

# Using Regions of Homozygosity to Evaluate the Use of Dogs as Preclinical Models in Human Drug Development

Sandra Paulina Smieszek (✉ [sandra.smieszek@vandapharma.com](mailto:sandra.smieszek@vandapharma.com))

Vanda Pharmaceuticals Inc

Mihael Polymeropoulos

Vanda Pharmaceuticals Inc

---

## Research article

**Keywords:** ROH, genomics, consanguinity, dog genome, deleterious variants, CYPD6, ADME genes

**Posted Date:** May 15th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-27728/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**BACKGROUND** Animals are used as preclinical models for human diseases in drug development. Dogs, especially, are used in preclinical research to support the clinical safety evaluations during drug development. Comparisons of patterns of regions of homozygosity (ROH) and phenotypes between dog and human are not well known. We conducted a genome-wide homozygosity analysis (GWHA) in the human and the dog genomes.

**RESULTS** We calculated ROH patterns across distinct human cohorts including the Amish, the 1000 genomes, Welllderly, Vanda 1 k genomes, and Alzheimer's cohort. The Amish provided a large cohort of extended kinships allowing for in depth family oriented analyses. The remaining human cohorts served as statistical references. We then calculated ROH across different dog breeds with emphasis on the beagle - the preferred breed used in drug development. Out of five studied human cohorts we reported the highest mean ROH in the Amish population. We calculated the extent of the genome covered by ROH ( $F_{ROH}$ ) (human 3.2 Gb, dog 2.5 Gb). Overall  $F_{ROH}$  differed significantly between the Amish and the 1000 genomes, and between the human and the beagle genomes. The mean  $F_{ROH}$  per 1 Mb was ~ 16 kb for Amish, ~ 0.6 kb for Vanda 1 k, and ~ 128 kb for beagles. This result demonstrated the highest degree of inbreeding in beagles, far above that of the Amish, one of the most inbred human populations. ROH can contribute to inbreeding depression if they contain deleterious variants that are fully or partially recessive.

**CONCLUSIONS** The differences in ROH characteristics between human and dog genomes question the applicability of dog models in preclinical research, especially when the goal is to gauge the subtle effects on the organism's physiology produced by candidate therapeutic agents. Importantly, there are huge differences in a subset of ADME genes, specifically cytochrome P450 family (CYPs), constituting major enzymes involved in drug metabolism. We should use caution when generalizing from dog to human, even if human and beagle are relatively close species phylogenetically

## Background

Animals are used as preclinical model for humans/human diseases in drug development. While central to drug development, there is lack of validity for many of these tests. Evidential weight provided by animal data was assessed in a recent study, evaluating the probability that new drugs may prove toxic (or not toxic) to humans<sup>1</sup>. The authors calculated likelihood ratios (LR) for a large set of drugs. Their major finding showed that lack of toxicity in animals provides little or no evidential weight to the probability of a lack of adverse drug reactions in humans, as well as large inconsistencies in LR across different species and classes of drugs human<sup>1</sup>.

Dogs, especially, are used in preclinical research to support the clinical safety evaluations during drug development. However, dogs are a particularly inbred species. Recently, the EMBARK project conducted large scale characterization for the dog regions of homozygosity (ROH) density maps for 2500 dogs comparing a range of breeds<sup>2</sup>. The presented results recorded ROH down to 500 kilobases. Authors report

