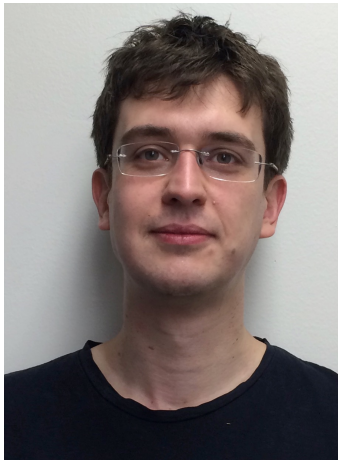
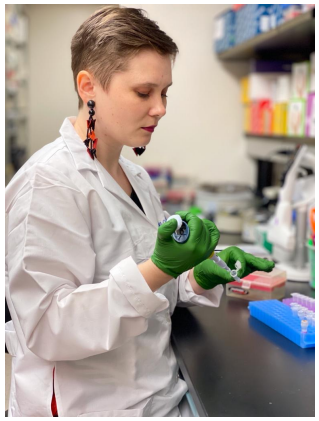
Update from July 14, 2021: Search for the gene and mutation causing inherited cataract(s) in American cocker spaniels



Leonardo Murgiano, PhD

August 5, 2022

Gustavo Aguirre, VMD, PhD
gda@vet.upenn.edu



Jessica Niggel, MSc

Anil Sigdel, DVM, PhD

Update since last July 14, 2021

- Grant from the spaniel club ended June 30, 2022.
- Decision was made not to continue the traditional funding/grant program as before because:

  - progress has been slow

  - cataract in the ACS are complex, not only in terms of inheritance but also in how the clinical appearance relates to a specific genetic trait/locus.

  - could not 'promise' that within the next 3 year period we would identify the gene/mutation and implement a genetic test.

- However, **DON'T** despair. We will continue to work on this project until we crack the problem.
- Alternative approach worked out with the ACSF !!!!!

Background

Genome wide association studies (GWAS) combined with whole genome sequencing (WGS) are used to map traits/diseases of interest to a genomic region (using GWAS) followed by careful analysis of the sequences in the candidate region using WGS. It is not necessary to use family groupings for GWAS, and unrelated or not closely related dogs having the same inherited disease will be effective along with unaffected controls. These tools were not existent or prohibitively expensive 20-25 years ago. In fact, WGS was something we dreamed about then, but the ~$100,000 cost for one sample was prohibitive. Approximately 10-15 years ago, WGS costs dropped to ~$10,000/sample, still an unaffordable price when considering that multiple samples were necessary. Now costs are ~$1,000 [40x pass and ~$375 (low pass)] and will likely decrease as technology improves and the number of samples analyzed can be increased exponentially at a much reduced cost.

The reduction in costs for SNP array genotyping, required for GWAS (now $100/sample in an array with 220,000 individual SNPs) permits us to carry out studies at a very reasonable cost. As an example, for studies of conditions which are inherited as simple autosomal traits (which are the majority of canine diseases but NOT cataracts in ACS), the sample costs now would be $5,000. This would include high-density genotyping (~220k markers) of a family of 10 dogs (2 non-affected parents and 6 pups that are non-affected, and 2 affected dogs; 10 samples = $1,000), and WGS on 4 dogs (2 non-affected parents, and 2 affected puppies; 4 samples = $4,000). Much higher density SNP chips are now available (~720k SNP markers), but are more costly (~$140 per sample) and require a larger number of samples (batches of 96) to be submitted for each run. As with WGS, this cost will decrease as will be the required number of samples submitted for each genotyping run.

So what is the problem? While getting GWAS and WGS is easy, relatively inexpensive and not hard to do once the appropriate dogs are selected for the studies, that is only the first step of the research. *The problem is that the amount of sequence information received, especially from WGS, is so large that it requires specialized training in bioinformatics, sequence analysis, and computational biology, along with large central computer facilities (such as available at Penn's supercomputing cluster) to effectively utilize and analyze the sequence data*.

Solving the Problem: Establishing the *Sylvia M. Van Sloun Laboratory for Canine Genomic Analysis*.

