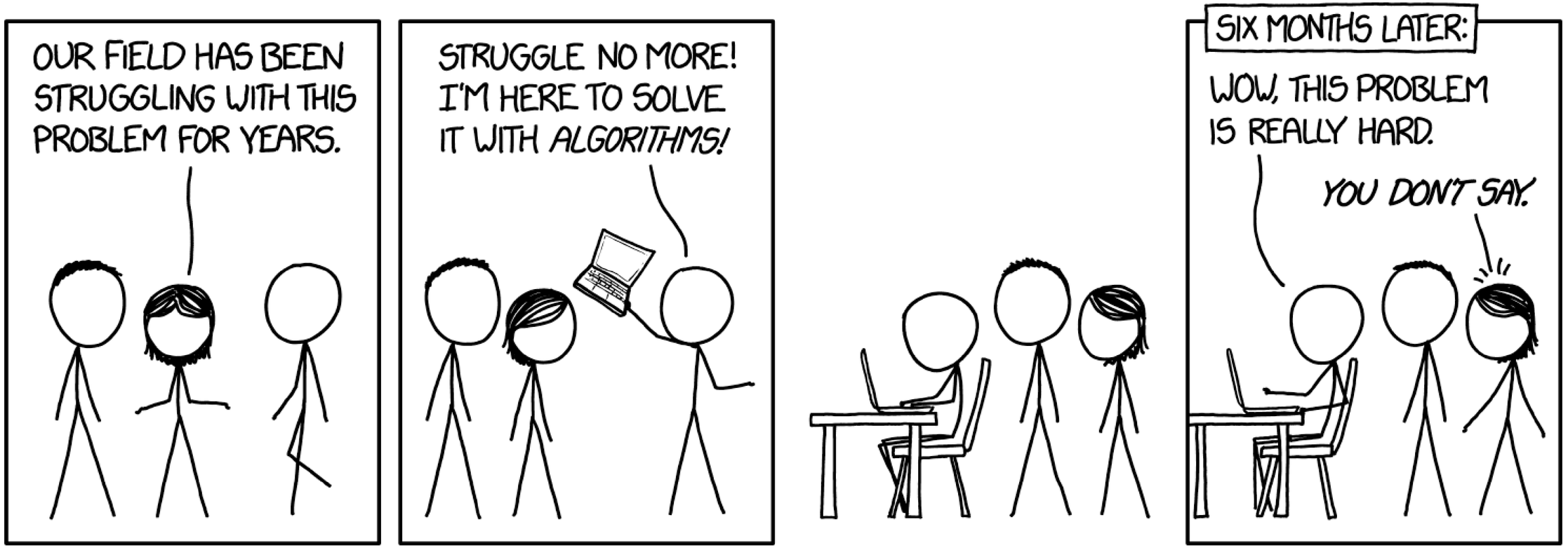
Improving the Performance and Health of All Working Dogs Through Large-Scale, Collaborative Genetics

Elinor K. Karlsson
University of Massachusetts Medical School & Broad Institute of MIT and Harvard

BROAD INSTITUTE

UMASS University of Massachusetts Medical School

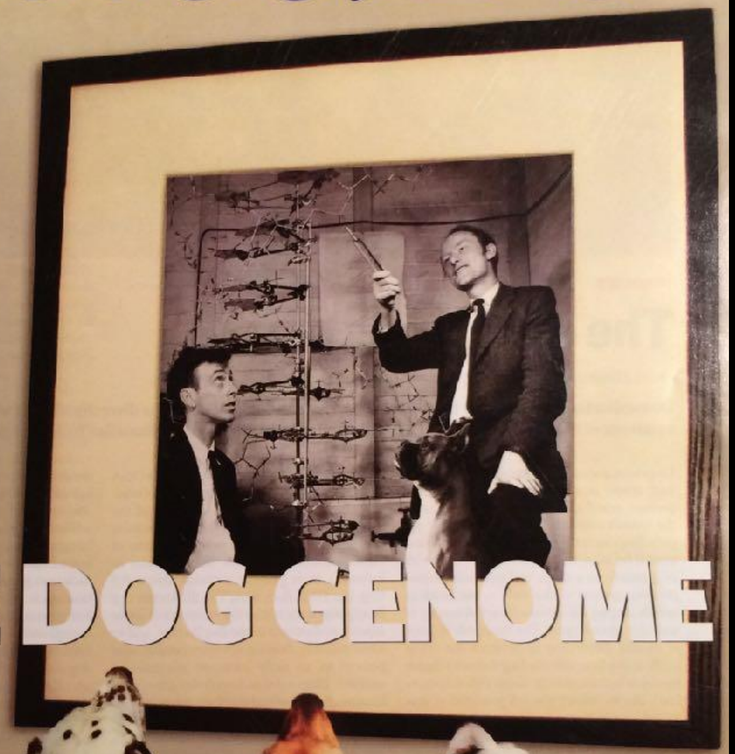


https://xkcd.com/1831/

8 December 2005 | www.nature.com/nature

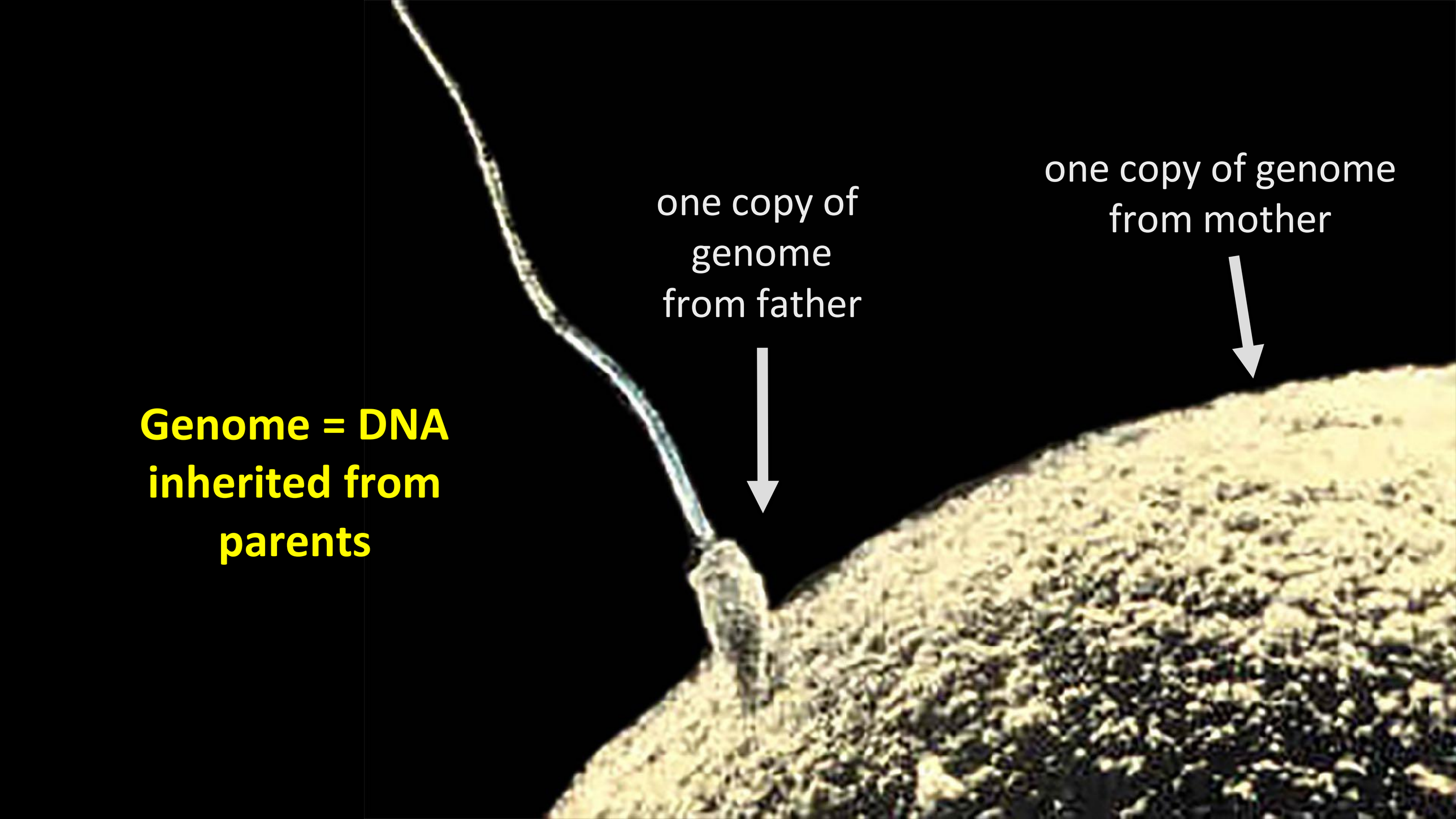
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

nature



THE DOG GENOME

“GENOMICS”

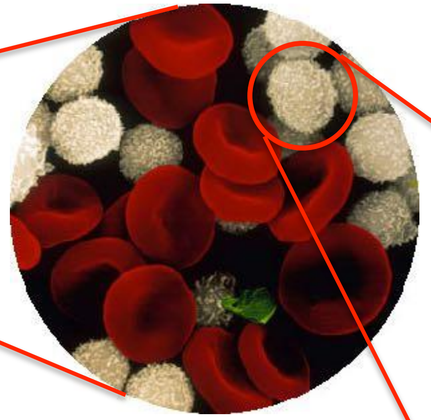

one copy of genome from father
one copy of genome from mother
Genome = DNA inherited from parents

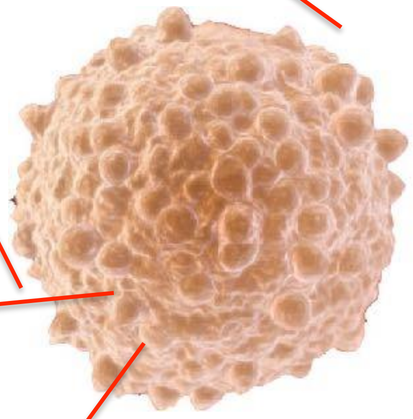

Fertilized egg =
1 cell =
1 genome

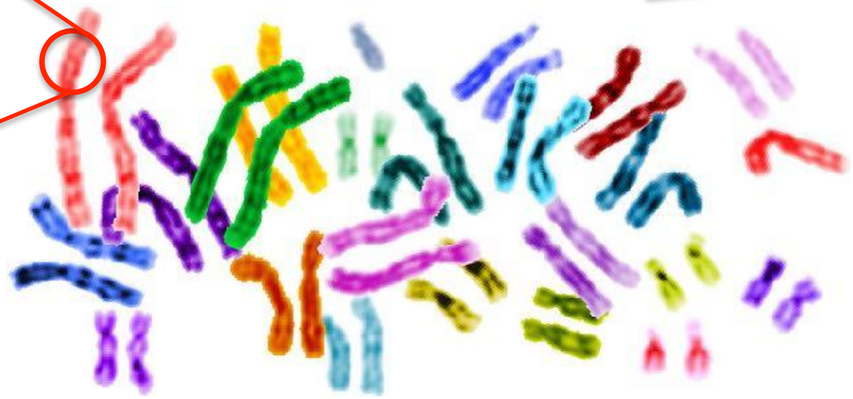
~40,000,000,000,000 cells*

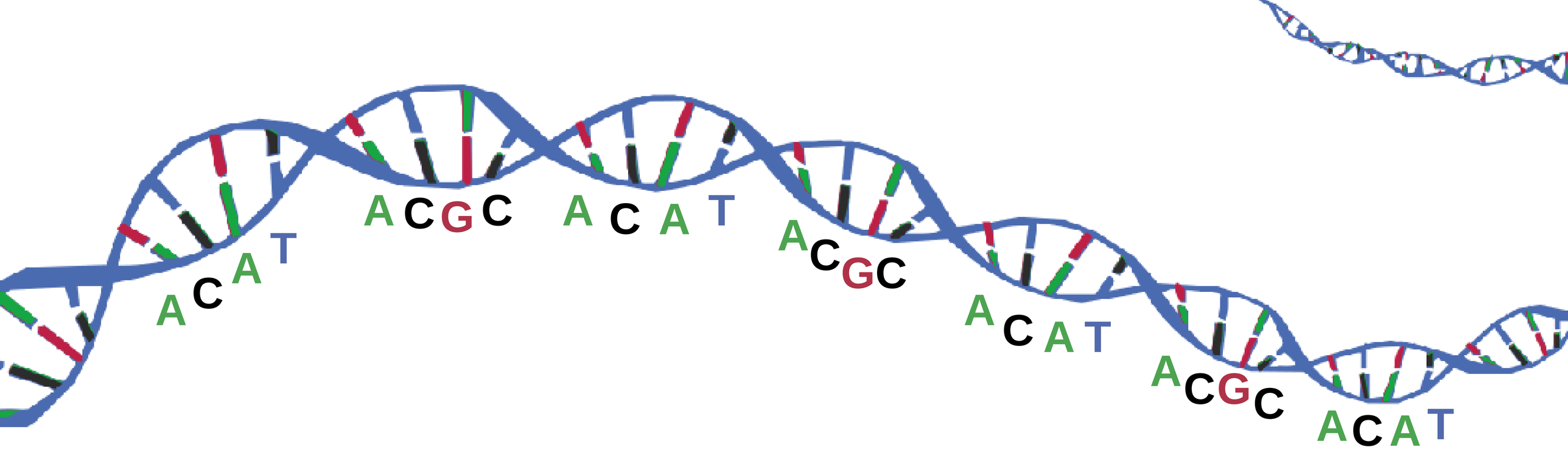
** very rough estimate*

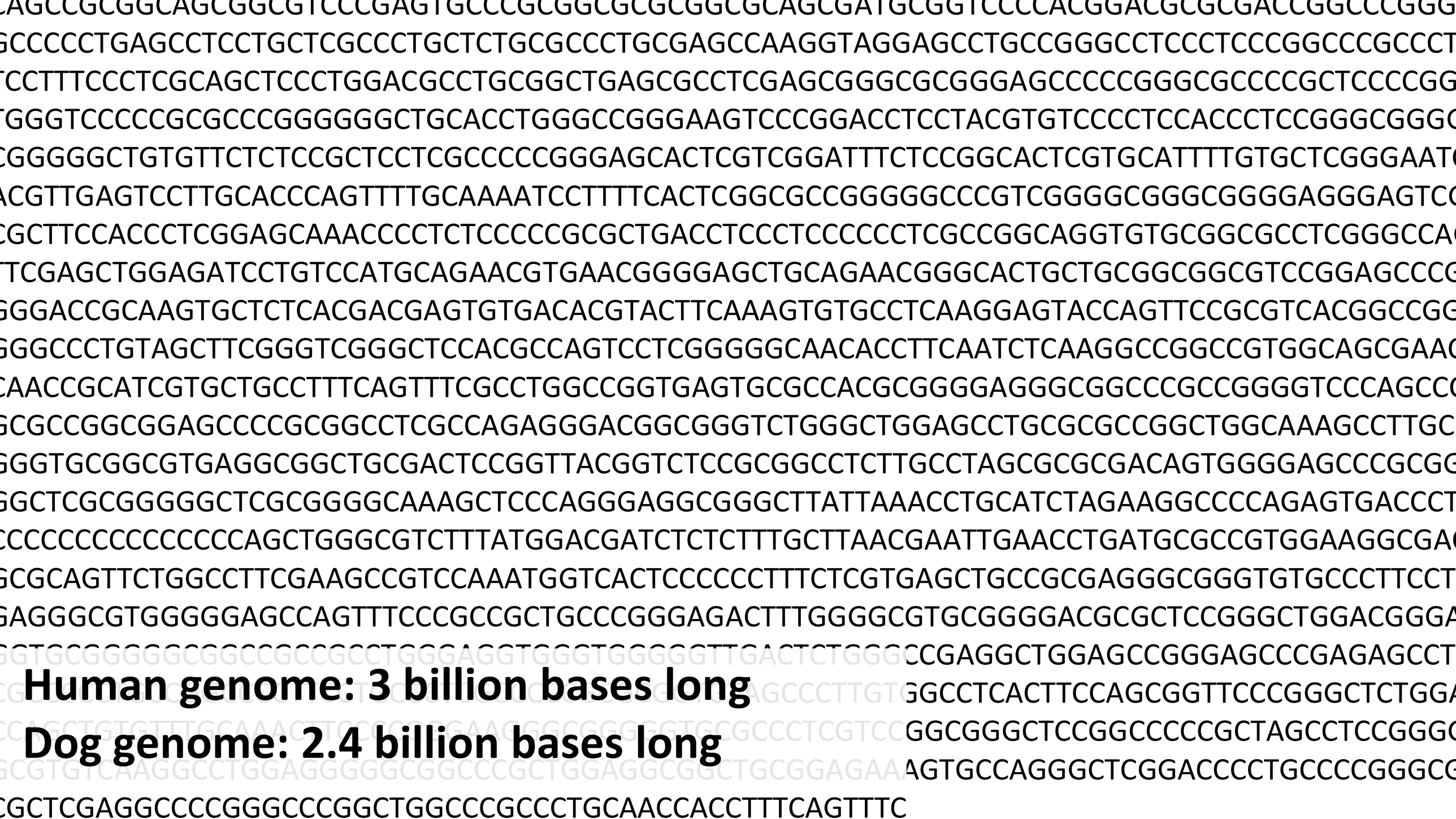
(nearly) every cell carries copy of genome



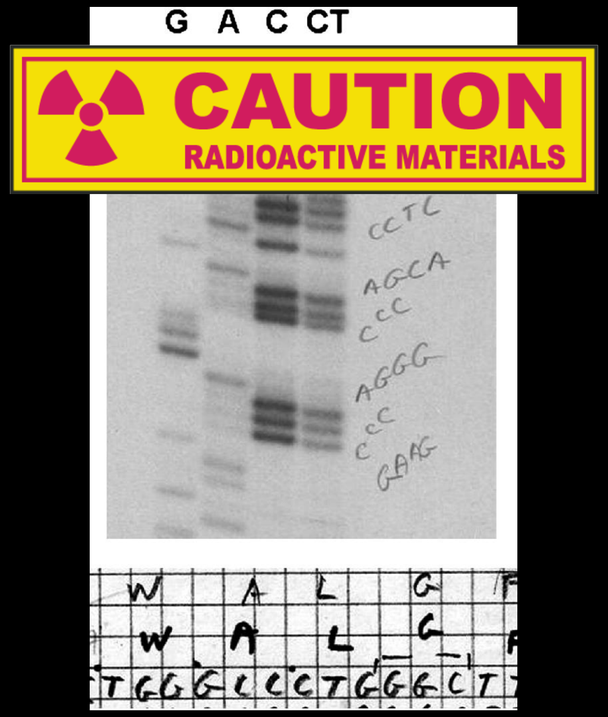





ACAT
ACGC
ACAT
ACGC
ACAT
ACGC
ACAT


Human genome: 3 billion bases long
Dog genome: 2.4 billion bases long

ancient times



Maxam-Gilbert sequencing

Cost per genome: impossible

late 1990s

Automated Sanger sequencing

$5 billion+ to sequence the first human genome ($50-100 million after that)