more than 678 homozygous deleterious recessive genotypes in the panel across 29 loci, 90% of which overlapped with ROH<sup>2</sup>. The study constitutes one of the most comprehensive evaluations of dog genome across several breeds of dogs, showing a great degree of variation. Patterns of ROH in dogs seem to imply a high degree of inbreeding across multiple breeds of dogs, regions furthermore associating with deleterious variants<sup>2</sup>. Long ROH are enriched in deleterious variation<sup>3</sup>. It is well established that inbreeding reduces reproductive fitness, as discussed initially by Charles Darwin in his initial observations in plants and later on in humans<sup>4,5</sup>. It is hence well established that inbreeding increases the incidence of recessive disease, exemplified long ago in 1902 in the case of alkaptonuria<sup>6</sup>, and in many cases of Mendelian disorders such as Tay-Sachs<sup>7</sup>. Moreover, the implications of extensive inbreeding were recently demonstrated in a study of grey wolves in Isle Royale, where inbreeding depression has brought wolves to the brink of extinction, likely via increased homozygosity of the deleterious variants<sup>8</sup>. The comparison of regions of homozygosity patterns between dogs and humans has not been characterized. In the present study, we focus on exactly that comparison. An ROH is defined as a continuous stretch of DNA sequence without heterozygosity in the diploid state (min ROH  $\geq$  1.5 Mb). We calculated ROH patterns across distinct human cohorts: the Amish, IGS 1000 Genomes<sup>9</sup>, Welllderly<sup>10</sup>, Vanda 1K genomes using both microarray and sequencing

We have incorporated several datasets in the analysis to have a broader view of the potential differences in and between the two species, as well as across breeds and human subpopulations. The Amish tend to have large families, with well-documented family history accessible via Anabaptist Genealogy Database, and constitute a valuable resource for genetic studies<sup>11</sup>. They constitute an example of a highly inbred population, and in fact ROH were first described in them. The 1000 genomes serve as mixed control representation of individuals. The Vanda 1000 genomes is a control set of whole genome sequencing samples with extensive phenotyping. The Welllderly constitutes a well-being well aged cohort of mixed ethnicity. The Alzheimer's cohort is a mixed ethnicity cohort offering yet another point of comparison. Using the same settings across datasets we calculated ROH across different dog breeds (EMBARC project) with emphasis on the beagle, as it is the preferred breed in drug development. We evaluate the degree of inbreeding in dogs compared to human populations. We also focus on the Cytochrome P 450 (CYP) family, as they are the major class of enzymes involved in drug metabolism, accounting for 75% of total metabolism<sup>12</sup>. These results should aid our understanding of the genetic validity of dogs as preclinical models for human drug development.

## Results

A ROH is defined as a continuous stretch of DNA sequence without heterozygosity in the diploid state (min ROH = established usually between 500 kb and 1.5 Mb). We estimated ROH using PLINK<sup>13</sup>. We hypothesized that at-risk genotypes would be highly enriched in ROH regions compared to the non-ROH genomic background. This enrichment can be used to evaluate the sensitivity and specificity of ROH-calling methods, and can provide a direct test of whether longer ROH tracts are more or less enriched for these recessive disease variants. We characterized the distribution of ROH across breeds and between

humans and dogs. Figure 1 displays the length and the number of ROHs identified across numerous dog breeds.

Vanda 1 k and the Welllderly ROH analysis was conducted on whole genome sequencing data. When we considered both the location and the allelic form of the ROHs, we were able to separate the populations by PCA, demonstrating that ROHs contain information on the demographic history and structure of a population. We calculated the extent of the genome covered by ROH ( $F_{ROH}$ ) (human 3.2 Gb, dog 2.5 Gb) as displayed in Table 1.

	Amish	Welllderly	ADNI	1000g	Vanda1k	Beagles
total ROH (n)	13165	2404	5534	1977	1012	4896
mean ROH per individual	14.74	4.7	6.72	3.14	1.03	61.97
mean length (kb)	6343	2203	3067	2445	1863	5203
median length (kb)	3577	1858	1856	2309	1746	3137
mean $F_{ROH}$	0.016057	0.003233	0.006532	0.002402	0.000604	0.128986
median $F_{ROH}$	0.013029	0.002821	0.004122	0.001553	0.000469	0.094472

**Table 1 Displays the total, mean ROH and ( $F_{ROH}$ ) across human cohorts compared to dogs**

In the Amish population, the  $F_{ROH}$  was 12x smaller in comparison to the beagle population.  $F_{ROH}$  differed significantly between the Amish and the 1000 genomes, and between the human and the beagle genomes. The mean  $F_{ROH}$  per 1 Mb was ~ 16 kb for Amish, ~ 0.6 kb for Vanda 1 k, and ~ 128 kb for beagles. This result demonstrated the highest degree of inbreeding in beagles, far above that of the Amish, one of the most inbred human populations, also displayed in Fig. 2. Figure 2 exemplifies the significantly higher observed ROH as compared to several human populations, including the Amish. Figure 3 further compares the human populations and the specific regions of high ROH across the human genome.

