American Cocker Spaniel Cataract – brief progress report

University of Pennsylvania, January 22, 2024

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 1013 to 1020, and we received updates on 4 dogs.

In our last report, we described the issues we faced due to the complex inheritance of the disease and of the low informativeness of the SNP data, and we explained out choice in investing heavily in whole genome sequencing for the type of population studied. In two subsequent waves we sequenced 20 very informative early onset cases, and we filtered all the variants these shared. Once these variants 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. 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 one we are looking for.

The high number of sequenced cases was chosen because of the following factors:

(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.

While the data was ultimately delivered by a prior collaborator, due to unforeseen technical issues, primarily poor scheduling and failure to keep to scheduled timetables, the turnover time suffered significant delays, which slowed down our progress. Ultimately, we gathered a list of small number of structural variants we could use as good candidates. Previous GWAS results were encouraging and pushed us toward focusing on three main regions detected by this association analysis. These regions, once analyzed, also contained genes that were very plausible candidates, and we were (and still are) excited with the results.

Unfortunately, the accurate identification of good markers for each region became less easy than expected once we searched in the general population. Currently, while we aren't excluding any

possibility, we are not at the stage in which we can declare with confidence a marker for early onset cataract.

In light of the results of the early onset cataract, we decided to proceed with a technique that was very useful for other projects by our group which also involve complex inheritance. As often stated in previous reports, two of the major issue in our search for causative markers were (high quality) sample numbers, and information density of the SNP chip. The latter means that for the project, the SNP data had too few markers to have a significant signal in our associations.

For this reason, we opted to genotype again our dataset, this time using the low pass whole genome sequencing a technique carried out by the company Neogen under the name Skimseek. The principle of this technique is to use whole genome sequencing on a given sample, but at a fraction of the cost. This low cost is because only a few sequences of DNA per position are read instead of the 30+ used for "proper" WGS.

This information is not used (and cannot be used) for direct variant discovery: the aim of this data, which is then put together and imputed by the company carrying out the analysis (therefore integrating the raw data and making it even more complete and informative), is to create a dataset comparable to a very "dense" SNP chip.

This denser marker dataset has many orders of magnitude more markers than a regular SNP chip - in the millions, in fact, instead of the "usual" hundreds of thousands of the old technique. The cost of this newer method is low (less than $95 per sample) well below that of the normal SNP chip . The samples used for this experiment were 27 late onset, and 80 controls. The results obtained by the analysis are very encouraging, with a very stark signal in one chromosome, and a few minor ones in others (Fig 1).

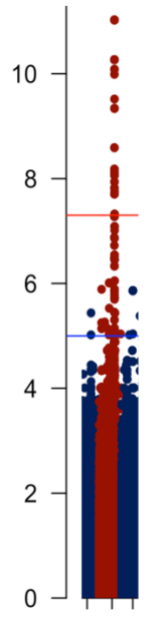

Fig 1 - Low-pass whole genome sequencing GWAS of late-onset cataract in American cocker spaniel. The peak shown is the "major" one discovered with the low-pass Skimseek technique. The strength of the association is the strongest we have found so far.

In the next step, we looked at the candidate regions and in the late-onset cases for whole genome sequencing, using dogs sequenced in both as reference. Low-pass Skimseek for 40 early onset selected cases is now in the works. All this data is planned to be integrated together.

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	1052
Total Informative dogs	614
Skimseek sequenced dogs	147
WGS Dogs	24 (20 Early onset, 10 Late onset)
Dogs to sequence in the future	0-5

Steps to be taken in coming 6 months:

- Analyze the early onset low-pass data and run all the GWAS with that new data.
- Integrate the early onset data with the early onset whole genome sequencing.
- Merge the early and later onset Skimseek datasets and run a general analysis (albeit we predict later and early onset to be two separate problems).
- Analyze the later onset high-coverage whole genome sequencing.
- Restart of the variant genotyping in the general ACS population in light of the new results.
- Proposal of a marker for early onset cataract in ACS.