



Opportunities and challenges for managing data to drive translational discovery

Adam Boyko

Associate Professor, Cornell University

Co-Founder/CSO, Embark Veterinary

Dogs as a model species

- Hundreds of phenotypically differentiated, closed populations
 - Facilitates genetic association and fine-mapping of causal mutations
 - 50-fold difference in body size; 2-fold difference in aging rate
- US population of ~80 million dogs (mostly pets)
 - More similar environment and clinical diagnoses as humans than other model species
- Global population of ~1 billion dogs (mostly village dogs)

Rapid adaptation

Genetic load

Personalized genomics



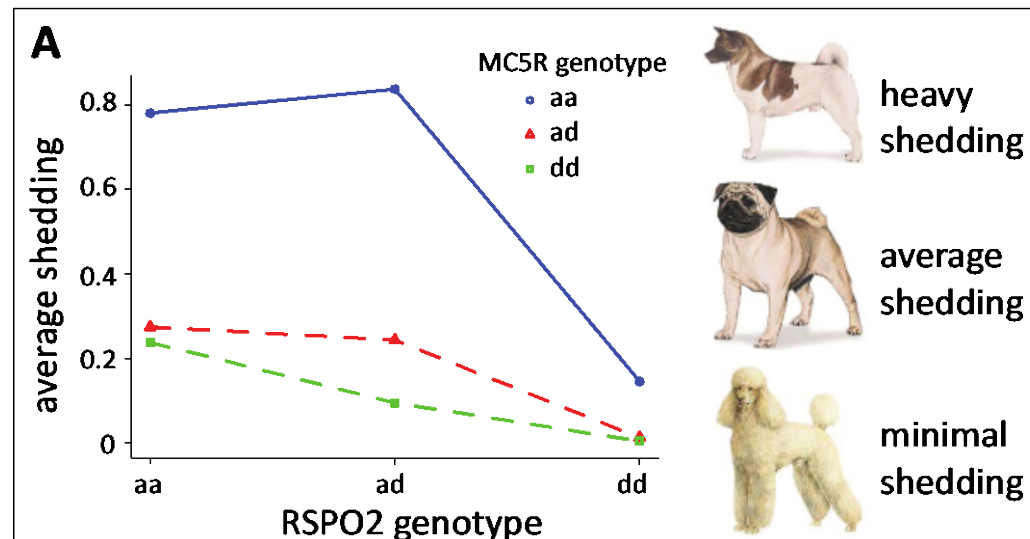
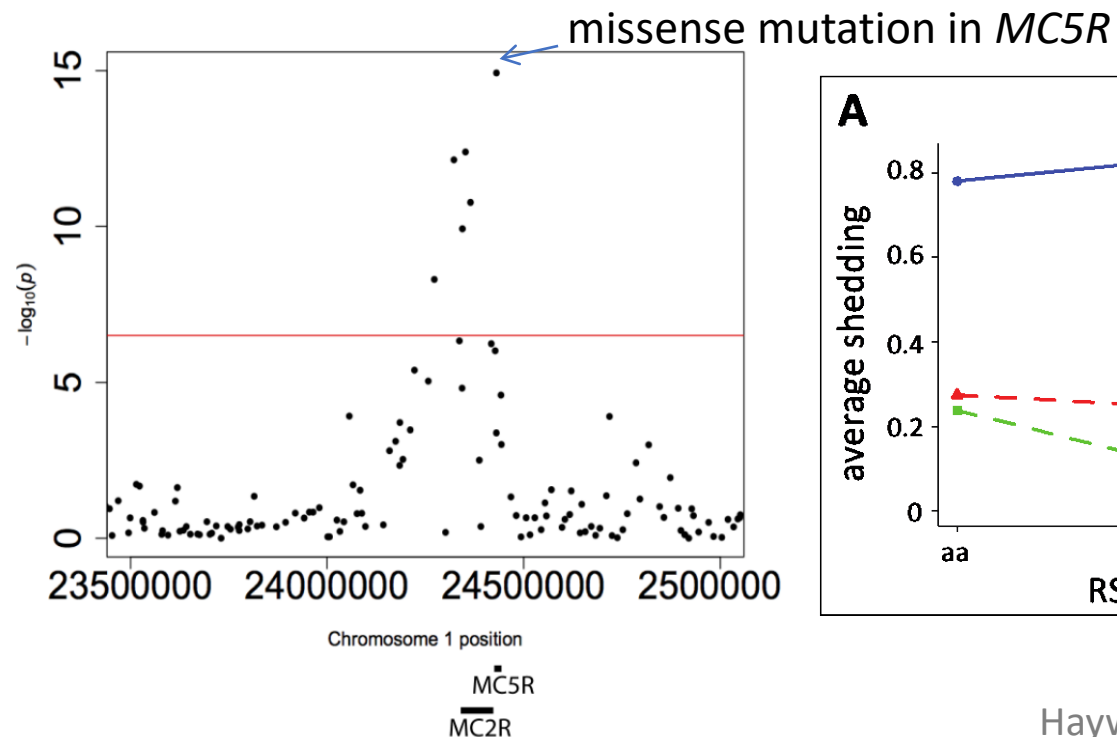
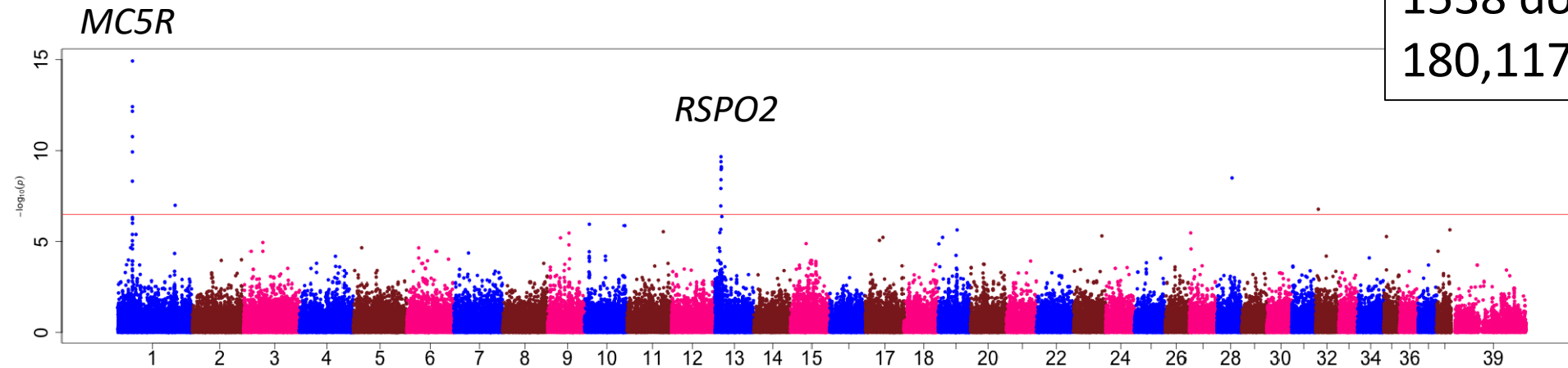
Dogs have been a fantastic model species for identifying Mendelian loci and large-effect QTLs



	dog	cattle	cat	pig	sheep	horse	chicken	rabbit	goat	Other	TOTAL
TOTAL TRAITS/DISORDERS	<u>782</u>	<u>549</u>	<u>358</u>	<u>282</u>	<u>257</u>	<u>240</u>	<u>222</u>	<u>98</u>	<u>89</u>	<u>679</u>	3646
Mendelian trait/disorder	<u>362</u>	<u>257</u>	<u>114</u>	<u>89</u>	<u>112</u>	<u>59</u>	<u>131</u>	<u>58</u>	<u>18</u>	<u>268</u>	<u>1533</u>
Mendelian trait/disorder; likely causal variant(s) known	<u>293</u>	<u>164</u>	<u>81</u>	<u>41</u>	<u>58</u>	<u>46</u>	<u>51</u>	<u>11</u>	<u>14</u>	<u>140</u>	<u>915</u>
Likely causal variants	<u>430</u>	<u>223</u>	<u>129</u>	<u>49</u>	<u>74</u>	<u>98</u>	<u>66</u>	<u>14</u>	<u>25</u>	<u>123</u>	<u>1249</u>
Potential models for human disease	<u>468</u>	<u>224</u>	<u>223</u>	<u>130</u>	<u>117</u>	<u>132</u>	<u>51</u>	<u>54</u>	<u>40</u>	<u>354</u>	<u>1826</u>

