Progress Report Summary

Time period 7/1/2020 through 12/31/2020

• Constant revision of the phenotypes and repetition of the GWAS. Cases (~44) and controls (~85) were marked for their haplotype and the haplotypes defined. Several combinations were used, adjusting for sub-population, sub-phenotype and the like.

• 2 major regions were identified, plus others were taken into account as co-present with a given phenotype or sub-phenotype.

• During the pandemics, we selected of a number of suitable samples to submit them for whole genome sequencing (WGS). These included 2 high quality cases (2-5 yr), 2 older category cases (5-8 yr) and 4 top-tier controls.

•Once the WGS data was received, mapping and analysis of the variants carried out. Variants were checked whether they were present in the same haplotype but exclusively for the cases; or present in the cases and controls and therefore probably causative due to incomplete penetrance. These were the haplotypes pointed out above.

• In order to be sure we were focusing on the right region, we opted for genome wide association study (GWAS) using a higher density SNP-chip. We selected the best cases and controls and genotyped them on a 712k SNP chip, as opposed the usual 220k. We sent 27 cases and 38 controls for the high density chip analysis.

• When the data came back, we ran a 712k only GWAS and noted the results. Then we merged all the data, imputed the missing dog, and ran additional GWAS with ALL the dogs genotyped so far. We also explored different GWAS and homozygosity mapping with different software, compared the results, annotated any discrepancy (not many). Results with 712k chip were in accordance with the 220k chip, and suggested that we could increase power of the the study by genotyping more dogs with the lower cost chip.

• We used the new data to go back to the WGS data. Looking at the data we realized that too many variables were in play as shown by too large of a pool of candidate genes/loci and we need more WGS and samples.

• We identified more samples to genotype again on 220k chip and an additional 12 samples are now being genotyped.

• We communicated with University of Bern in Switzerland to confirm that their sequencing platform is now back in operation after pandemic related closure. This will be important in next phase as cost is much less expensive and quality is comparable or better.

Goals for the coming time period: 1/1/2021 – 6/30/2021/

• Data for the new batch of SNP genotyped dogs (GWAS) has been received. It is being processed to reanalyze and incorporate the new genotypes. As well, all the phenotypes will be re-assessed to ascertain that we have the "cleanest" data possible.

• Samples have been selected for the additional WGS. 6 cases and 6 controls. This time we will focus on the top-tier cases and no older onset dogs. We had planned in doing this late December/early January but decided to wait to include additional dogs from a separately funded and unrelated study as we received a discounted rate for a larger sample set.

• Once this new SNP chip dataset is analyzed, it will confirm the candidate regions and, and hopefully reduce the size of the genomic region. This will save time and resources during the mutation sequencing and filtering future phase apt to find the gene/markers.

• We keep looking for more samples to genotype again on 220k, building up more batches of 12. On this part, breeders' contribution is essential. More dogs = better statistics.

• Once the new WGS data is available, we will re-map it. The mapping will be carried out on more than one canine reference (canfam 3 and newly released canfam4) using a revised alignment method we just optimized to reduce "noise". Then the variant hunting will be carried out in the candidate regions. Additionally, this greater dataset will allow us to attempt other comparison methods using WGS.

• The results will be used to hunt for markers that we will then test with sanger sequencing on our whole ACS dataset.

Molecular Genetic Studies of Inherited Cataracts in the American Cocker Spaniel - progress report

University of Pennsylvania, Feb 11, 2021

Personnel: ***University of Pennsylvania***
Gustavo Aguirre, VMD, PhD; Principal Investigator
Leonardo Murgiano, PhD; Senior Research Investigator
Jessica Niggel, M.Sc.; Research scientist

Objectives

The principal objective is the identification of the genes and genetic variants responsible for inherited cataract in American Cocker Spaniels (ACS). Our group was granted funding and support by the ACSF to map and characterize the genetic origin of the cataract in American Cocker Spaniel. The ultimate goal of our project is to detect a suitable genetic marker to be used for a breeding test that would help prevent the production of ACS puppies with this condition.

In the initial stages of our investigation, we were assuming a simple disease scenario for the cataract in ACS, with an easily identifiable inheritance mechanism and therefore easily detectable markers. However, further studies have revealed this to be incorrect. Within this report we describe our findings, the progress achieved, and the challenges faced, as well as explain our recent proposal and future goals.

