Protocol for technical assessment of H3ABioNet nodes

Background

The H3ABioNet 2012 General Assembly in Cape Town decided to implement an assessment exercise designed to evaluate the preparedness of the network nodes in dealing with genome-wide datasets produced by H3Africa research projects. They created a Node Assessment Task Force and gave it a mandate to design and implement this assessment, with the help of an external scientific advisory board. The present document results from the work of the Task Force during its first year of existence and outlines the modalities under which the assessment will be administered starting in October 2013.

The initial brief of the Task Force was to evaluate the ability of the nodes to manage and analyze two types of datasets:

1. Whole genome or whole exome high-throughput sequencing data, typically produced from an Illumina HiSeq 2000 sequencer. The output of the analysis should be a list of sequence variants present in the genome of the individual from whom the DNA was extracted, relative to a reference human genome sequence.
2. Data from microarray-based genotyping platforms, typically Affymetrix SNP 6.0 or Axiom arrays, collected from a stratified population (e.g. affected vs normal individuals). The outputs of the analysis should be a list of properly called alleles, and a report on the population structure and association of alleles with phenotypes.

While it is likely that other types of datasets will be added in the future, e.g. microbial 16S rDNA amplicons from environmental microbiome samples, the Task Force has concentrated on the two initial types described above. This does not preclude a widening of the exercise's scope in the future. Any changes in scope will be announced to the H3ABioNet community in a timely manner.

Preparation of candidate nodes

Candidate nodes are strongly encouraged to prepare in advance of taking part in the exercise, to ensure a maximal likelihood of successfully completing it. Steps in this preparation include:

1. Choosing which of the types of data they would like to analyze, and downloading the relevant Standard Operating Procedures (SOPs) prepared by the Task Force. The SOPs include advice as to the hardware and software requirements for the analysis of the datasets.
2. Ensuring that they have access to a computational platform that is adequate to perform the analysis and on which the necessary software has been installed.
3. Ensuring that they fully understand how to run the software, including input and output formats and processing options.
4. Running some practice analyses on datasets to which they already have access, which could be downloaded from public repositories, obtained from collaborating research groups, or provided by the Task Force as practice data.
5. Ensuring that they fully understand the SOPs and the form in which the results should be delivered.

Members of the Task Force will be available at all times to help clarify technical issues as well as expectations. Once a node determines that it is ready, it should formally announce its candidacy.

Declaration of candidacy and initiation of the assessment

An *ad hoc* node assessment team of 2-3 members of the Task Force will be nominated for each evaluation, and a corresponding mailing list will be created and communicated to the candidates. The team shall not include any individuals from the candidate node. Candidacy of a node shall be declared in an e-mail to the team. The time stamp on this e-mail shall be considered the start point of the period allowed to complete the exercise. All further correspondence during the exercise should be sent to the same address. The declaration of candidacy should explicitly specify which type of dataset they wish to analyze, as assessments will be carried on a single data type at a time. Nodes are free to initiate overlapping assessments if they so wish.

Within 24 hours of receipt of the candidacy e-mail, the team will send to the candidate node, at the address used to declare its candidacy, a link to a secure site (with username and password) where the datasets to be analyzed can be downloaded. Additional files prepared by the Task Force and providing supplementary information about the test datasets may be included, as may checksum values to verify their integrity. It will be the responsibility of the candidate node to notify the Task Force of any problems encountered during the downloading of data files. The site where the datasets can be downloaded will also be used to upload the results of the analyses.

Performance of the assessment exercise

Each candidate node shall have six weeks in which to complete the exercise. During this period, they will be free to consult any available documentation provided by the Task Force or publicly posted on the Web. However, they are strictly prohibited from seeking help from other H3ABioNet nodes, and particularly those who have already successfully completed the assessment. Seeking such help during the formal assessment period shall be considered equivalent to cheating and shall result in the node being barred from taking the exercise for at least one year. Providing help, particularly if such help includes running analyses and/or writing reports, shall also result in sanctions against the offending node.

Once the analysis has been completed, the candidate node shall upload all relevant information, including a detailed log of the analyses performed, the relevant output files and a report detailing the conclusions drawn from the results, to the secure site provided by the assessment team. Updates can be uploaded until the end of the six week period, and nodes are encouraged to start uploading documents well ahead of this deadline. This will ensure that even if they should experience last minute problems, the Accreditation Board will have ample evidence available to document the work performed during the assessment period.

Evaluation of the results

The evaluation of the results will be performed by the Accreditation Board. The Board is composed of the Chair of the Task Force and external specialists knowledgeable in the types of data analysis requested during the assessment exercise. Currently (September 2013) the external specialists are Drs Brad Chapman (Harvard School of Public Health), Fran Lewitter and George Bell (Whitehead Institute, MIT) and Noah Zaitlen (UC San Francisco).

The reports produced by the candidate nodes will be made available expeditiously to the members of the Accreditation Board, who will be asked to provide their comments and evaluations within four weeks of having received them. Evaluation criteria include the accuracy of the results obtained, the soundness of the protocols used to obtain them, and the quality and clarity of the reports produced by the candidate nodes. The Board members will be asked to provide recommendations as to whether the candidate node should be accredited by the H3ABioNet network as being capable of analyzing the type of data that they received. Whether the final evaluation is positive or negative, the Board will provide detailed comments to the candidate node, with the aim to help them improve the quality of service offered to their scientific collaborators.

Awarding of accreditation certificates

Certificates will be awarded within a week of receipt of the Board's comments by the H3ABioNet network, under the responsibility of its coordination office and taking into account the recommendations provided by the Accreditation Board. The certificates will carry no official value beyond endorsement by the H3ABioNet scientific leadership, and shall not be construed to do so. In particular, they do not constitute a guarantee that a node will produce accurate, high quality results in the future.

If a node should fail to be certified after having undergone an assessment exercise, it will have the opportunity to try again, taking into account the comments of the Accreditation Board. However, it shall wait for at least six months before doing so, and it is expected that during this period it will review its operations and thoroughly prepare itself for a second exercise. The identities of nodes that have failed to obtain accreditation shall remain confidential.