Report from the inaugural PomBase Scientific Advisory Board meeting
Date: 14 October 2022
Location : online

Attendees
SAB
Li-Lin Du (chair) National Institute of Biological Sciences, Beijing, China
Kathleen Gould Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN, USA
Sabina Leonelli Exeter Centre for the Study of the Life Sciences & Department of Sociology, Philosophy and Anthropology, University of Exeter, Exeter, UK
Samuel Marguerat MRC London Institute of Medical Sciences, London, UK & Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK
Sophie Martin Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland
PomBase
Valerie Wood, Project Manager/Curator Co-PI (Cambridge University)
Kim Rutherford, Software developer (Cambridge University)
Manuel Lera-Ramirez, Curator (UCL)
Jurg Bahler Co-PI (UCL)
Juan Mata Co-PI (Cambridge University)
Jaqueline Hayles, Community Curator (The Crick Institute)

Website Suggestions/Improvement

JBrowse
The RNAs in the JBrowse context view outnumber the proteins and can be a little distracting.
  ● Suggest: Make RNAs less prominent **Action item**

Gene Page, top of the page focus
The proposal to change the focus at the top of the page from JBrowse sequence context view and to include more functional display options (structures, pathways and decorated protein views) was well received.

Gene Page, literature section
The SAB felt that the literature should include all known literature on a gene, not only curated literature.
  ● Suggest: Load all PubMed articles for a gene **Action item**
  ● Suggest: Change the default literature from 'alphabetical author" to lowest "gene number" so that gene focussed publications are displayed first **Action item**

Gene Page, navigation
  ● Suggest: Improve the menu bar contents and functionality to make it clearer what the data sections are **Action item**

Gene page, summaries.
- Suggest: try to solicit gene page summaries from the community.
  - PomBase began this a few years ago, and a few summaries were collected, but did not follow up very forcefully.
  - An automated summary first pass was proposed with an option for researchers to adopt and improve and maintain the automated summaries **Action item**
    - Would need exemplar templates to show what is required.

Community Curation suggestions
Better instructions for community curation. There was a general feeling that the initial learning curve is steep but once people have done it once it is much easier. Need more detailed instructions so that lab heads do not need to train lab members.
- Suggest: Embed how-to video on front page of Canto and add link in emails from Canto (make videos and look into JoVE Peer Reviewed Scientific Video Journal for publication) **Action item**: Manu is making Canto videos as part of an Elixir data steward fellowship

Novel/upcoming areas to pursue
Solicit collaborations with external groups working in AlphaFold and AI/ML generally
- Contact groups doing interaction/complex prediction/ docking with AlphaFold to collaborate
- Try to foster collaborations with groups like 'DeepMind' whose next focus is towards modelling all aspects of biological systems. **Action item**: Contact the Deepmind at the Crick to see if our extensive datasets are useful for modelling projects

Datasets to host
- Long read transcriptome (Jeff Boake lab) would complement other transcriptome data (NOTE I Can't find this paper Action item

Metadata and release notes improvements
      • A web page explaining the relationships between PomBase version number, NCBI version number, and Ensembl version number would be useful (Note: need to distinguish between sequence version, and annotation version, I am not sure that Ensembl has an annotation version number)
      • Report sequence feature changes accompanying each release Action item

Quality Control
      • SAB generally agreed that focusing on fixing problems before including new data was a good idea

Nomenclature
      • General agreement about preceding nc-RNAs with "nc" (**Action item** check this proposal with the GNC)

• Improving guidelines for genetic nomenclature reporting was supported. General consensus that we should mandate how alleles should be labelled and described in PomBase (non-conforming syntax would be retained as synonyms). PomBase will pursue publication of updated guidelines in consultation with the GNC.

Sequence features
      • Consider using published Deep CAGE data to represent 3'UTRs (**Action item**: In progress, almost complete)