Current knowledge shows that 298 genes encode Phase I and II drug metabolizing enzymes, transporters, and modifiers (ADME genes), with CYPs constituting a major family. To further evaluate the ADME related consequences of inbreeding, we focused on the CYP family, including most common CYPs involved in metabolism of human drugs (including CYP2D6, CYP2Bs, CYP2Cs, CYP3As). We mapped orthologues between dogs and humans using BIOMART<sup>14</sup> (orthologues, from human to dogs). We report that beagles alone,  $F_{ROH}$ , would be 25% so even higher than our average in beagles, 12.9%. For humans that estimate would be 1/1000 on the same parameter. The results are presented in Fig. 4.

## Discussion

The study found a high degree of homozygosity across the genomes of beagle dogs, which is an indication of a high degree of inbreeding. This creates an evolutionary bottleneck, and exposes

homozygous deleterious variants. Furthermore, due to heterozygote advantage as outlined by overdominance hypothesis, the long term loss of diversity further reduces heterozygote advantage. For example, the risk for autosomal recessive disease is proportional to the degree of parental relationship in consanguineous families<sup>15</sup>. In humans, the total length of ROH per individual shows considerable variation across individuals and populations, with higher values occurring more often in populations with known high frequencies of consanguineous unions. Even across the genome, the distribution is far from uniform<sup>16</sup>: variation is correlated with recombination rate, as well as with signals of recent positive selection. Importantly, long ROHs are more frequent in genomic regions harboring genes associated with autosomal-dominant diseases<sup>17</sup>. Interestingly, the top-ranked ROH hotspot in the human genome (from our data) is located on chromosome 2p, in a region with low recombination rates and containing CYP26B1<sup>17</sup>.

We compared the degree of genetic inbreeding in beagle dogs using the measure of Runs of Homozygosity (ROH). This analysis identifies contiguous regions of the genome that are inherited from both parents. The study showed that unrelated humans show these regions of homozygosity in about 1/1000 of their genomes, while beagle dogs do so in 12.8 percent of their genome. In other words, beagle dogs are on average 100-times more inbred than humans. The Amish population in Pennsylvania were about 12-times more inbred than the general population. It is well accepted that drug efficacy and safety studies in genetic isolates such as the Amish cannot be generalized for the population at large.

The ADME genes constitute a highly polymorphic set of genes involved in Absorption, Distribution, Metabolism, and Excretion of drugs. This was exemplified in a recent effort to characterize the distribution of 95 polymorphisms in 31 core ADME genes in 20 populations worldwide<sup>18</sup>. The present effort to evaluate CYPs *between* species is displayed in Fig. 4, with unsurprising results. CYP2D6, another gene with allelic variants and encoding enzymes with variable degrees of activity, is associated with the development of hepatotoxicity after use of certain pharmaceutical agents. The CYP2D6 gene is polymorphic with over 150 allelic variants (<https://www.pharmvar.org/gene/CYP2D6>)<sup>19</sup>. The existing variation in the frequencies of alleles of polymorphic pharmacogenes among different ethnic groups may be responsible for severe adverse reactions to, or altered efficacy of, a wide variety of drugs.