2007

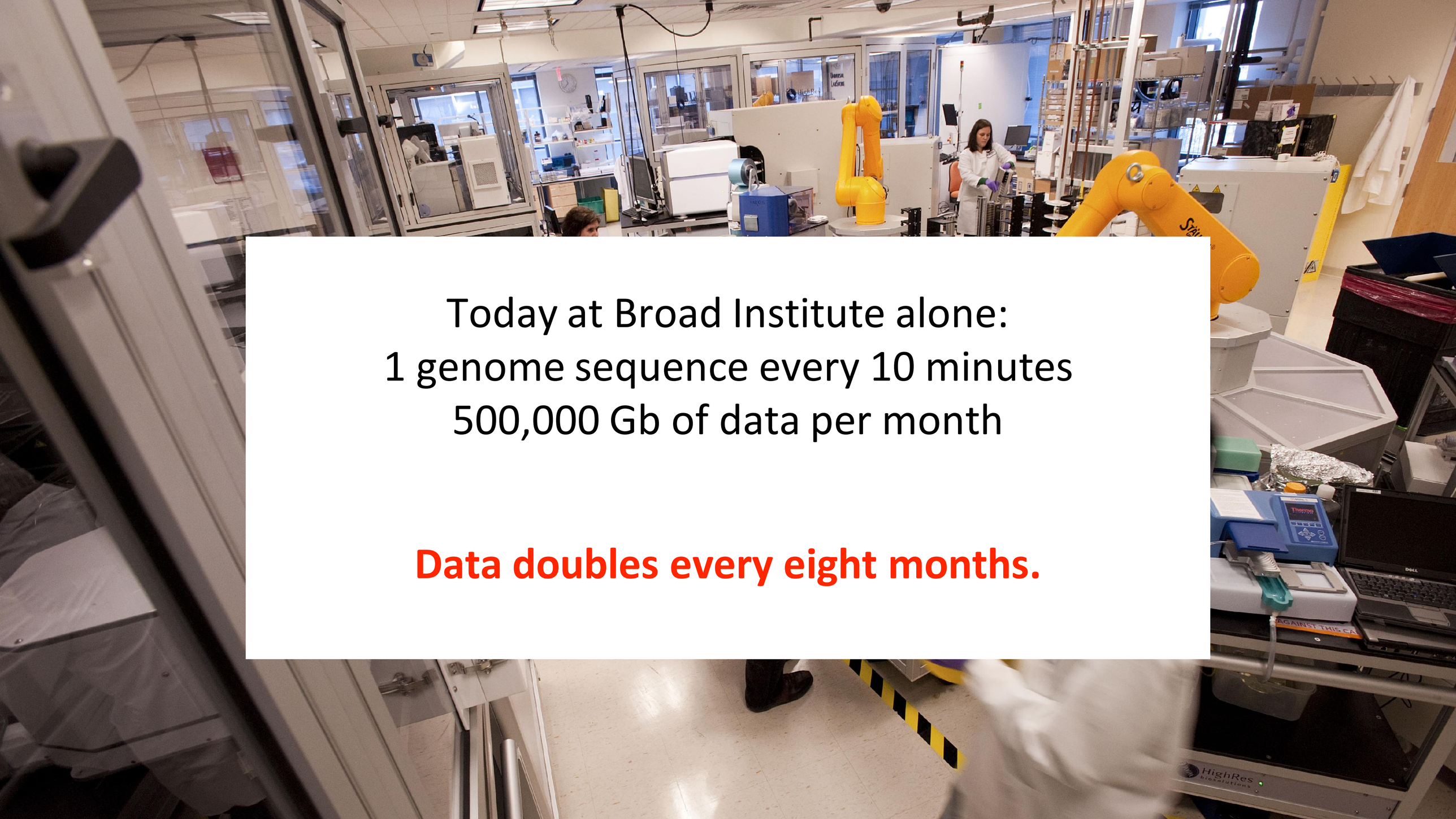
Illumina HiSeq 2000
$5000-$10,000 per genome

2014

Illumina HiSeq X
$750 / genome
$100 -> most info

NEXT GENERATION SEQUENCING


STÄUBLI
DANGER
THIS MACHINE STARTS AND STOPS AUTOMATICALLY
HighRes



Today at Broad Institute alone:
1 genome sequence every 10 minutes
500,000 Gb of data per month

Data doubles every eight months.

UK

biobank uk

goal: 500,000

NIH (USA)

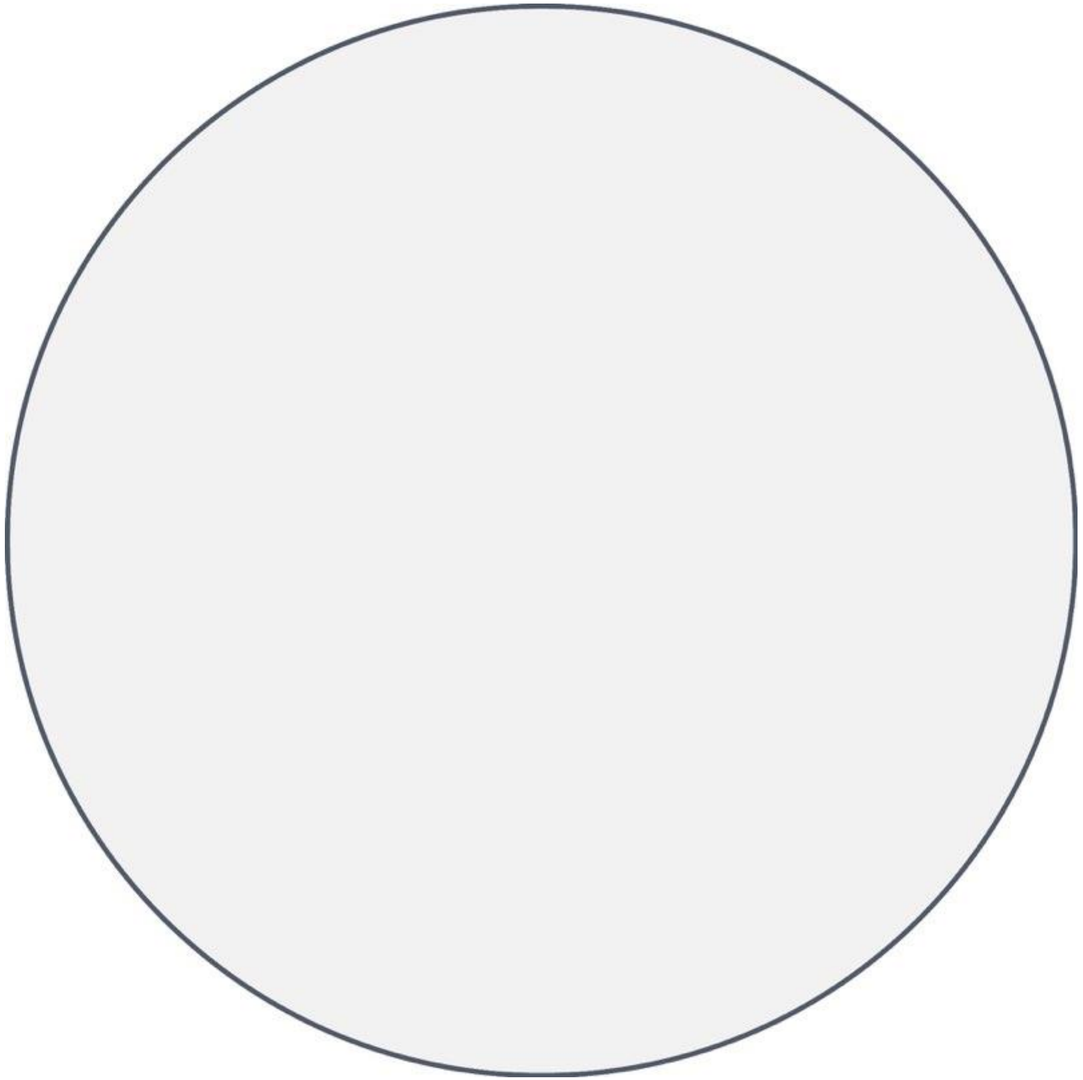
All of Us℠

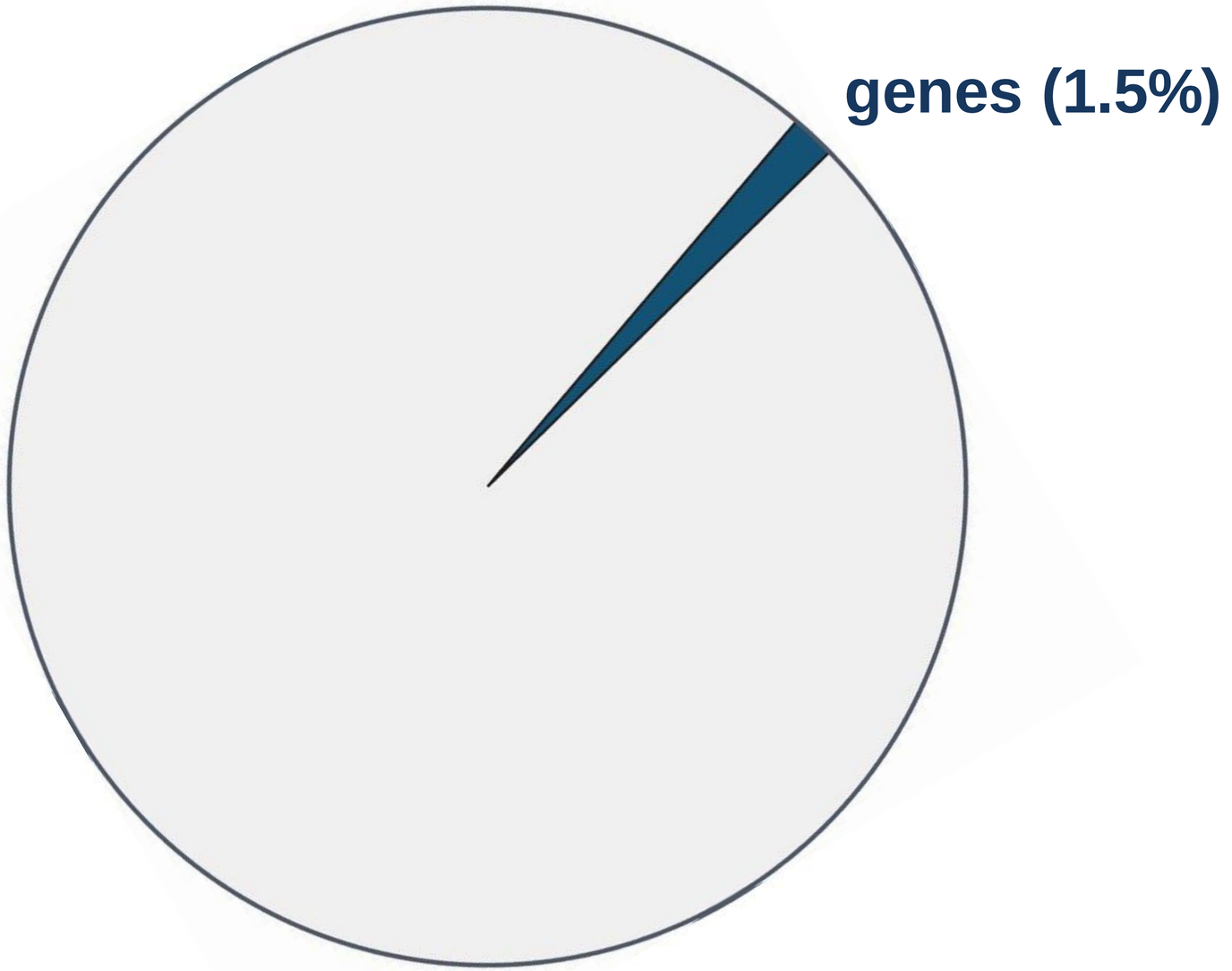
THE FUTURE OF HEALTH BEGINS WITH YOU

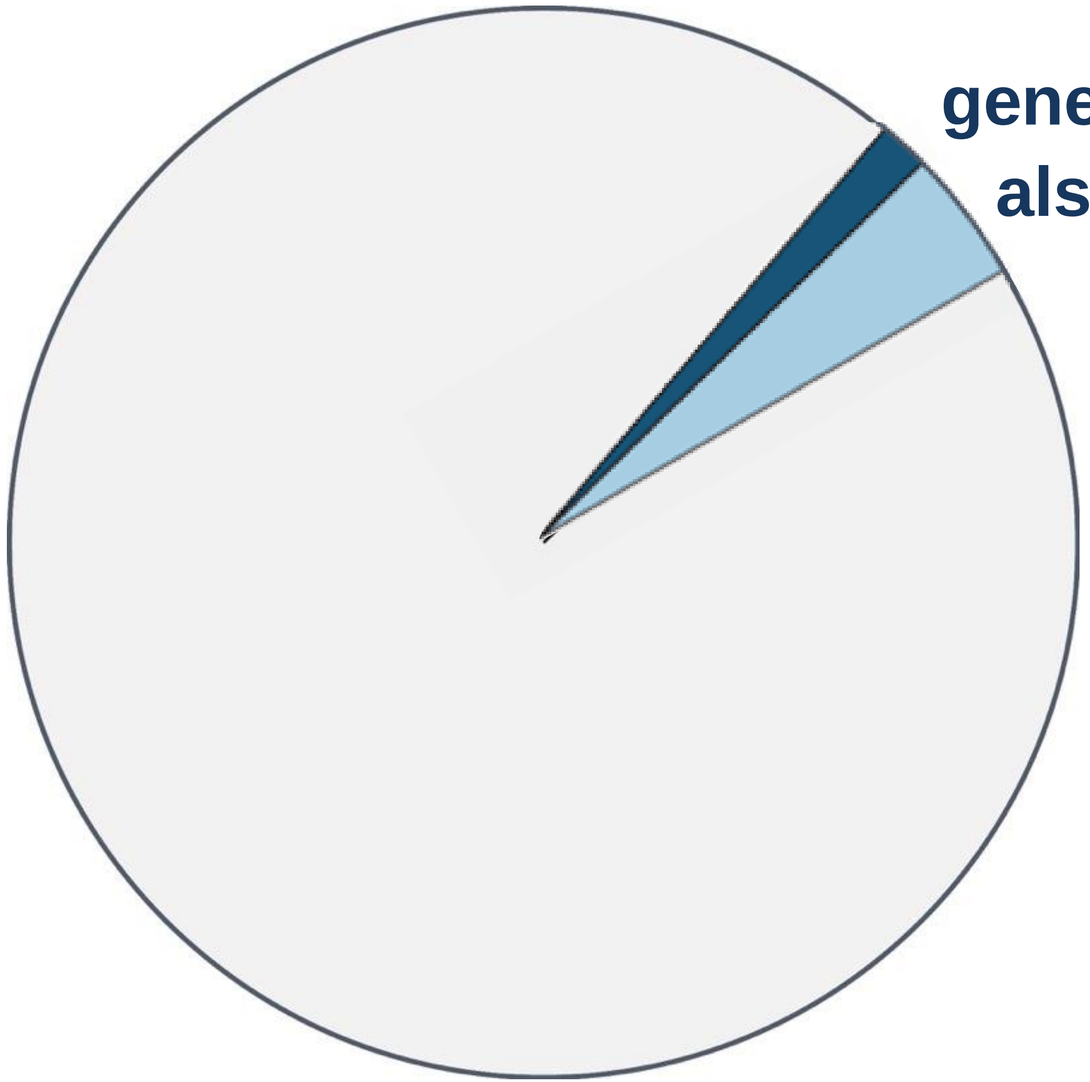
The Precision Medicine Initiative®

goal: 1,000,000

< 1000 dogs - but that is starting to change




genes (1.5%)

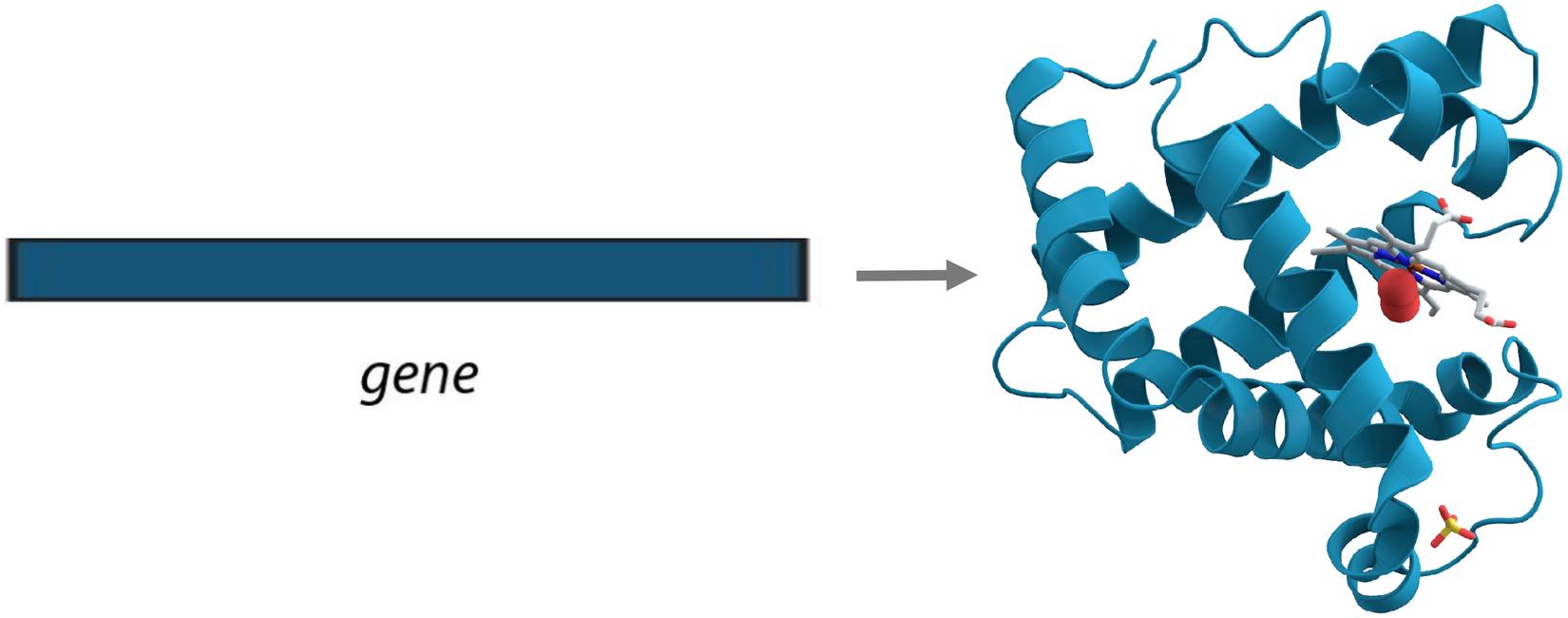


genes (1.5%)

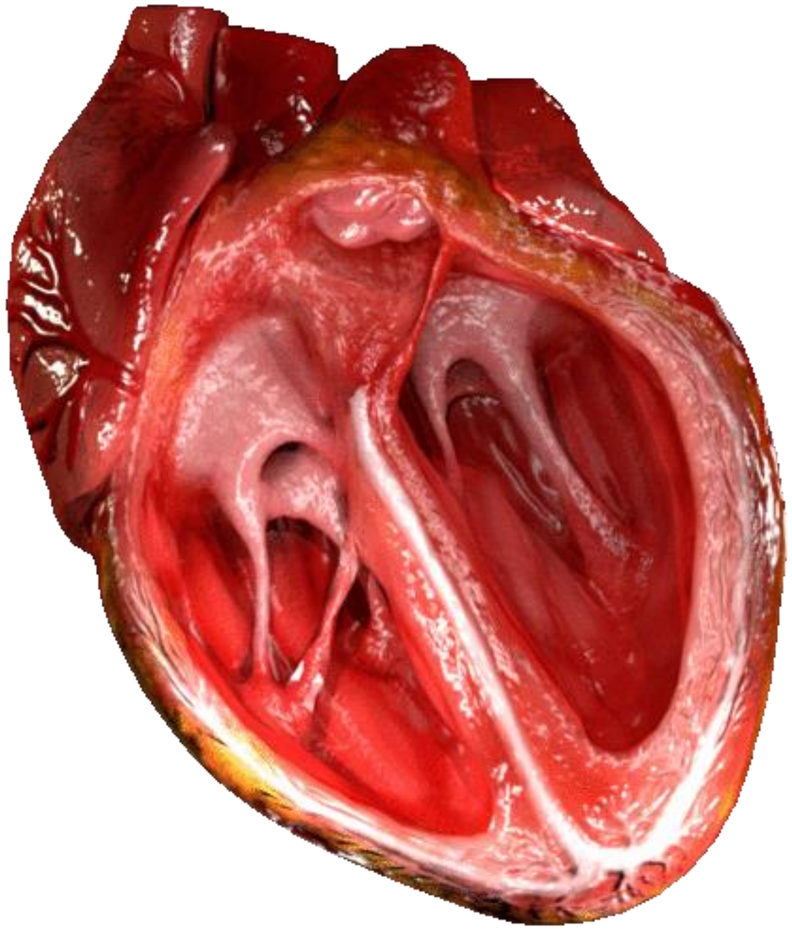
also important (7-10%)

(regulatory)

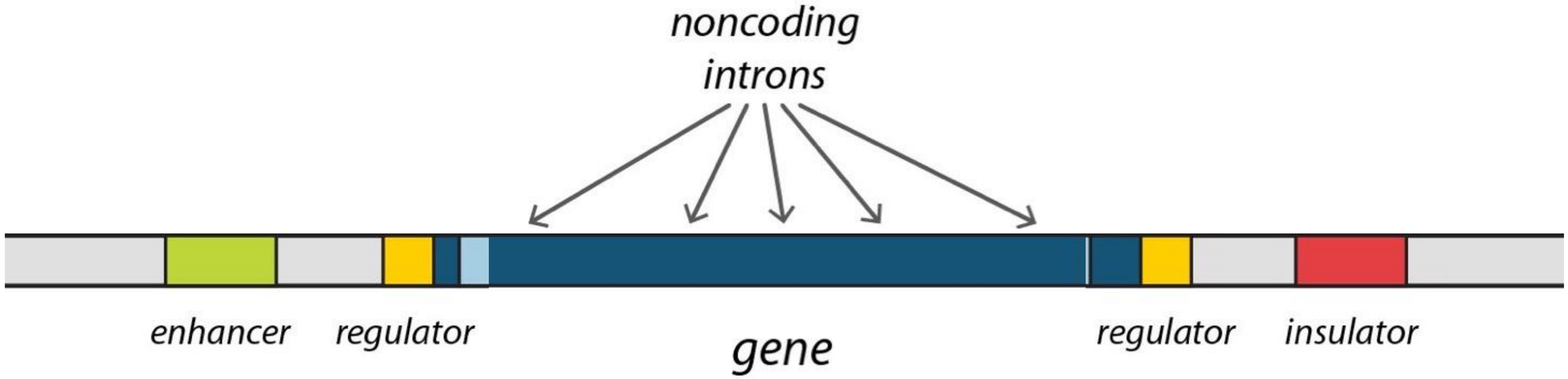
If every cell in your body has the same DNA (your genome) …

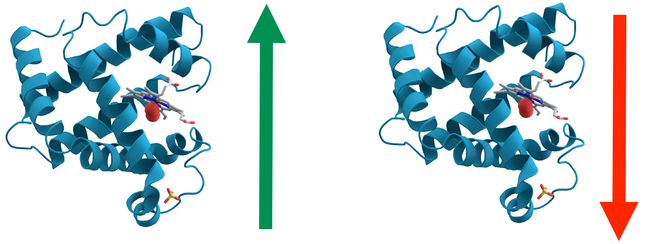


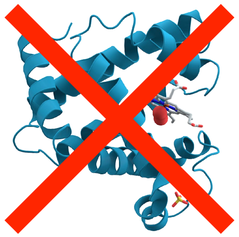
Why isn’t every cell the same?