Background

Cataracts are the most common cause of vison impairment in humans and other mammals, and are a frequent ophthalmic disease also seen in dogs. Several breeds are affected by such condition, included the American Cocker Spaniel (ACS), with an estimated prevalence of 8-11%. Such percentages describe both acquired and inherited cataracts: the latter category contains cataract phenotypes that are clinically similar but may have a different genetic etiology and only a superficial clinical similarity. Data gathered to this point support such interpretation for the ACS. ACS dogs with inherited cataracts are born with normal lenses, which then proceed to opacify over time, leading to blindness by 2-10 years of age.

The mechanism of inheritance in ACS has been previously proposed as being

autosomal recessive, but our subsequent observations suggested a situation more complex than the one predicted in the preliminary phase of the project. We have observed that the most likely mechanism involves the presence of potential risk factors based on the sub-populations studied.

As stated previously, a significant element in the progress of our project is the thorough classification of suitable and verified samples in the ACS population. Our constant re-analysis and update of the cases and controls present in our database would not be possible if not for the outstanding cooperation of the ACS owners and breeders. This allowed us to pinpoint specific areas of the genome associated in varying degrees with the condition, and to refine such association with each iteration of the analysis (excluding false positives).

Our final aim remains the identification of gene(s) and vulnerability loci associated with the most common form of cataract in ACS and on validating its inheritance mechanism. We achieved such analysis of the database through tight communications with the owners and the breed club. After reaching a sufficient number of samples, we planned and executed the use additional resources and techniques in order to move the project forward. Nonetheless, we concluded that the complexity of the issue requires additional data in order to pinpoint the exact genetic cause of the condition.

Cataract in ACS – nature of the samples

Cataracts are often inherited conditions. They are characterized by opacity/cloudiness of the lens, arising due to lens protein misfolding, solubility changes and aggregation leading to vision impairment of progressive severity, occasionally demanding surgical intervention. American Cocker Spaniels are among the most commonly cataract-affected dog breeds.

As previously reported, we acknowledged a spectrum of cataract phenotypes differing in location, progression rate, whether they are unilateral or bilateral, genetic background and age of onset. We considered the latter parameter, above the rest, as a most crucial factor for the classification and grouping of our samples. Specifically, inherited cataracts in ACS are thought to appear sometime around 2-5 years of age and progress. Nonetheless, we have found a subset of cases where cataracts, presumably inherited, begin between 5-9 years of age.

We stressed for a correct gathering of information about the affected and unaffected dogs and for a precise assessment of the phenotype and the selection of a good control sample group. As stated previously, this is essential in order to select candidate cases for cataracts predictable as having a genetic etiology.

Cataracts can be caused by environmental effects such as UV light exposure, mechanical trauma, poor nutrition, or exposure to toxic substances. They can also occur as secondary effects of other ophthalmic diseases, such as uveitis or glaucoma. We used the maximum care in excluding any possible secondary cataract phenotype with a high likelihood of not having a genetic etiology, and thus lowering the quality of the dataset.

Research on genetic diseases in companion animals

Current research in genetic diseases in domestic animals is based on three main principles: (I) Construction of a suitable dataset, obtained through the identification of cases and valid controls (II) Mapping of the variants associated with the condition studied (III) Validation through sequencing.

The importance of (I) is described and explained in the above paragraph, and we previously described the steps that have been made thanks to this approach, included the addition of new cases.

(II) is generally achieved using SNP genotyping. The method uses purified DNA, preferably obtained from blood samples of cases and controls, that is placed on 'chips'-specific platforms scanned for strategically selected genetic variation markers, called single nucleotide polymorphisms (SNPs). Thorough the information obtained by such experiments, we can explore the presence of common (and ideally, exclusive) shared regions among the cases. Such region could be, as an example, common homozygous intervals (as it happens in recessive diseases). Analysis of markers inherited from parents and identical by descent can even pinpoint shared linked interval in heterozygous regions of the chromosome (as in dominant diseases). Research is constantly trying to improve such technology with denser chips, that increases the amount of information contained.

Another common type of analysis is the Genome Wide Association Study

(GWAS) that uses the SNP chip platform. Such study pinpoints higher frequency of certain SNPs in cases vs. controls, associating these variations with the disease. GWAS can be implemented on a wide population of dogs with reasonable computation time, and regardless of the family information about the samples. Moreover, GWAS can better predict variable degrees of association of a locus with the condition, giving away vital information in the investigation of a more complex inheritance mechanism. In fact, GWAS has been a vital part of our approach, since there is no perfect segregation of the markers between cases and controls. Often, the dataset generated for GWAS analysis is also used to search shared homozygous regions among the cases, which we routinely do in our analysis albeit so far, we did not detect any detectable trends.