Fur shedding

123 breeds
1538 dogs
180,117 SNPs



The Village Dog Project

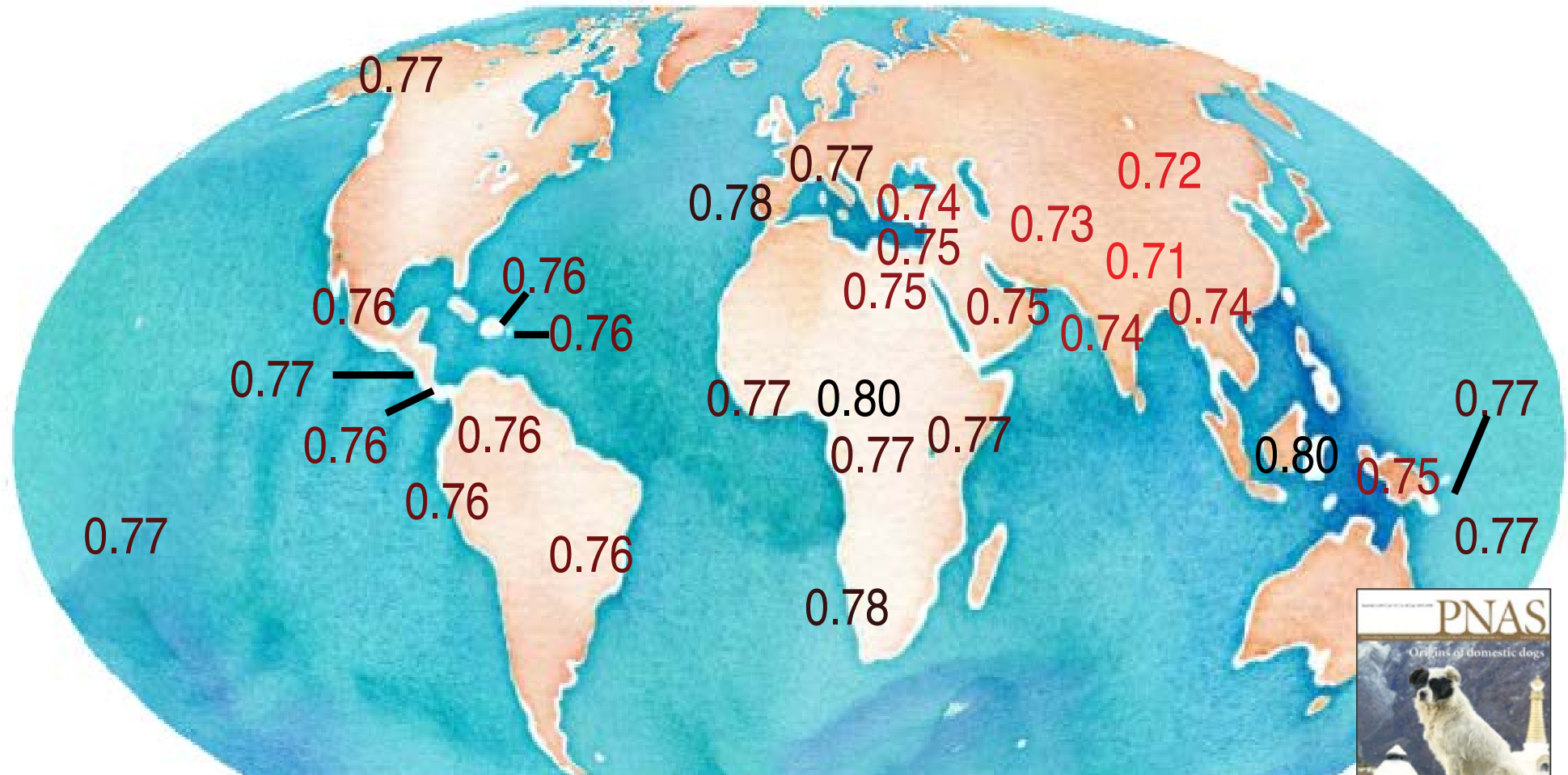


Cornell University
Center for Vertebrate Genomics



NATIONAL
GEOGRAPHIC™

Short range LD in village dogs consistent with Central Asian domestication origin



Great progress identifying Mendelian and large-effect QTLs underlying morphology and disease, but...

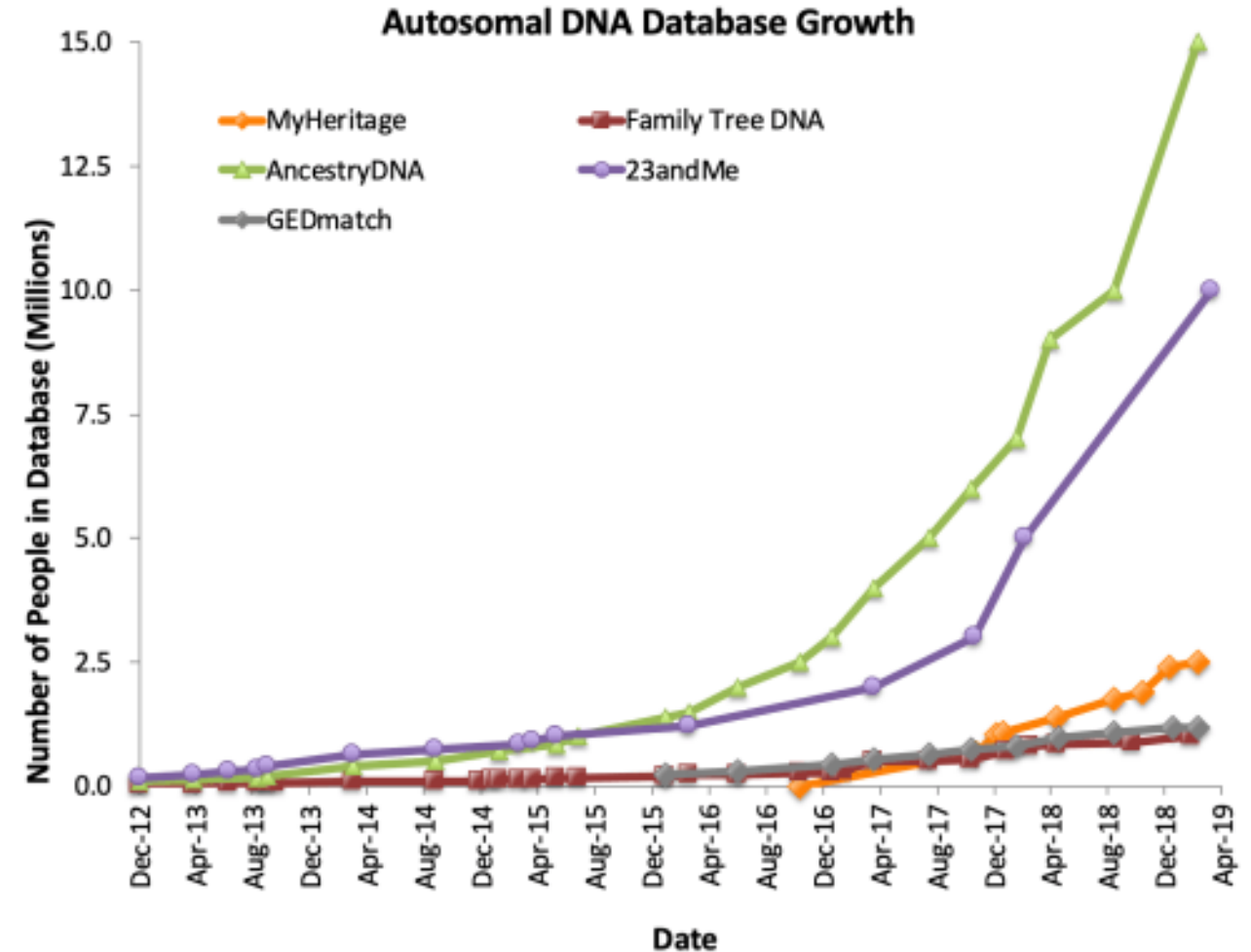
Slower progress using genetics to predict and reduce complex diseases (cancer, autoimmune disease, orthopedic disease)