Regulatory sequence controls how the DNA is used

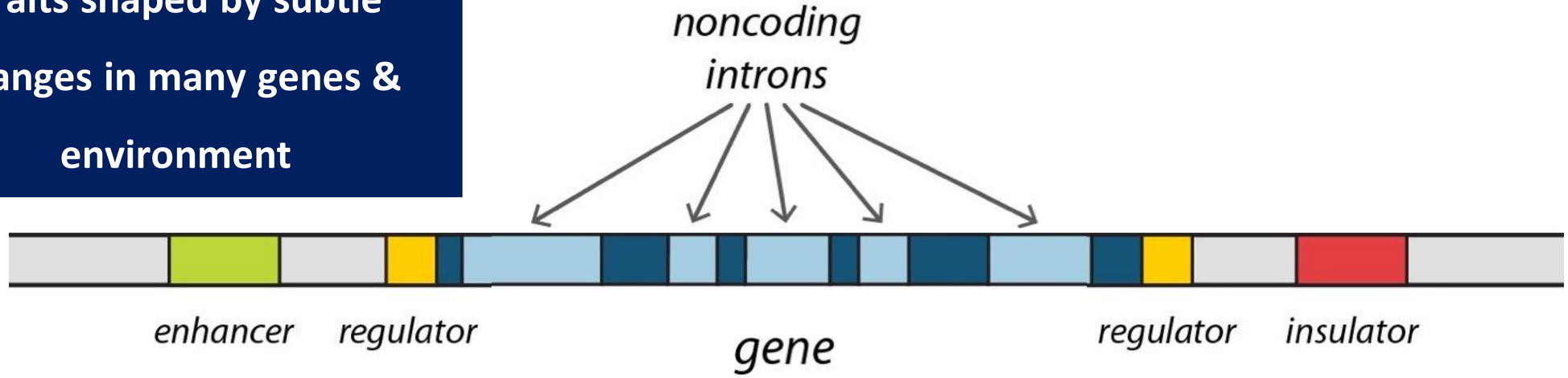


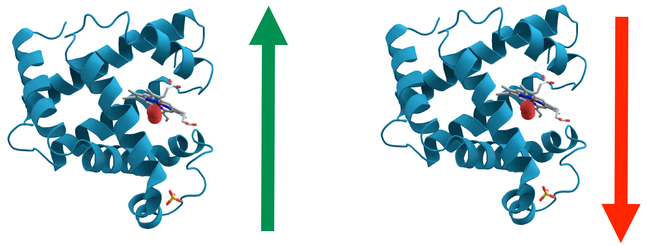


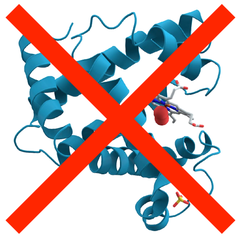


Complex traits, like behavior, result from changes in gene regulation

Traits shaped by subtle changes in many genes & environment







How do we find important parts of the genome?

How do we figure out what they do?

Why IS dog genetics a thing?


Herding
Herding
Guarding
Sight
Speed
Retrieving
Guarding
Retrieving

Aggression

Anxiety

Cognitive disfunction

Compulsive disorders

Noise & storm phobia

Impulsivity

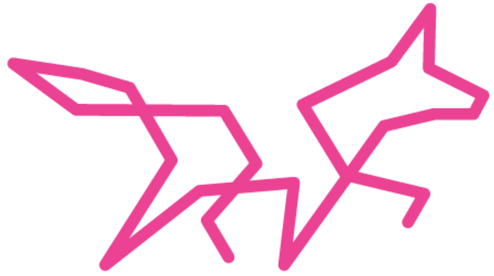
Allergies

Cancer

Others

How do changes in DNA lead to changes in dog behavior & health?

Can this help us find better treatments for diseases?



DARWIN'S DOGS

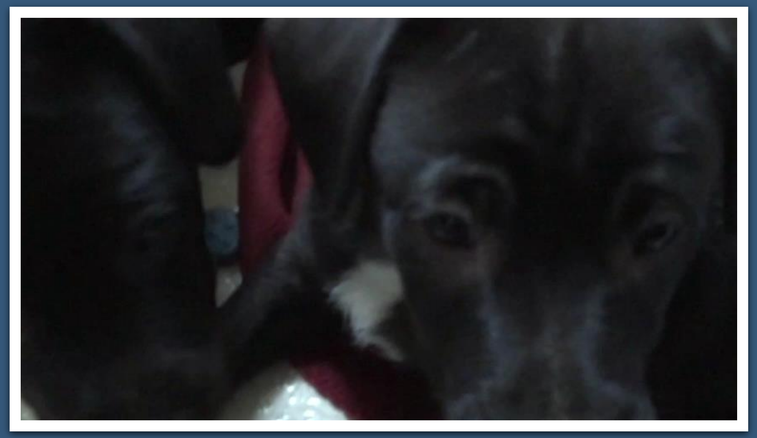
UMASS University of Massachusetts Medical School + BROAD INSTITUTE + IAABC INTERNATIONAL ASSOCIATION OF ANIMAL BEHAVIOR CONSULTANTS

Can we use genomics to:

Help *predict* whether a puppy will be successful

Increase the rate of selection in breeding programs

Put dogs into the right jobs



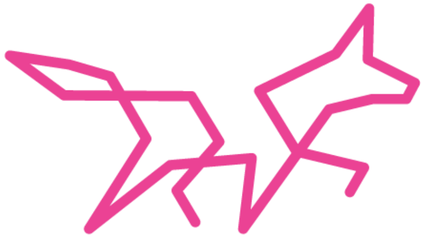
PLUS: Can be done at young age

CAUTION: Will never be perfect (environment matters)

We’re building a shared data resource to accelerate research

No one owns the data
We’ll share share any genetic tests we develop

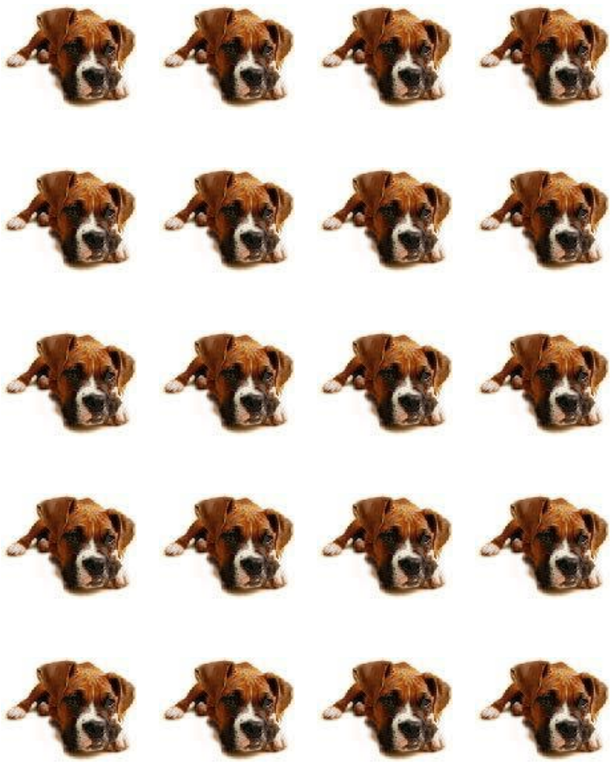
WORKING DOG
PROJECT



DARWIN’S DOGS

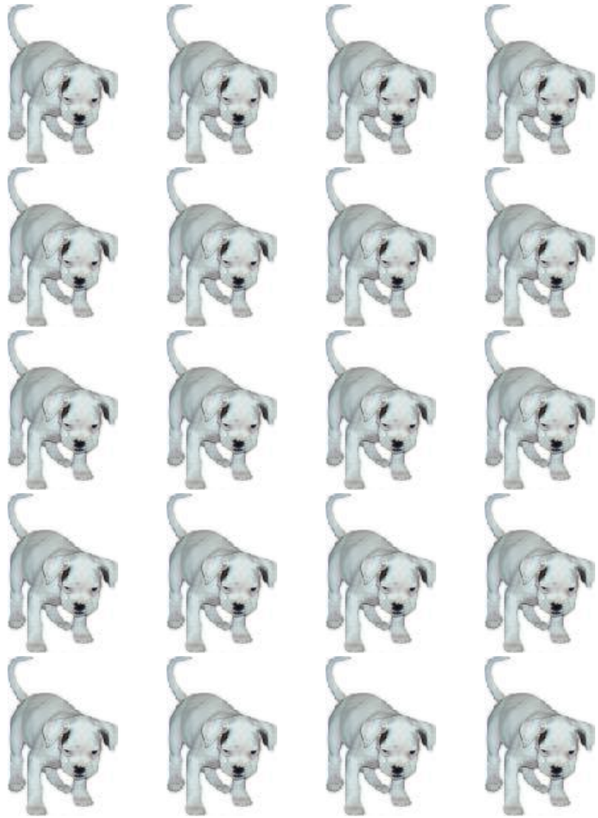
Mapping genes in dogs:
the science

controls
(healthy dogs)



Where are they different?

cases
(dogs with disease or trait)



Coat Color in Boxers

9 Solid Boxers

$S^{}S^{}$

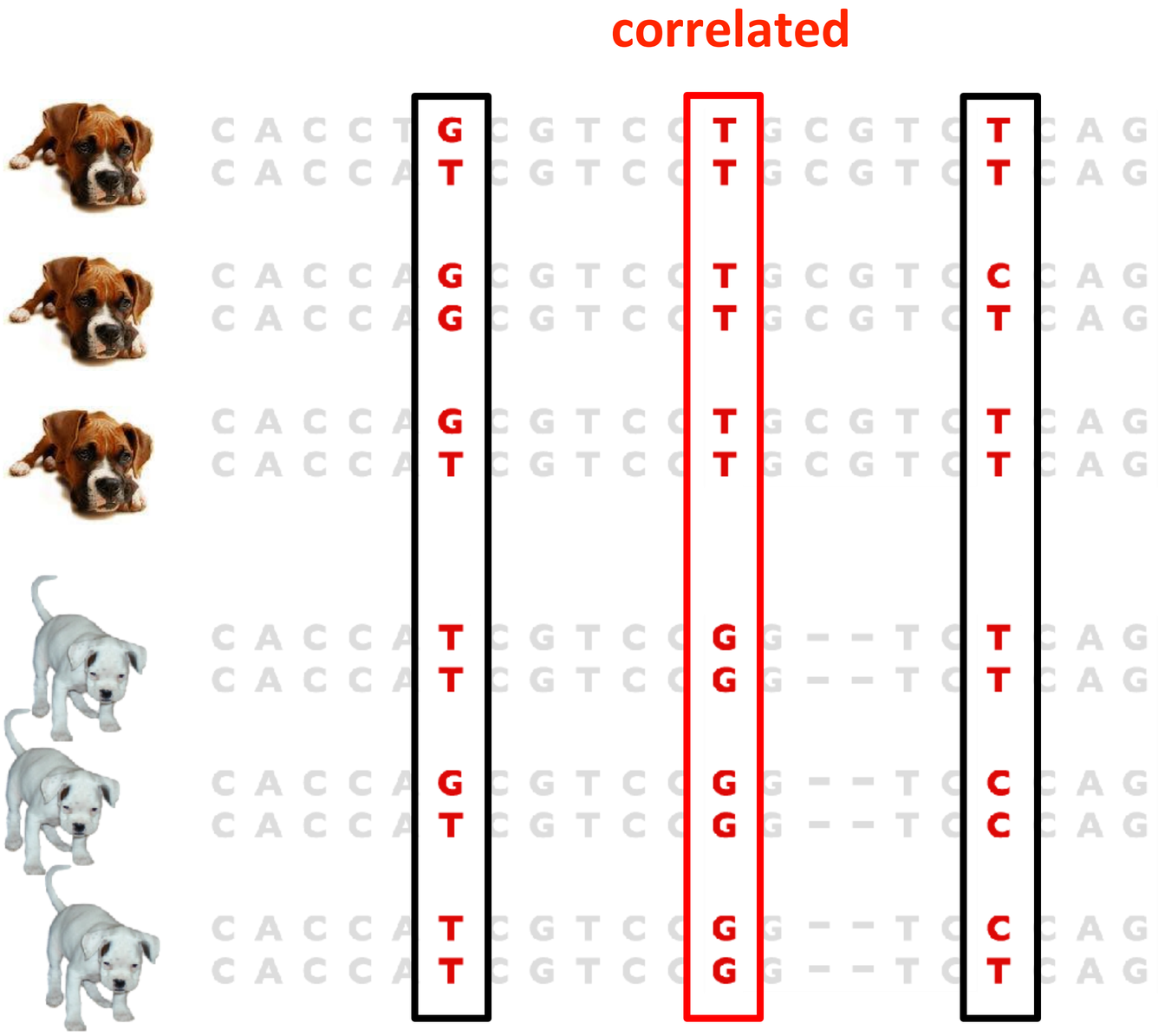
$S^{}S^{w}$

10 White Boxers

$S^{w}S^{w}$

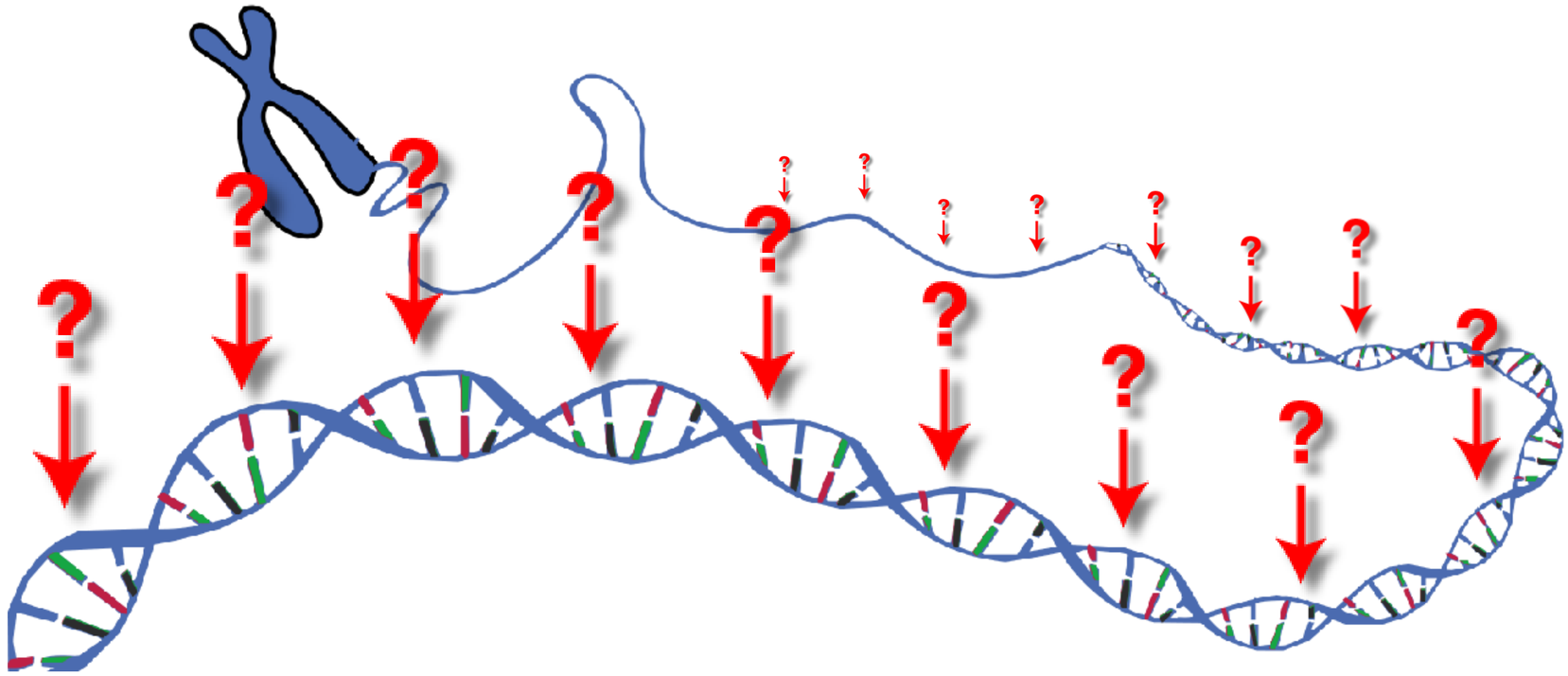
~2% deaf

“Is this marker correlated with my trait?”