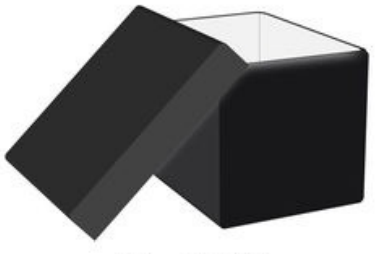
Benefit to ACS cataract project:

-a large part of research costs are personnel and covered by Van Sloun.

-adding another highly trained investigator to the project provides 'more eyes" to review data, develop new mathematical models in search of the cataract locus in the breed.

Research will continue to identify the gene/mutation and develop a DNA test.

Genetic dissection of cataract



open the black box:

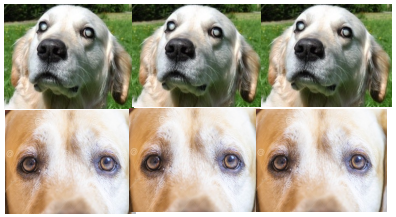
identify genomic regions, candidate genes and causal variants

Why we want to decipher the genetic basis of cataract?

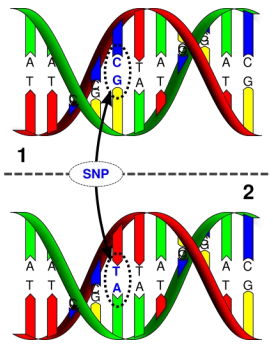
- ❑ to better understand the biology of this complex phenotype
- ❑ to generate information that can lead to the development of new drugs/therapies
- ❑ to develop novel genomic strategies for reducing cataract via selective breeding

How to decipher the genetic basis?

phenotypic variation

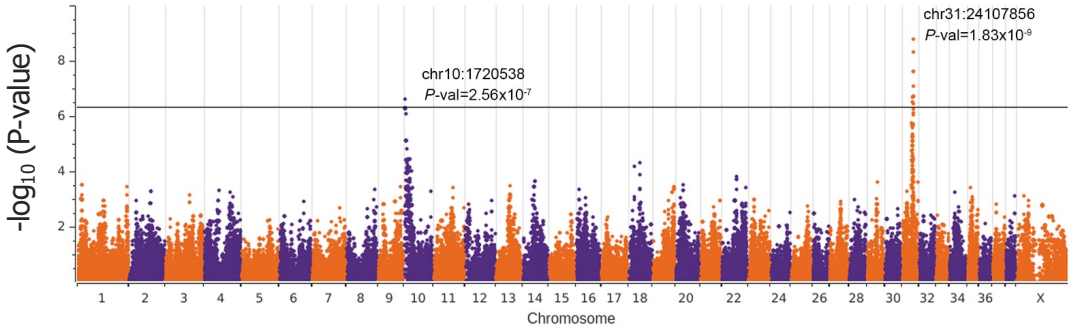


+



genome-wide SNP information

identify regions/SNPs
associated with the trait
under study



genome-wide association study
whole-genome scan

ACS in the study

Total dogs	1004
Total of Informative dogs	609
Potential cases	143
Bilateral	92
Asymmetrical	51
Controls	466
Too young to be properly assessed for study inclusion at this time	299
Total # of Excluded dogs (this includes dogs with lack of follow up, no pedigrees or clinical records or other eye/systemic diseases that prevent assessment of the cataracts; they are 'excluded' from the specific initial studies but later on will be 'included' once gene(s)/mutation(s) are identified)	395

