

Discussion Topics

1. A gold standard TE annotation: is there a need?
 - a. For improved annotation of genomes.
 - b. For better benchmarking of annotation tools and approaches. Would benchmarking be an important component of TE annotation? How?
2. What regions would be annotated, and how?
 - a. Which genomes? Model systems only? Which populations?
 - b. Is good gene annotation prerequisite? What other prerequisites?
 - c. Number of regions, length, choose randomly or include categories (e.g. TE-dense vs. gene-dense)?
 - d. How large does the benchmark need to be, in terms of species and amount of annotated DNA per genome?
3. Reaching a consensus:
 - a. How many groups should annotated a given region?
 - b. How should differences be resolved or represented?
4. What data would be included in the annotation? E.g.,
 - a. Evidence, certainty
 - b. TE type, copy-#, family
 - c. Nesting, transduplication/transduction, etc.
 - d. Age, lineage
 - e. Gene annotation errors
5. Annotation tools: what is needed?
 - a. Existing tools per group, common tools, or new tools?
 - b. Data integration:
 - i. Just merge using, say, BED files?
 - ii. But how to display annotation details such as evidence?
6. Crowd sourcing?
 - a. Sufficient expert time not available?
 - i. How much would be needed?
 - b. Could non-experts produce reliable results?
 - i. E.g. students or general public.
 - c. Approaches
 - i. Test users on previously annotated regions.
 - ii. Cross-validate different participants' annotations.
 - d. Tools

TE Annotation Workshop, 18-25 April 2014

McGill University, Bellairs Research Institute, Barbados

- i. What additional tools are needed to harness crowd sourcing?
 - e. Connect with more groups with crowd-sourcing experience for ideas.
- 7. White paper (workshop report)
 - a. Goal: to inform broader genomics community on state of TE annotation, need for gold standard (if there is one).
 - b. Plan for drafting paper.