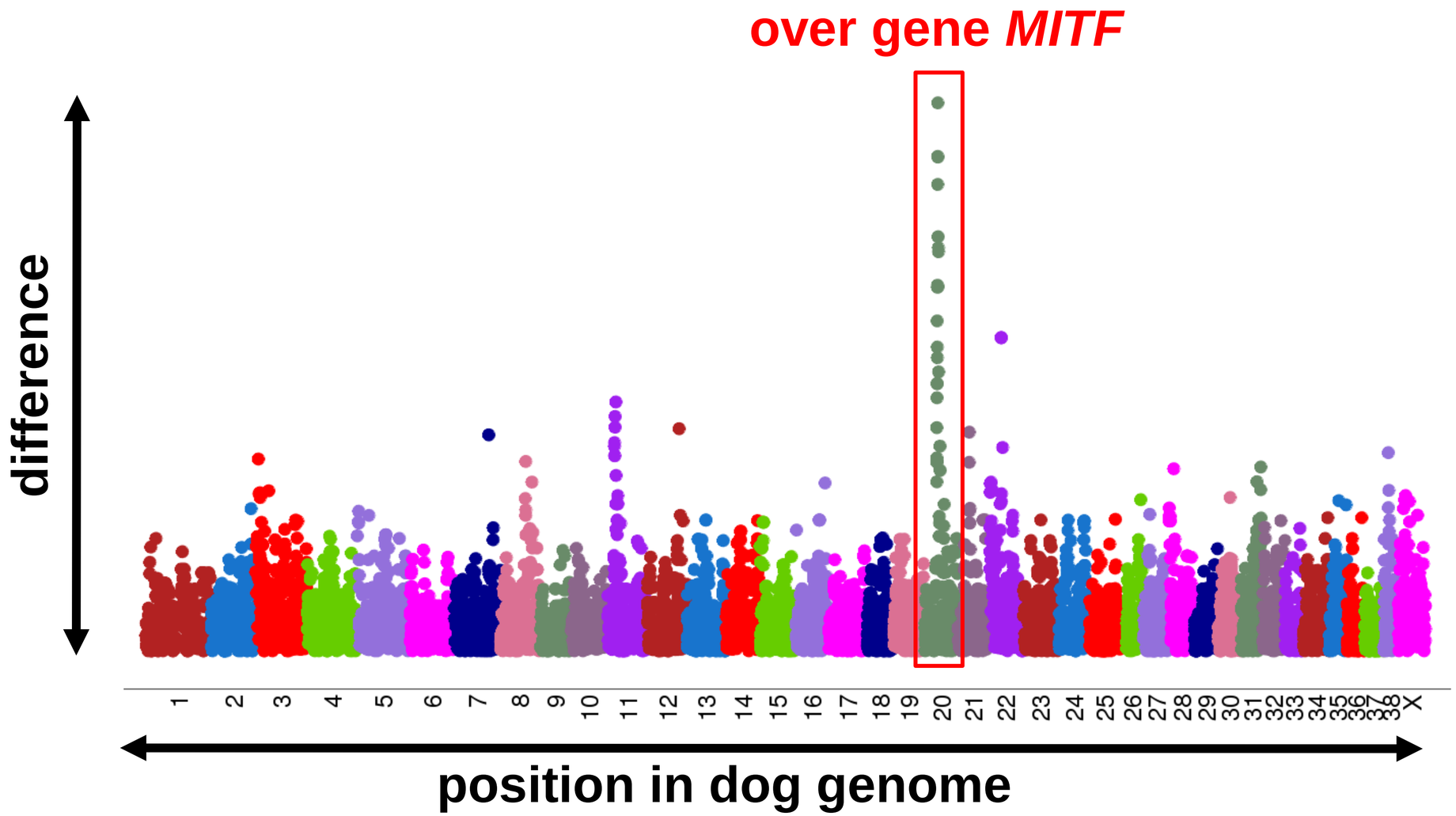


Genome-wide association study (GWAS)

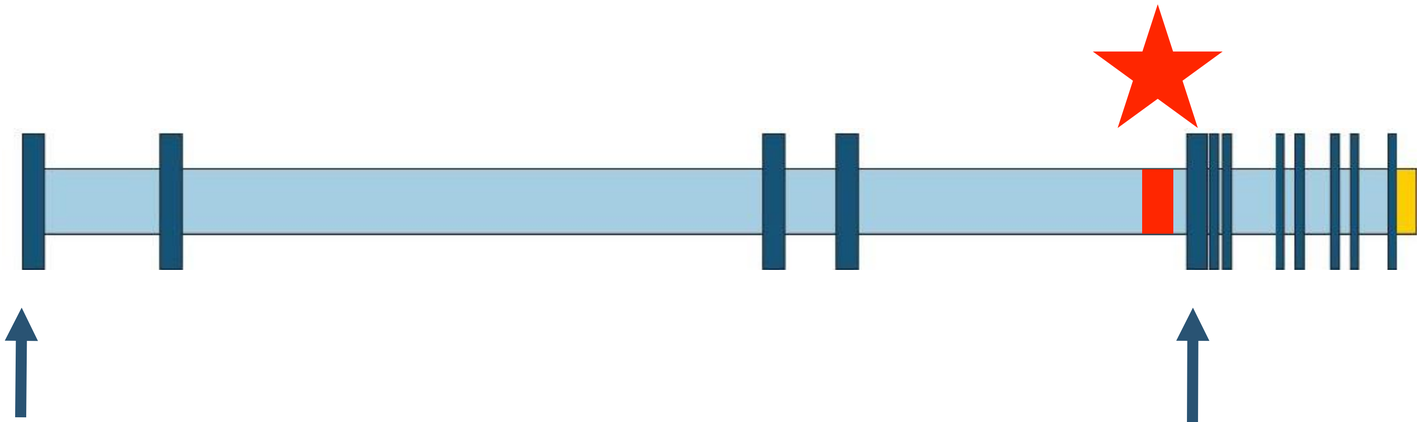
For each SNP, ask “Is this SNP correlated with my trait?”



Find region with most correlated SNPs



Karlsson et al, Nature 2008



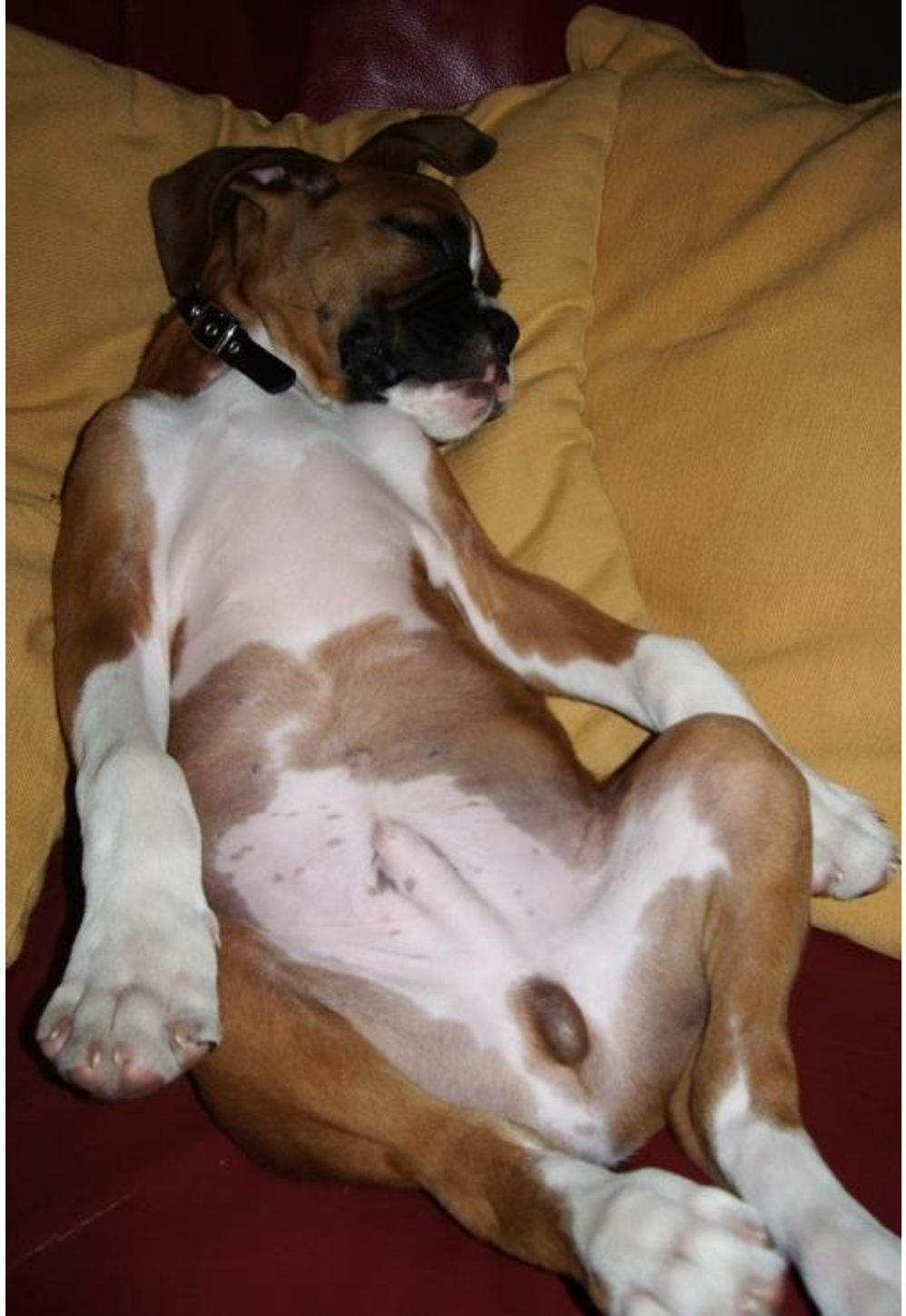
MITF

start of “A”: version:
needed for healthy eyes

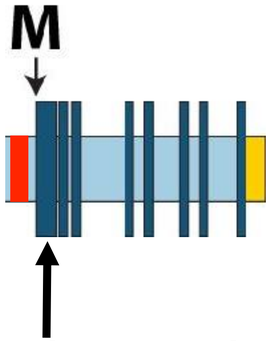
start of “M” version:
needed for normal
pigmentation

makes protein

regulatory



hypothesis: regulation of M is disrupted



makes protein needed for pigment cell migration

Karlsson et al, Nature 2008

Illustrates power of dog genetics

Mice with broken MITF protein:

White

Deaf

Blind

Dead

Dogs with MITF regulatory variants:

White

Rarely deaf

Normal eyes

Not dead

Compulsive disorder in dogs

normal behavior done too much

distressing, time-consuming and impairing

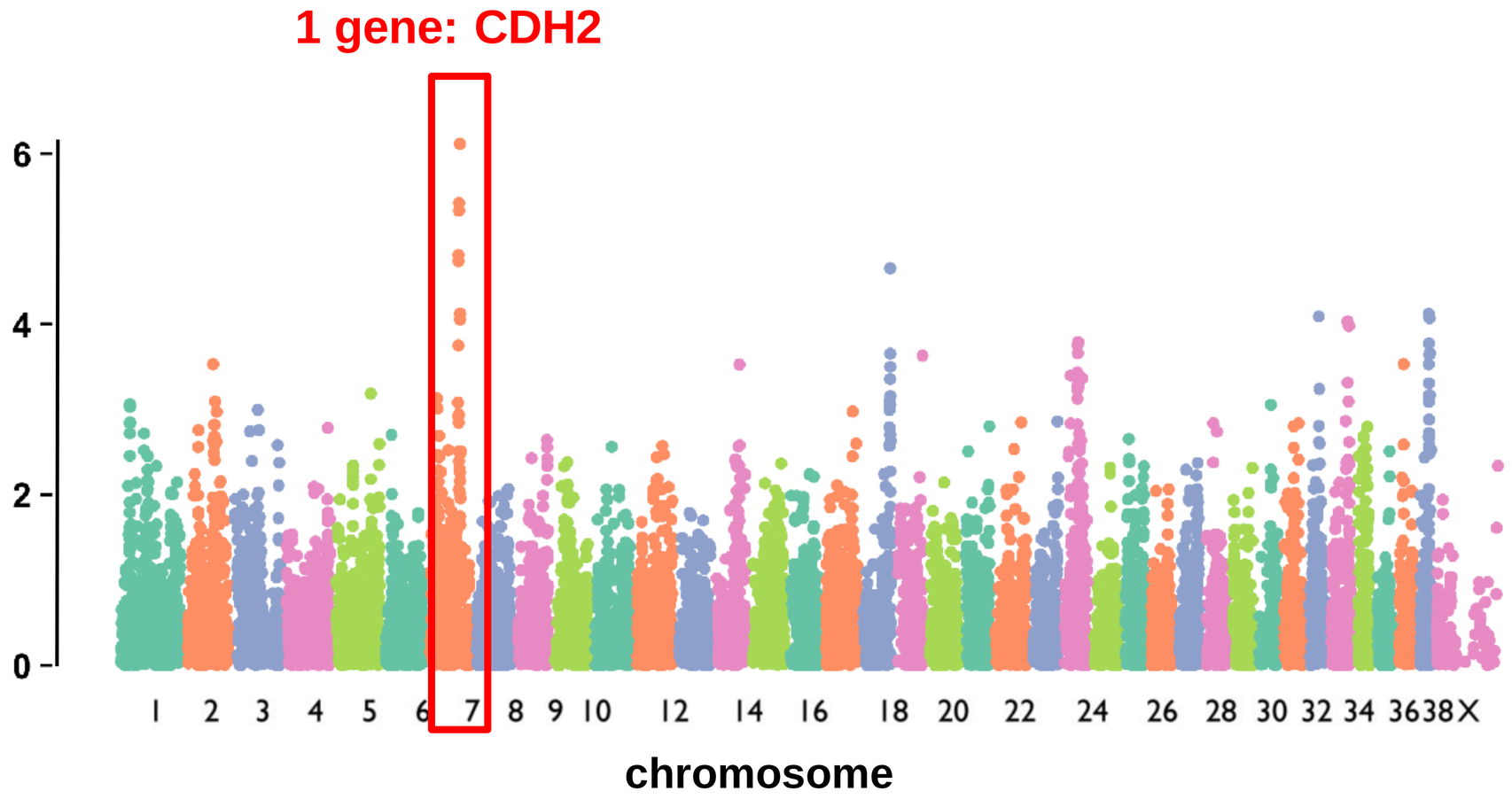
onset in adolescence

highly heritable

poor response to treatment (including SSRIs)

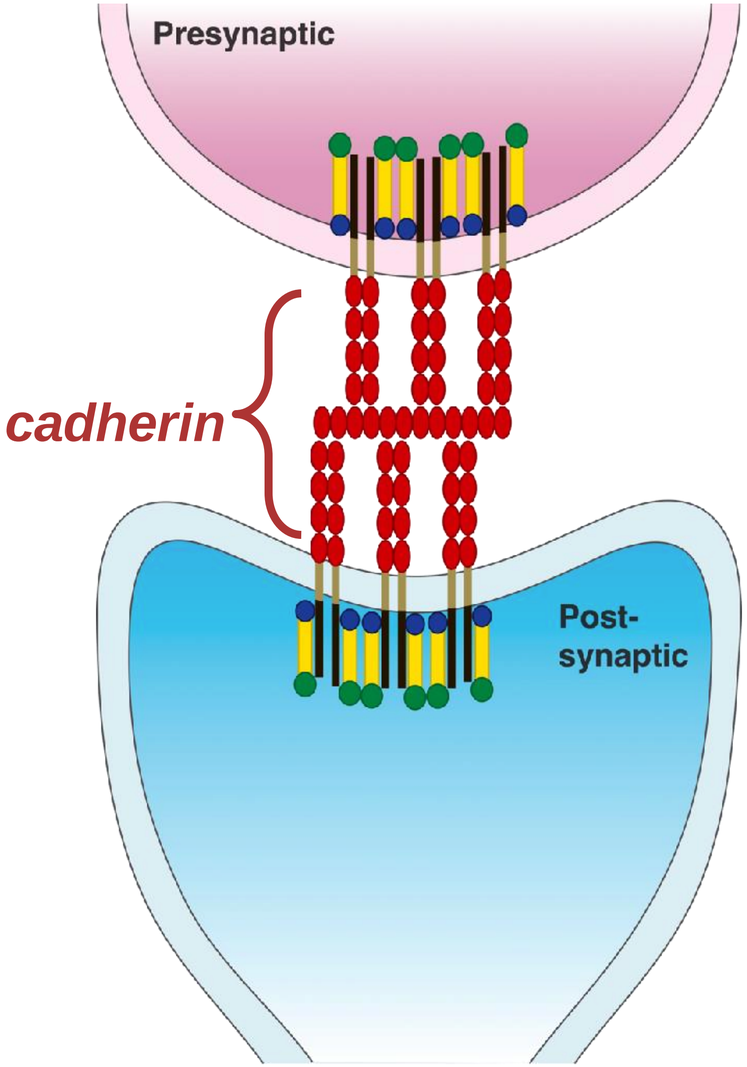
Movies by Alice Moon-Fanelli

Compare genomes of 92 affected and 67 healthy dobermans



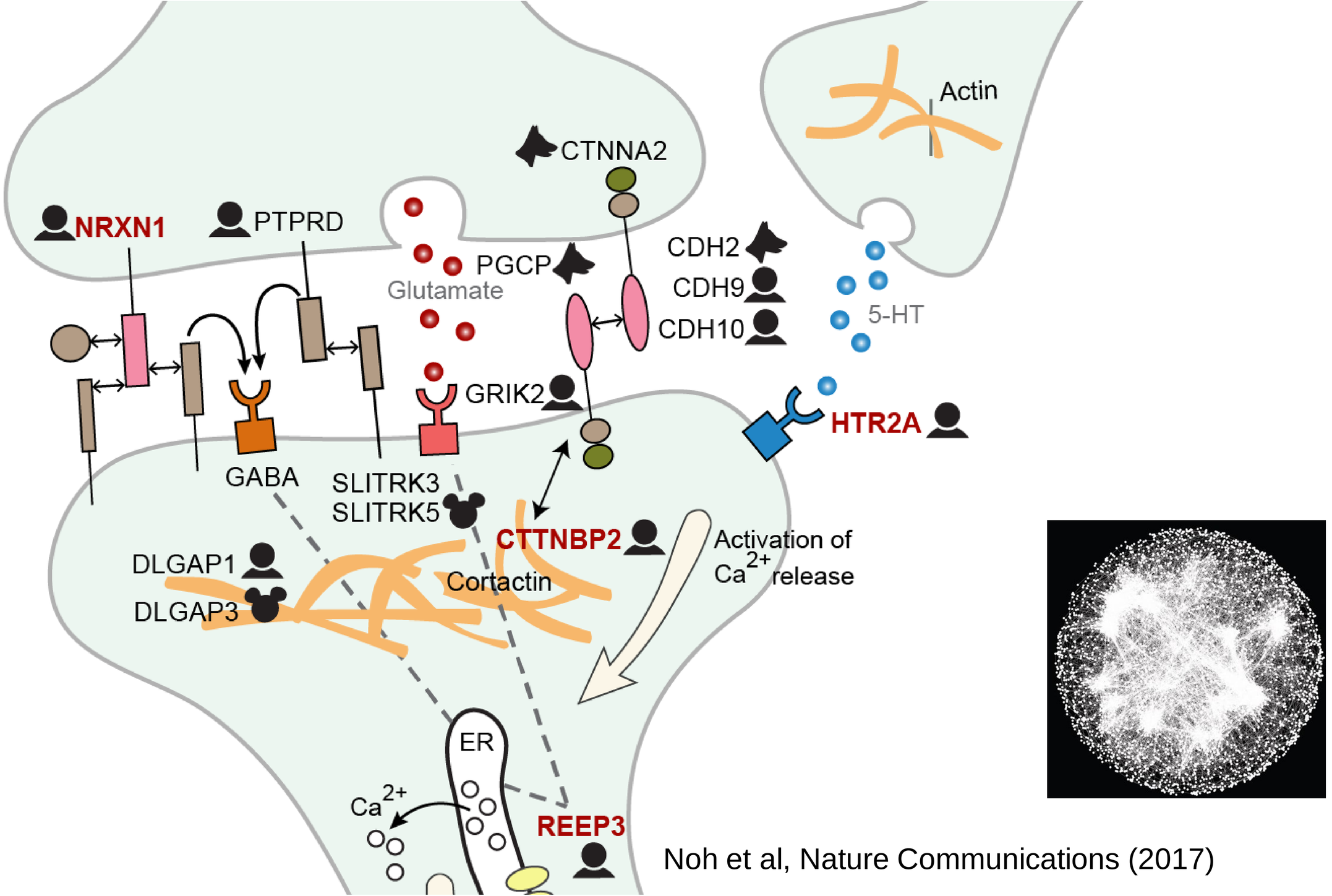
Dodman et al, Molecular psychiatry 2010

Function of CDH2



- **Neural cadherin**
- **Found in glutamatergic synapses**
- **Synaptic plasticity (learning and memory)**

Dodman et al, Molecular psychiatry 2010



Noh et al, Nature Communications (2017)

Can we predict which dobermans will get OCD?

