Variants within the SP110 nuclear body protein modify risk of canine degenerative myelopathy

Emma L. Ivansson[a,b,1,2], Kate Megquier[a,b], Sergey V. Kozyrev[a], Eva Murén[a], Izabella Baranowska Körberg[c,3], Ross Swofford[b], Michele Koltookian[b], Noriko Tonomura[b,d], Rong Zeng[e], Ana L. Kolicheski[e], Liz Hansen[e], Martin L. Katz[f], Gayle C. Johnson[e], Gary S. Johnson[e], Joan R. Coates[g], and Kerstin Lindblad-Toh[a,b,1]

[a]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, 751 23 Uppsala, Sweden; [b]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142; [c]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden; [d]Department of Clinical Sciences, Cummings School of Veterinary Medicine at Tufts University, North Grafton, MA 01536; [e]Department of Veterinary Pathobiology, College of Veterinary Medicine, University of Missouri, Columbia, MO 65211; [f]Mason Eye Institute, School of Medicine, University of Missouri, Columbia, MO 65201; and [g]Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of Missouri, Columbia, MO 65211

Canine degenerative myelopathy (DM) is a naturally occurring neurodegenerative disease with similarities to some forms of amyotrophic lateral sclerosis (ALS). Most dogs that develop DM are homozygous for a common superoxide dismutase 1 gene (*SOD1*) mutation. However, not all dogs homozygous for this mutation develop disease. We performed a genome-wide association analysis in the Pembroke Welsh Corgi (PWC) breed comparing DM-affected and -unaffected dogs homozygous for the *SOD1* mutation. The analysis revealed a modifier locus on canine chromosome 25. A haplotype within the SP110 nuclear body protein (*SP110*) was present in 40% of affected compared with 4% of unaffected dogs ($P = 1.5 \times 10^{-5}$), and was associated with increased probability of developing DM ($P = 4.8 \times 10^{-6}$) and earlier onset of disease ($P = 1.7 \times 10^{-5}$). SP110 is a nuclear body protein involved in the regulation of gene transcription. Our findings suggest that variations in SP110-mediated gene transcription may underlie, at least in part, the variability in risk for developing DM among PWCs that are homozygous for the disease-related *SOD1* mutation. Further studies are warranted to clarify the effect of this modifier across dog breeds.

degenerative myelopathy | amyotrophic lateral sclerosis | ALS | SOD1 | SP110

Amyotrophic lateral sclerosis (ALS) is the most common adult-onset motor neuron disorder, with 50% of patients dying within 2–3 y of the onset of clinical signs (1). Despite significant progress in the mapping of genetic risk loci, development of successful therapeutic strategies has remained elusive, in part, due to the heterogeneity of the disease both genetically and phenotypically. Further genetic dissection will facilitate the discovery of modifier genes, which influence disease onset and severity, and may point the way to new therapeutic approaches. In this study, we detail the use of a comparative approach to identify a genetic modifier that affects disease penetrance and age of onset in degenerative myelopathy (DM), a canine model of ALS.

The dog is a particularly powerful comparative disease model for genetic studies of complex traits, combining aspects of the tractability of a model organism with the advantages of genetic trait mapping in population isolates, enabling the mapping of genetic risk factors using modest sample sizes (2). Dogs are predisposed to many of the same complex diseases that humans are, share an environment with their human owners, and receive a sophisticated level of medical surveillance and care (3).

ALS and canine DM are similar at a phenotypic, clinical, and genetic level. ALS is characterized by progressive loss of motor function and is characterized by stiffness and slowing of movements, difficulty in speaking and swallowing, muscle atrophy, and severe weakness culminating in paralysis. Mortality is typically secondary to failure of the respiratory muscles. Familial forms of the disease account for 5–10% of cases; the most common age of onset is 47–52 y for familial ALS and 58–63 y for sporadic disease (1). There are

several clinical subtypes with variable phenotypic presentation and prognosis.

Over 20 y ago, a mutation in the superoxide dismutase 1 gene (*SOD1*) was the first genetic risk factor to be identified (4). To date, more than 160 *SOD1* mutations involving all five exons have been identified in patients with ALS (alsod.iop.kcl.ac.uk/) (5). *SOD1* has been followed by a growing list of ALS-associated genes (6–8), including an intronic repeat expansion in chromosome 9 open reading frame 72 (*C9ORF72*) present in patients

Significance

Degenerative myelopathy (DM) is a canine disease very similar to amyotrophic lateral sclerosis (ALS) in humans. We previously showed that DM is a promising model for ALS, because genome-wide association identified a mutation in superoxide dismutase 1 gene (*SOD1*), a known ALS gene. This mutation found in many dog breeds increases the risk of DM, and the pathological findings and clinical progression of the two diseases are similar. In this study, we identify a modifier gene, SP110 nuclear body protein (*SP110*), which strongly affects overall disease risk and age of onset in Pembroke Welsh Corgis at risk for DM. Dissecting the complex genetics of this disease in a model organism may lead to new insights about risk and progression in both canine and human patients.

GENETICS

with sporadic ALS (9, 10) and 38% of patients with familial ALS (11). A recent study reported that some patients with familial ALS harbor mutations in more than one of the recognized ALS genes, including *SOD1*, *C9ORF72*, *TARDBP*, fused in sarcoma (*FUS*), and *ANG* (12). A large-scale exome sequencing study identified *TBK1* as an ALS susceptibility gene (13). The same genes associated with familial ALS have been found to harbor mutations in patients with sporadic ALS (7, 8). In summary, the current knowledge supports genetic heterogeneity in ALS etiology and suggests that genetic factors may play a role in patients with apparently sporadic disease.