Samples received since last webinar

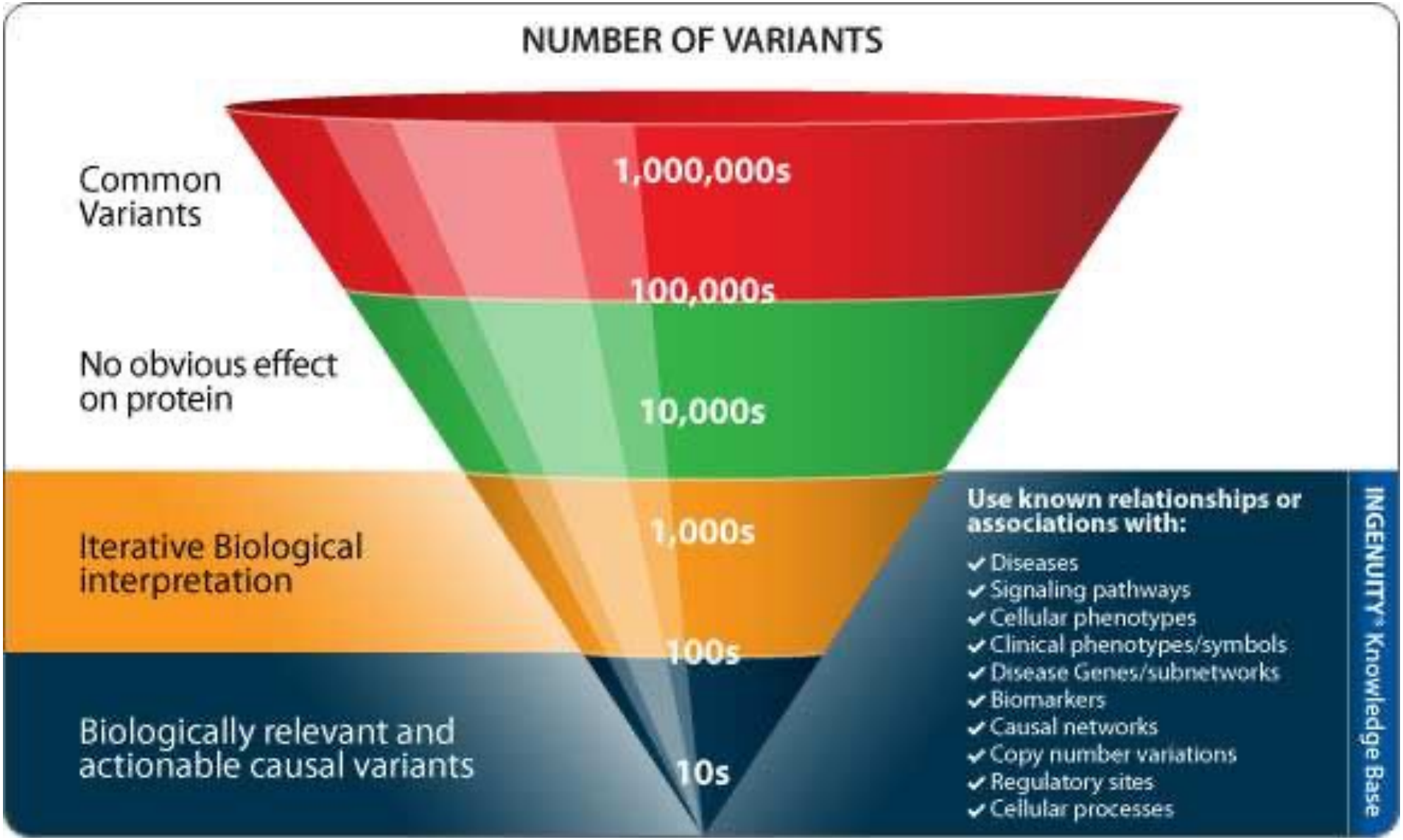
- We have noted a decrease in # samples and/or updates received from breeders in general.
- Effort with CAER/OFA to recruit samples from >8 yr old normals now discontinued. Very few received!!
- Nonetheless, with Aguirre's June clinic, there was a rebound with several cases and controls of excellent quality collected and records updated.
- Overall, 21 dogs were sent to SNP chip genotyping this year, 11 since the last report.

SNP chip

Total ACS on SNP chip	230
Total cases	109
Asymmetrical	38
Bilateral	71
Early Onset	62
Late Onset	47
Total controls	121
Total Excluded (previously SNP'd but do not fit groups in terms of rigorous criteria now used)	31