Slower progress finding genes underlying behavior

Slower progress understanding the role of genetics on longevity

Large-scale studies (hundreds of thousands of dogs) are needed!

- WTCCC
- The GenomeAsia 100K Project
- UK Biobank
- DiscovEHR
- GERA
- Million Veteran Program
- CHARGE Consortium



© 2019 by Leah Larkin, www.theDNAgeek.com/dna-tests

No genome-wide dog DNA test existed in 2015

CANINE

Wisdom Panel

Paw Print
Genetics
A division of Genetic Veterinary Sciences, Inc.

VetGen
Voluntary Genetics Services

HUMAN

myriad
WHEN DECISIONS MATTER

Quest
Diagnostics

ancestryDNA

23andMe

Helix

NARROW TEST

GENOME-WIDE TEST

Personalized genetics for dogs

- Predict disease predispositions
 - Identify relatives and ancestry
 - Participate in citizen science research
 - Explain traits
-
- 8 million dogs born in the US every year are potential research subjects for important genetic and longitudinal study

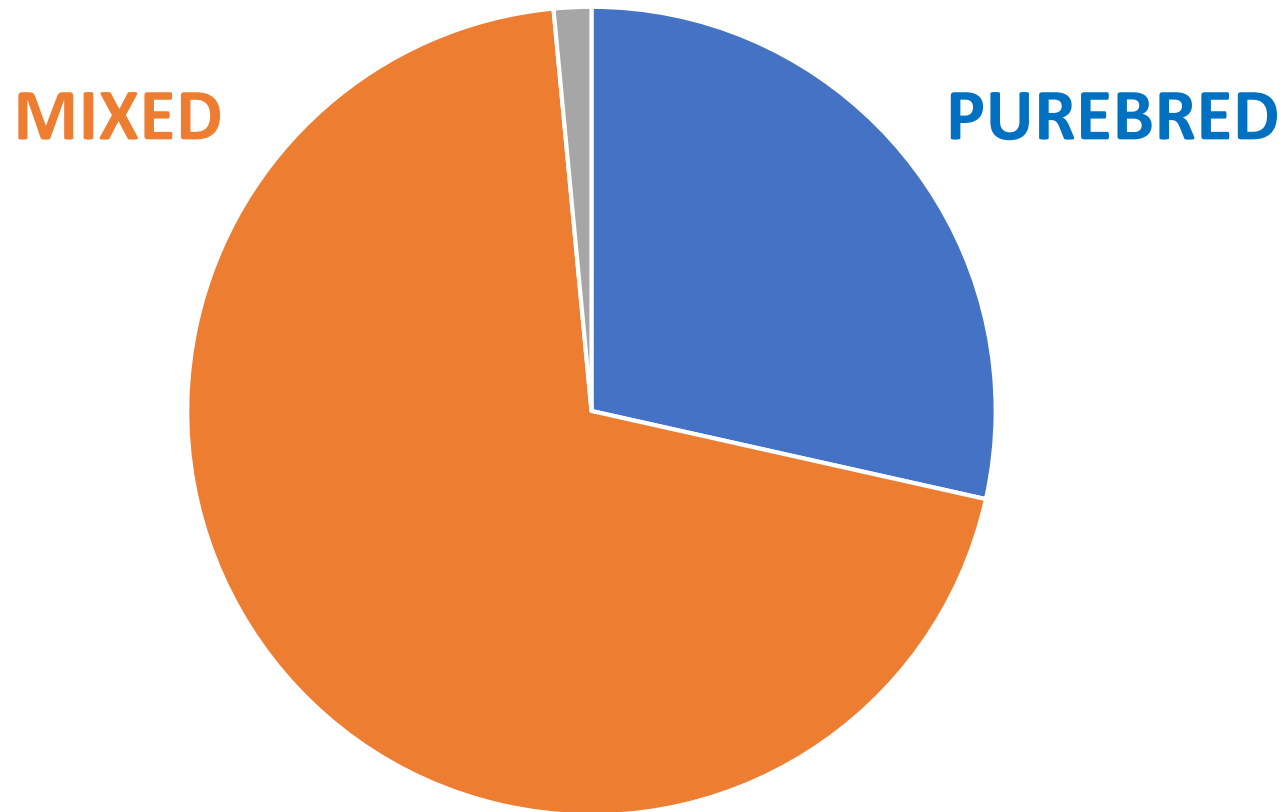




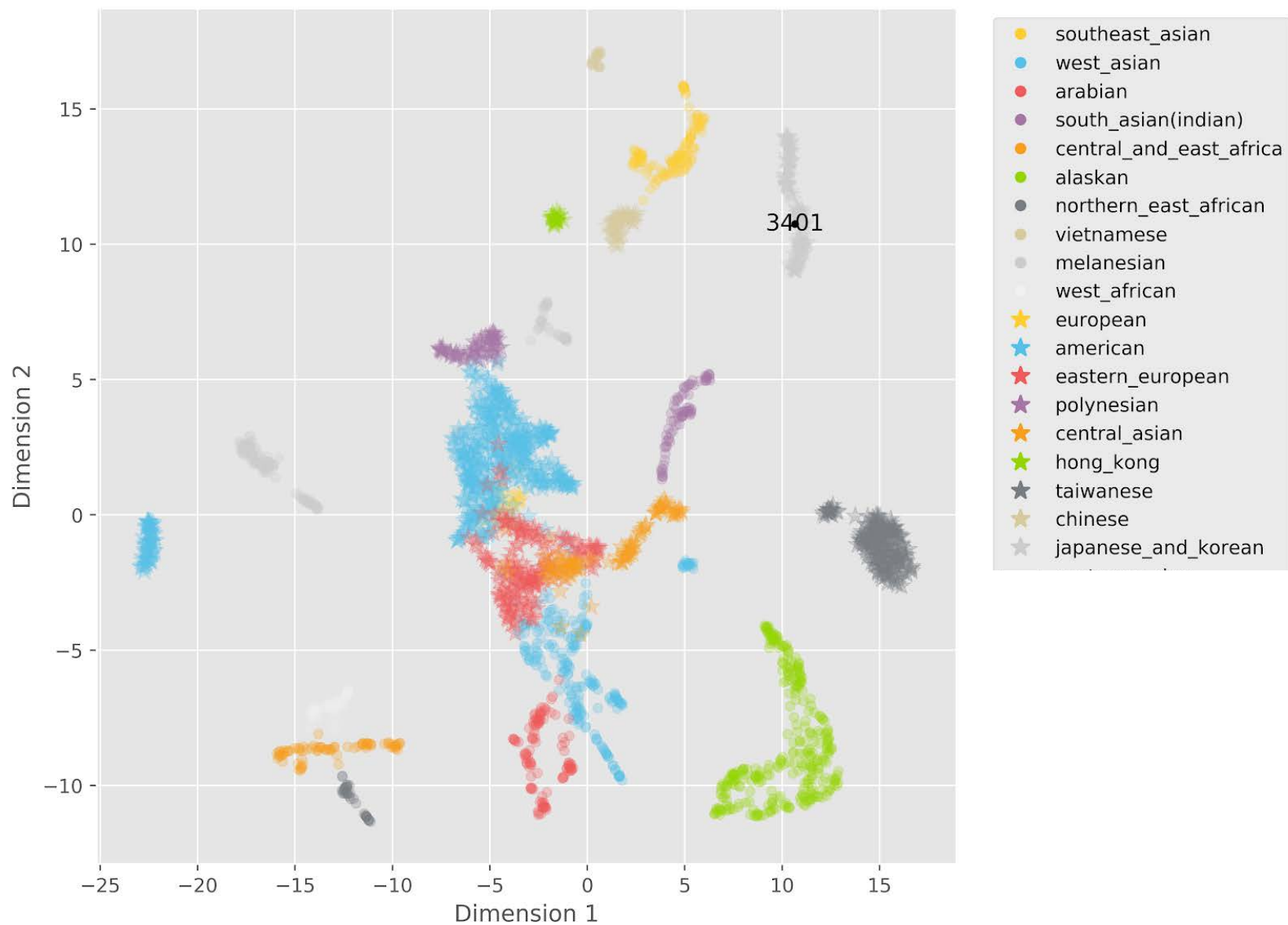
Company started in 2015,
Consumer kit launched 2016



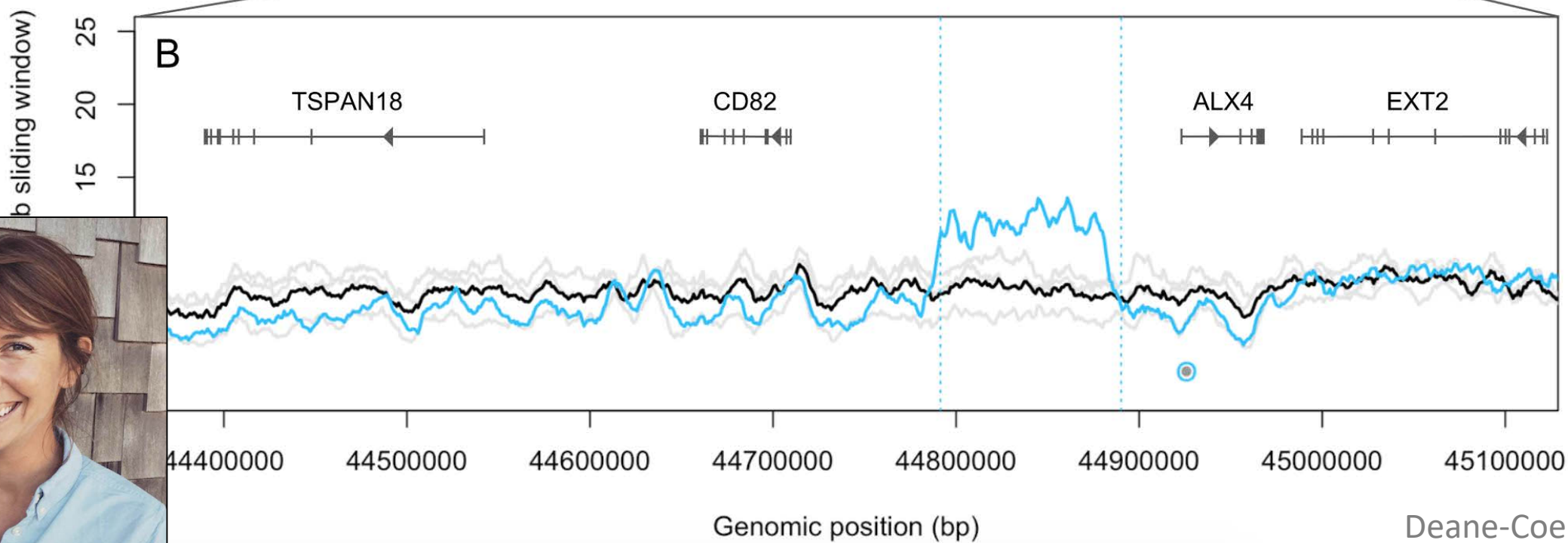
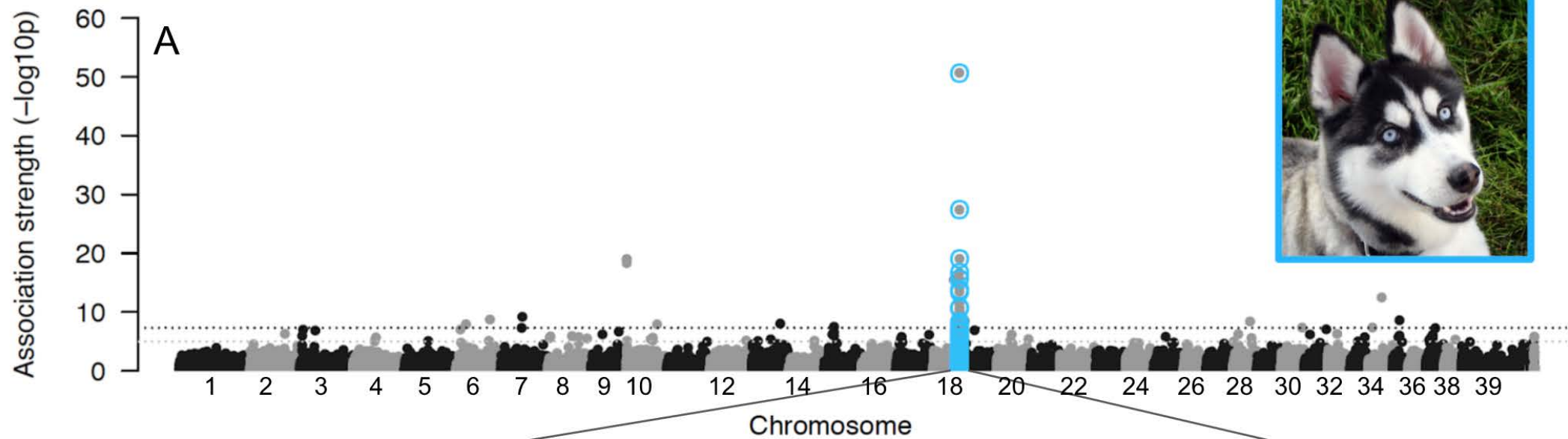
**900,000+ dogs in database x 230,000
SNPs = 200+ billion genotypes**