Like ALS, canine DM is a naturally occurring, progressive adult-onset disease that leads to paralysis and death (14). The first clinical signs usually occur after 7 y of age and include general proprioceptive ataxia and asymmetrical spastic weakness of the hind limbs. Signs then progress to paraplegia, thoracic limb weakness, and, ultimately, flaccid tetraplegia (15). A presumptive clinical diagnosis is made by ruling out potential causes of compressive myelopathy; however, confirmation requires histopathological examination of the spinal cord (16). The pathological features of DM are similar to the pathological features of ALS (17–22). DM has been confirmed in over 24 breeds (16, 23) and presumptively reported in another nine breeds (16). In a previous study of canine *SOD1*, a *SOD1:c.118G > A* transition was identified that leads to a nonsynonymous substitution (E40K) in the homologous codon to the human E40G mutation (17). Homozygosity for the variant allele was associated with risk of developing DM in five dog breeds (17). A separate *SOD1* missense mutation has been discovered in Bernese Mountain Dogs, but has only been detected in that specific breed so far (23, 24). Biochemical characterization of the two canine SOD1 mutant proteins indicated the increased propensity to form protein aggregates with retained enzymatic activity, supporting a toxic gain-of-function role in canine DM similar to that role in human ALS (25). Taken together, these findings indicate that DM has potential as a naturally occurring disease model for human *SOD1*-related ALS.

Since the initial study was published (17), more than 35,000 dogs of multiple breeds have been genotyped for the *SOD1: c.118G > A* transition. Of the tested dogs, 49% were homozygous for the ancestral allele (G), 24% were homozygous for the risk allele (A), and 27% were heterozygous (GA), but the frequency of the *SOD1* risk allele was highly variable between breeds (23). In the Pembroke Welsh Corgi (PWC) breed, we have been able to confirm DM through histopathology in 53 phenotypically affected dogs; all of these dogs were homozygous for the *SOD1* risk allele (23). We noted that among PWCs with two copies of the *SOD1* risk allele, there were examples of dogs developing DM at a relatively early age (7–9 y), whereas others reached 15 y of age without any signs of DM. Similarly, all 42 Boxers with DM confirmed through histopathology were homozygous for the *SOD1* risk allele (23); however, in contrast to the PWC, there were few examples of Boxers homozygous for the *SOD1* mutation that reached old age (>11 y) without developing signs of DM.

Among dogs homozygous for the common DM-associated *SOD1* mutation, the variable prevalence and age of onset within and between breeds suggest that additional genetic factors play a role in DM. We hypothesized that the variability in penetrance of the disease phenotype could result from variations at additional genetic loci that modify disease risk, and could be detected by performing genome-wide association (GWA) analysis comparing affected and unaffected dogs homozygous for the *SOD1* risk allele. Identification of modifier loci will likely aid in understanding the etiology underlying DM, and might also provide insight into the pathogenesis of ALS. In the current study, we report a modifier locus within the SP110 nuclear body protein (*SP110*) on canine chromosome 25 (cfa25) that affects risk and age of onset of DM in the PWC breed, and is associated with altered expression

and changes in the gene isoform ratio of *SP110* that may be relevant to disease development.

Results

GWA Analysis Detects a Modifier Locus on cfa25. We performed a GWA analysis in at-risk PWC dogs homozygous for the *SOD1* risk allele to detect genetic modifier loci that differentiate between dogs that developed the disease early and those dogs that did not develop disease even at an advanced age. By comparing cases with a confirmed diagnosis and early onset of DM signs with older dogs without any signs of the disease, we obtained phenotypes that were clearly defined and well separated. GWA analysis was performed in a final dataset of 15 affected and 31 unaffected PWCs. Quality control left 119,768 SNPs at a total genotyping rate of 99.9% for analysis. There were no outliers in the dataset according to the multidimensional scaling plot (*SI Appendix*, Fig. S1) and the lambda (genomic inflation factor) was 0.99, indicating successful control of population stratification. The analysis revealed a single locus of association on cfa25, with the strongest associated SNP (BICF2G630104165, located at position canFam2 25:45,443,320) reaching genome-wide significance ($P = 2.7 \times 10^{-8}$) (Fig. 1). The association in this region was well outside of the 95% confidence interval based on the distribution of effect size beta values (Fig. 1). Removal of the five most strongly associated SNPs on cfa25 and any SNPs tagged by these SNPs ($r^2 > 0.4$) extinguished the association, demonstrating that the inflation in the quantile–quantile plot reflected association from this region only (*SI Appendix*, Fig. S2).

Fine-Mapping of the GWA Locus Reveals a Haplotype Associated with Risk. Whole-genome sequencing of three PWCs, two affected by DM at the age of 9 y and one without signs of DM at the age of 14 y, was performed to generate comprehensive information on genetic variants present in this breed. In 10 Mb surrounding the GWA locus, we identified a total of 35,050 SNPs and 10,740 small insertion or deletion of bases (INDELS). To pinpoint the location of the association signal on cfa25, we selected 101 SNPs in the region cfa25:42,181,379–46,659,998 from the whole-genome sequencing data for genotyping in the GWA sample set. Adding the 101 SNPs to the GWA analysis revealed another four variants in the vicinity of the top GWA SNP significantly associated with risk of disease (Fig. 2 and Table 1). Genotype data from these five significantly associated SNPs were used for haplotype phasing. The results indicated that the five SNPs form four haplotypes in the PWC breed, with the most common haplotype at an overall frequency of 80% and the second most common at an overall frequency of 14% (Table 2). The second most common haplotype contained the risk alleles from the five associated SNPs and was carried in at least one copy by the majority of cases (nine of 15 cases, 60%) but only one control (one of 35 controls, 3%) in the GWA dataset (Table 3). We designated this haplotype the "PWC risk haplotype."