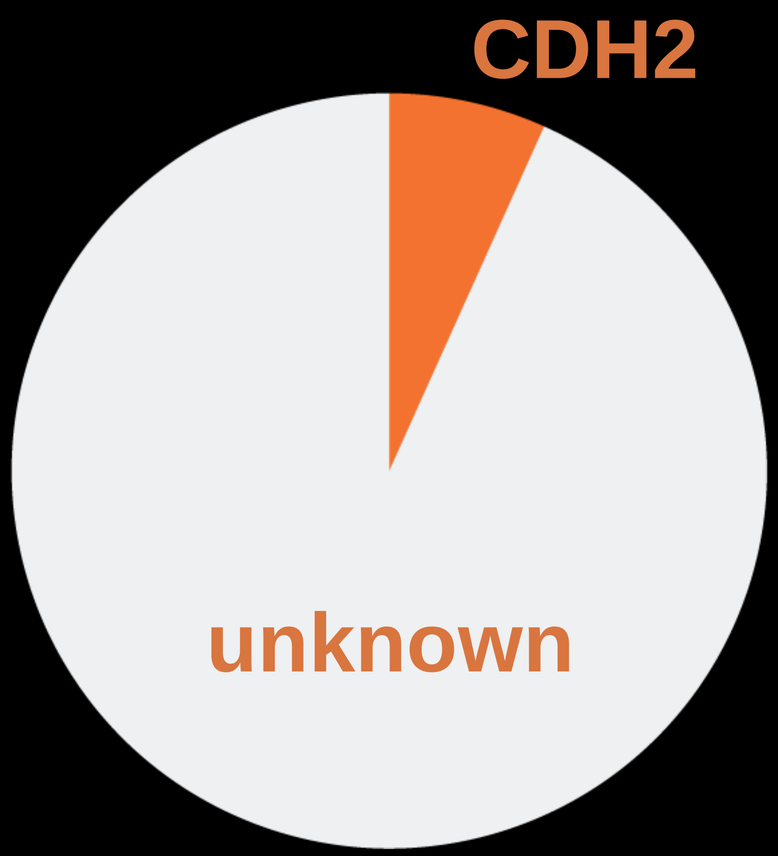
NO

Read more:

karlssonlab.org/pets/

Problem #1: we've only figured out part of the story

How much of the OCD risk in dobermans did we explain?

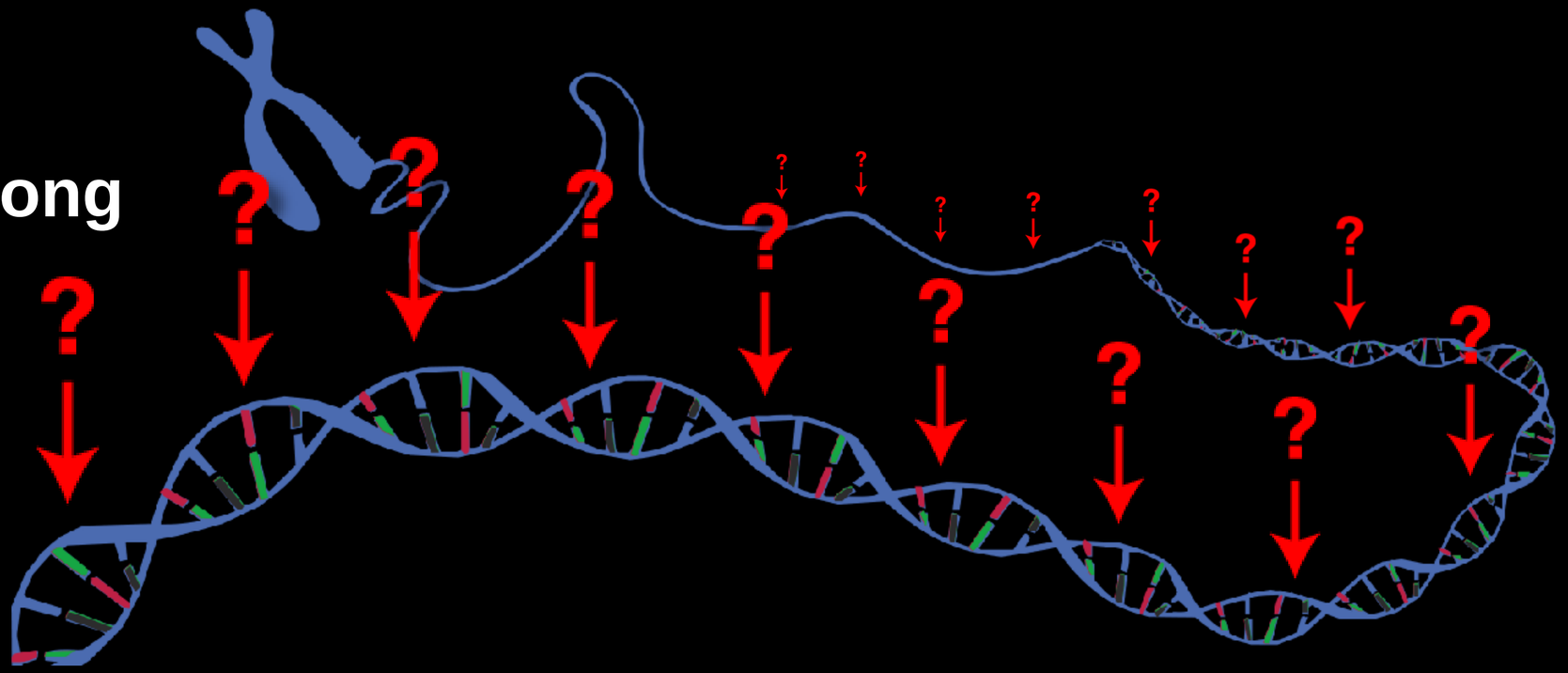


Problem #1: we've only figured out part of the story

Problem #2: We didn't find the "causal" mutation

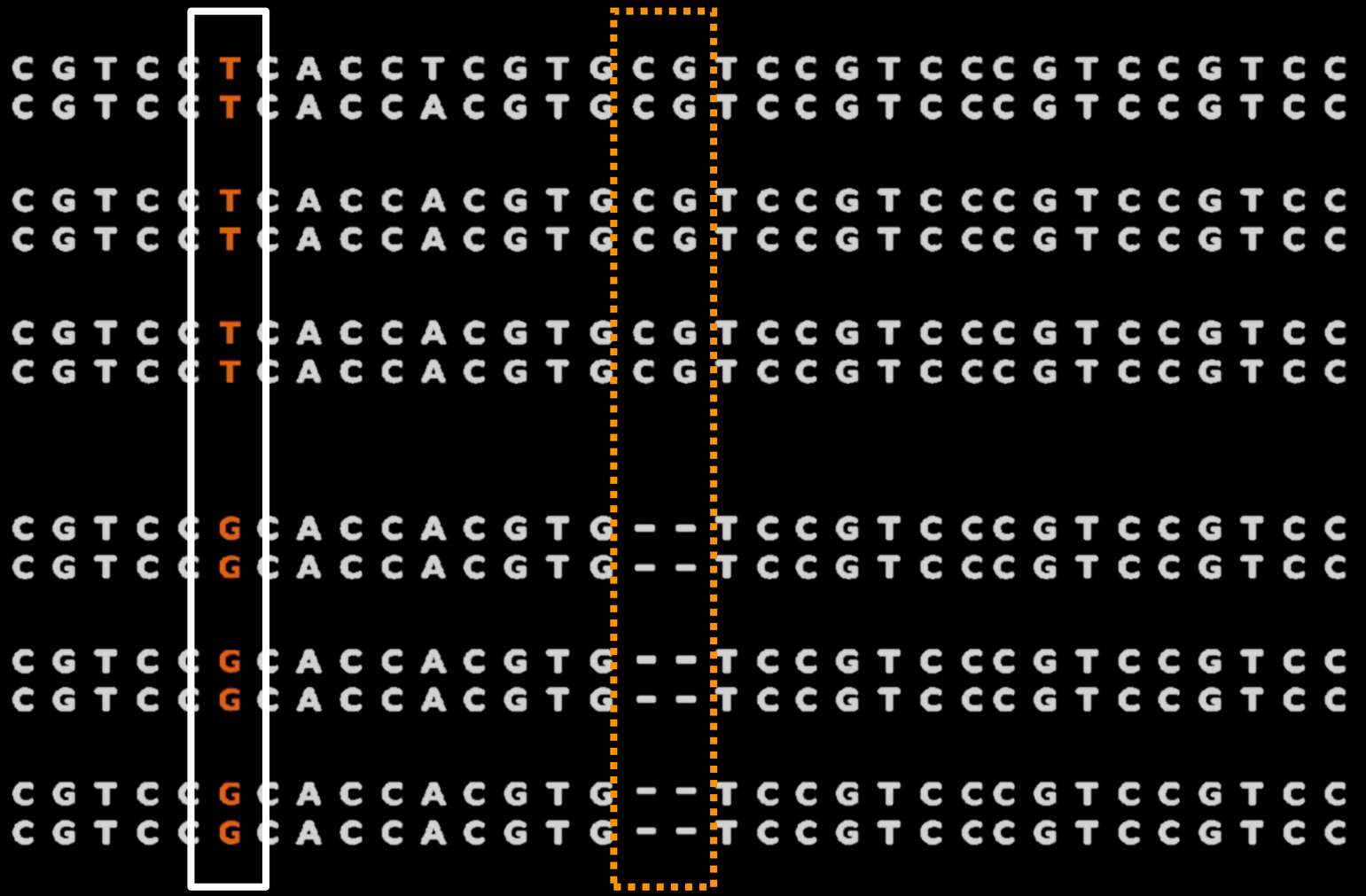
2008: 20,000 SNPs
2014: 150,000 SNPs
2017: 600,000 SNPs

Whole genome:
2,400,000,000 bases long

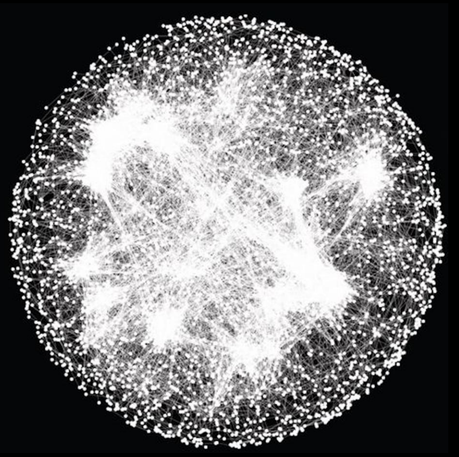


Problem #2: We didn't find the "causal" mutation

A good marker in dobermans may not work in other breeds (or mutts)



Problem #3: Genes interact with one another



The effect of the CDH2 variant may be different in different dogs (with different personalities?)

Problem #4: We never tested predictive power

Scientists: Can we find a change in DNA connected to this disease?
Can it help us understand the biology of the disease?

Breeders: How do I breed healthier, more successful dogs?

Owners: Will my dog get sick?

Solution:

We need BIG sample sizes

ARTICLE

doi:10.1038/nature13595

Biological insights from 108 schizophrenia-associated genetic loci

Schizophrenia Working Group of the Psychiatric Genomics Consortium*

Schizophrenia is a highly heritable disorder. Genetic risk is conferred by a large number of alleles, including common allele[...]hizo-phre[...] asso-ciatio[...] been previ[...]ty for the findings. Many findings have the potential to provide entirely new insights into aetiology, but associations at *DRD2* and several genes involved in glutamatergic neurotransmission highlight molecules of known and potential therapeutic relevance to schizophrenia, and are consistent with leading pathophysiological hypotheses. Independent of genes expressed in brain, associations were enriched among genes expressed in tissues that have important roles in immunity, providing support for the speculated link between the immune system and schizophrenia.

36,989 cases and 113,075 controls

The goal:

Predict whether a puppy is likely to be successful
Put dogs into the right jobs

Step 1. Correlation

Get DNA for a group of dogs → For each dog, find out if they have disease or trait → find genetic variants correlated with trait

Step 2. Develop a predictive test

Get DNA for a NEW group of dogs → For each dog, find out if they have disease or trait → Develop test to predict which dogs will have trait from DNA

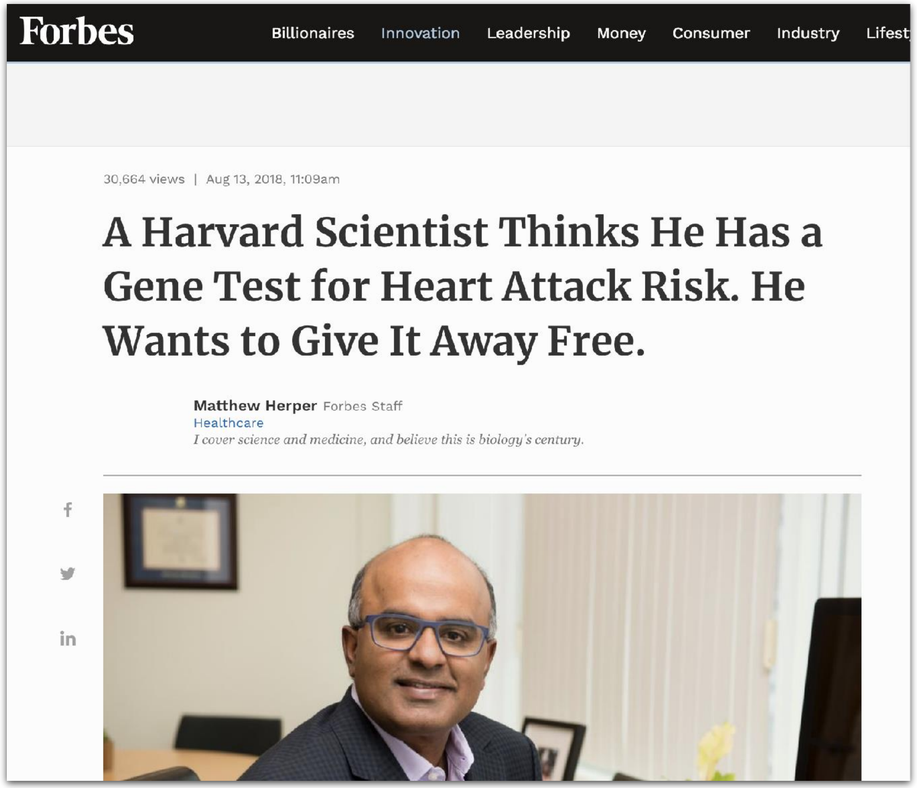
Step 3. Validate predictive test

Get DNA for a NEW group of dogs → For each dog, find out if they have disease or trait → Validate genetic test

Does genetic prediction work?

New genetic test for heart disease risk in humans

For each person: combines information from 6 million SNPs into 1 “polygenic risk score”



Forbes

Billionaires Innovation Leadership Money Consumer Industry Lifest

30,664 views | Aug 13, 2018, 11:09am

A Harvard Scientist Thinks He Has a Gene Test for Heart Attack Risk. He Wants to Give It Away Free.

Matthew Herper Forbes Staff
Healthcare
I cover science and medicine, and believe this is biology's century.

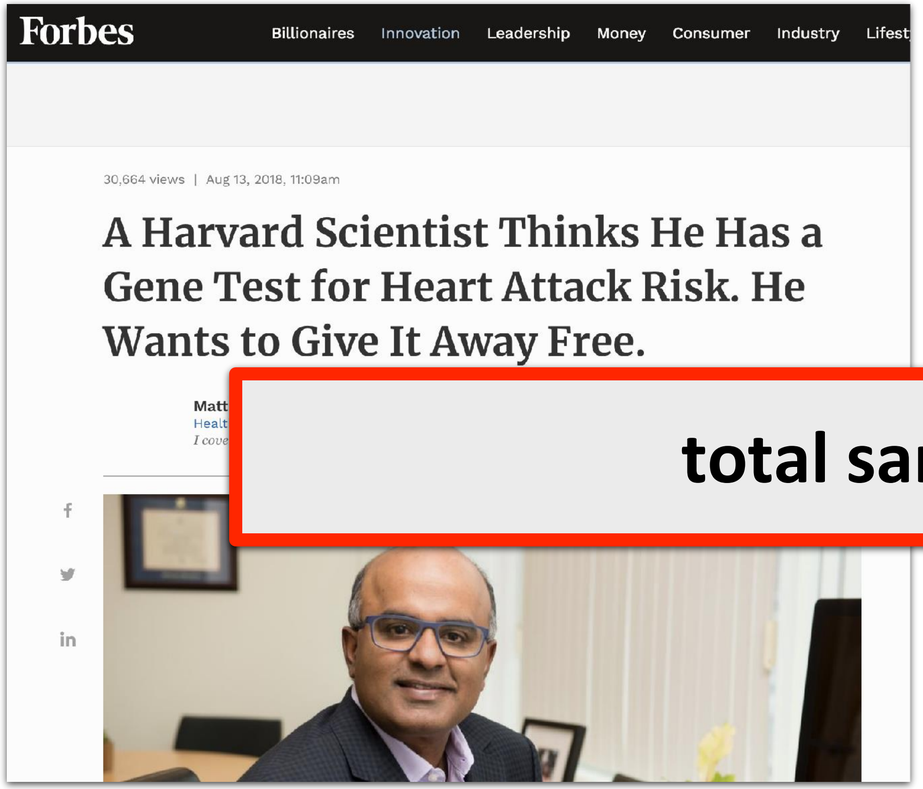
High Risk Score + unhealthy lifestyle

→ **high risk of heart disease**

High Risk Score + healthy lifestyle

→ **average risk of heart disease**

Does genetic prediction work?



Step 1. Correlation

60,801 cases +123,504 controls

Step 2.Develop predictive test

23,119 patients from the UK

total sample size: 496,402 people

288,978 patients from the UK

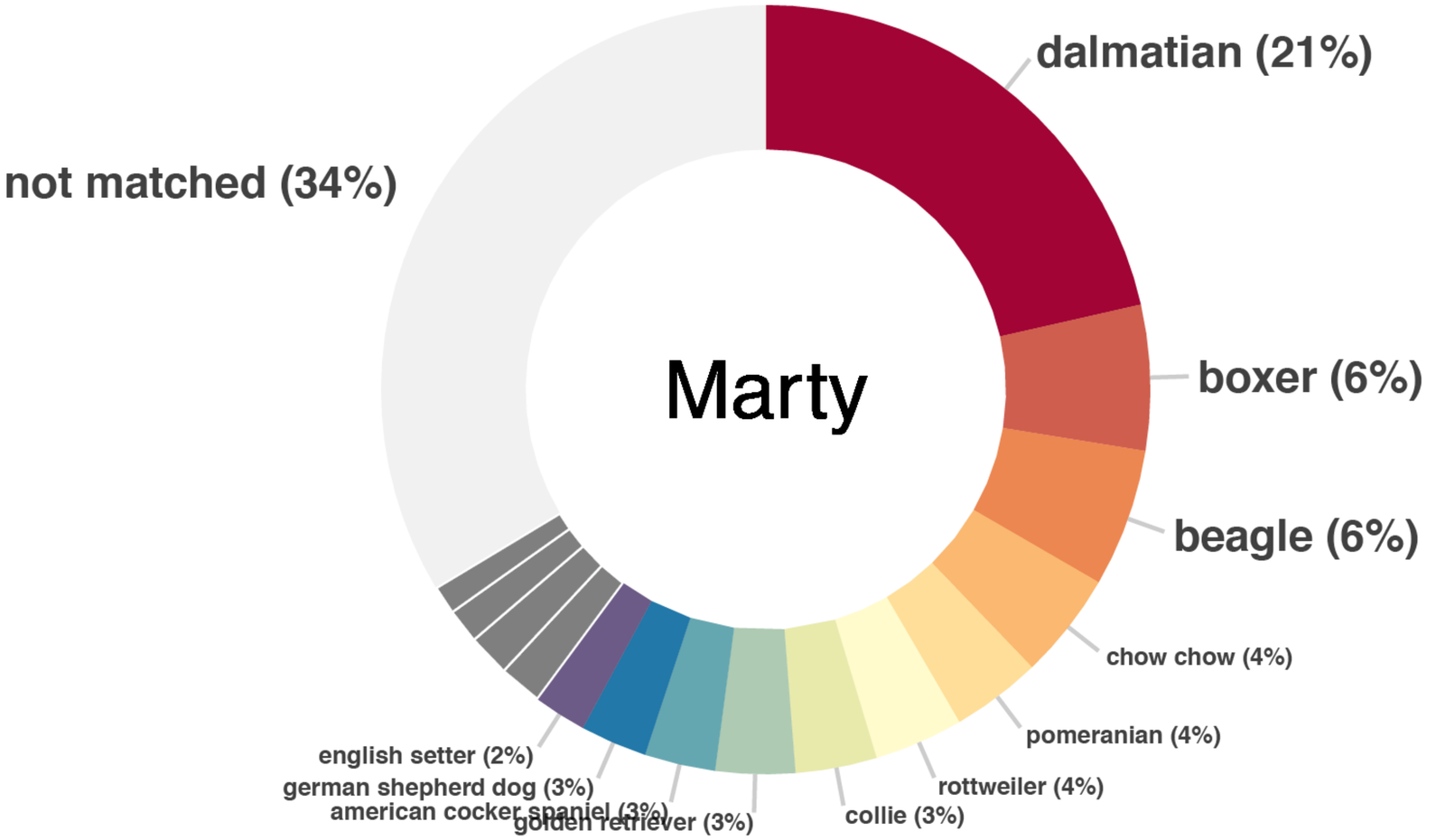

Even if behavioral genetics is easier in dogs, we need lots of dogs

We’re combining information from complementary studies

DARWIN’S DOGS

Pet dogs	Purpose-bred working dogs	Wolf-dog hybrids
very large sample size	moderate sample size	small sample size
very diverse ancestry	breeds & intentional mixes	spectrum of wolf/dog mixes
owner reported phenotypes	phenotyped by professionals	phenotyped by researchers
no / weak selection on behavior (some historical selection)	strong selection on behavior	mixes of behaviorally distinct populations