VILLAGE DOGS



Brett Ford





Complete Heterochromia



Sectoral Heterochromia



Blue



Amber



Light Brown



Brown



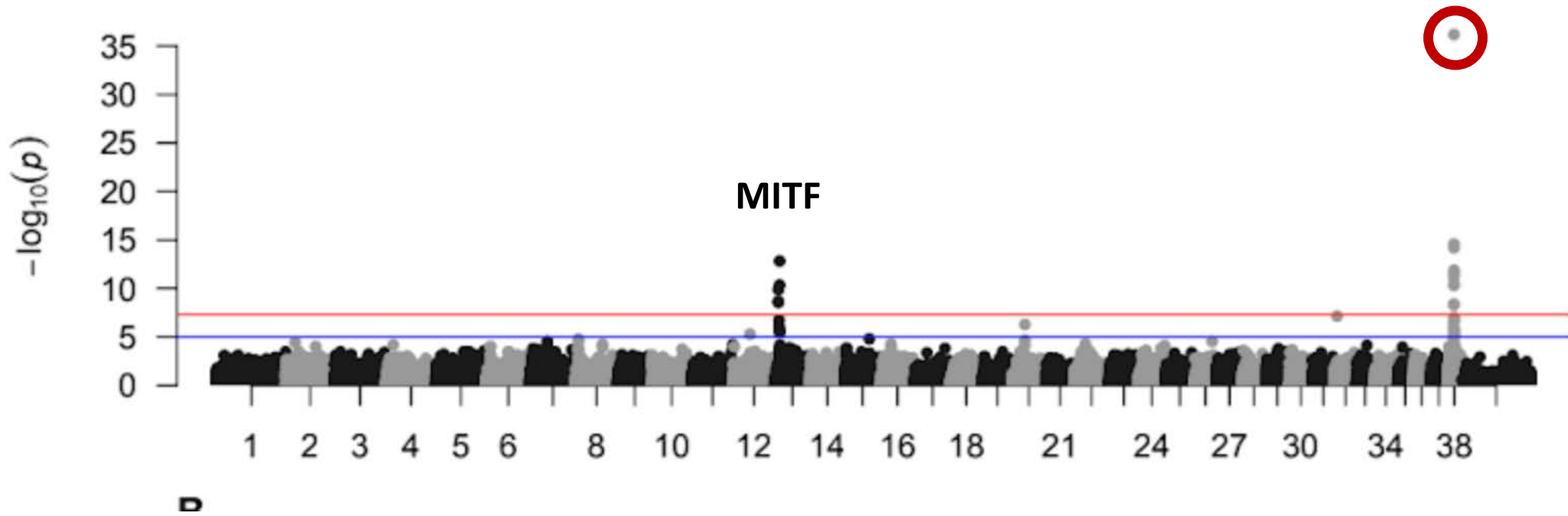
Dark Brown / Black

Contingency table comparison for the duplication vs. the top associated GWAS SNP, in our validation panel of 2,769 mixed-breed and purebred dogs.

	Marker absent	Marker present
Blue	Duplication: 24 GWAS allele: 20	Duplication: 84 GWAS allele: 88
Brown	Duplication: 2,615 GWAS allele: 2,435	Duplication: 46 GWAS allele: 226

GWAS SNP: $P = 4.9 \times 10^{-120}$

Duplication: $P = 5.2 \times 10^{-290}$





Embark's annual health survey

- Detailed health information sent to every Embark customer that has opted to participate in research
- Over 180,000 surveys completed so far in 2021
- Survey results help guide and prioritize research projects for the coming year

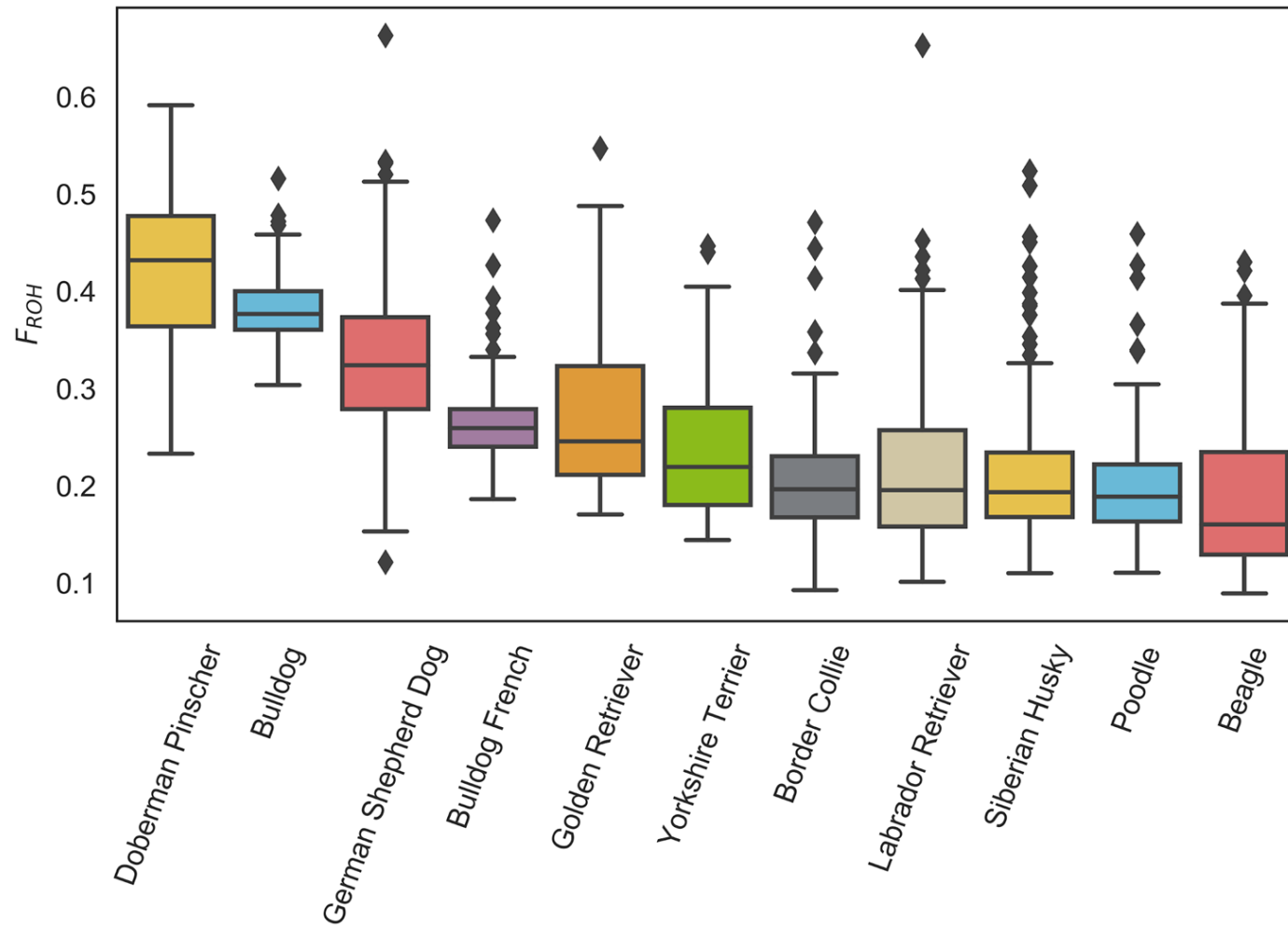
How would you rate Dara's health, currently? *