The samples included in the GWA analysis were selected to represent the extreme phenotypes: early-onset cases (n = 15, mean age of onset = 9.0 y, SD = 0.7) and older healthy dogs (n = 35, mean age at ascertainment = 13.2 y, SD = 1.33). In this set of samples, the frequency of cases carrying the risk haplotype in at least one copy was significantly different from controls ($P = 1.7 \times 10^{-5}$, Fisher's exact test). We next evaluated the frequency of the risk haplotype in an additional set of PWCs with confirmed DM but less strict age of onset (n = 32, mean age of onset = 11.6 y, SD = 1.3), as well as in additional unaffected PWCs (n = 13, eight older than 11 y and five without exact age information), to replicate the association. Again, the risk haplotype was present at a higher frequency in the affected dogs (10 of 32 cases, 31%) compared with the unaffected dogs (one of 13 controls, 8%), but the difference in frequencies was not statistically significant ($P = 0.14$). Because the phenotypes in the replication dataset were
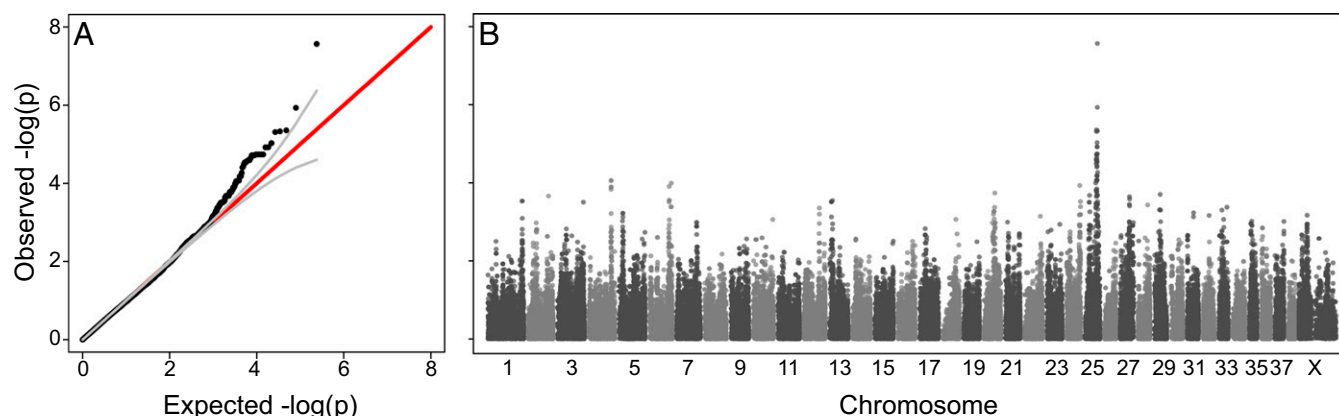
Fig. 1. GWA analysis identified a modifier locus on cfa25 associated with risk of canine DM in the PWC breed. The final dataset consisted of 15 affected and 31 unaffected dogs. The SNP at cfa25:45,443,320 reached genome-wide significance ($P = 2.7 \times 10^{-8}$). A quantile–quantile plot, $\lambda = 0.99$ (A) and Manhattan plot (B) are shown.

less stringent, it was expected that the effect would be less strong. Merging the discovery and replication datasets resulted in a significant difference between the frequencies of affected (40%) and unaffected (4%) dogs carrying at least one copy of the risk haplotype ($P = 1.5 \times 10^{-5}$), supporting that the haplotype was associated with risk (Table 3). We noted that of all DM-affected PWCs in the present study ($n = 47$), 16 were heterozygous for the risk haplotype and three were homozygous, whereas among the unaffected PWCs ($n = 48$), the two carriers were heterozygous for the risk haplotype. The population haplotype frequencies based on 273 PWCs without phenotype information (Table 2) illustrated that the population frequency of the risk haplotype was between the frequencies of affected and unaffected PWCs.

The Modifier Affects Age at Onset. Due to the study design, the mean age of onset was lower in the discovery dataset than in the replication dataset (9.0 vs. 11.6 y; $P = 4.7 \times 10^{-11}$). To assess whether having the risk haplotype at the modifier locus affected age of onset, we performed Kaplan–Meier analysis using the development of DM signs as the event and age of onset or age at ascertainment as the time to event. The analysis incorporated both the discovery and replication datasets (47 cases and 48 controls). Comparing individuals with and without the risk haplotype revealed a difference in the probability of developing signs of DM over time ($P = 4.8 \times 10^{-6}$, log-rank test) (Fig. 3). The individuals were all predisposed to DM through the *SOD1* risk genotype, but at the age of 11 y, the probability of showing signs of DM was 0.77 in dogs with the *SP110* risk haplotype and 0.18 in dogs without the risk haplotype.

The Haplotype Associated with Risk in PWCs Is Common in Boxers. We evaluated whether the PWC risk haplotype also influenced risk of DM in the Boxer breed. Haplotype data were available from 25 Boxers homozygous for the *SOD1* risk allele: 15 affected by disease with histopathology confirming DM (mean age of onset = 9.5 y, SD = 2.0) and 10 without signs of DM at the age of 11 y. The haplotype was common in Boxers; all 25 dogs studied carried at least one copy (*SI Appendix*, Table S4), and 21 (including all unaffected dogs) were homozygous, indicating reduced variability in Boxers for this region of the genome.