We’re studying all dogs



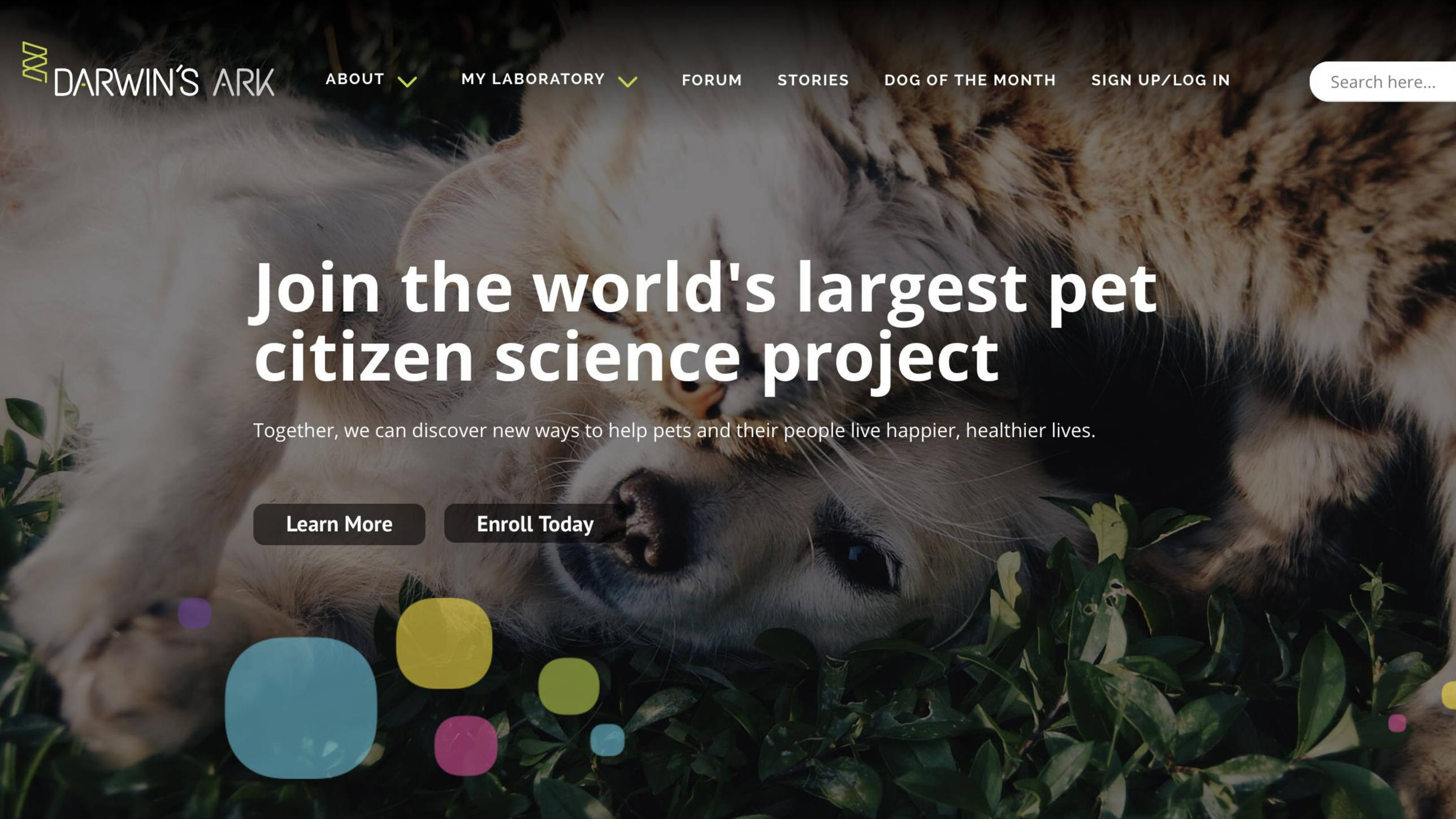
We’re using low-pass (1x) sequencing rather than arrays

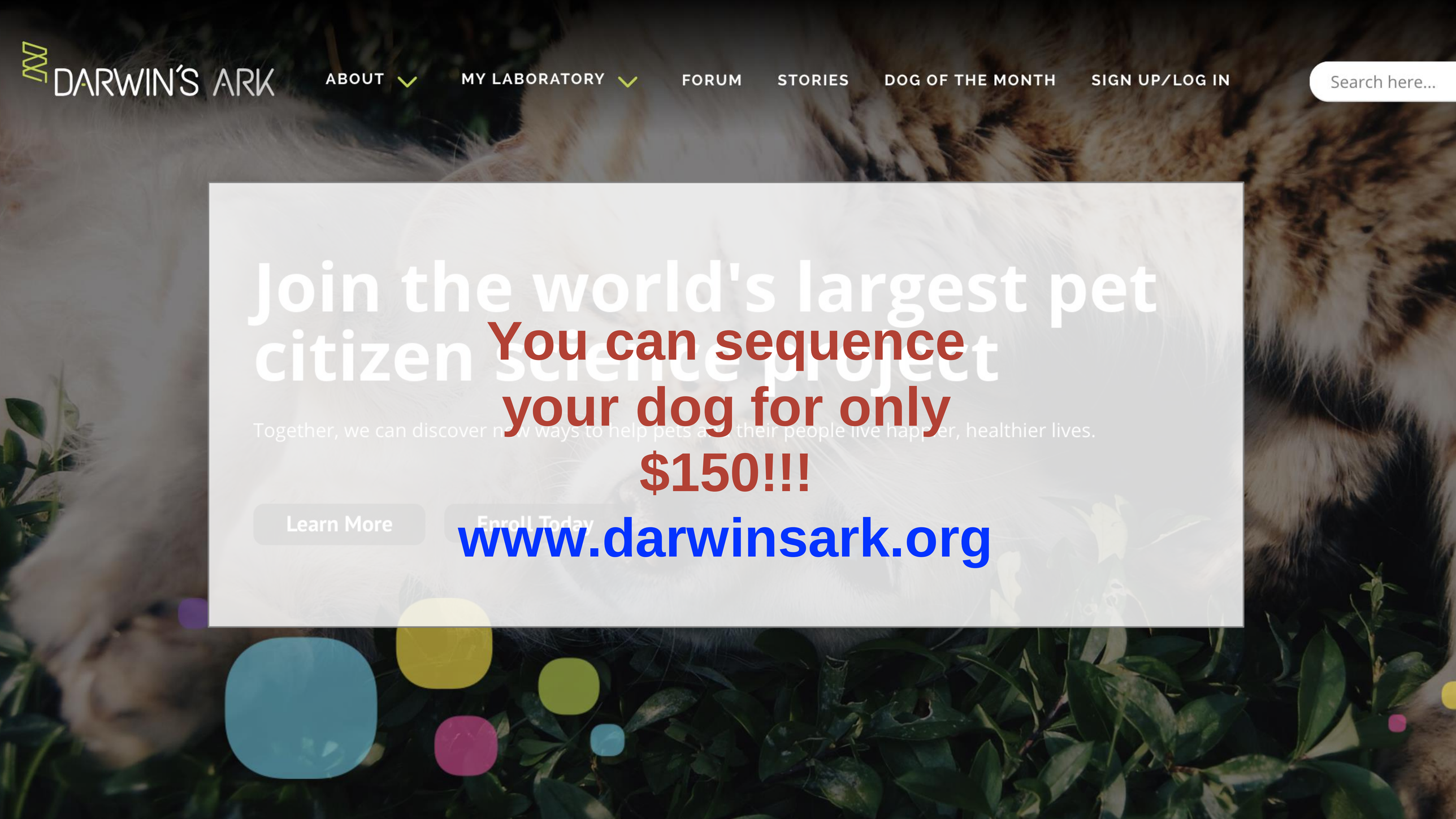
Validated in 10 very mixed breed dogs with 30x WGS data

gencove

	30x Whole Genome Sequencing	Genotyping	Low-pass sequencing*
Cost	~$1,400	$100-$200	**<$100**
# SNPS typed	>30 million	<800,000	**7-9 million**
Amount of DNA required	>500 nanograms	>500 nanograms	**2-5 nanograms**
Average accuracy vs 30x	N/A	98.75%	**99.14%***

** Higher accuracy possible with more stringent parameters (fewer SNPs called)*


DARWIN'S ARK
ABOUT
MY LABORATORY
FORUM
STORIES
DOG OF THE MONTH
SIGN UP/LOG IN
Search here...
Join the world's largest pet citizen science project
Together, we can discover new ways to help pets and their people live happier, healthier lives.
Learn More
Enroll Today


DARWIN'S ARK
ABOUT
MY LABORATORY
FORUM
STORIES
DOG OF THE MONTH
SIGN UP/LOG IN
Search here...
Join the world's largest pet citizen science project
Together, we can discover new ways to help pets and their people live happier, healthier lives.
Learn More
Enroll Today
You can sequence your dog for only $150!!!
www.darwinsark.org

The “Science” of Citizen Science

Darwin’s Ark lets ordinary citizens become scientific partners. Our projects, such as Darwin’s Dogs, combine genetics and behavior to advance the understanding of complex diseases.

We’re a non-profit that subscribes to the open science model. This means ***we do not own and will never sell your pet’s data***. You contribute to an open source database and we share all the new discoveries we find freely with you and researchers around the world. Together we can advance health care for pets and their humans. For details on our process, see our Frequently Asked Questions.

22,567

DOGS
REGISTERED

2,487,298

ANSWERS
COLLECTED

~5000 DNA samples collected

~1500 sequenced

Elinor's Laboratory

>EDIT DESCRIPTION

CITIZEN SCIENTISTS

Primary Owner

Elinor Karlsson

Member Since 2016-10-23
elinork@gmail.com

>EDIT

Other Members Invite People

MY DOGS + Add a Dog

0 Likes

Darwin

> EDIT
> CONTACT FOR MORE HELP

Born 2003-11-24

Order DNA Kit Free Kit Waitlist

Survey Status

2 / 21 Surveys

Next Surveys

>Communication
>The Great Outdoors

View All >

Question 1/10

Play

Darwin enjoys life

- Strongly Agree
- Agree
- Neither Agree Nor Disagree
- Disagree
- Strongly Disagree

Next

200+ questions per dog

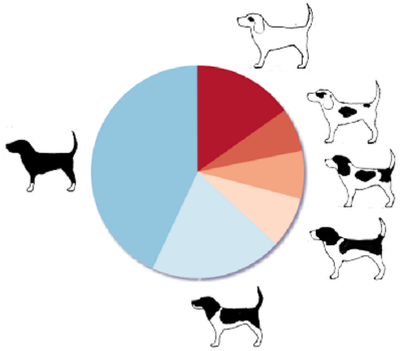
1. Morphology

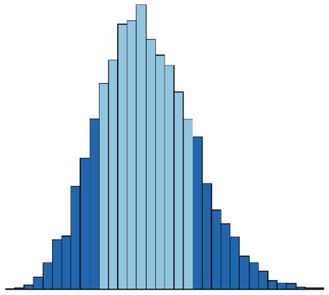
2. Selected behaviors

3. Complex behavioral traits & disorders

4. Traits suggested by dog behavior experts (IAABC)

5. Food & food allergy related traits





Food Allergy Science Initiative

GOdogs

Returning results: how does your dog compare?

Today: Simple trait comparisons

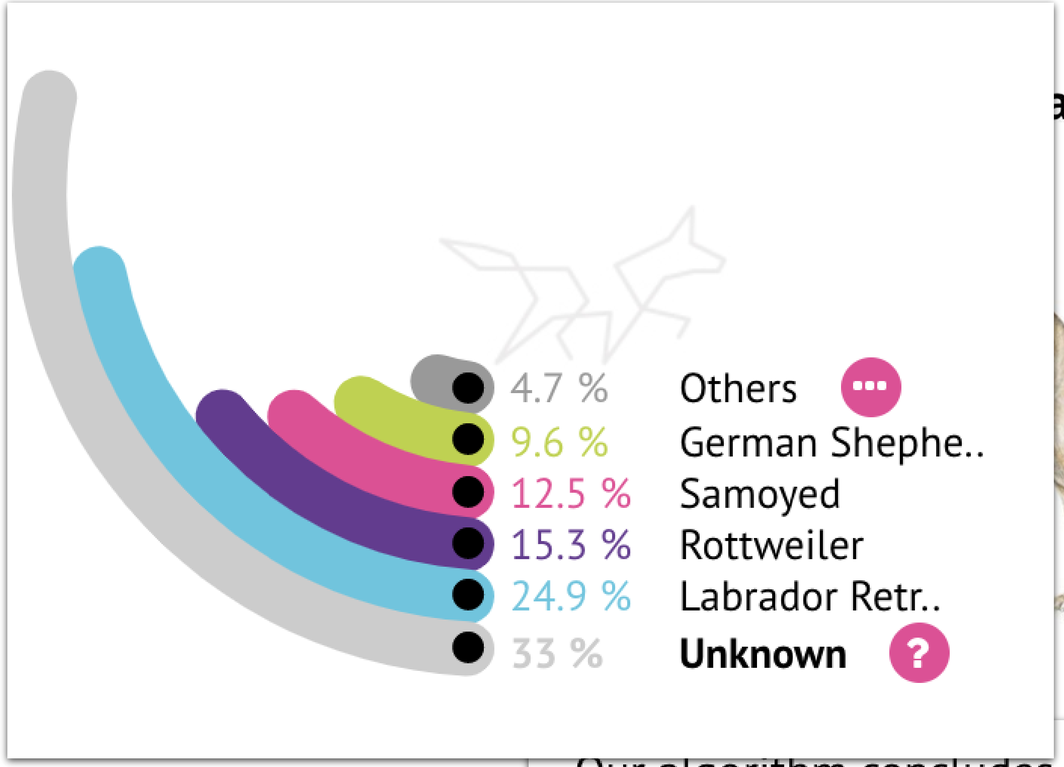
Your dog enjoys life

Your dog wants to play

Your dog becomes aggressive when excited

Returning results: what is your dog's breed ancestry?

Today: Breed results

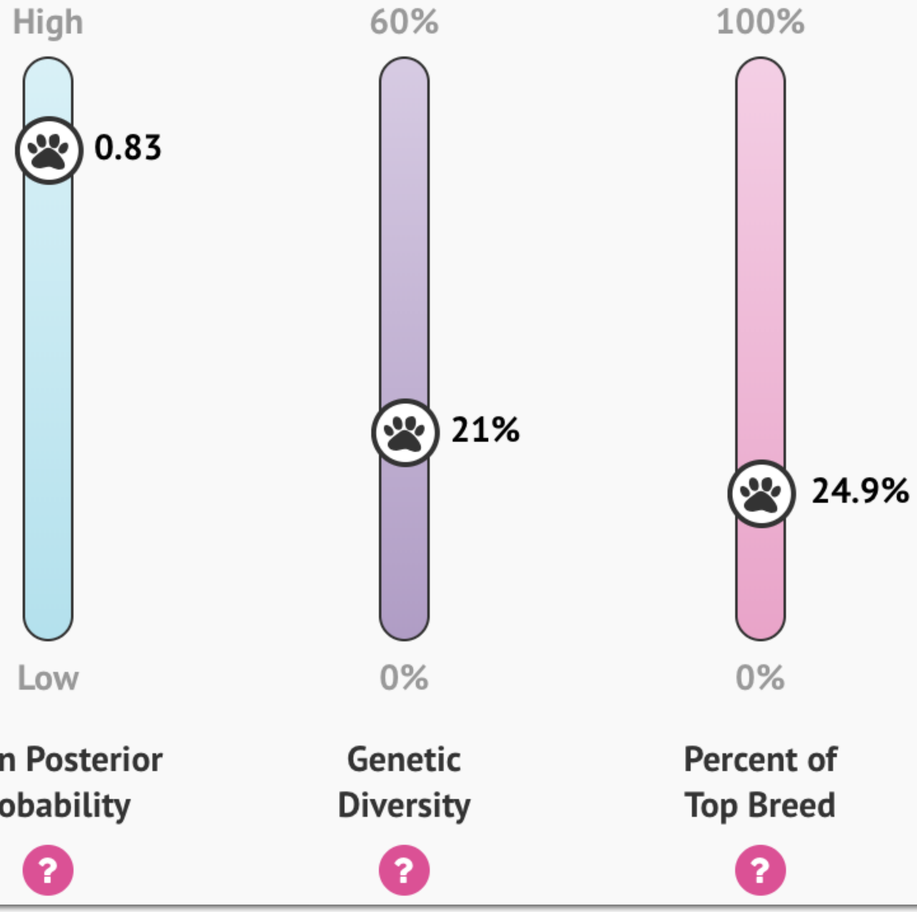


ador Retriev

Our algorithm concludes Jenny is a mix
biggest portion (24.9%) of her ancestry
Retriever, but we find other breeds as w

How did we get these results?

We used a combination of three statistics and Jenny's DNA results to decide that she was a mixed breed dog.



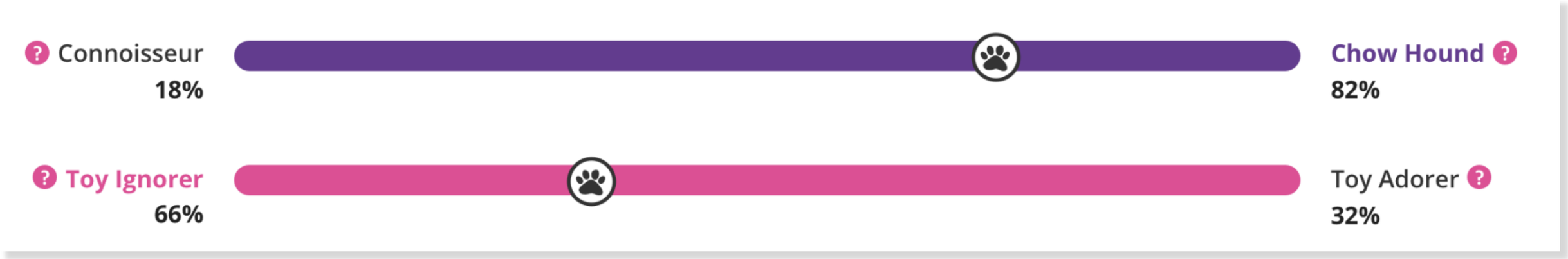
Do you think we got Scruffy's breed mix right?