What has been found so far

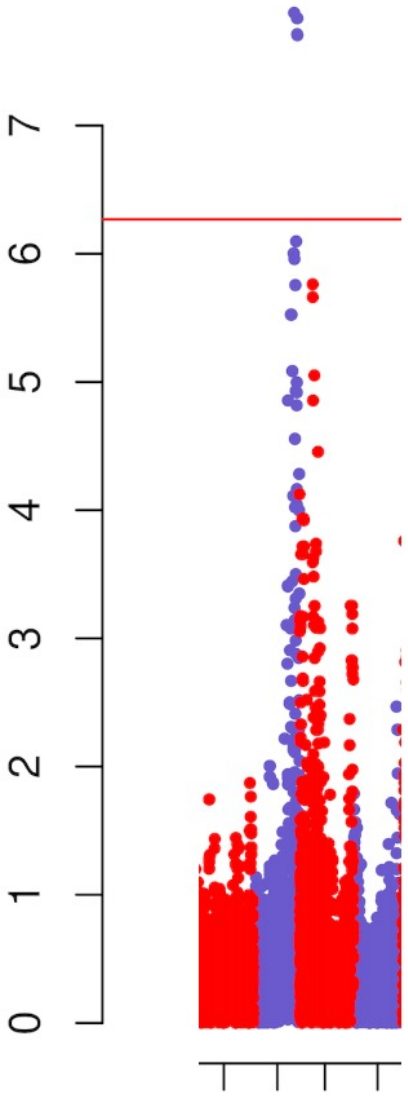
- At least 2 haplotype that can explain 50% + of the cases each
- The two regions had >9000 and >5000 variants respectively
- A preliminary genotyping of the most promising was carried out
- Genotyping didn't lead to a definitive discovery – better strategies need to be implemented


NUMBER OF VARIANTS
Common Variants
1,000,000s
100,000s
No obvious effect on protein
10,000s
1,000s
Iterative Biological interpretation
100s
Biologically relevant and actionable causal variants
10s
Use known relationships or associations with:
✔ Diseases
✔ Signaling pathways
✔ Cellular phenotypes
✔ Clinical phenotypes/symbols
✔ Disease Genes/subnetworks
✔ Biomarkers
✔ Causal networks
✔ Copy number variations
✔ Regulatory sites
✔ Cellular processes
INGENUITY® Knowledge Base