The PWC Risk Haplotype Is Present in Dogs of Other Breeds and May Influence DM Risk. Whole-genome sequencing data of dog pools (26) showed variation at all five associated sites across dog breeds, indicating that these variants were not unique to the PWC and Boxer. To investigate the presence of PWC haplotypes in other breeds, we genotyped representatives from 85 dog breeds for the five SNPs. Complete genotyping was achieved for 265 dogs, and the data were used to phase haplotypes. The four haplotypes observed in the PWC were the most common in the other breed dataset, representing 82% of all haplotypes (*SI Appendix*, Table S4). The remaining 18% consisted of 10 haplotypes, each with an overall frequency of less than 5%. The PWC risk haplotype was detected in 38 breeds with an overall frequency of 23%. The DM status of these dogs was unknown, except for a subset of confirmed DM cases in other breeds with known mutations in *SOD1* ($n = 36$). In this subset of cases, 64% carried the PWC risk haplotype compared with 30% of unphenotyped dogs that did not carry the *SOD1* mutation ($n = 183$).

The Associated Haplotype Resides Within SP110. The associated haplotype encompassed 12.5 kb of exonic and intronic sequences within the gene *SP110* on cfa25 that encodes the SP110 nuclear body protein. There were two coding substitutions in the five SNPs with GWA; the SNP with the strongest association (cfa25:45,443,320) was a synonymous substitution, and the variant
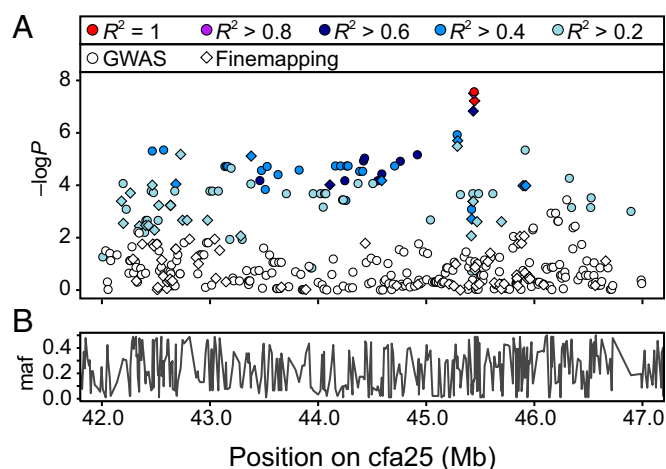
Fig. 2. Fine-mapping of the cfa25 region associated with DM in PWCs. (A) Analysis of the GWA data, together with an additional 101 SNPs identified through whole-genome sequencing, allowed fine-mapping of the associated region. Besides the top GWA SNP, another four SNPs reached genome-wide significance. These SNPs were tightly linked and located in close proximity to the top GWA SNP. (B) Minor allele frequency (maf) in PWCs across the cfa25 region. In the PWC, there was no evident drop in heterozygosity across the fine-mapped region.

Table 1. SNPs with genome-wide significant association in fine-mapping analysis

| SNP | OR | P_{emmax} | Base pair location | Minor allele frequency | | PWC alleles | | | | LD with topsnp (r^2) | Annotation |
				Affected	Unaffected	Minor (risk)	Major	canFam2	hg19		
cfa25:45435040	1.8	1.5×10^{-7}	25:45,435,040	0.42	0.02	A	G	A	A	0.74	Intron of *SP110*
cfa25:45437568	2.0	3.1×10^{-8}	25:45,437,568	0.38	0.02	T	C	T	T	0.82	Intron of *SP110*
BICF2G630104165	1.9	2.7×10^{-8}	25:45,443,320	0.40	0.03	G	A	A	A	1	Synonymous coding *SP110*
cfa25:45445891	1.9	6.0×10^{-8}	25:45,445,891	0.42	0.03	G	A	G	—	1	Intron of *SP110*
cfa25:45447628	1.9	6.0×10^{-8}	25:45,447,628	0.42	0.03	T	A	A	T	1	Nonsynonymous coding *SP110*

Five SNPs displayed genome-wide significant associations in analysis performed by EMMAX incorporating two principal components to adjust for population structure. The SNPs were in strong LD and located in a 12.5-kb region within the gene *SP110*. The five SNPs in the table were used to construct the haplotypes displayed in Table 2. OR, odds ratio.

at cfa25:45,447,628 was a nonsynonymous substitution (Table 1). The effect of the nonsynonymous variant on the dog protein iso-forms was predicted to be neutral using scale-invariant feature transform (SIFT) (27), Polymorphism Phenotyping version 2 (PolyPhen-2) (28), screening for non-acceptable polymorphisms (SNAP) (29), and the consensus classifier PredictSNP (30).

To detect any additional variants on the haplotype, we deep-sequenced the 12.5-kb region in 34 PWCs from the GWA dataset with available DNA. This analysis revealed another 32 SNPs and six INDELS that were merged with the GWA and fine-mapping data and analyzed for association. Three of the additional vari-ants were in perfect linkage disequilibrium (LD) with the top SNP and showed equally strong association: cfa25:45,444,053, cfa25:45,444,120, and cfa25:45,445,768 ($P = 2.7 \times 10^{-8}$). *SI Appendix*, Table S1 lists association results for all variants detected in the haplotype region. Fig. 4 illustrates the final association re-sults, including the sequence-detected variants, lifted over to the corresponding region of the human genome hg 19 (GRCh37). In addition to the two coding variants mentioned above, the var-iants at cfa25:45,444,053 (translating to hsa2:231,067,960) and cfa25:45,445,768 (translating to hsa2:231,072,975) were inter-esting functional candidates because they overlap a hotspot for transcription factor binding.