Beagle dogs are routinely used for studying the safety of new pharmaceuticals. This practice has been in place for a hundred years, and is followed (some might suggest dogmatically and without question) by pharmaceutical developers and regulators. The present study is significant in that it questions the predictive validity of these commonly conducted dog studies. The very high degree of inbreeding we report suggests that beagles must be a poor model to predict toxicity even for other dog breeds, and of course an inappropriate model to do so for humans. Reliance on beagles to predict human drug safety is not warranted, especially if metabolism of a drug is known to be driven by the CYP family. The results of this study should lead to the abandonment of the routine use of dogs in human drug toxicology studies, and urge researchers and regulators to instead adopt appropriate and relevant scientific approaches in order to ensure human drug safety.

## Conclusion

The fraction of the genome covered by ROH ( $F_{ROH}$ ) was significantly different between human populations (Amish and 1000 genomes) and between humans and beagles. In the Amish population, one of the most inbred human populations, the  $F_{ROH}$  was 12-times smaller in comparison to the beagle population. ROH patterns in beagles increase their susceptibility to inbreeding depression, and may introduce deleterious or recessive traits. Due to the high degree of inbreeding in beagles, preclinical research should use caution when extrapolating from dogs to humans (especially when the enzymes metabolizing the drug are known), despite any physiological similarities between the species.

## Methods

### Datasets

EMBARC canine dog genome project

Vanda 1000 Genomes

1000 Genomes Consortium

### ROH detection

#### Genomic data analysis ROH

ROH values were calculated in human cohorts including the Amish, IGSR 1000 genomes, Welllderly, Vanda 1 k genomes. We detected ROH using PLINK. ROH scores reflect the probability of a stretch of SNPs being homozygous due to LOH and they are determined via the homozygosity frequency for each SNP in the genome. We also detected ROH using PLINK on the EMBARK dog data. We next calculated the extent of the genome covered by ROH ( $F_{ROH}$ ). ROH were defined as runs of at least 50 consecutive homozygous SNPs spanning at least 1500 kb, with less than a 1000 kb gap between adjacent ROH and a density of SNP coverage within the ROH of no more than 50 kb/SNP, with one heterozygote and 5 no calls allowed per window. Detection of ROH, depletion with confidence scores, LD in ROH, region analysis and comparison across species was performed on both human and dog datasets.

### DNA quantification

Incoming nucleic acid samples were quantified using fluorescent-based assays (PicoGreen) to accurately determine whether sufficient material is available for library preparation and sequencing.

### DNA integrity

DNA sample size distributions were profiled by a Fragment Analyzer (Advanced Analytics) or BioAnalyzer (Agilent Technologies), to assess sample quality and integrity.

### Genotyping

At the NYGC, we ran the HumanCoreExome 24v1.3 array for all human DNA samples sequenced.

## **WGS library preparation and sequencing, Truseq PCR-free (450 bp)**

Whole genome sequencing (WGS) libraries were prepared using the Truseq DNA PCR-free Library Preparation Kit

## **WGS Germline analysis part I**

Whole genome data were processed on NYGC automated pipeline. Paired-end 150 bp reads were aligned to the GRCh37 human reference (BWA-MEM v0.7.8) and processed with GATK best-practices workflow (GATK v3.4.0).

The mean coverage is 35.8, it reflects the samples average. The coverage for the FLG region tracks with the coverage for the whole genome. The coverage information per FLG region per sample is provided in the supplementary file.

All high quality variants obtained from GATK were annotated for functional effects (intronic, intergenic, splicing, nonsynonymous, stopgain and frameshifts) based on RefSeq transcripts using Annovar (<http://www.openbioinformatics.org/annovar/>)<sup>20</sup>. Additionally, annovar was used to match general population frequencies from public databses (Exac, GnomAD, ESP6500, 1000 g) and was used to prioritize rare, loss-of-function variants.

## **Abbreviations**

ROH Region of homozygosity

CYP cytochrome P450 family

## **Declarations**

## **Ethical approval and consent to participate**

Not applicable

## **Consent for publication**

All authors consent for publication

## **Availability of data and materials**

Not applicable

## Competing interests

SPS, MP are all employees of Vanda Pharmaceuticals,  
There are no other interests to disclose

## Funding

The study was funded by Vanda Pharmaceuticals Inc.