Sequencing (III) consists, in general terms, in the determination of the exact DNA sequence of a given genomic region (of variable size, included a genome in its entirety). A common and fast sequencing method is the Sanger sequencing, used for the comparison of candidate mutations in cases and controls (that is, to validate whether a given mutation is associated with the condition, thus possibly being the causative one). Sanger is often used even for the development and execution of a genetic test for a disease. We are working on its complete implementation on specific candidate markers in order to assess their frequency and segregation (between cases and controls) of several candidate variants in our population.

A limited, targeted use of Sanger sequencing is relatively cheap, but the exploration of a whole genome sequence would make it unfeasible and too expensive. On the other hand, Whole Genome Sequencing (WGS) methods have brought a whole new level in the exploration of genetic defects, because they allow us to obtain detailed information about the genome of a sequenced animal. WGS is particularly useful when the sequencing of a high amount of candidate variants in one or more cases would be time and cost prohibitive if done using more conventional approaches. Additionally, the cost of WGS is decreasing over time – the analysis of a single genome dropped by $1/5^{th}$ or even $1/10^{th}$ (depending on the specific methods) in the last 10 years. This is especially crucial when a lot of data is required due to the complexity of the inheritance.

An ideal scenario in the study of a genetic defect involves the use of SNP chip for the mapping the disease to a specific chromosomal region and sequencing a putative candidate gene(s) for the validation of the data once the genomic region is identified. Even in case of more than one associated/implicated region, a careful evaluation of the samples selected for WGS, a consistent dataset and a high number of controls can finally unveil the genetic etiology of the disease.

Summary of the previous work (and progress to date):

COVID-19 and ACS cataract research:

While the current COVID-19 pandemic situation unfortunately is still ongoing, thanks to the careful planning of the University of Pennsylvania our working schedule is now restored. Although research has resumed, there are some restrictions observed due to the pandemic that do somewhat slow us down. While our team consists of 3 people, our in-person interactions are limited due to COVID-19, and all of our meetings are done through video conferencing like “Zoom” or “Skype”. Shortages and backorders in PPE, certain chemicals, as well as in single use labware do prove to be challenging sometimes (although the situation has improved considerably since the start of the pandemic). Waiting times for some of the backordered items are long-sometimes up to a few months. However, with our determination to push the project forward while still observing the pandemic health and safety rules we do manage with making our own reagents and buffers (even making our own PPE in the early stages of the pandemic), and we have scanned all the incoming paperwork and classified it while meeting through Zoom. Even during the initial lockdown phase where most research activity diminished, we were able to send samples to be run in the high density SNP chip and continued some of the WGS analysis on samples already collected and run. Additionally, we renewed our contact with the University of Bern and their sequencing facility, now back in function.

We implemented several strategies during this period of the study. As stated previously, the choice of a given approach was done based on the quality of the dataset available at the moment, and the reliability of the information. The recent batch of new samples (or updated information) improved the dataset on each subsequent iteration.

Candidate genes and Pedigree analysis

While in the ongoing process of collecting sufficient samples needed for detailed genomic studies, we carried out a preliminary candidate gene analysis in order to exclude more obvious genes. As reported, the results were negative, and we found no associated variant in those selected genes with the cataract phenotype (for more details about these results, see the previous Progress Reports). Nonetheless, the functional role of genes showing to be mutated in WGS data is indeed taken into account when

the variant has to be considered as a candidate for testing.

While studying the family trees, we hypothesized that an autosomal recessive inheritance is at play, and that such model would explain at least a significant part of our cases. Nonetheless, a deeper analysis of the data suggested that a common, shared genetic variants causing *all* the genetic cataracts in the ACS population is unlikely.

Samples received

Compared to the previous report, the number of dogs participating the study increased to 869 from the 831 reported last time. A short breakdown of the samples follows:

Total dogs	**871**
Total of Informative dogs	**565**
Potential cases	***112***
Bilateral	80
Unilateral or very Asymmetric	32
Controls	***453***
Too young to be properly assessed for study inclusion at this time	225
Total of Excluded dogs	**287**

Table 1 –Total of dogs included in the dataset. Count of dogs that are sufficiently informative, type of cases, potential controls and dogs not suitable for the study. Causes for exclusion: co-morbidity with another eye condition, doubts about diet, the dog prematurely deceased (especially if DNA/blood is missing), lack of feedback on updates (fortunately, this now is a very rare occurrence), lack of an official diagnosis by a certified veterinary ophthalmologist (or of monitoring post diagnosis), inconsistent records (very rare occurrence). Of the dogs shown above, only the ones with consistent records over time can be genotyped!