- (•) Yes
- ( ) No
- ( ) Not Sure

Next

Returning results: personality profiles

Soon: Personality profiles from surveys



Returning results: morphology predictions

Future: Morphology predictions

X-Small

under 15 lbs

Chihuahua, Pomeranian, Shih Tzu, Papillon

Small

15-30 lbs

West Highland White Terrier, Boston Terrier, Dachshund, Jack Russell Terrier

Medium

30-50 lbs

Corgi, English Springer Spaniel, English Bulldog

Large

50-75 lbs

Golden Retriever, Labrador Retriever, Standard Poodle, German Shorthair Pointer

X-Large

75 lbs and Greater

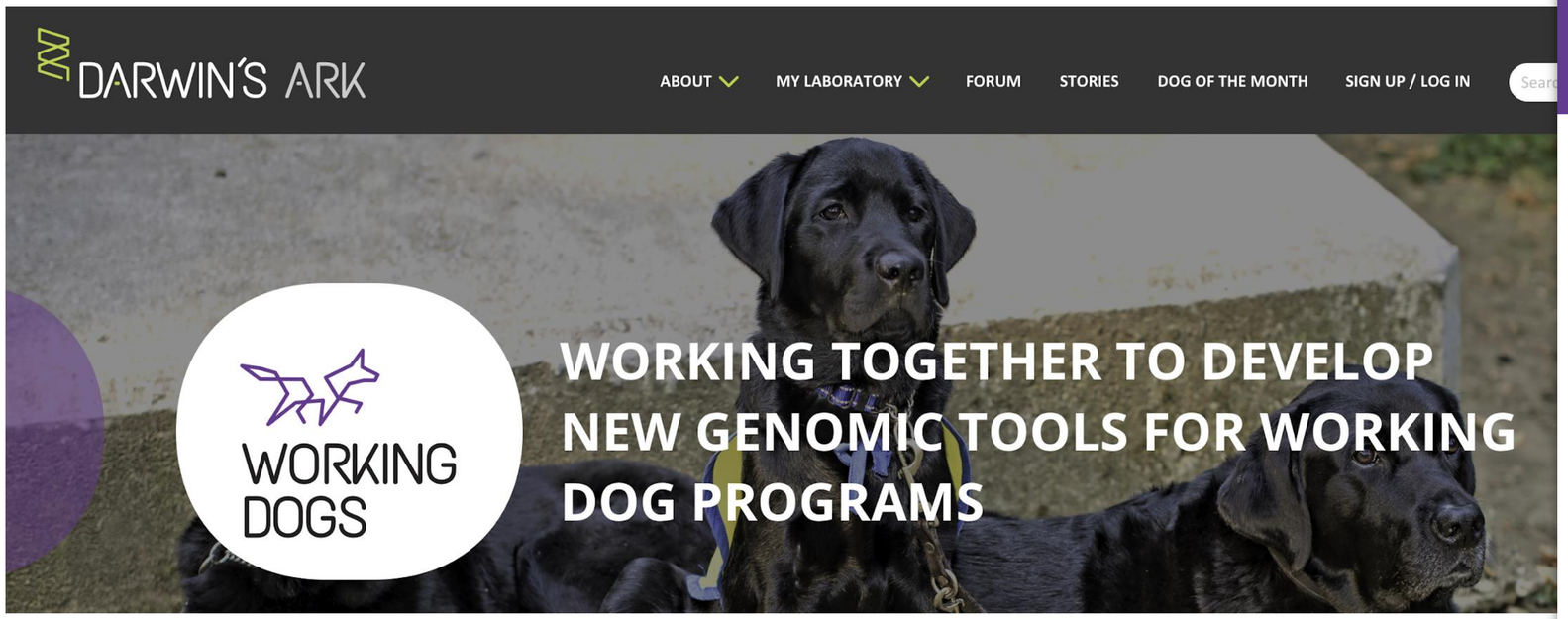
Leonberger, Irish Wolfhound, Great Dane, Newfoundland

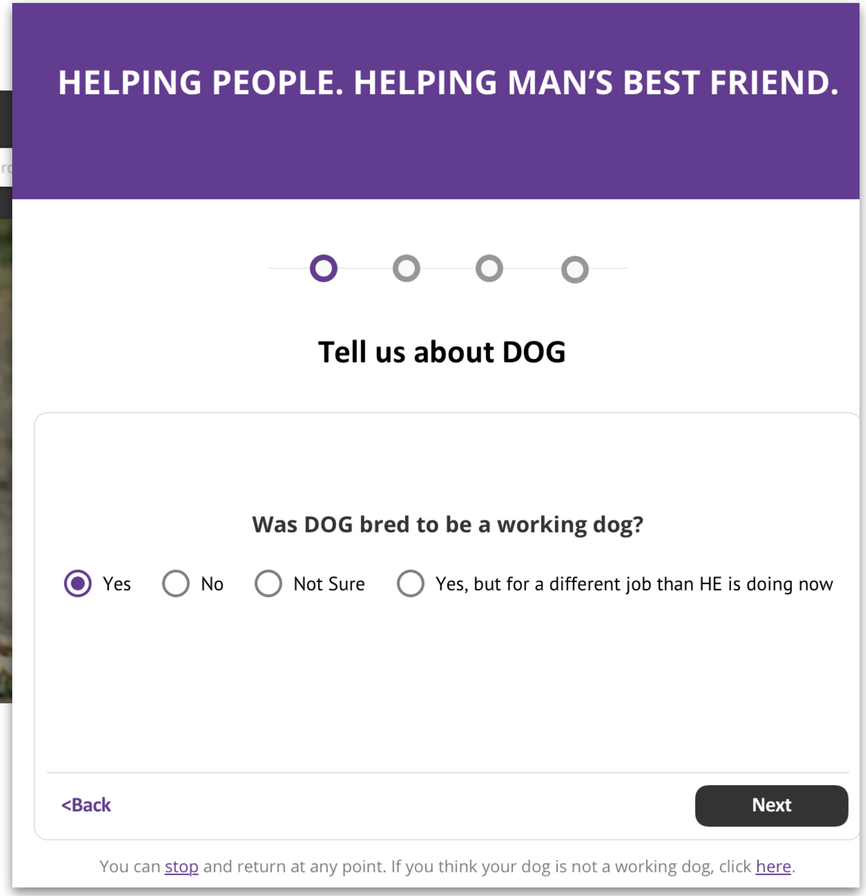
We predicted **Pirl** with size 2.7 (±0.5), Larger than 68% of dogs.

Xue Li

Coming soon: “Working Dog” portal on Darwin’s Ark site

Any dog owner can enroll their working dog.





Our approach: ask same question in different ways

DARWIN'S DOGS

Pet dogs	Purpose-bred working dogs	Wolf-dog hybrids
very large sample size	moderate sample size	small sample size
very diverse ancestry	breeds & intentional mixes	spectrum of wolf/dog mixes
owner reported phenotypes	phenotyped by professionals	phenotyped by researchers
weak selection on behavior & health (some historical selection)	strong, ongoing selection on behavior & health	mixes of behaviorally distinct populations

Focus on getting HUGE datasets

guide dogs

scent dogs

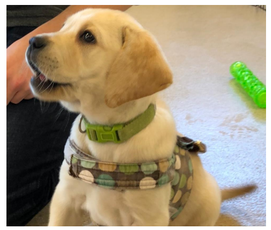
assistance dogs

other working dogs

DARWIN'S DOGS *pets*

Open resource for science

Tens of thousands of dogs
10 million DNA markers per dog
Behavioral traits



?

Statistical power & scientific expertise to develop useful genetic tests

guide dogs

scent dogs

assistance dogs

other working dogs

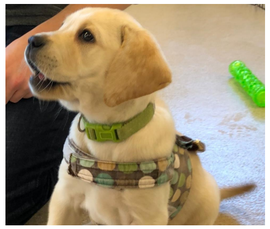
DARWIN'S DOGS

pets

de-identify data

Open resource for science

Tens of thousands of dogs
10 million DNA markers per dog
Behavioral traits



?

Statistical power & scientific expertise to develop useful genetic tests

Why test your dog's DNA?

SHORT TERM: Type variants linked to diseases

MEDIUM TERM: Accelerate breeding program progress

LONG TERM: Develop new tools for genetic prediction

WORKING DOG PROJECT

IWDBA DOGS SERVING HUMANITY

Why test your dog's DNA?

SHORT TERM: Type variants linked to diseases

Why test your dog's DNA?

SHORT TERM: Type variants linked to diseases

Work with organizations to figure out what tests are useful in their dogs now

Caveat: need to assess low-pass sequencing accuracy & speed

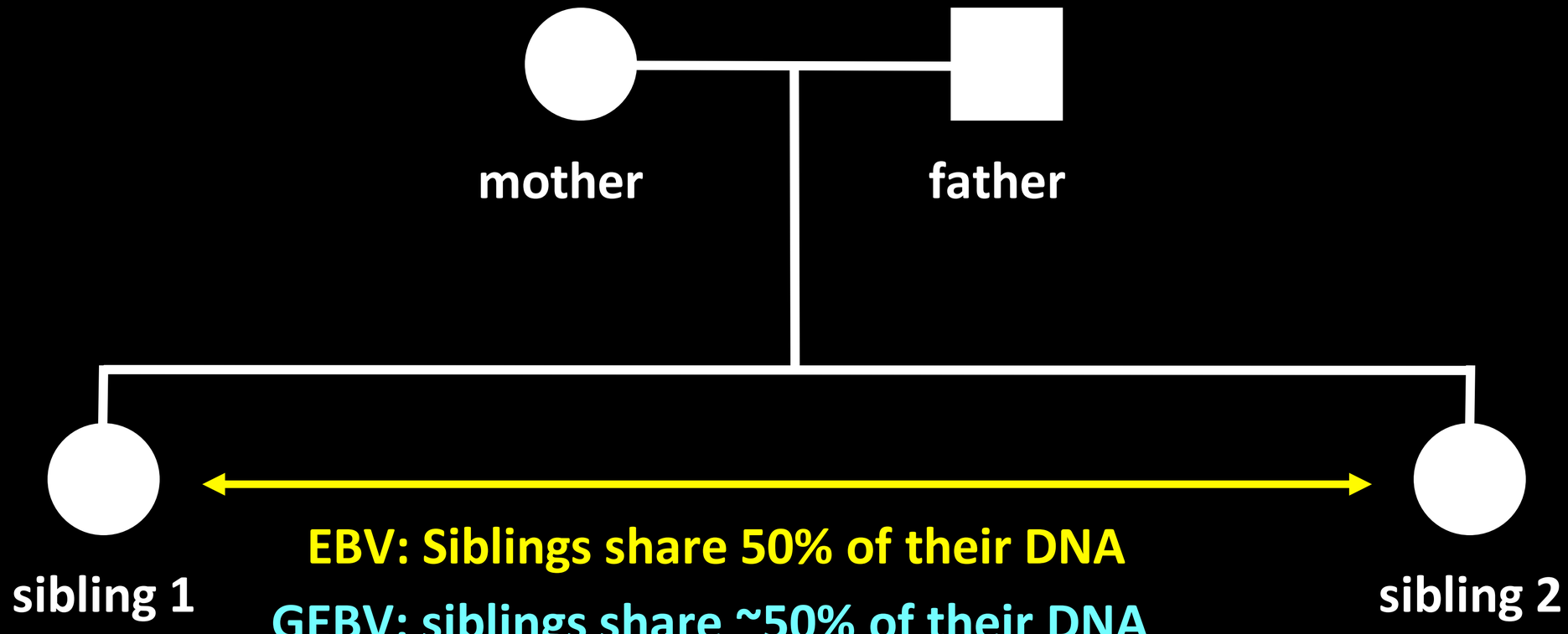
Why test your dog's DNA?

MEDIUM TERM: Accelerate breeding program progress

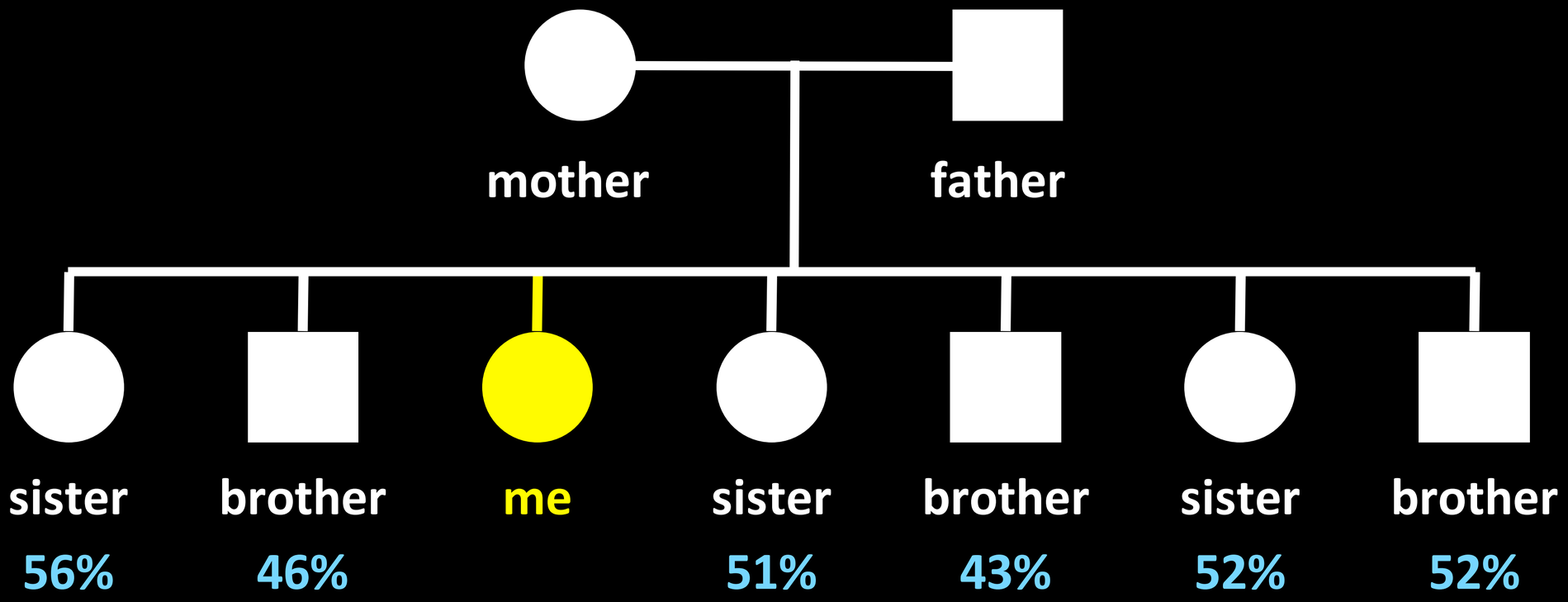
Breeding for better dogs through genomic selection

Select dogs for breeding based on their performance *and* the performance of their relatives

No genetic testing: Estimated Breeding Value
With genetic testing: Genomic Estimated Breeding Value (GEBV)


mother
father
sibling 1
sibling 2
EBV: Siblings share 50% of their DNA
GEBV: siblings share ~50% of their DNA


KIWI


mother
father
sister
56%
brother
46%
me
sister
51%
brother
43%
sister
52%
brother
52%

Why test your dogs' DNA?

MEDIUM TERM: Accelerate breeding program progress
with more accurate GEBVs
(even for programs with smaller pedigrees)

Why test your dogs' DNA?

LONG TERM: Develop new tools for genetic prediction

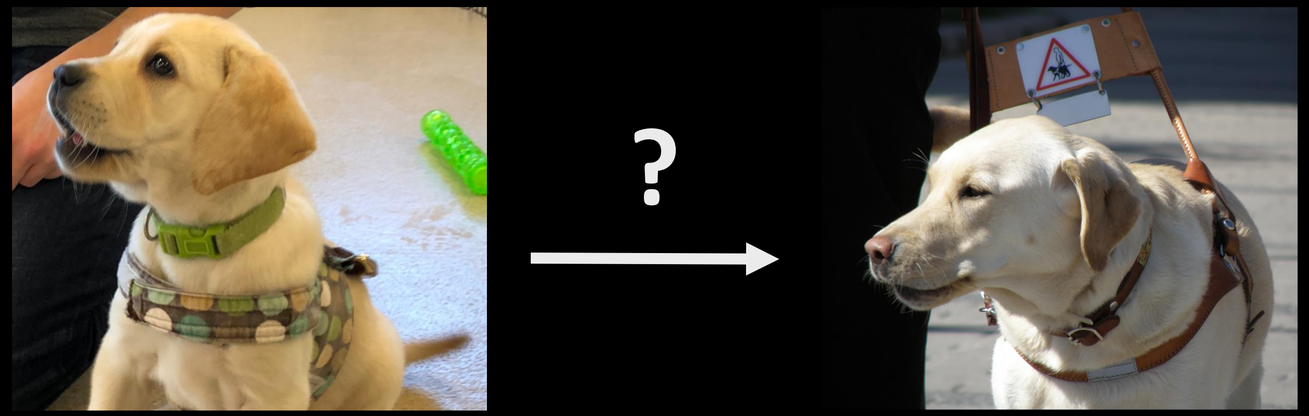
LONG TERM: Understand & treat diseases better -
in dogs and their humans

Dog genomics is in its infancy

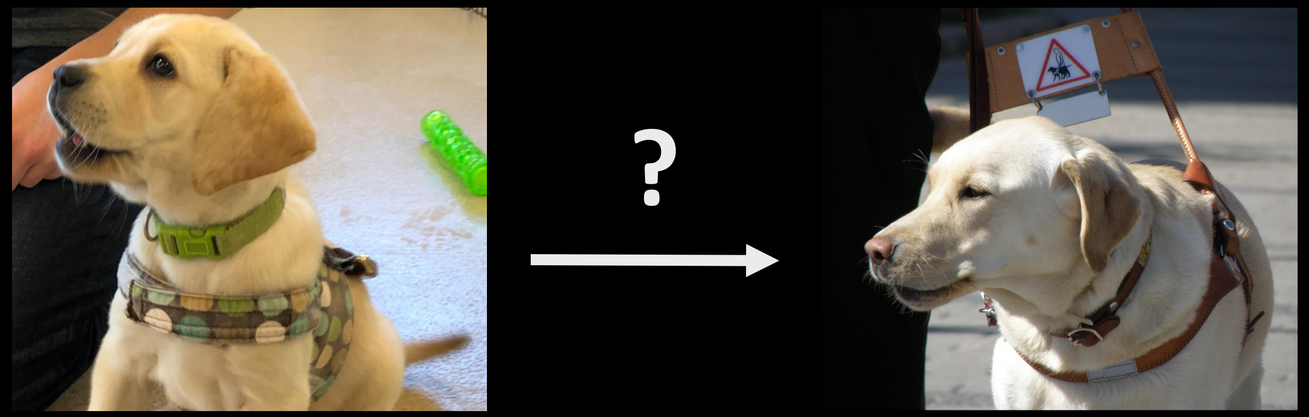
We can’t yet predict diseases or traits (except for a few easy ones)

We’ll never be perfect (environment matters a lot!)

Genomics can be powerful tool - if we think big



We need new approaches



Is this puppy worth the cost of training?

We need new approaches

How do we breed more dogs that succeed?