- ☐ Excellent
- ☐ Good
- ☐ Fair
- ☐ Poor
- ☐ Deceased

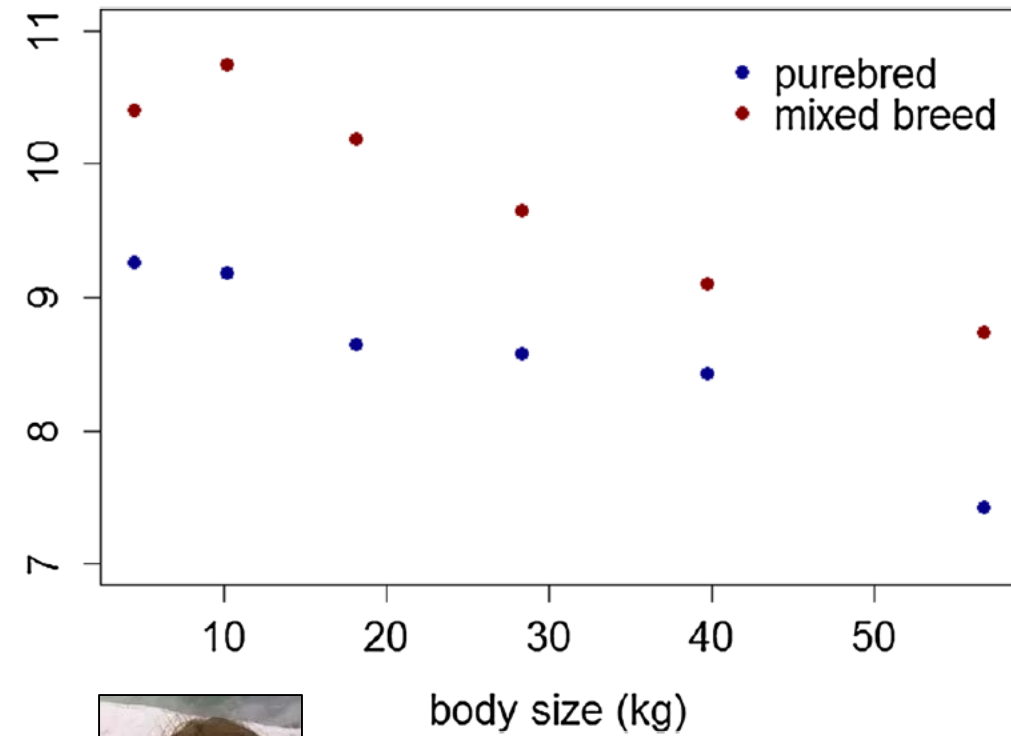
Has Dara been diagnosed with any conditions or disorders in any of the following categories? Select all that apply. (Note: please include categories for new diagnoses as well as pre-existing conditions; questions about the specific conditions will follow.) *

- | | |
|---|--|
| ☐ Airway or lungs (respiratory conditions) | ☐ Hormones (endocrine conditions) |
| ☐ Allergies or conditions affecting the skin, nose, ears, or fur | ☐ Immune system (immunological conditions) |
| ☐ Blood (hematologic conditions) | ☐ Infectious disease (bacterial or viral infections) or Parasites |
| ☐ Bones or joints (orthopedic conditions) | ☐ Kidney or bladder (urinary conditions) |
| ☐ Brain, spinal cord, or nerves (neurological conditions) | ☐ Liver or gallbladder (hepatic conditions) |
| ☐ Cancer of any organ system (skin, blood, bone, etc.) | ☐ Reproductive system |
| ☐ Dental/Oral | ☐ None of these |
| ☐ Eyes (ophthalmic conditions) | ☐ I don't know/Not sure |
| ☐ Gut (gastrointestinal conditions) | ☐ Other - Write In (Required) <input type="text"/> |
| ☐ Heart (cardiovascular conditions) | |

