragraphs of this section of this BioRes Digest we will summarize these ten simple rules for creating a DMP.

Rule 1: Determine the Research Sponsor Requirements
Different funding agencies have different policies - some ask for specific details, others ask for broad plans. Funding agencies usually provide DMP requirements in either the public request for proposals (RFP) or in an online grant proposal guide. Keep in mind that the principle objective should be to create a DMP that will be useful for your project and should be treated as a "living document". Although funding agencies constrain the length of a DMP to a certain number of pages, a more detailed DMP can be submitted as an appendix or supplementary file.

Rule 2: Identify the Data to Be Collected
Research Data: defines the data and materials generated and covered in a DMP.
Types of data: data and materials may take the form of physical samples (biospecimens), digital formats e.g. genome sequences, electrocardiograms, clinical measurements (phenotypes).
Sources of data: where is the data coming from? e.g. is it from human participants, public databases, or is it propriety data which cannot be shared due to licensing?
Volume of data: the amount of data expected to be collected and stored, could be physical collections such as biospecimens or digital data.

Rule 3: Define How the Data Will Be Organized
The types, sources, and volume of data influences data organization, for example, human participant information should be stored, de-identified, in a secure database system like REDCap or OpenClinica, and not as text files on a communal laboratory PC. If the data will be submitted to an access controlled repository it will need to be formatted and suitably organized for submission e.g. EGA.

Rule 4: Explain How the Data Will Be Documented
Metadata: additional data that provides contextual details of research data generated. Provides standardized, structured information on how the actual data was collected, processed and interpreted e.g measurement types, instruments used, analytical methods, software versions.

Follow us:

#H3ABioNetResearch
#H3ABioNet @H3ABionet

Ontology: common vocabulary used to describe various data attributes by a community that needs to share data.

- Where there are no ontologies available, widely used community adopted controlled vocabularies should be used.
- Use of ontologies and community adopted controlled vocabularies assists in the searching and finding of data and are required by most data repositories.

Rule 5: Describe How Data Quality Will Be Assured

Some RFPs do indicate what requirements for Quality Controls/ Quality Assurances are expected. Depending on the nature of the study and the types of data being collected and /or generated, the focus on each of the above activities within the DMP will differ.

The QC/ QA of data fidelity is linked to the data types being collected and also for analysis. In terms of data entry, verification and the use of specifically constrained fields should be used e.g. "@" for a valid email address, range of dates.

Other QC could be physical sample inspection, testing for purity. In the case of analyses each type of data analyses does have its own QC steps e.g. trim bases that have Phred scores less than 35. Visual inspection of the data during analyses helps in the QC process e.g. QQ plots in GWAS to determine confounding effects within population structure between cases and controls.

Rule 6: Present a Sound Data Storage and Preservation Strategy

Collecting, generating, and analyzing data incurs a significant amount of cost and personnel time. A lack of data storage and preservations strategy could see 3 - 4 years of work lost in the amount of time it takes for a computer to crash or a hard drive failure.

A DMP should take into account the following questions:

- How long should the data be stored and made available for?
- How will the data be secured and shared amongst collaborators before being made available to the wider community?
- How will the data be archived and made retrievable in future?

Data will need to be stored while being collected and analyzed, which should be estimated beforehand from the actual grant application. Provision for the storage of data in two locations and one off site location for disaster recovery should be articulated. Measures to protect the data should be included - the level of data security and access is dependent on the nature of data.

Once the analysis is completed and results published, the data might need to be made publicly available according to the funding organization's policies.

Rule 7: Define the Project's Data Policies

Many funding organizations require that DMPs include explicit policy statements about how data will be managed and shared. Such policies should include, if appropriate:

1. Licensing or sharing arrangements that pertain to the use of preexisting materials.
2. Plans for retaining, licensing, sharing, and embargoing (i.e., limiting use by others for a period of time) data and other materials.
3. Legal and ethical restrictions on access and use of human subject and other sensitive data.

Rule 8: Describe How the Data Will Be Disseminated

Providing access to research data enables the reuse of data and adds value. A plan for the dissemination of data is usually required by the funding organization. The data dissemination plan should be as concise and specific as possible stating how, when and what data will be released. A data dissemination plan that has the minimum possible restrictions to the release and access of data when the project is completed and results published is the most preferred where applicable.

BioRes Digest

H3ABioNet

Main Article Summary

Follow us:

#H3ABioNetResearch
#H3ABioNet @H3ABionet

Main Article Summary

Certain types of data of the same study can be submitted to different repositories. For example, sequence data from a microbiome study can be submitted to the publicly accessible European Nucleotide Archive (ENA) while the phenotype data may be submitted to European Genome Archive (EGA).