We need new approaches

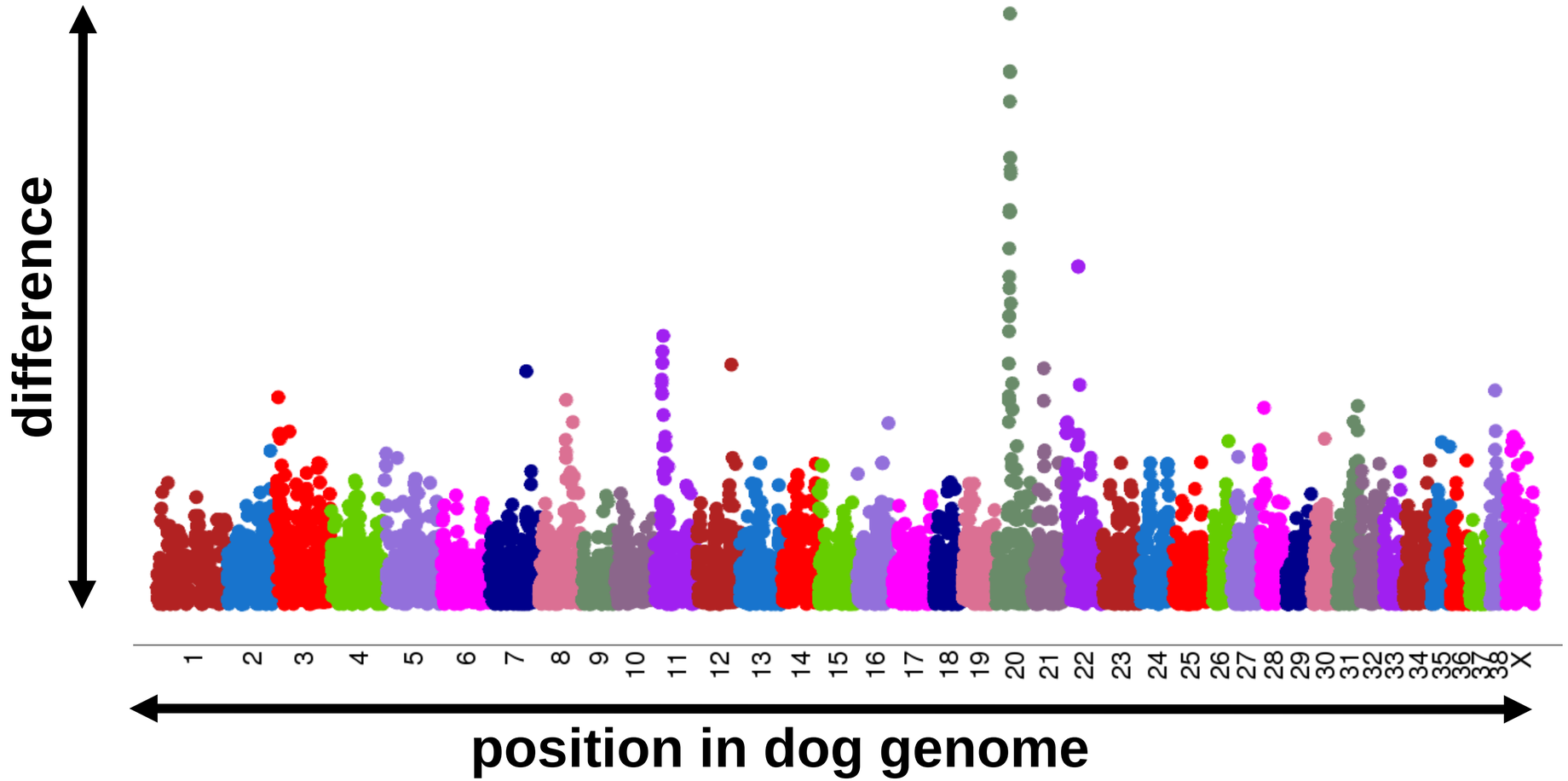
Which job is this dog suited to?

Will it work?

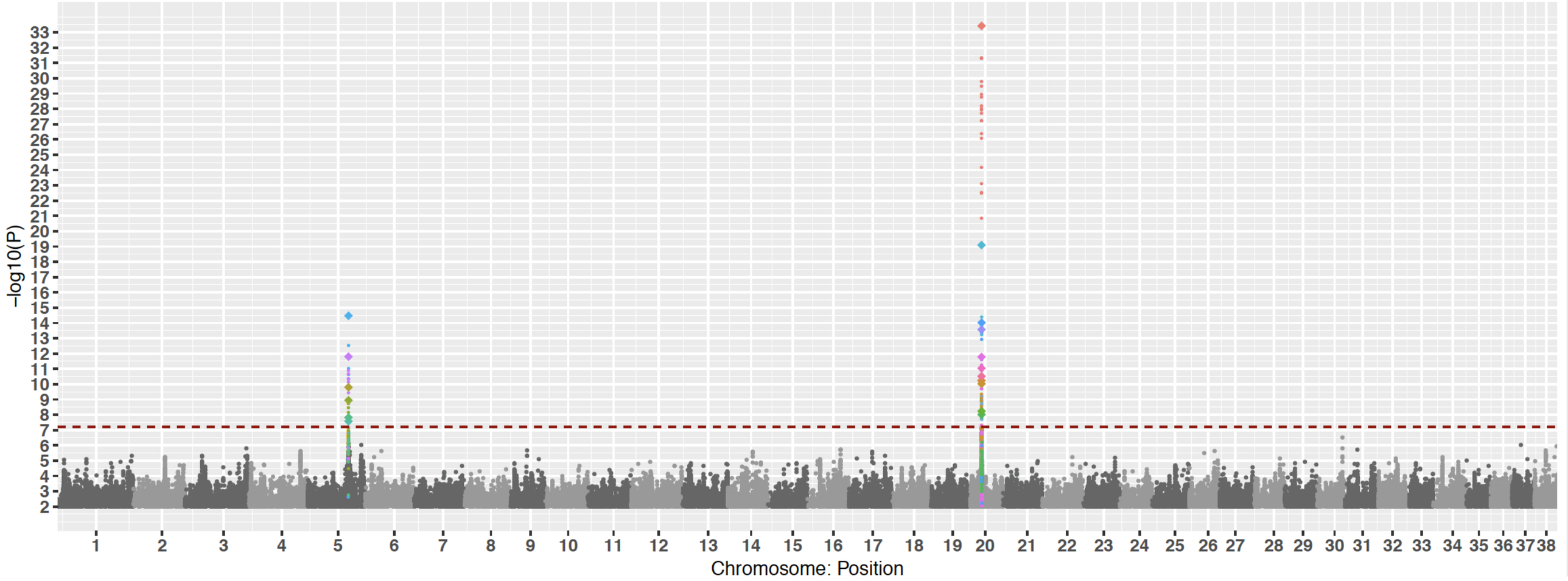
YES!

with lots of dogs
AND
lots of genetic data from each dog

White coat color with 19 dogs & 20,000 markers

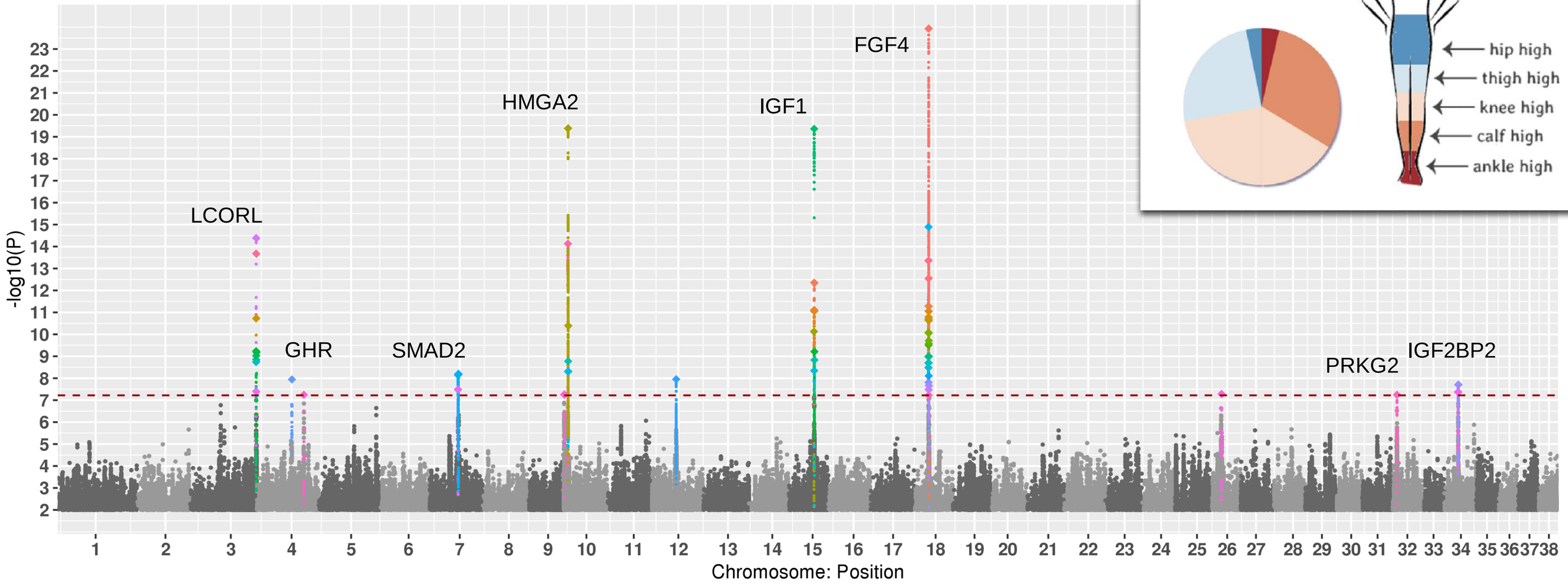


White coat color with 1000 dogs and 5-10 million markers



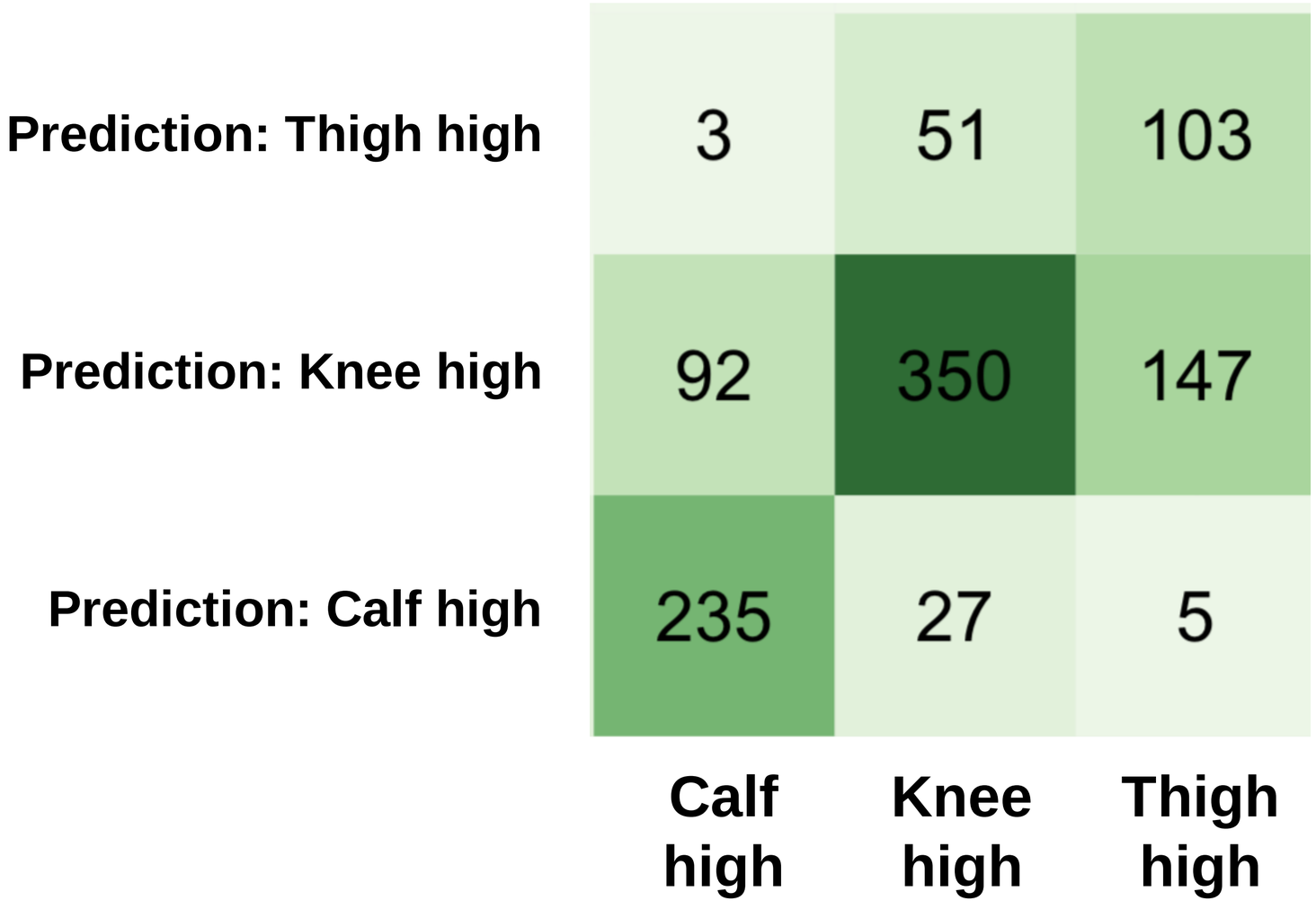
Size GWAS with 1000 dogs

Kathleen Morrill

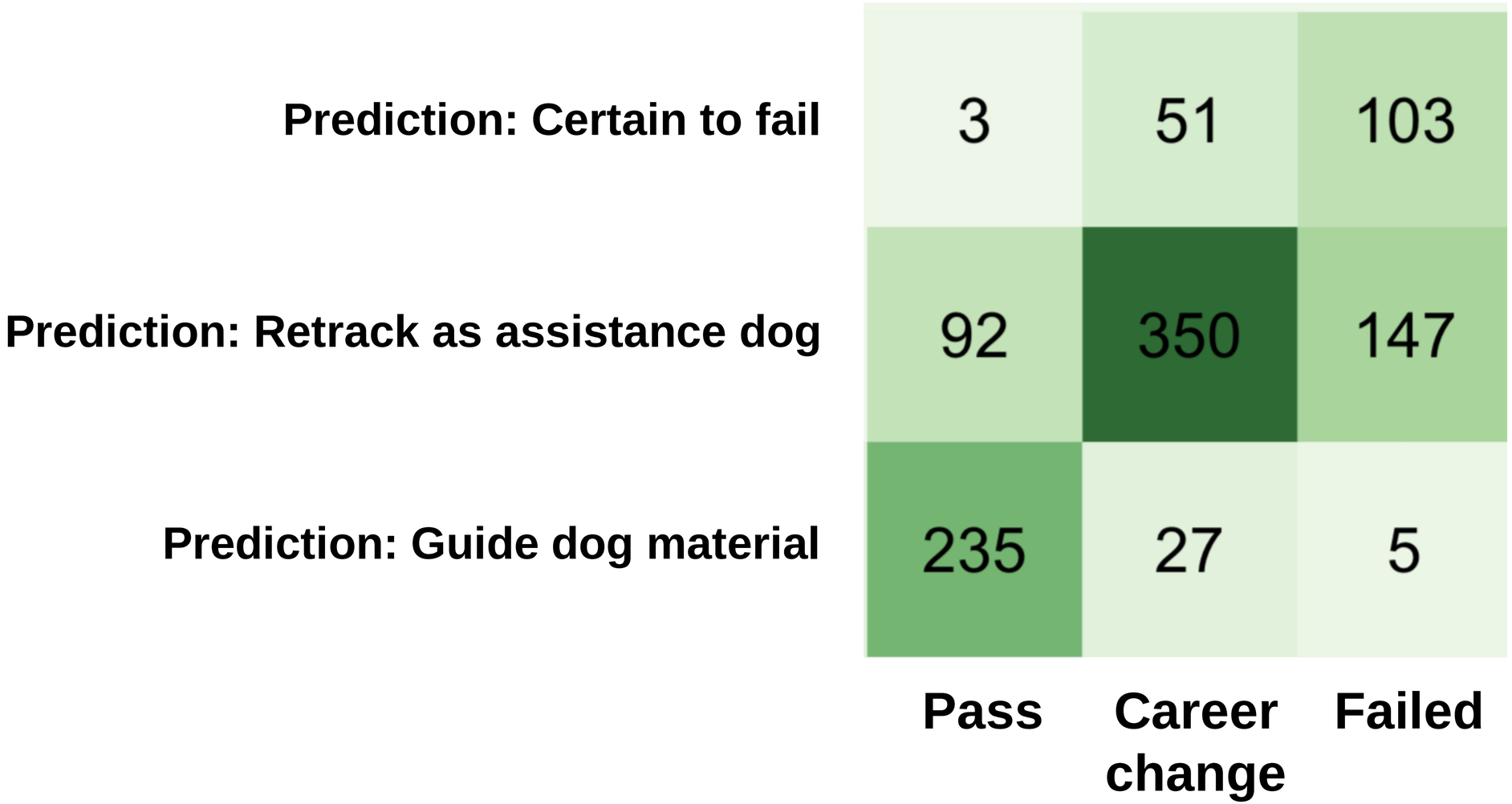


Sutter *et al* 2007 | Koltes *et al* 2009 | Rimbault *et al* 2013 | Metzger *et al* 2013 | Aniek *et al* 2018 | Plassais *et al* 2019

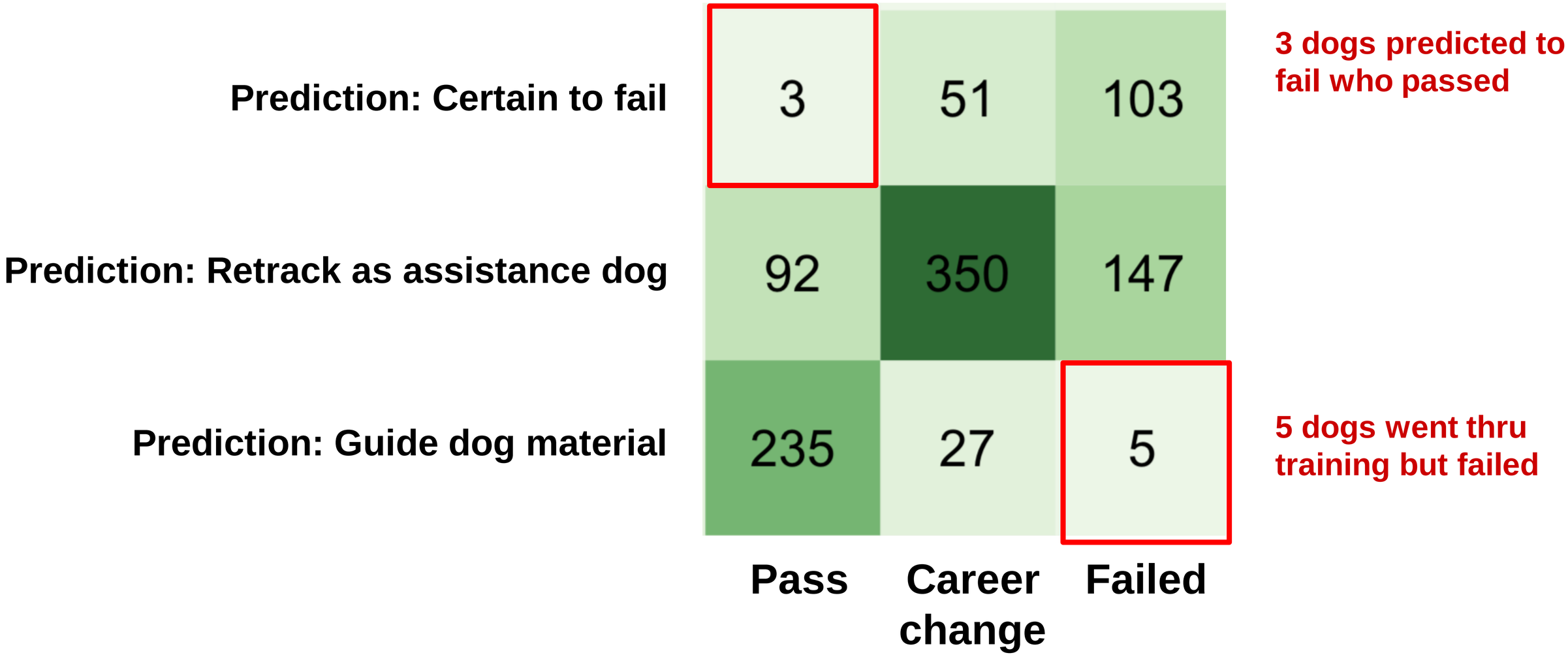
Can we predict how big a dog will be?



Hypothetical: What if we could predict behavior?



Hypothetical: What if we could predict behavior?



Pilot project ongoing

- Inhibited with stress
- Activated with stress
- Harness sensitivity

- Thunderstorm phobia
- Hip dysplasia
- Cancer
- Other diseases

We'll start with most heritable traits (collab. with GEB and IWDBA)

We're looking for new collaborations as Working Dog Project grows

Have your own dog? Sign up today! darwinsark.org

@UMass
Kathryn Lord
Jesse McClure
Brittney Logan
Shirley Xue Li
Gaurav Chauhan
Kathleen Morrill

Darwin's Ark Foundation
Charlie Lieu
Marjie Alonso

@Broad
Jessica Hekman
Linda Boettger
Hyun Ji Noh
Michele Koltookian
Jeremy Johnson
Ross Swofford
Jacob Alonso
Diane Genereux

Jane Russenberger
Eldin Leighton

theriogenology foundation
The Future of Animal Reproduction

The Manton Foundation

AKC REUNITE

Food Allergy Science Initiative

BROAD IGNITE