## Author contributions

Authors read and approved the manuscript

SPS wrote the manuscript and conducted the analyses

MP mentorship, provided oversight for the analyses and is the PI for the study

## Acknowledgements

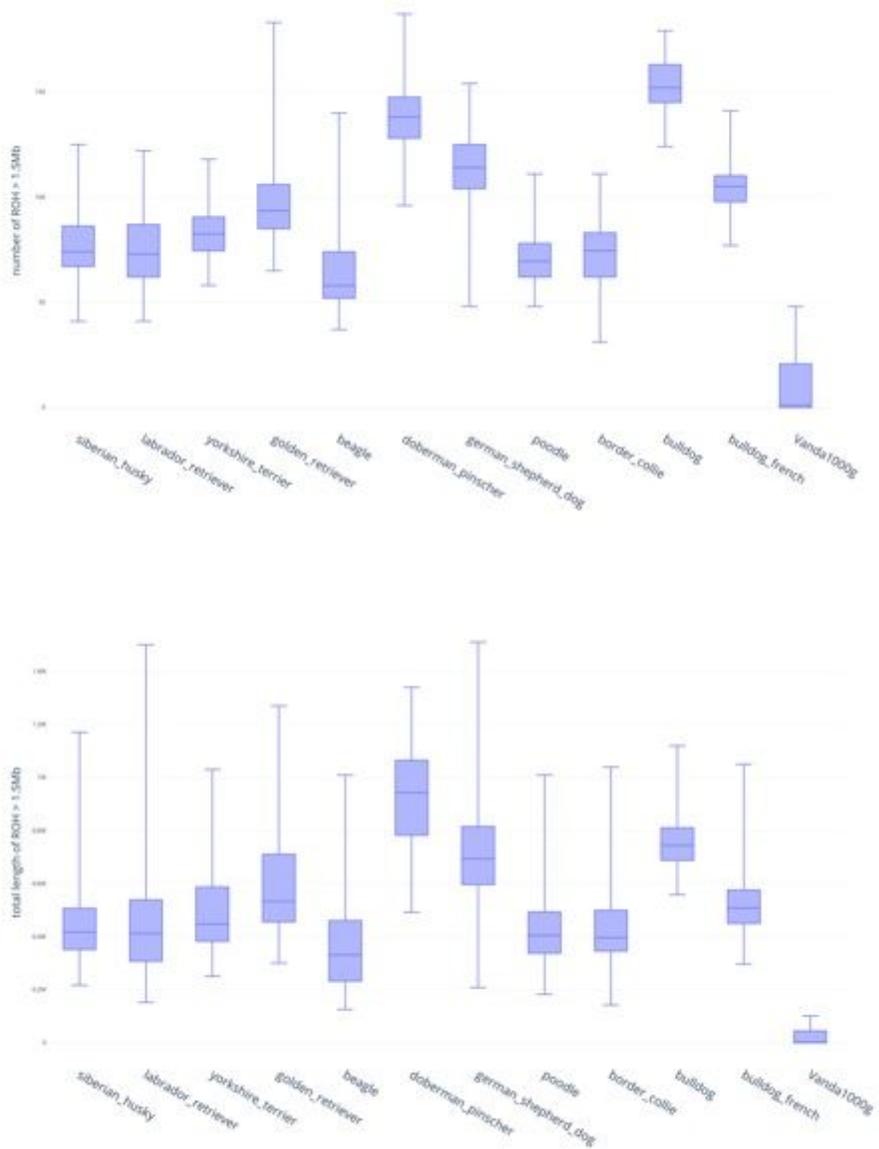
Acknowledgements: all participants of our clinical trials