Inbreeding and longevity vary by breed and size



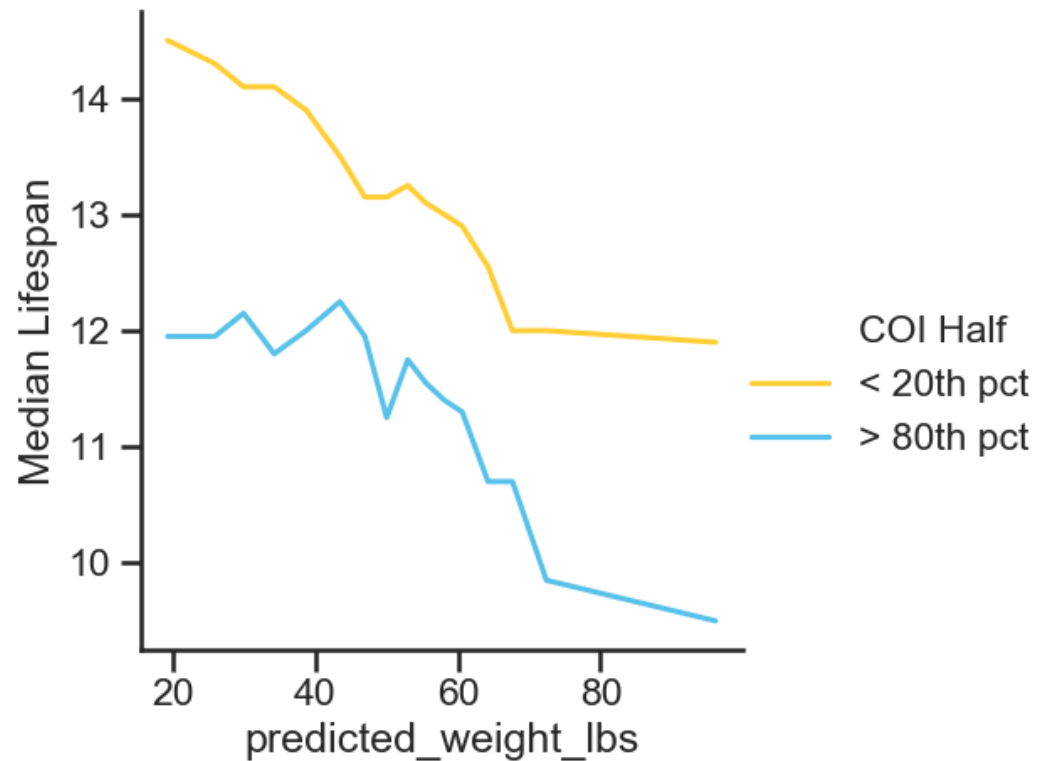
Sams & Boyko, G3, 2019



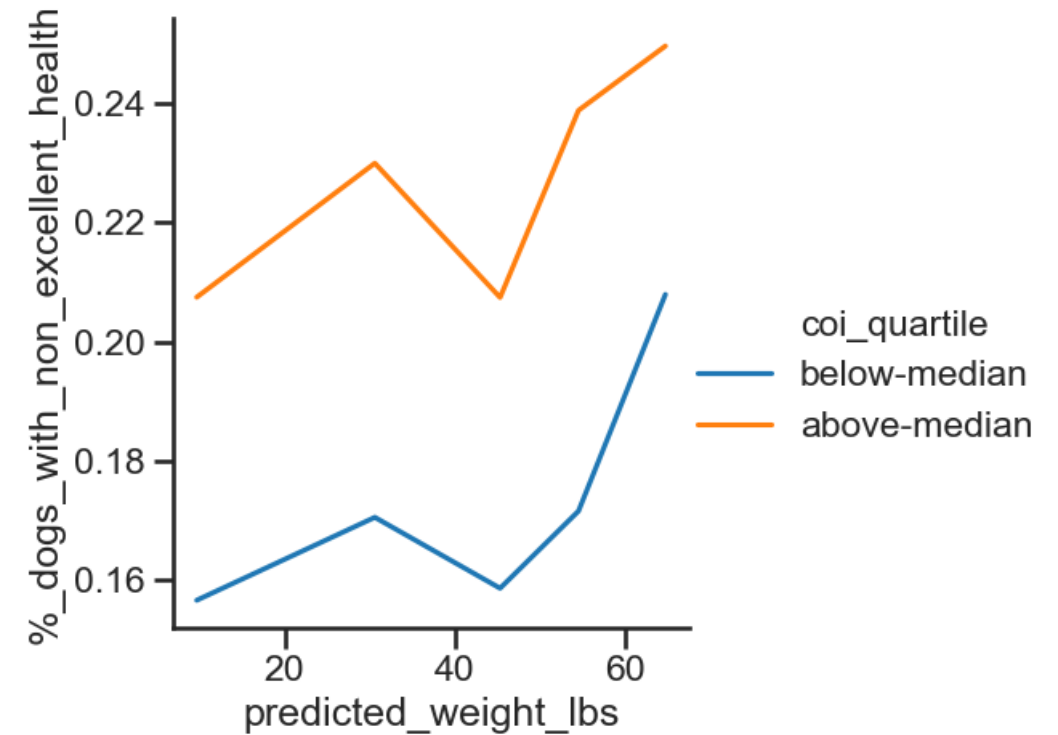
Yordy et al, Conserv Genet, 2019

COI is significantly associated with lifespan and health

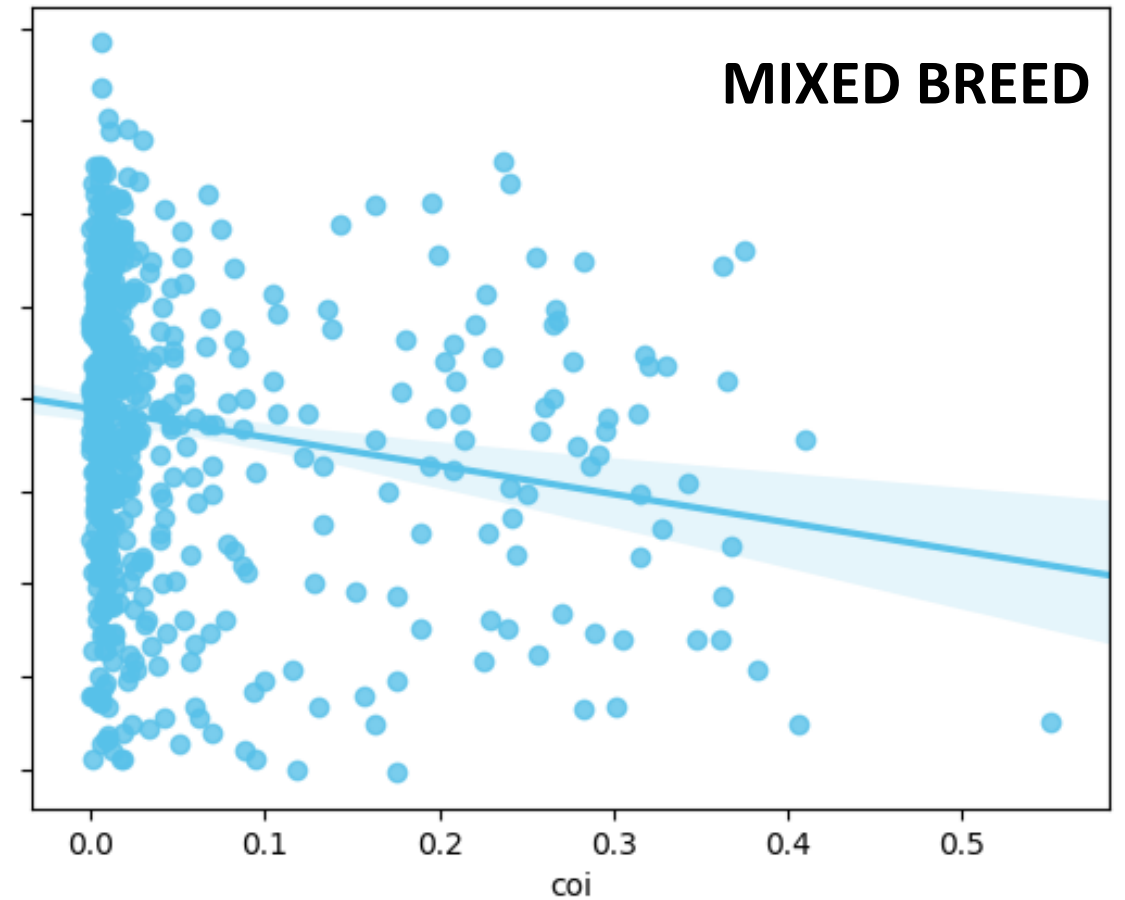
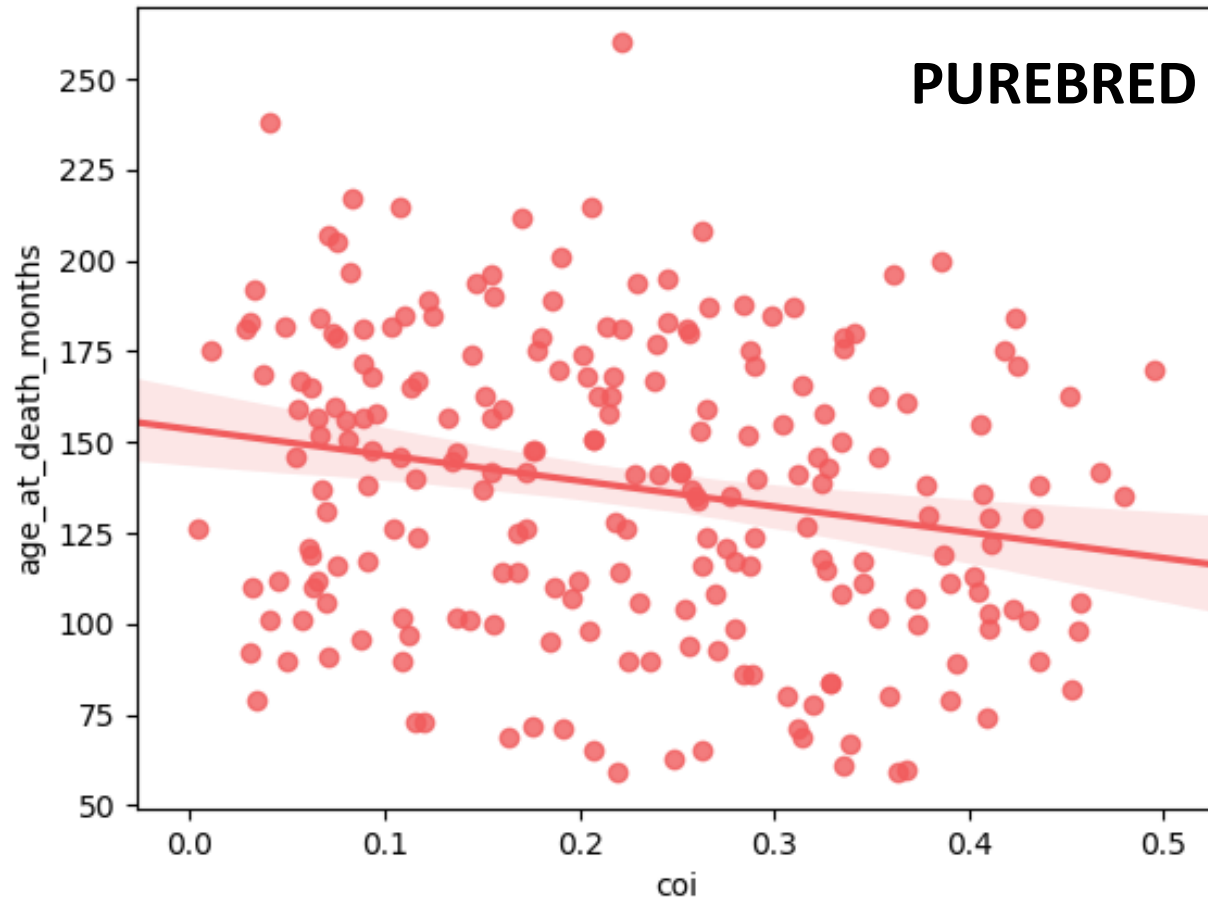
Longevity data from 1,682 deceased dogs shows nearly 3-year difference in lifespan