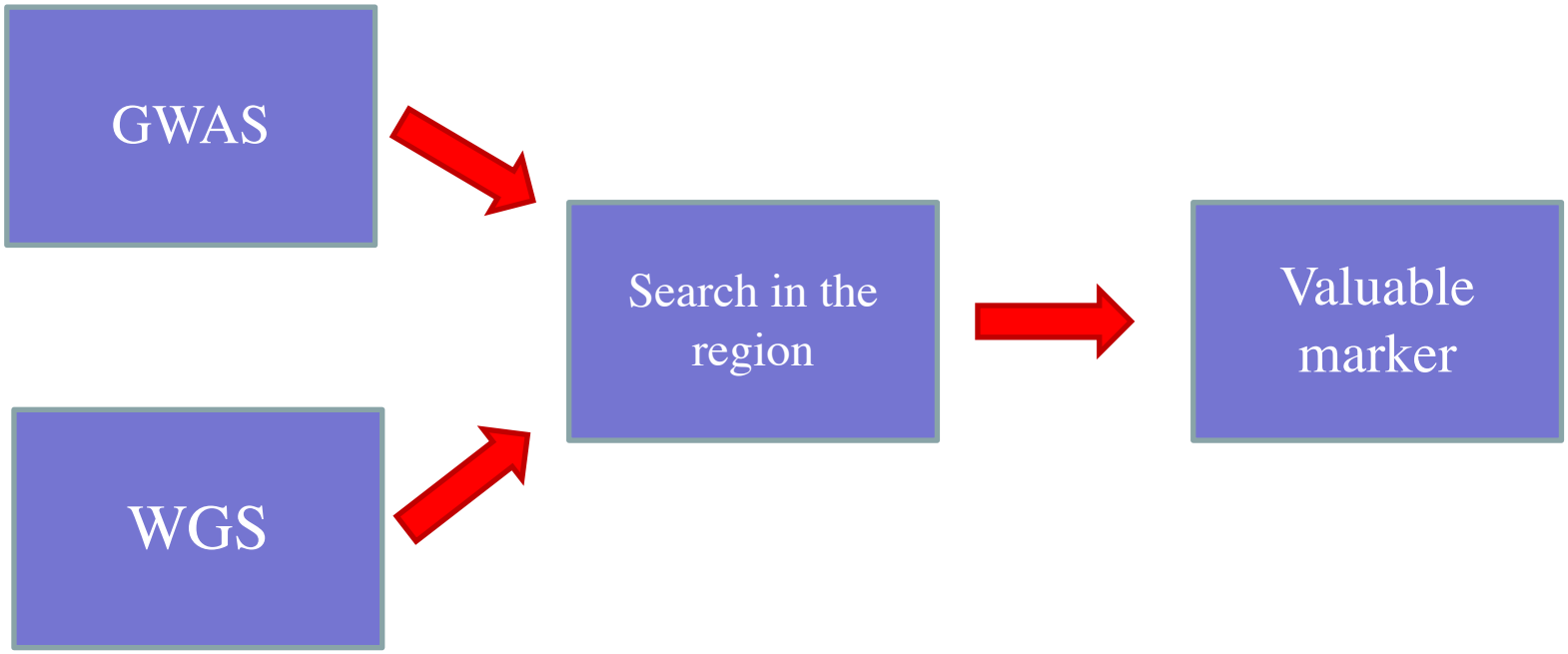
New GWAS models and math

Anil Sigdel

- New models to better take into account variables like sex, **age** of onset…
- To be repeated on larger sample pool
- GWAS limits still present

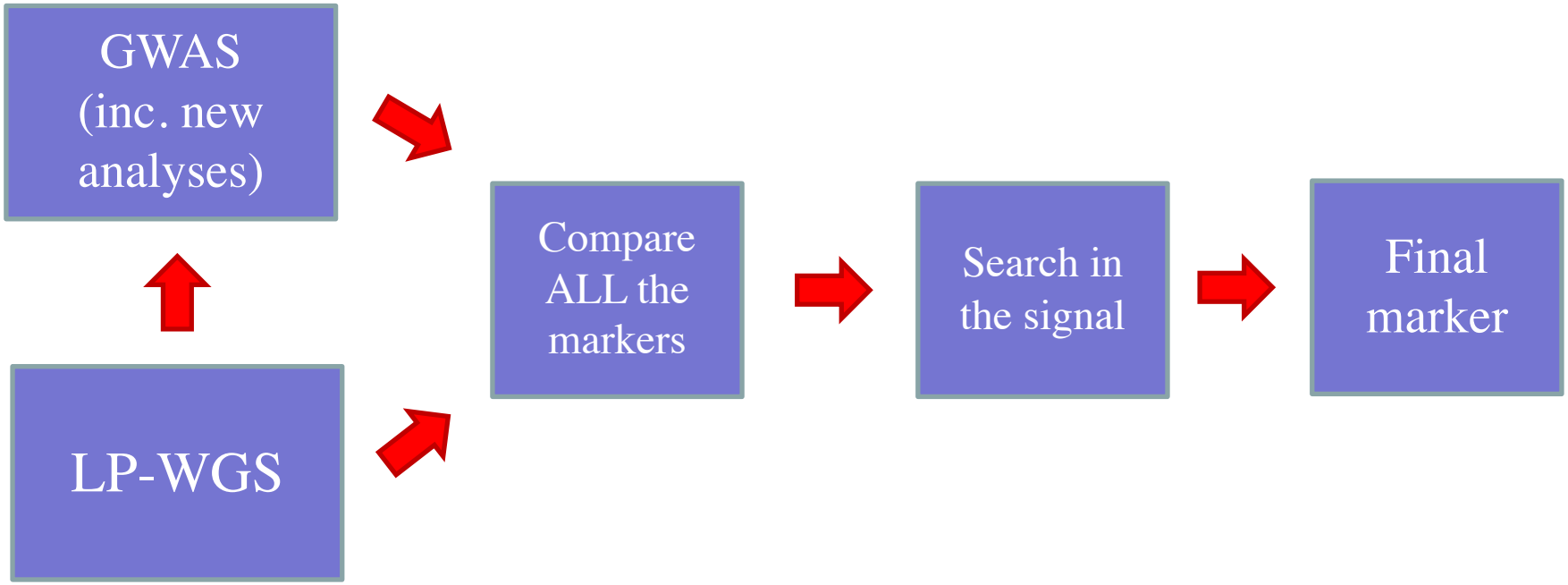


Classic approach



Approach now discontinued because with additional samples analyzed the results change so it is difficult to focus on a region of interest.

New Approach using LP-WGS



LP-WGS-GWAS approach

- 20 dogs just sent
- 10 early onset cases, 10 controls
- Data integrated with the genotyped dogs
- Previous WGS dogs will be also included in the analyses
- Depending on results with early onset cases, 10 late onset cases will be sent.