Rule 9: Assign Roles and Responsibilities

A comprehensive DMP should clearly designate the roles and responsibilities of every named individual and organization associated with the project. Roles may include; data collection, data entry, QA/ QC, metadata creation and management, backup, data preparation and submission to an archive, and systems administration. Large multi-investigator projects may benefit from having a dedicated staff person(s) assigned to data management. Treat your DMP as a living document and revisit it frequently (e.g quarterly basis). Assign a project team member to revise the plan, reflecting any new changes in protocols and policies.

Rule 10: Prepare a Realistic Budget

A common error when developing a DMP is forgetting to budget for all activities involved in the data management life cycle for the activities. Data management is time consuming and costs money in terms of software, hardware, and technically skilled personnel that are highly sought after by industry. Review your DMP and make sure to link components in your DMP with specific line items with the budget proposal and the budget justification to support the people that manage the data as well as to pay for the requisite hardware, software, and services. Check with your IT department and the preferred data repository department so that requisite fees and services are budgeted appropriately.

In summary, a good DMP takes a lot of thought and engagement with various stakeholders such as the University's IP office, IT department, financial officer - start early. At first, developing a DMP is hard work and time consuming, but once done it can be reused with information updated from various sources that will enable the subsequent DMPs to be written up relatively quickly with most successful parts of the DMP being incorporated into subsequent projects.

A data management plan should provide your project members, funders, and others with an easy-to-follow road map that will guide and explain how data are treated throughout the life of the project and after the project is completed.

A DMP provides a vehicle for conveying information to and setting expectations for your project team during both the proposal and project planning stages, as well as during project team meetings later when the project is underway. The best plans are living documents that are periodically reviewed and revised as necessary according to needs and any changes in protocols, policy, technology, and staff.

Back to Contents for BioRes Digest

Follow us:

#H3ABioNetResearch
#H3ABioNet @H3ABionet

Back to H3ABioNet Newsletter: Back to Contents

BioRes Digest

H3ABioNet

Latest Articles

Follow us:

#H3ABioNetResearch
#H3ABioNet @H3ABionet

Latest Articles

1. Jeremy Leipzig. **A review of bioinformatic pipeline frameworks**. Brief Bioinform 2016 bbw020. *doi: 10.1093/bib/bbw020*.

Abstract

High-throughput bioinformatic analyses increasingly rely on pipeline frameworks to process sequence and metadata. Modern implementations of these frameworks differ on three key dimensions: using an implicit or explicit syntax, using a configuration, convention or class-based design paradigm and offering a command line or workbench interface. Here I survey and compare the design philosophies of several current pipeline frameworks. I provide practical recommendations based on analysis requirements and the user base.

2. L. Schaeffer, H. Pimentel, N. Bray, P. Melsted, L. Pachter. **Pseudoalignment for metagenomic read assignment**. arXiv 2015. *doi: arXiv:1510.07371v2*.

Abstract

We explore connections between metagenomic read assignment and the quantification of transcripts from RNA-Seq data. In particular, we show that the recent idea of pseudoalignment introduced in the RNA-Seq context is suitable in the metagenomics setting. When coupled with the Expectation-Maximization (EM) algorithm, reads can be assigned far more accurately and quickly than is currently possible with state of the art software.

Motivation: Read assignment is an important first step in many metagenomic analysis workflows, providing the basis for identification and quantification of species. However ambiguity among the sequences of many strains makes it difficult to assign reads at the lowest level of taxonomy, and reads are typically assigned to taxonomic levels where they are unambiguous. We explore connections between metagenomic read assignment and the quantification of transcripts from RNA-Seq data in order to develop novel methods for rapid and accurate quantification of metagenomic strains.

Results: We find that the recent idea of pseudo alignment introduced in the RNA-Seq context is highly applicable in the metagenomics setting. When coupled with the Expectation-Maximization (EM) algorithm, reads can be assigned far more accurately and quickly than is currently possible with state of the art software, making it possible and practical for the first time to analyze abundances of individual genomes in metagenomics projects.

Availability: Pipeline and analysis code can be downloaded *here*.

3. Yun Zhang, David J. Topham, Juilee Thakar, Xing Qiu. **FUNNEL-GSEA: FUNctioNal ELastic-net Regression in Time-course Gene Set Enrichment Analysis**. Bioinformatics 2017 btx104. *doi: 10.1093/bioinformatics/btx104*.

Abstract

Background: Gene set enrichment analyses (GSEA) are widely used in genomic research to identify underlying biological mechanisms (defined by the gene sets), such as Gene Ontology terms and molecular pathways. There are two caveats in the currently available methods: i) they are typically designed for group comparisons or regression analyses, which do not utilize temporal information efficiently in time-series of transcriptomics measurements; and ii) genes overlapping in multiple molecular pathways are considered multiple times in hypothesis testing.