The Associated Noncoding Variants Show Regulatory Potential. To validate the predicted regulatory potential of SNPs cfa25:45,444,053 and cfa25:45,445,768, we cloned genomic DNA fragments with the variants in a luciferase reporter vector and measured the effect of these alleles on luciferase gene expression after transfection into the Jurkat human T-cell line (Fig. 5 *A* and *B*). In addition to cfa25:45,445,768, the second DNA fragment contained variants at cfa25:45,445,751, cfa25:45,445,837, and cfa25:45,445,891 that were in LD with cfa25:45,445,768, and we thus measured the total effect of four SNPs. The choice of cell line was based on the fact that the highest levels of *SP110* gene expression were reported in immune cells, including T cells, at BioGPS (31). We found that for both cloned fragments, the risk allele provides lower expression levels compared with the nonrisk allele. The DNA fragment with the cfa25:45,444,053 variant showed re-pressive properties compared with the vector (Fig. 5*A*) but, upon cell stimulation, induced the reporter gene expression almost twofold compared with the nonstimulated cells. The fragment with the cfa25:45,445,768 variant enhanced expression over the control vector (Fig. 5*B*) but did not show a large induction of gene expression upon cell stimulation. This effect may be due to an inducible cell-specific enhancer located in the first fragment, which is more active in, for example, B cells, dendritic cells, or natural killer cells than in T cells. The fact that risk alleles from

both fragments are associated with lower gene expression indicates that both SNPs may exert a cumulative effect on the levels of *SP110*. Because the variant at cfa25:45,444,120 is located in a region that does not translate to the human genome, its regulatory potential could not be predicted by conservation of regulatory marks; therefore, it was not analyzed in the functional experiment.

To investigate the roles of the intronic SNPs at cfa25:45,444,053 and cfa25:45,445,768 as potential regulatory binding sites, we per-formed electrophoretic mobility shift assays (EMSAs). Assay of the risk allele at cfa25:45,444,053 yielded two stronger bands compared with the nonrisk allele, whereas assay of the nonrisk allele yielded a band that was not seen in the risk allele assay (Fig. 5*C*). These findings may indicate that the risk allele increases binding affinity for yet unidentified transcription factors while eliminating binding of a different factor. EMSA of the SNP at cfa25:45,445,768 revealed one stronger band in the risk allele assay compared with the nonrisk allele assay (Fig. 5*D*). In addition, we evaluated the top GWA SNP, the synonymous substitution at cfa25:45,443,320. This evaluation revealed two stronger bands in the risk allele assay compared with the nonrisk allele assay, suggesting that a higher binding affinity may be created by the SNP at this location (*SI Appendix*, Fig. S3). Further experiments will be needed to evaluate the results of the EMSAs quantitatively and to identify the factors involved in differential binding.

Risk Alleles Correlate with Alternative Splicing of SP110 and Change the Balance Between Isoforms. We next measured expression of *SP110* and the closely related gene *SP140* in the blood cells of healthy Nova Scotia Duck Tolling Retriever (NSDTR) dogs genotyped for cfa25:45,444,053. The *SP140* gene is located 5 kb upstream of *SP110* in a head-to-head position, and thus could share

Table 2. Haplotype frequencies in PWC

| | GWAs | | Replication | | Population |
Haplotype	Affected	Unaffected	Affected	Unaffected	Unknown
	($n = 15$)	($n = 35$)	($n = 32$)	($n = 13$)	($n = 273$)
GCAAA	0.57	0.97	0.78	0.85	0.79
ATGGT	0.33	0.01	0.19	0.04	0.14
GCGGT	0.07	0.01	0.02	0.04	0.04
ACAAA	0.03	0.00	0.02	0.04	0.02

The four haplotypes present in PWC, and their frequencies across the GWA and replication datasets, as well as the population estimate based on unphenotyped PWC, are shown. Haplotypes were constructed by phasing genotype data for the variants at cfa25:45,435,040, cfa25:45,437,568, cfa25:45,443,320, cfa25:45,445,891, and cfa25:45,447,628 using PHASE (62, 63).

Table 3. Frequency of PWC carrying haplotype ATGGT differed between affected and unaffected individuals

	Discovery		Replication		Merge	
Carrier status	Affected, n (%)	Unaffected, n (%)	Affected, n (%)	Unaffected, n (%)	Affected, n (%)	Unaffected, n (%)
One or two copies of ATGGT	9 (60)	1 (3)	10 (31)	1 (8)	19 (40)	2 (4)
No copy of ATGGT	6 (40)	34 (97)	22 (69)	12 (92)	28 (60)	46 (96)
P using Fishers two-sided test	1.7×10^{-5}		0.14		1.5×10^{-5}	

The frequency of individuals at least heterozygous for the haplotype ATGGT differed between affected and unaffected PWC; haplotype ATGGT was associated with risk because it occurred in 40% of the cases but in only 4% of the unaffected cases.

regulatory regions with *SP110*. The NSDTR breed has a higher frequency of the minor allele compared with the PWC breed, which facilitated study of the SNP effect on gene expression.