DNA samples were isolated from blood or buccal swabs by personnel at OptiGen LLC who previously collaborated in the study (in this regard, we wish to thank the breeders for the fact that the overwhelming majority of samples are blood samples, easier to work with and generally bringing better DNA yield). All of the blood samples have been sent to us in EDTA lined tubes, to prevent clotting. We extracted the DNA from the blood samples of cases and controls considered suitable for the study. In the

recent years, >95% of the received samples were blood samples. We have to report that this is appreciated since the DNA yield and quality from blood extraction is drastically superior and is useful if multiple genetic tests have to be carried out.

Phenotype reassessment

We previously reported the development and use of a standardized eye exam research form. We wish to stress again that the forms are extremely useful and important to the study, we have noticed that still not every veterinary ophthalmologist will use them. This has been a problem as the forms used-OFA-CAER-are inadequate for consistent diagnosis. A proper form can be downloaded through the following link:

https://drive.google.com/file/d/1N-oFJUM3kCjGBQsCPvsuY2xId13Cd5ni/view?usp=sharing

Clicking on this link will direct to a page with the document. It can be downloaded (top right) and/or printed, and a pdf copy is included with this report. **Please note: This is an updated version of the link and form (Jan 2021).**

Each time new samples are added, and a sufficient number of updates is gathered, we analyze the new information and re-classify the dogs. We make use of our carefully organized archive and classify the samples as Cases, Controls, Excluded (due to the phenotype being probably explained by a non-genetic etiology) and samples simply too young to be evaluated with certainty (therefore the assignment is to temporarily not use the samples awaiting future clinical updates).

As previously stated, we have discovered that ACS seem to exhibit distinct sub-types of phenotypes of inherited cataract. Primarily, we registered (I) a possible stratification of the phenotypes in regard of the age of onset. We also (II) noted that there seems to be a second type of classification of the cataract phenotype, where one eye develops a cataract at an early age and several years later a second cataract appears in the fellow eye. We also (III) took into account the anteroposterior position of the cataract onset for the classification of the phenotype.

Our principal means of classification of the phenotypes was on the basis of age (I). In fact, since we started to carefully re-assess the phenotypes of the dogs, this element was our primary concern in order to include a sample in the “Cases” or

“Controls” groups, and more importantly, asses the quality of the “Case” with a relevant score. Such subdivision is distinct and both groups consist in a high amount of samples.

In case of (II) and (III), we considered the conditions separately (sub-phenotypes, so to say) in the initial iterations of the analysis, but we were unsure about our preliminary results because of the lower amount of samples for a given subset (e.g. “anterior unilateral cataracts samples”). After the last iteration of genotyped data, with a higher number of samples in hand, we now will use strategies that can allow us to explore the possibility of association of a genomic region with a specific phenotype.

Importantly, we did not ignore the possibility of taking in account the phenotype sub classes (I-II-III) in light of the population structure of the dataset after our PCA analysis (Figure 1). However, we previously reported that the data gathered so far do not seem to indicate a strong effect of the sub-phenotypes indicated above compared to the stronger sub-population effect (see below).

SNP genotyping and data analysis

In our last report, we outlined a considerable improvement in our SNP chip dataset. We were able to review our records and improve the amount of the second-best quality cases (57 to 62). The total of excluded dogs dropped from 57 to 48. The total of high quality controls dropped by one. Because of the new surge of samples and updates, we are going through phenotype assessment and exam record analysis of dogs previously not included in the study. The review has been initiated, but this has yet to be completed since we need specific updates on a few critical samples already genotyped that could significantly improve the current dataset; request for updates on these critical samples has been made. Nonetheless, thanks to the contribution and feedback of the breeders in the previous iteration we were able to genotype 12 additional high quality samples (5 cases 7 controls). We are looking forward to this precious addition to our dataset.

As stated in our previous report, we took advantage of the new, higher density (220k vs 170k) of the current canine SNP chips. The new chip is ~30% more informative, with no information loss compared to the older one (that is, more SNPs were added to the new version but with full compatibility with the older one). Specific computational techniques were used to raise the information density of the old dataset at the level of the new one (“imputation”, through the popular software Beagle,

extensively used by our group in other projects).