Latest Articles

Follow us:

#H3ABioNetResearch
#H3ABioNet @H3ABionet

Results: We propose an inferential framework for GSEA based on functional data analysis, which utilizes the temporal information based on functional principal component analysis, and disentangles the effects of overlapping genes by a functional extension of the elastic-net regression. Furthermore, the hypothesis testing for the gene sets is performed by an extension of Mann-Whitney U test which is based on weighted rank sums computed from correlated observations. By using both simulated datasets and a large-scale time-course gene expression data on human influenza infection, we demonstrate that our method has uniformly better receiver operating characteristic curves, and identifies more pathways relevant to immune-response to human influenza infection than the competing approaches.

Availability: The methods are implemented in R package FUNNEL, freely and publicly available *here*.

4. Ji, T. and Chen, J. **Statistical models for DNA copy number variation detection using read-depth data from next generation sequencing experiments**. Aust. N. Z. J. Stat. 2016, 58: 473-491. *doi:10.1111/anzs.12175*.

Abstract

In this 'Big Data' era, statisticians inevitably encounter data generated from various disciplines. In particular, advances in bio-technology have enabled scientists to produce enormous datasets in various biological experiments. In the last two decades, we have seen high-throughput microarray data resulting from various genomic studies. Recently, next generation sequencing (NGS) technology has been playing an important role in the study of genomic features, resulting in vast amount of NGS data. One frequent application of NGS technology is in the study of DNA copy number variants (CNVs). The resulting NGS read count data are then used by researchers to formulate their various scientific approaches to accurately detect CNVs. Computational and statistical approaches to the detection of CNVs using NGS data are, however, very limited at present. In this review paper, we will focus on read-depth analysis in CNV detection and give a brief summary of currently used statistical analysis methods in searching for CNVs using NGS data. In addition, based on the review, we discuss the challenges we face and future research directions. The ultimate goal of this review paper is to give a timely exposition of the surveyed statistical methods to researchers in related fields.

5. Xavier Watkins, Leyla J. Garcia, Sangya Pundir, Maria J. Martin. **The UniProt Consortium; ProtVista: visualization of protein sequence annotations**. Bioinformatics 2017 btx120. *doi:10.1093/bioinformatics/btx120*.

Abstract

Summary: ProtVista is a comprehensive visualization tool for the graphical representation of protein sequence features in the UniProt Knowledgebase, experimental proteomics and variation public datasets. The complexity and relationships in this wealth of data pose a challenge in interpretation. Integrative visualization approaches such as provided by ProtVista are thus essential for researchers to understand the data and, for instance, discover patterns affecting function and disease associations.

Availability: ProtVista is a JavaScript component released as an open source project under the Apache 2 License. Documentation and source code are available *here*.

Back to Contents for BioRes Digest

Back to H3ABioNet Newsletter: Back to Contents

BioRes Digest

H3ABioNet

Most Cited Articles

Follow us:

#H3ABioNetResearch
#H3ABioNet @H3ABionet

Most Cited Articles

***'Most Cited Articles'* is updated monthly. Rankings are based on citations to online articles from *HighWire Press*.**

1. Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration**. Brief Bioinform 2013, 14 (2): 178-192. *doi: 10.1093/bib/bbs017*.

Abstract

Data visualization is an essential component of genomic data analysis. However, the size and diversity of the data sets produced by today's sequencing and array-based profiling methods present major challenges to visualization tools. The Integrative Genomics Viewer (IGV) is a high-performance viewer that efficiently handles large heterogeneous data sets, while providing a smooth and intuitive user experience at all levels of genome resolution. A key characteristic of IGV is its focus on the integrative nature of genomic studies, with support for both array-based and next-generation sequencing data, and the integration of clinical and phenotypic data. Although IGV is often used to view genomic data from public sources, its primary emphasis is to support researchers who wish to visualize and explore their own data sets or those from colleagues. To that end, IGV supports flexible loading of local and remote data sets, and is optimized to provide high-performance data visualization and exploration on standard desktop systems.

Availability: IGV is freely available for download *here*, under a GNU LGPL open-source license.

2. P. Librado, J. Rozas. **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data**. Bioinformatics 2009, 25 (11): 1451-1452. *doi: 10.1093/bioinformatics/btp187*.

Abstract

Background: DnaSP is a software package for a comprehensive analysis of DNA polymorphism data. Version 5 implements a number of new features and analytical methods allowing extensive DNA polymorphism analyses on large datasets. Among other features, the newly implemented methods allow for: (i) analyses on multiple data files; (ii) haplotype phasing; (iii) analyses on insertion/deletion polymorphism data; (iv) visualizing sliding window results integrated with available genome annotations in the UCSC browser.

Availability: Freely available to academic users *here*.

BioRes Digest Editorial Team

1. Editor in Chief: Prof. Faisal M. Fadlelmola, Chair of H3ABioNet Research Working Group

2. Editors:
- Dr. Amel Ghouila, Co-Chair of H3ABioNet Research Working Group
- Dr. Sumir Panji, H3ABioNet Network Manager
- Kim Gurwitz, Training Officer H3ABioNet

Back to Contents for BioRes Digest

Back to H3ABioNet Newsletter: Back to Contents