2021-11-16-21 Best Practices Meeting Notes

Date

18 Oct 2021

Attendees

Committee Member	Present	Absent
Kuffel, Gina (NIH/NCI) [C]	X	
Musk, Philip (NIH/NCI) [C]	X	
Pihl, Todd (NIH/NCI) [C]		X
Unknown User (hechtt)		X
Kim, Erika (NIH/NCI) [E]	X	
Sommers, Connie (NIH/NCI) [E]		X
Elaine Ostrander	X	
Deborah Knapp	X	
Jeff Trent	X	
Dawn Duval	X	
Anju Singh		X
Amy Leblanc	X	
Heather Gardner	X	
Cheryl London	X	
Christina Mazcko	X	
Paula Jacobs	X	
Renee Chambers	X	
Shaying Zhao	X	
William Hendricks		X

Outstanding Action Items

- [x] Elaine to send link to new GSD reference genome.
- [ ] Elaine to schedule Reuben Buckley to present at a future meeting
- [x] Kuffel, Gina (NIH/NCI) [C] to send out Buckley paper.
- [x] Kuffel, Gina (NIH/NCI) [C] to set up presentation with Shaying's group for next meeting.

Discussion items

Item	Who	Notes
Welcome & Introductions	Gina Kuffel, Shaying Zhao	

Pan Cancer Study Recap	Shaying Zhao	• Breed prediction pipeline • The 15TB harmonized reads are generated from cases with matched tumor and normal samples both passing our comprehensive quality control measures, including breed validation and prediction using our software provided
Canine MHC-I Genotyping Software	Yuan Feng	
Canine Breed Prediction Pipeline	Kun-Lin Ho	

Committee Goals	Gina	• Variant filter file
		  ◦ Some of the filters are not well understood by the group
		  ◦ A discussion is needed to identify overlap and differences in variant calling filter files being used by various groups
		  ◦ Dependent on genome best practices for calling germ-line mutations
		• Best Practices publication
		  ◦ Initial Outline on Box: https://nih.app.box.com/folder/137612017625
		  ◦ Elaine to potentially take the lead on this initiative
		  ◦ A milestone driven outline with dates and assigned tasks is needed
		  ◦ Set to begin work on this effort by mid August
		  ◦ Target date for submission is October 1, 2021

Meeting Minutes (Not Verbatim)

EO- I sent an email out to Laurna Kenby at the University of Manchester who does a ton of work with Class II DLA, no clinical data, but population data is available. Massive numbers of alleles that are not found unless you have very large cohorts. We could put together a breed portfolio on class I. Scope of the dataset is extraordinary.

SZ- Connecting us with Dr. Kenby could be really helpful.

Previous Meeting Minutes (Not Verbatim)

EO- Link to Michigan pipeline will be available soon. I can share with group when I have it.

RC- About to do sequencing on a chunk of dogs, the imputation paper could be a great reference.

EO- Helpful with low-pass sequencing.

EO- There will be a release of all variants. Everything will be re-called since they aligned to the new GSD reference.

EO- We can have Reuben Buckley attend the meeting next time and give a Q & A. He is also in the alignment group.

CL- Most dogs are not purebred, have we looked at alleles in mixed breeds?

SZ- The software does not care if the sample is from a purebred or a mixed breeds, it only requires fastq files from RNA-Seq.

2nd Presentation
EO- If you are looking at more closely related breeds it becomes harder and harder to discern variants unique or enriched in a certain breed. Have you explored the published clades?

KH- We haven't yet explored closely related breeds specifically.

EO- It's all published and I can send along some of those publications.

CL- Trying to wrap my head around how to interpret WGS and WES data and variant call files? Do we need to have tumor matched normal and how deep do we need to go with sequencing to ensure we aren't calling variants that are just breed specific.

KH- Tumor and normal samples are included in our pipeline. Without it we can't rule out sequencing and alignment errors. Without it we would also have computational expenses. Sequencing depth must be at least 30X.

Action items