2020 GA4GH Connect Virtual Meeting

Sequence Annotation Breakout Agenda

Details subject to change.

Meeting Goals: Converge strategy on sequence annotation

Relevant Work Streams: GKS, LSG

Chairs: Sean Upchurch, Andy Yates

Notetaker:

Report Back Slide

Thursday, March 26, 2020 at 12:55 - 13:50 UTC

Approx Time	**Discussion Topic**	**Related Materials**	**Speakers**
1 min	Designate notetaker	None	Sean, Andy
25 min	RNAseq team plans on sequence annotation - Expanding Use Cases documentation - API design features	Master Slides	Sean
25 min	GKS's plan on sequence annotation - Suitability (or lack thereof) of existing formats (e.g. GFF) - Scope of annotation (genes? Splicing? etc.) - Integration with VA and VR components where possible	GitHub Issue from VA	Andy
5 min	Project governance - Which Work Stream will manage this product?		Sean, Andy

Minutes:
RNASeq discussion
YF: We needed for bulk RNA and functional annotation
LB: Writing valid GFF with additional constraints on order and implementation might be useful
YF: Phenotypes (suggested as metadata in original discussion) might be better off as a separate service
- SU: Probably viable as a future update than a initial item
- YF: Connection between sequence and the function

GKS Discussion

Michael Stromberg - when working with RefSeq transcripts, need to have a use different sequence than underlying genomic sequence. Requires a transformation. GTF/GFF3/BED are not ideal. Would suggest baking this into the format.
Mark Diekhans: Yes, fundamentally what is happening is a pairwise alignment problem. They annotate RNAs and project them onto the genome. Is it evidence behind the annotation, or with the annotation itself? They often construct the sequence from GenBank in the RNA Space, not DNA. To me this is evidence behind the annotation rather than the annotation itself.
AY: Is this proposal going to be useful here?
MD: We lose information as we can't represent everything. e.g. Perfect genome is false assumption. But on the flip side 90% of users probably get what they need.