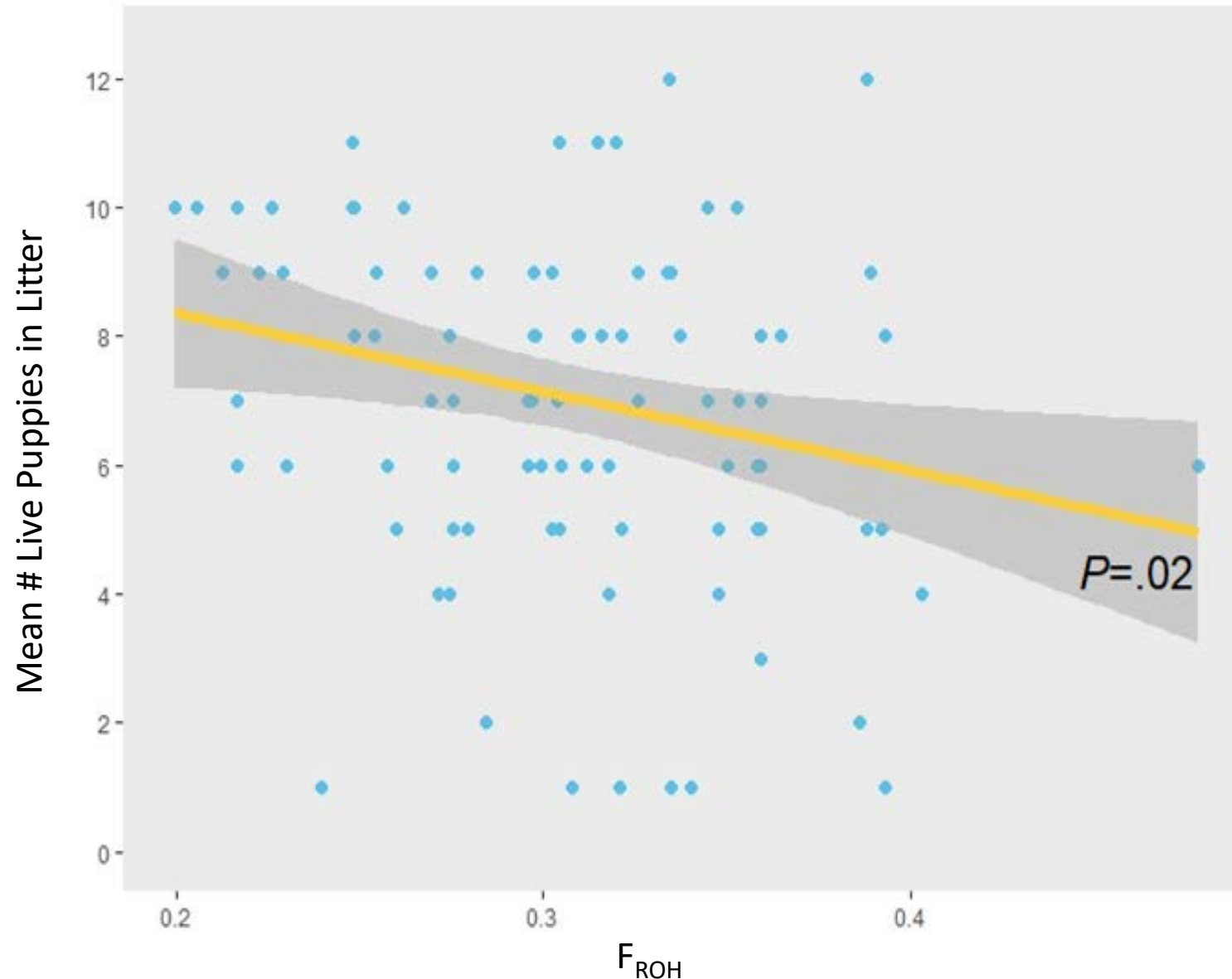
Higher probability of excellent health in low COI dogs based on owner-reported health of 28,660 dogs



Age at death vs inbreeding



Fecundity vs inbreeding in Golden Retrievers



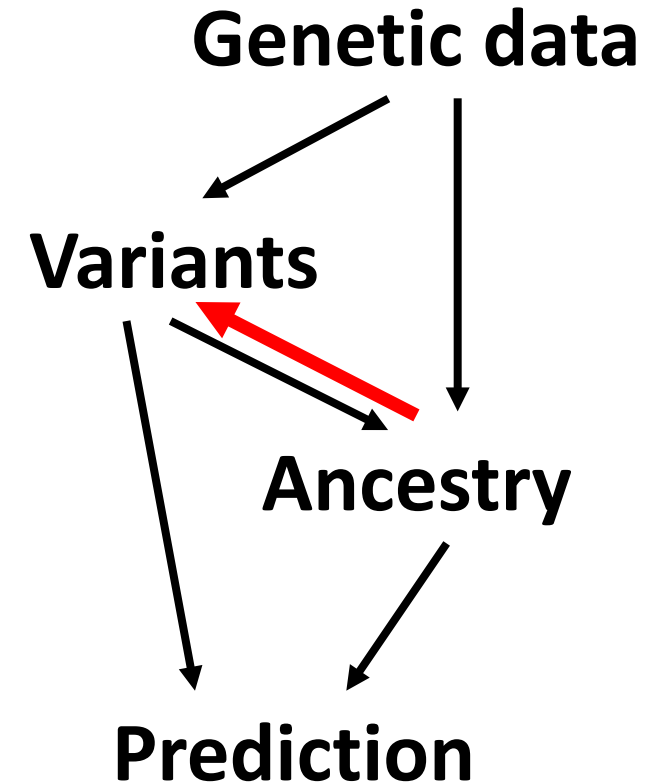
~ 1 less puppy per
10% increase in
inbreeding.



Chu et al, Mamm Genom, 2019

900,000+ dogs in, lessons learned

- Clean data >> dirty data; big data >> small data
 - Goal is not perfect data; it's balancing sample size and signal:noise to maximize statistical power
 - Standardization, tracking, quality control
- Storage is cheap; analysts are expensive
 - Maximize data availability and ease of use
 - Store disparate types of data together
 - Automate discovery
- PLINK works great for 9000 dogs; not for 900k
- SQL works great for 9000 dogs; not for 900k



Dogs are a great model, but they aren't mini-humans

Factor	Humans	Dogs
Exercise/obesity	Very Important	Very Important
Stature	Less Important	Very Important
Age	Very Important (always known)	Very Important (often unknown)
Reproductive history	Important	Very Important
Smoking history	Very Important	Less Important
Inbreeding	Less Important	Very Important

THANKS!



Collaborators

Marta Castelhana (Cornell)
Daniel Promislow (UWash)
Rory Todhunter (Cornell)
Jess Hayward (Cornell)
Laura Sams (Cornell -> UMinn)
Jennifer Yordy (Cornell)
Michelle White (Cornell -> Broad)
Missy Simpson (Morris)

Embark

Erin Chu
Petra Deane-Coe
Brett Ford
Meghan Jensen
Taki Kawakami
Aaron Sams
Frank Andrescavage
Calvin Leather

