

IHEC Ecosystems & Assay Standards Meeting 02-06-20

Attendees:

Thomas Juettemann (EMBL-EBI)
David Brownlee (McGill)
Paul Stretenowich (McGill)
Ksenia Zaytseva (McGill)
Sinead Aherne (CEMT)
Simon Heath (CNAG)
Daniel Zerbino (EMBL-EBI)
Garth Ilsley (EMBL-EBI)
Ted Perkins (uOttawa)
Sita (CEMT)
Martin Hirst (CEMT)
David Bujold (McGill)
Idan Gabdank (ENCODE)
Emilio Palumbo (CRG)
Jonathan Steif (CEMT)
Jose C. Marugan (EMBL-EBI)

May agenda & notes:

[IHEC Ecosystems & Assay Standards Meeting 05-05-20](#)

Agenda:

Data Ecosystems:

1. Metadata spec modifications
 - a. Review of action points
 - Thomas & Sita: test new error reporting on EpiRR
 - *Done*
 - DZ: update version number to 2.0
 - *Done*
 - DZ: Check EpiHK status
 - *Submitted*
 - b. Matters arising:
 - Error reports to partners: https://www.ebi.ac.uk/~juettema/ihec_validation/
 - New stats: <http://t.ly/j8To>

- c. Open issues (online discussion only):
 - UMI ChIP support: <https://github.com/IHEC/ihec-ecosystems/issues/82>
 - d. Long term issues
 - Unified submission: <https://github.com/IHEC/ihec-ecosystems/issues/50>
 - User testing of submission process: <https://github.com/IHEC/ihec-ecosystems/issues/35>
 - Outreach plan: <https://github.com/IHEC/ihec-ecosystems/issues/61>
 - e. Open Pull Requests (online discussion only):
 - Update Readme: <https://github.com/IHEC/ihec-ecosystems/pull/102>
2. Data Submissions
- a. Any upcoming data updates for EpiRR or the IHEC Data Portal <https://www.ebi.ac.uk/vg/epirr/summary> ?
 - b. EpiHK has been submitted.
 - c. McGill's hg38 aligned data is coming along (hg38 realigned version of original hg19 aligned processed data).
3. AOB
- a. Errors in past submissions:
 - CEEHRC/Blueprint largely associated to molecule attribute being placed in one sample/experiment
 - Many reports point to genuine issues.
 - ACTION: DZ investigate missing link between epigenomes IDs and samples (e.g. in KNIH 37 EPIRR IDs != 111 samples)
 - ACTION: TJ to separate Experiment and Sample counts in the error report (since they are separate sections of the IHEC data hub).

Assay Standards:

- 1. Sharing of processed data via sFTP.
 - a. The Vancouver centre has begun processing data through the ChiP container and will have the first batch of data ready for sharing in the coming weeks.
 - i. Update on input control sharing
 - b. Please review and comment on the following doc:
 - i. [IHEC Integrative Analysis Data Sharing CEMT_v3.pdf](#)
 - c. Martin goes through the document and highlights the decision tree.
- 2. Sign off on all 3 IHEC standardized pipelines:
 - a. [IHEC Standard Pipelines Code Review](#)
 - i. ChIP:
 - 1. Status - IHEC container complete & released.
 - 2. Paul Stretenowich, Sita, Ted Perkins -> benchmarking update
 - a. We are separating the singularity container issues from the pipeline itself in order to let us move ahead with the EpiAtlas analysis.
 - b. Documentation needs to be updated.
 - c. We will discuss how to go about making the containers more user friendly on an upcoming call (Action Item).
 - 3. Update on compute requirements -> Can the container be run on a PC?

4. Paul has some outstanding issues running the container in a compute Canada environment but has not had much time to dedicate to this.
 - ii. GemBS:
 1. Status - Pipeline complete, IHEC container to be finalized & released.
 2. Feedback - CEMT, EMC
 3. The final steps ahead of sign off on this pipeline are for CEMT and McGill to compare their md5 checksums and choosing whether the bigwig files should be set as a single file or split out the strand information.
 - iii. RNA:
 1. Status - Pipeline still in development.
 2. Feedback - CEMT, EMC
 3. Emilio published a new update for the pipeline this morning, the CEMT and EMC groups will continue testing and providing feedback (**Action Item**).
3. Hi-C standards:
 - a. Hi-C core metrics approved by the working group in February:
 - i. [IHEC Hi-C QC Metrics](#)
 - ii. A detailed description of the adopted metrics: Section - II.a. Quality Control <https://www.encodeproject.org/documents/75926e4b-77aa-4959-8ca7-87efcba39d79/>
 - iii. Calculation code: <https://github.com/aidenlab/juicer/blob/encode/SLURM/scripts/juicer.sh#L1105>
<https://github.com/aidenlab/Juicebox/blob/master/src/juicebox/tools/ct/old/LibraryComplexity.java>
 - iv. Next steps?
 1. The next step should involve considering how to incorporate Hi-C into an updated version of the 'IHEC Assay Standards Working Group Recommendations Nov2012-3' document which is housed on the IHEC website: <http://ihec-epigenomes.org/research/reference-epigenome-standards/>
 - a. The best way to do this would be for someone from the ENCODE or 4DN group to begin drafting a Hi-C section to add to the doc. There was no one on the call to ask so Martin will raise the issue on the next ISSC and reach out to Erez and Ananda.
 2. We should also consider how to integrate this data into the Integrative Analysis efforts.