As mentioned above, before the lockdown period we managed to select dogs from our best samples (60 dogs in total - 26 cases and 34 controls of the highest quality, see above) and send them to be processed for a third type of SNP chip using a new technology. Such technology allows the genotyping of the selected samples for 712k SNPs, more than three times the original information. This would allow us to identify, if possible, areas of the genome with poor coverage in the older chip versions. These newly analyzed areas might harbor candidate gene(s) that require greater scrutiny. The older SNPs are still present and therefore can be used to impute (see above) this new information in the rest of the dataset. The new batch of genotyped dog (220k, see below) would be added to this dataset and imputed as well.

Each case and control subset was classified based on the age of onset, laterality, anterior-posterior side of development of the cataract, and reliability of the sample (generally age-related). We checked whether there was some sex or age bias in the ratio of bilateral and unilateral cataract. Examples follow. Note that the phenotypes and sub-phenotypes do not seem to diverge significantly according to age and sex.

GWAS: We carried out a whole new series of Genome Wide Association Studies (GWAS). Additionally, another series of analyses was done using all the cases (62) and controls (70) within the whole population. We used the R package GenABEL (used in numerous animal genetics studies) to aid in this analysis. The aim of such studies is to associate a specific genomic region and its markers to a cohort of study cases. Thus, we also used association analysis packages from plink 2.0 in order to validate the findings and check whether the association found is consistent with one carried out with a different program.

Since we accumulated a greater number of controls, updated the cases, we repeated the population structure analysis as in the previous report: Principal Component Analysis (PCA) of the dataset (created by the same GenABEL software). As previously reported, roughly 80% of the total individuals would fall within one of the two sub-populations of uneven size. The two sub-populations were used for separate analysis, each time using as cases only the ones falling into one or another of the two sub-population.

For the next round of GWAS, we need some critical updates on a few important samples, as well as the SNP chip data of the new dogs. Once this data is available, we

will implement this new GWAS round.

Phasing: Also, we are running the haplotype phasing with the software Beagle that is in our experience very reliable (it has been tested in different ongoing and concluded projects carried out by our group – it's also widely reported in literature). We focused primarily on the regions reported above, and on any suggestive peak identified by GWAS for populations A, B and for the total population.

We counted cases and controls with the suspected haplotypes in order to identify trends. This is a vital step for our research in order to know "where" to search for candidate markers. Thus, we identified two major candidate regions in two different chromosomes- two 2-3 Mb genomic intervals with a high likelihood to contain genetic variants associated with cataract. When the new batch will be available, the phasing software will be used to impute. As always, we will consider possible low penetrance of a risk factor, selecting alleles with good signal but frequent in the controls as well. We cannot exclude anything at this stage, marker genotyping will tell!

Homozygosity mapping: In the previous reports, we speculated that since it's possible that the cataract condition (or at least some of these, if we are dealing with more than one within the population) is recessive, but not in a single autosomal manner. Furthermore, it's also possible that two regions apparently identical between cases and controls are in fact distinct at the fine molecular level. We count on whole genome sequencing data also to elucidate this possibility. For this reason, we are still considering the possibility that regions in which most dogs (cases and controls) are homozygous searching for exclusive markers to add to the pools of the one to be tested. This is in addition to the results obtained from GWAS.

Whole genome sequencing: Cases and controls of the best quality were selected, 2 cases, 2 cases selected from the older category, and 4 of the best controls. As described above, the state of the pandemic and of the lockdown initially prevented us from using the planned facility in Europe, but we quickly found a local workaround. Now the facility in Bern, operated by our colleagues in Switzerland, is again running. Their equipment was restarted and tested in late December 2020/January 2021 with success and is now ready for an intake of new samples. Thus, we are now able to access a cheaper and more reliable whole genome sequencing pipeline once again.

Future prospects and plans

As stated in our previous proposal, in this last part of the last funded period, we focused on the following aspects of the cataract in ACS research; we re-analyzed and confirmed our case-control Genome Wide Association Studies and genetic mapping. The first was cleaned up and realigned. Additionally, the data was mapped to an updated canine genome reference. This is because more publicly available canine genetic data has been made accessible. Such data was used to compare the candidate genetic markers found in the cataract-affected ACS with the general canine population. WGS data was re-analyzed in light of the updated phenotypes.

