American Cocker Spaniel Cataract Study – brief progress report

University of Pennsylvania, August 21, 2023

Personnel: ***University of Pennsylvania***
Gustavo Aguirre, VMD, PhD; Principal Investigator
Leonardo Murgiano, PhD; co-Principal Investigator
Anil Sigdel, DVM, PhD; Postdoctoral Researcher
Jessica Niggel, M.Sc.; Research Scientist

In this installment of the American Cocker Spaniel (ACS) Inherited Cataract study report, we would like to again thank all the ACS breeders and owners for their continued support and participation of our study. Compared to the previous report, the number of dogs participating in the study increased from 1005 to 1013. Additionally, we received updates on nine dogs.

In our search for the detection of the causative variant(s) for cataract in ACS, we encountered issues as the study progressed. These were very likely related to the complex inheritance of the disease, and of the low informativeness of the SNP data for the type of population studied. or this reason, we opted for a radical approach putting whole genome sequencing (WGS) as a priority and this approach has recently brought very significant results.

In our last report, we stated that we sequenced 12 very informative early onset cases, focusing on all the variants these shared. Once these were detected, these results were filtered against two different datasets – the canine WGS files generated by our lab, and against a recently accessible WGS dataset that comprises now of more than two thousand dogs (databases increase in size over time because scientists keep adding dogs to this coordinated effort). After this first round of WGS of early onset cases, we analyzed the variants detected, we filtered them, remaining with variants in the order of tens (which is, for a whole genome of millions of variants, extremely low). After interpolating the WGS with the last GWAS mapping info obtained, three were selected as the most promising.

Previously, the cases were added as we went, analyzing the preliminary output and then evaluating if more were required, and this was done on an iterative manner, but the new lower costs for WGS allowed us to send several additional samples at once.
As an expansion to what planned in the last report, we opted to send eight additional cases. This was done for several reasons:

(I) Reduce the number of variants to be tested by mass Sanger sequencing, now to be carried out on the whole available population (hundreds of dogs, manually intensive).

(II) Avoid any false positive – we wanted a stronger association with the mapped region and confirm that analysis.

(III) Avoid any false negative – we wanted to be sure that other regions deemed less likely, but still possible, weren't excluded too early in the analysis.

This was done through a collaboration with a colleague able to sequence several samples at a time with his high throughput sequencing instrument platform. Our collaborator also has access to a specialized exclusive dataset that would make the filtering even more effective. While the data was ultimately delivered, due to unforeseen technical issues the turnover time suffered significant delays, so we were able to carry out our side of the data process only very recently, therefore delaying the subsequent Sanger sequencing and this report.

Nonetheless, the data from 20 early onset dogs were extremely informative and the results are excellent. After a first round of WGS of early onset cases, we analyzed the variants detected, we filtered them, and we grouped them by impact and gene function. The results were again filtered against two different datasets – the canine WGS files generated by our lab, and against the WGS dataset comprised of more than two thousand dogs. Basically, every marker shared by all the cases was checked on this large population of dogs. If it was present in high numbers, especially in non-cocker dogs, it was excluded because unlikely to be the disease causing one.
What was detected in our last iteration and pointed out in the previous report was mostly confirmed and additional "noise" was removed.

In the previous report, we planned to carry out a similar approach for the same dataset but checking for structural variants. These larger, structural variants are modifications that affect a longer strip of DNA, and some don't affect a gene directly, but only how its function is regulated.

These require a more complex analysis to detect and filter. We managed to run this variant detection tools and we added 3 large structural variants to the small pool of polymorphisms detected. Even in this case, we are confident about these results due to the large pool of dogs used.

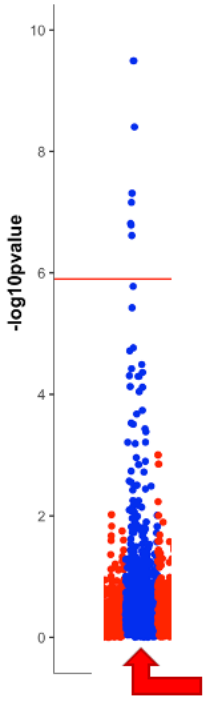

Fig 1 - Whole genome sequencing and filtering for markers shared by all cases gave us a short list of variants to explore. Our list is overall (but not 100%) consistent with the most associated regions detected by GWAS. In the figure, a GWAS peak is shown with a red arrow indicating a close, extremely rare variants filtered with WGS.

For what concern early-onset cataract, the only action forward will be Sanger sequencing of the entire available ACS population databank for all these candidate variants. As always, the order of priority is based of the predicted functional role and impact – more "suspicious" variants will be tested (or will continue to be tested if the sequencing already started because of previous, partial results).

Due to the delays experienced with the collaborator, unfortunately we weren't able to repeat the whole procedure described above with later-onset cataract ACS dogs – we needed to look at the new WGS first. Now that we have the data and it is being analyzed, and that we assessed that this strategy is correct approach, we will send for WGS the selected later onset cases suitable for this high-throughput process. We will search for shared regions, for region exclusive for the later onset, or for a combination of both – as an example, any putative regulatory region that could have a role in determining the age of onset.

To avoid delays for technical reasons, we will submit the new samples to a newly available sequencing platform from a company specialized in animal genetics, at a still reasonable cost. The hoped for collaboration with the colleague has not worked out to our expectations as the reporting of the data has taken weeks-months, and we are unable to carry out the studies with such delays.

Other techniques offered by the company in support of complex-inheritance diseases which could be used for the search of any additional locus if present; these are currently under evaluation.

On the breeders' part, we once again stress the critical importance of updates and new samples. Lack of updates endangers the quality of the dataset and therefore the results. We also ask any owner of non-affected, older dogs to send samples for the study. We think that due to the nature and inheritance of the condition, non-affected controls are vital. Positive feedback and large numbers are the keys to success.

Total Dogs	1013
Total Informative dogs	610
SNP-Genotyped Dogs (informative)	175
WGS Dogs	24 (20 E.O.)
Dogs to sequence in the future	10-15 (estimate)

Steps to be taken in coming 6 months:

- Confirm the WGS analysis, also taking into account allele frequency and possible multiple exclusive loci – the strategy has been already planned for that.
- Complete the Sanger sequencing of the whole sample pool.
- Proposal of a marker for early onset cataract in ACS if the Sanger sequencing results pan out.
- Whole genome sequence of the later onset cases.
- Evaluate additional strategies for the discovery of the later-onset locus and marker (new techniques are available).