The splicing of the canine *SP110* gene is very complex, and many aberrant alternative transcripts can be detected at low levels (32). We found two major *SP110* isoforms: a full-length transcript and a previously unidentified transcript with in-frame skipping of exon 8 (Δ exon 8 transcript) (Fig. 4C and *SI Appendix,* Fig. S4). Both transcripts include alternative splicing of exon 16, bringing the total of highly abundant isoforms to four (both the full-length and Δ exon 8 transcripts plus or minus exon 16). Although all four isoforms are constitutively coexpressed, we found that the levels of exon 8 were dependent on the genotype at cfa25:45,444,053 (Fig. 6 A and B and *SI Appendix,* Fig. S4A). The risk allele (G) at cfa25:45,444,053 correlated with down-regulation of the full-length transcript and up-regulation of the Δ exon 8 transcript, whereas the nonrisk allele (T) was associated with the opposite trend. The inclusion of exon 16 was independent of genotype and occurred equally in the full-length and Δ exon 8 transcripts. The

Fig. 3. Kaplan–Meier analysis of time to onset of DM signs comparing carriers and noncarriers of the risk haplotype. Results of Kaplan–Meier survival analysis of PWCs with and without the risk haplotype using onset of DM as the event and age at onset as the time to event are shown. Individuals without an event were censored at the last time point when information regarding signs of DM was available. Carriers of the risk haplotype showed an increased probability of developing signs of DM over time ($P = 4.8 \times 10^{-6}$, log-rank test). At the age of 11 y, the probability of not showing signs of disease was 0.82 [SE = 0.05; 95% confidence interval (CI) = 0.74–0.92] in dogs without the risk haplotype and 0.33 (SE = 0.10; 95% CI = 0.18–0.61) in dogs with the risk haplotype.

total *SP110* gene expression measured by quantitative RT-PCR with primers common for all isoforms showed a weak trend toward gene down-regulation in the risk allele G, although this trend did not reach statistical significance ($P = 0.098$) (*SI Appendix,* Fig. S5A). The risk allele had no effect on *SP140* gene expression (*SI Appendix,* Fig. S5B). Interestingly, the two coding variants, nonsynonymous cfa25:45,447,628 and synonymous cfa25:45,443,320, are located in exons 6 and 9, and noncoding SNPs cfa25:45,445,768 and cfa25:45,444,053 are correspondingly located in introns 6 and 8 (Fig. 4C).

Careful examination of the genomic sequence brought two more variants to our attention: SNP cfa25:45,445,274 in intron 7 and synonymous SNP cfa25:45,445,209 in exon 8. These SNPs are in LD with the aforementioned four SNPs and also showed high association (*SI Appendix,* Table S1). Although these two SNPs do not alter the splicing sites directly (33, 34) (*SI Appendix,* Fig. S6), the variant at cfa25:45,445,209 may be involved in the correct splicing of exon 8 by stabilizing the binding of the SRp55 exonic splicing enhancer factor (35) (*SI Appendix,* Fig. S7).

The SP110 Locus in Human GWA Studies. The GRASP (36) tool was used to search published GWA studies for supporting associations in human patients. SNPs within or near *SP110* were searched for association with ALS specifically, or with the broader "Neuro" phenotype category, which includes other neurodegenerative disorders as well as neurodevelopmental or neuropsychiatric disorders. Twenty-nine (29) associations with a P value less than 10^{-3} were found in 27 unique SNPs, including one association with ALS (*SI Appendix,* Fig. S8 and Table S3). Due to the role of inflammation in the pathophysiology of ALS, we also searched GRASP for SNPs associated with diseases in the "Inflammation" category, finding five subsignificant associations with three unique SNPs in the region, one of which overlaps both the Neuro and Inflammation categories (*SI Appendix,* Fig. S8 and Table S3). Overall, the associations in the region within and near the *SP110* gene were enriched in the Neuro category, with 28 of 54 (51.9%) of the total associations with $P \leq 10^{-3}$ falling within this category.

The ALS-associated SNP (rs12162384) is located 74 kb downstream of *SP110* ($P = 3.8 \times 10^{-5}$). This association comes from a GWA study of 266 sporadic cases of ALS and 1,190 controls in an Italian population (37). A second, less significant ALS-associated SNP ($P = 3.1 \times 10^{-3}$) is located 6 kb downstream of *SP110* (38). The most significantly associated SNP in the Neuro category near *SP110* is associated with autism spectrum disorders ($P = 6.8 \times 10^{-6}$) (39).