## References

1. Bailey, J., Thew, M. & Balls, M. An Analysis of the Use of Animal Models in Predicting Human Toxicology and Drug Safety. *Altern. to Lab. Anim.* **42**, 181–199 (2014).
2. Sams, A. J. & Boyko, A. R. Fine-Scale Resolution of Runs of Homozygosity Reveal Patterns of Inbreeding and Substantial Overlap with Recessive Disease Genotypes in Domestic Dogs. *G3 (Bethesda)*. **9**, 117–123 (2019).
3. Szpiech, Z. A. *et al.* Long Runs of Homozygosity Are Enriched for Deleterious Variation. *Am. J. Hum. Genet.* **93**, 90–102 (2013).
4. DYER, W. T. T. The Effects of Cross and Self-Fertilisation in the Vegetable Kingdom. *Nature* **15**, 329–332 (1877).
5. Berra, T. M., Alvarez, G. & Ceballos, F. C. Was the Darwin/Wedgwood Dynasty Adversely Affected by Consanguinity? *Bioscience* **60**, 376–383 (2010).
6. Garrod, A. E. The incidence of alkaptonuria: a study in chemical individuality. 1902. *Mol. Med.* **2**, 274–82 (1996).
7. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* **19**, 220–234 (2018).

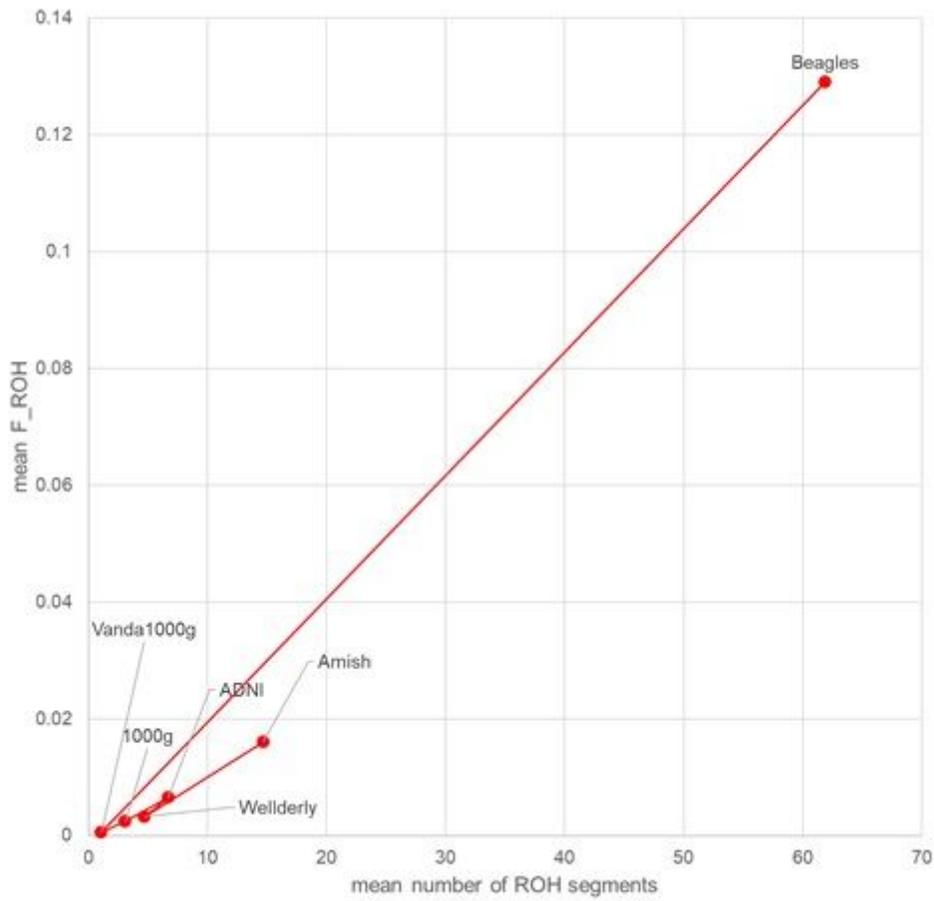
8. Robinson, J. A. *et al.* Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Sci. Adv.* **5**, eaau0757 (2019).
9. Clarke, L. *et al.* The 1000 Genomes Project: data management and community access. *Nat. Methods* **9**, 459–462 (2012).
10. Erikson, G. A. *et al.* Whole-Genome Sequencing of a Healthy Aging Cohort. *Cell* **165**, 1002–11 (2016).
11. Agarwala, R., Biesecker, L. G. & Schäffer, A. A. Anabaptist genealogy database. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **121C**, 32–37 (2003).
12. Wang, J.-F. & Chou, K.-C. Molecular Modeling of Cytochrome P450 and Drug Metabolism. *Curr. Drug Metab.* **11**, 342–346 (2010).
13. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
14. Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589-98 (2015).
15. Sund, K. L. *et al.* Regions of homozygosity identified by SNP microarray analysis aid in the diagnosis of autosomal recessive disease and incidentally detect parental blood relationships. *Genet. Med.* **15**, 70–78 (2013).
16. Pembroke, W. G., Babbs, A., Davies, K. E., Ponting, C. P. & Oliver, P. L. Temporal transcriptomics suggest that twin-peaking genes reset the clock. *Elife* **4**, 1–15 (2015).
17. Pemberton, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275–92 (2012).
18. Škarić-Jurić, T. *et al.* Characterization of ADME genes variation in Roma and 20 populations worldwide. *PLoS One* **13**, e0207671 (2018).
19. Gaedigk, A., Whirl-Carrillo, M., Pratt, V. M., Miller, N. A. & Klein, T. E. PharmVar and the Landscape of Pharmacogenetic Resources. *Clin. Pharmacol. Ther.* **107**, 43–46 (2020).
20. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).