A complex disease: We hypothesized in the previous report that the occurrence of cataracts in American Cocker Spaniel is likely a complex of 2 or more diseases. As shown, a greater number of cases and controls leads to better and more encouraging results. The selection of the appropriate sub-populations of cases and controls moved forward the analyses and the project, and we are now have been able to identify our candidate region and to implement whole genome sequencing (WGS).

Tackling the complexity: Even if we cannot show, at this moment, a simple and complete association of a single marker with the cataract in ACS, we have found that we can trace and identify trends and associations both under the assumption of a recessive disease, and under the assumption of a disease associated with loci of vulnerability not necessarily inherited in a recessive manner (we cannot, at this point, suggest a dominant inheritance – if such, the penetrance would be fairly low or dependent upon the co-existence of multiple factors, not necessarily all of them genetic).

We update our immediate and future objectives as listed below and compare them with what was stated in the last report.

. **A)** As always, we renew our stated intention to increase the sample number in the database: a greater number of cases means we will be able to enrich the specific sub-populations, and a greater number of controls allowing us to avoid false positives. The

Research Scientist dedicated to the project spends a significant amount of time in the management of the database and in the interaction with the breeders and owners to obtain samples and updates, and that our database improved in numbers and diversity. Additionally, in order to carry out the updates of the status of dogs already included in the study and genotyped, the Research Scientist will contact owner of dogs of known status that need an update.

. **B)** We previously stated that we would go through an in-depth analysis of the data output, never ignoring the slightest suggestive peak. Thanks to the contribution of the breeders we were able to add additional dogs to the analysis. The genotyped batch is currently being run by the company dedicated to this technology and the data will be made available to us soon after this report is submitted, greatly increasing the sample pool. This last iteration of SNP chip will be imputed using the available high-density SNP chip data (712k chip) already genotyped previously. If additional suitable candidates will be found, they will be inserted in additional batches.

. **C)** We stated previously how our preliminary cross-reference of the data did not point out any specific correlation between laterality, i.e. unilateral or bilateral, and age of cataract onset. Once a smaller pool of markers is available, we will observe their segregation with the sub-phenotypes once more. We will also re-do a GWAS analysis with these co-variates in mind once the new SNP chip data is available.

. **D)** At the moment whole genome sequencing data analysis will be the focus of our next steps. Through the data generated, we will select suitable candidate markers to be tested within the population. Validation will then happen in two steps – through further sequencing, investigating the segregation of a candidate variant within the population, and/or with further experiments confirming in vitro a supposed effect of the variant on gene expression, translation, splicing.

E) We plan to tackle the complexity and “noise” of the currently available WGS output with a radical approach – we will carry out Whole Genome Sequencing on a batch of 12 additional ACS. Our initial intention was to have half of these (6) as cases and the rest (6) as controls. Again as done before, half of the cases (3) are supposed to be early onset cataracts while the remaining dogs (3) would be later onset (but well-assessed) cases. In combination with the updated mapped data, this will give us a significant boost in our WGS dataset. This would enable us to immediately search and filter for shared markers among these cases and end up in a shorter amount of time with a list of candidate markers for the cataract in ACS.

The renewal of our relationship with the Swiss facility led us to plan for additional potential dogs to use in WGS. If a sufficiently high number of highly reliable cases and controls can be sequenced, additional new methods of analysis can be implemented. This would dramatically increase the chance of the project resolution.

F) The final strategy is to start with a large number of candidate markers, that would be then tested (Sanger sequencing) in small but significant groups of dogs belonging to the study. Once the marker “passes” this first selection, we plan a scenario in which fewer and fewer markers would be genotyped in a progressively greater number of dogs. Most markers in few dogs, to few markers in most dogs. The final candidates would be then tested on the whole ACS population and the suitable markers for a cataract test would be proposed to the ACSF.

Based on the data currently available (see our last report), we gather that a given marker will be predictive of a cataract risk factor in dogs belonging to a given ACS sub-population. We also count on the fact that new GWAS and WGS data will help minimize the number of markers to be explored first, and later proposed. This part of the strategy would be implemented after the WGS of the new dogs and the new mapping due to the new SNP data.

As in the last report, we share our excitement for the prospect of whole genome sequencing dog samples. We still think that acquisition of this informative dataset will allow us to move the project forward. At the moment, our main task is to carefully filter, test, and validate in order to refine what has been done so far. We are quite optimistic of the results and certain that we are using the correct approach going forward.