Discussion

DM shares clinical, pathological, and biochemical characteristics with upper motor neuron onset forms of ALS (16, 17, 20, 25). We have previously shown that a mutation in *SOD1* was associated with risk of DM in several dog breeds and that, as in patients with *SOD1*-related ALS, cytoplasmic aggregates containing SOD1 protein were present in the spinal cord motor neurons of affected individuals (17). No SOD1-containing aggregates were found in control spinal cords from wild-type homozygotes. The mechanism
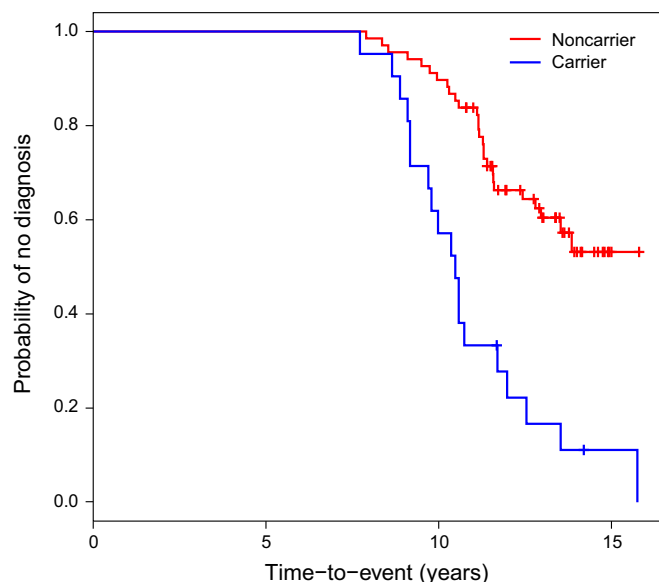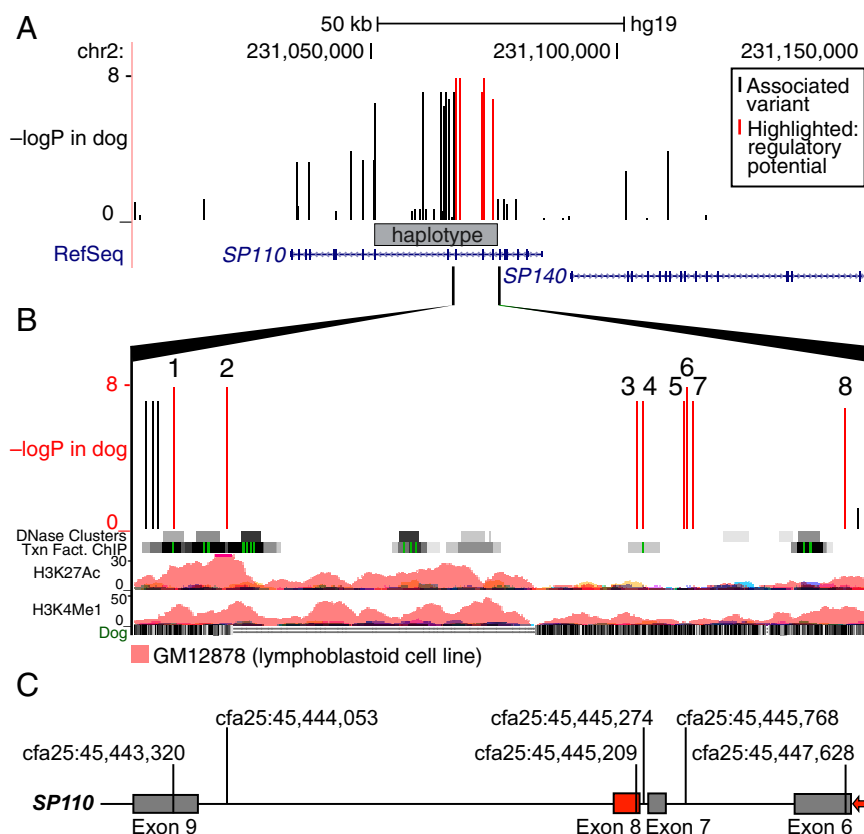
Fig. 4. Associated haplotype and potential regulatory variants lie within the *SP110* gene. (*A*) LiftOver of all variants in the associated region on cfa25 from the dog genome (Broad/canFam2) arrived at chromosome 2 in the human genome (GRch37/hg19), although not all variants had a corresponding site in the human genome (*SI Appendix*, Table S1). The PWC-associated haplotype is indicated by the gray bar. Association results (–logP) are shown in red and black, with SNPs of particular interest for their regulatory potential highlighted in red. Reference Sequence (RefSeq) genes are indicated in blue (genome.ucsc.edu) (67). (*B*) Zooming in closer on the associated region revealed that the haplotype resided within the *SP110* gene and harbored a number of strongly associated sequence variants. The figure shows tracks of ENCODE data (68) with digital DNaseI hypersensitivity clusters, transcription factor binding sites, H3K27Ac and H3K4Me1 histone marks, as well as the Multiz alignment of the dog genome in the lower track. The ENCODE tracks support the presence of regulatory elements in this area, particularly around the top SNP and toward the end of the haplotype. Highlighted SNPs (in red) include the following: (1) cfa25:45,443,320, a synonymous variant in dog exon 9, evaluated using EMSA; (2) cfa25:45,444,053, a dog intronic regulatory variant evaluated by luciferase assay, EMSA, and allele-specific PCR; (3) cfa25:45,445,209, within dog exon 8, potentially regulating splicing; (4) cfa25:45,445,274, within dog intron 7, potentially influencing splicing; (5) cfa25:45,445,751, a dog intronic variant, evaluated by luciferase assay; (6) cfa25:45,445,768, a dog intronic regulatory variant, evaluated by luciferase assay and EMSA; (7) cfa25:45,445,837, a dog intronic variant, evaluated by luciferase assay; and (8) cfa25:45,447,628, a nonsynonymous change in dog exon 6, predicted to be neutral (genome.ucsc.edu) (67). (*C*) Schematic structure of the *SP110* gene from exon 6 to exon 9 with variants found associated and functionally relevant to gene regulation. The exon (exon 8) undergoing alternative splicing is shown as a red box. The direction of transcription is shown by the red arrow.

behind SOD1 toxicity in ALS pathology is still unclear, although studies in transgenic rodent ALS models suggest that the expression of the mutant SOD1 in nonneuronal cells, such as astrocytes and microglia, has a definitive role in disease pathogenesis (40).

The current study aimed to reveal why some dogs homozygous for the *SOD1* mutation were susceptible to DM, whereas others seemed resistant. We identified a modifier locus within the *SP110* gene that was associated with an increased probability of developing signs of DM, and an earlier age of onset in PWCs. The risk haplotype usually occurred as a single copy, suggesting that one copy of the modifier allele was sufficient to affect function, thus supporting a dominant effect.

The fact that most Boxers were genetically similar across the *SP110* locus implies that the locus is unlikely to modify the risk of DM between individual Boxers homozygous for the *SOD1* risk allele, and illustrates that the genetics underlying susceptibility could differ between breeds. Because the haplotype associated with risk in PWC appears very common in the Boxer breed, the locus may contribute to the overall genetic predisposition in Boxers, but it is also possible that there are additional loci acting

as modifiers. The enrichment of the PWC risk haplotype among confirmed DM cases in other breeds suggests that the PWC modifier might affect multiple breeds. Future studies, including well-characterized samples from unaffected dogs with the *SOD1* mutation, are needed to understand fully the role of the PWC modifier in the Boxer as well as in other breeds.