## Figures



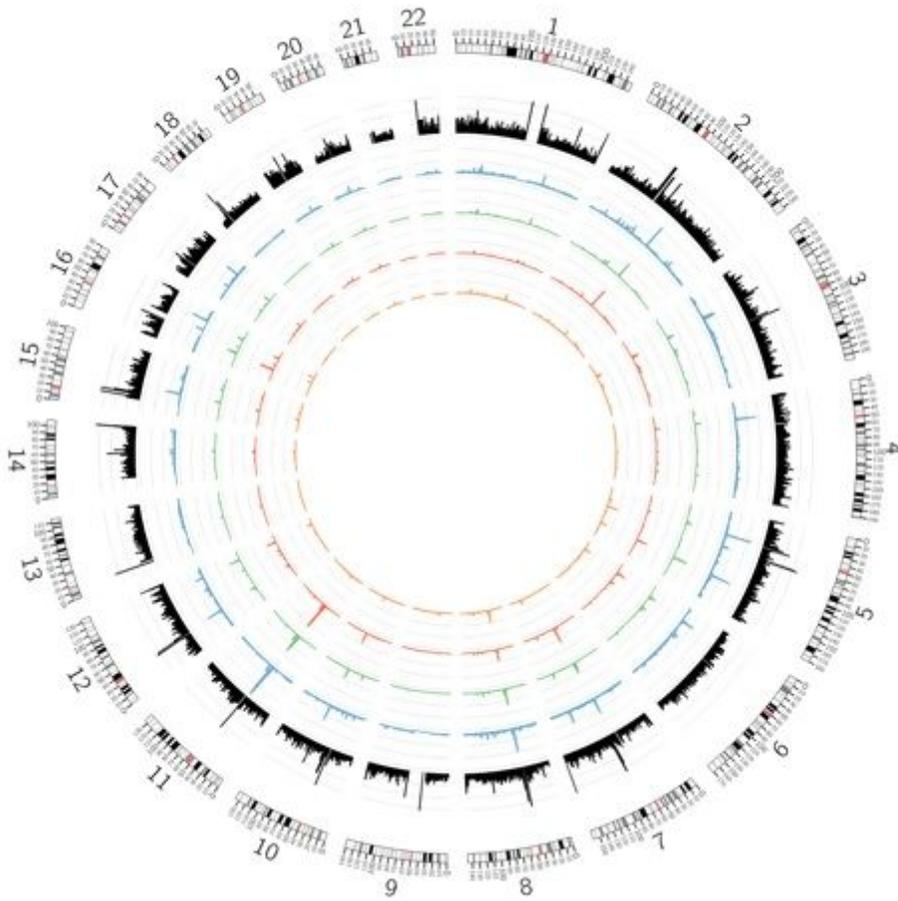
**Figure 1**

The total number and length of ROH identified across dog breeds as compared to mixed ethnicity human cohort: Vanda 1000 genomes.



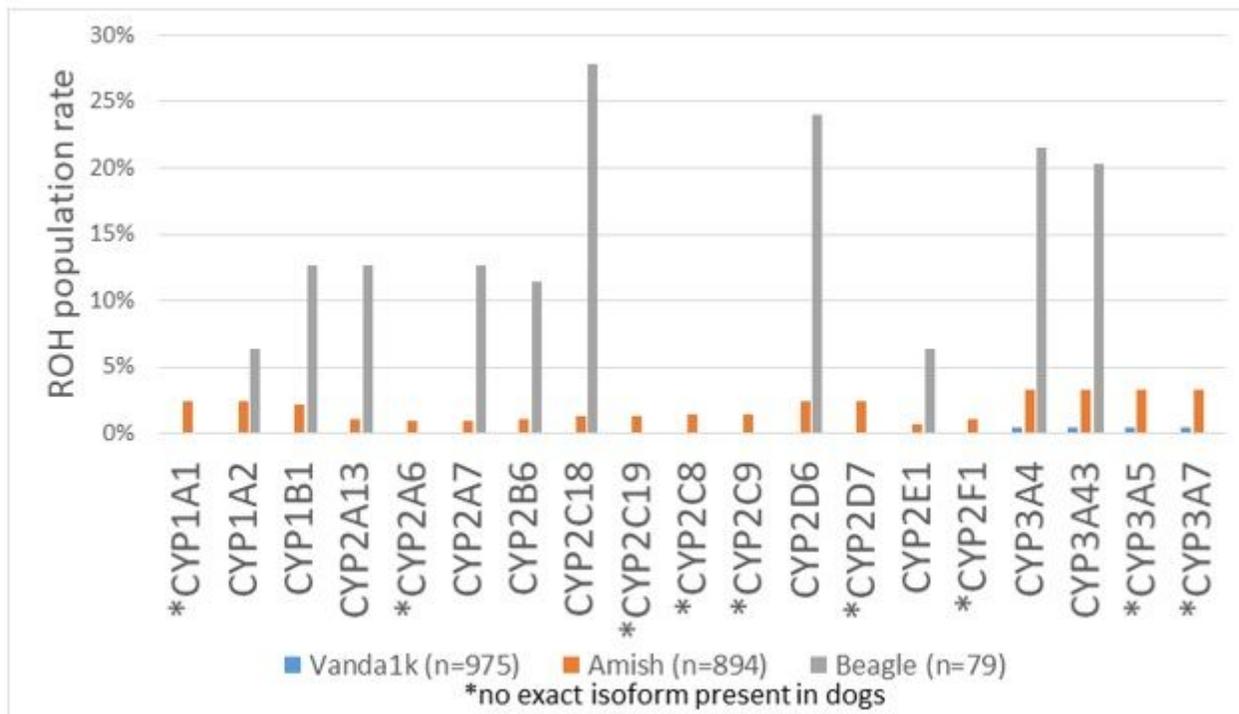
**Figure 2**

Mean F\_ROH on the y axis and mean number of ROH segments across several human and dog cohorts



**Figure 3**

ROH patterns across human cohorts: Human genome arranged circularly end-to-end: from inside to outside, the rings display ROH across different human cohorts starting from the innermost: 1000g, ADNI, Welllderly and Amish (blue). The outermost ring represents conservation score averaged per region of the human genome.



**Figure 4**

ROH patterns in CYPs compared between humans and dogs. Especially high signal is seen for CYP2C18, CYP2D6, CYP3A4, CYP3A43 amongst others.