SP110 is a member of the SP100/SP140 family of nuclear body proteins expressed in a variety of tissues, but most strongly in immune cells (32, 41, 42). It is a component of promyelocytic leukemia nuclear bodies, which form a part of the nuclear matrix and influence transcription, apoptosis, senescence, and response to DNA damage or infection (43).

Mutations in *SP110* have been reported in "familial hepatic venoocclusive disease with immunodeficiency," suggesting that *SP110* plays a role in the immune response (44–47). Recently, SP110 was identified as a regulator of the IFN-stimulatory DNA sensing pathway, an important part of the innate antiviral response (48). Interestingly, two other ALS genes, *FUS* and *OPTN*, have already been linked to this pathway (49, 50). This recurring connection suggests that the role of SP110 in DM may be
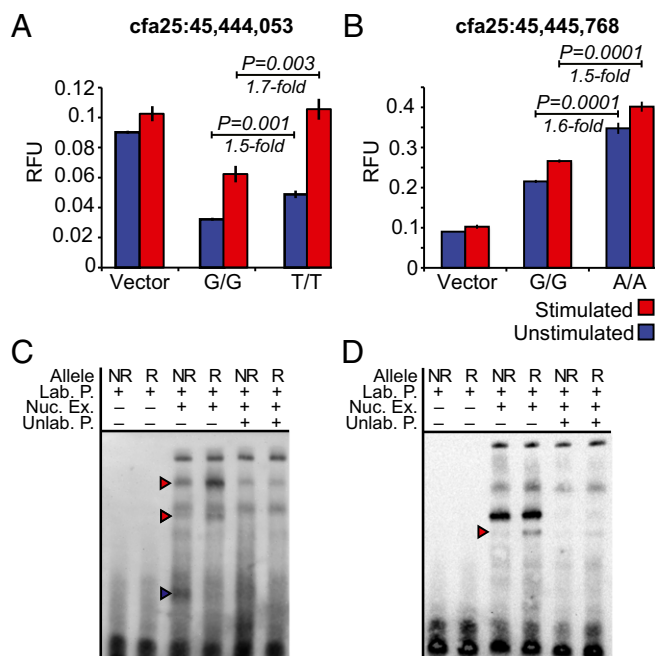
Fig. 5. Luciferase and EMSA analyses of variants assessing regulatory potential. The regulatory potential of intronic SNPs cfa25:45,444,053 (*A*) and cfa25:45,445,768 (*B*) was assessed by luciferase reporter assay in the Jurkat T-cell line. After transfection, cells were left unstimulated or stimulated with PMA and ionomycin for 12 h. The risk alleles for both SNPs correlate with lower luciferase levels in nonstimulated and stimulated cells. RFU, relative fluorescence units. Bars represent mean values ± SEM. Statistical analysis was done using an unpaired *t* test. An EMSA was performed to test for differential DNA binding of Jurkat-cell nuclear extract between the nonrisk and risk alleles at cfa25:45,444,053 (*C*) and cfa25:45,445,768 (*D*). (*C*) Binding to the risk allele at cfa25:45,444,053 appears stronger than the corresponding binding to the nonrisk allele at two locations (red arrowheads), and binding of one transcription factor may be lost in the risk allele (blue arrowhead). (*D*) At cfa25:45,445,768, binding to the risk allele at one location appears stronger than in the nonrisk allele (red arrowhead). Lab. P., labeled probe; NR, nonrisk allele; Nuc. Ex., Jurkat nuclear extract; R, risk allele; Unlab. P., unlabeled probe.

related to the same DNA sensing pathway. Neuroinflammation is a prominent feature in ALS and is characterized by a dialogue between microglia, T cells, and neurons, creating a balance between neuroprotection and neurotoxicity (51).

The functional analysis of selected associated variants indicates that they are involved in *SP110* gene regulation. The risk alleles of both intronic SNPs tested by the reporter assay were associated with gene repression. Indeed, when we measured the total expression of *SP110* in blood cells, there was a trend toward *SP110* down-regulation in the risk genotype. However, this trend did not reach statistical significance. Furthermore, we identified dramatic changes in the balance of different splice transcripts coding for proteins with or without an amino acid region encoded by exon 8. Whereas the full-length transcripts coding for two proteins of 720 and 705 amino acids (with or without exon 16) were repressed in the risk allele, the Δ exon 8 transcripts coding for 703 and 688 amino acids (with or without exon 16) were up-regulated, which may indicate a tightly controlled regulation of *SP110* functionality. The exon 8-encoded portion is located in the interdomain linker connecting the homogeneously staining region (HSR) and Sp100, AIRE-1, NucP41/75, DEAF-1 (SAND) domains (52), and it may potentially be involved in the protein–protein interactions, which may result in expression changes of target genes and/or execution of different transcriptional programs.

Interestingly, it was shown recently by yeast two-hybrid interaction screening that SP110 physically interacts with survival of motor neuron 1 (SMN1) and transthyretin (TTR) (53). This interaction is noteworthy because these two proteins are known to be involved in the degenerative neuromuscular disorders spinal muscular atrophy and TTR amyloidosis, pointing to the possibility of common or intersecting pathways affected in neurodegeneration that may result in the development of different yet related neuromuscular diseases.

In addition, SNPs with subsignificant associations with human neurodegenerative, neuropsychiatric, and inflammatory disorders are found within and around *SP110* (*SI Appendix*, Fig. S8). Although the associations in humans do not reach significance, the strong signal in dogs may support the involvement of *SP110* in the human disease pathogenesis and warrant a deeper analysis of ALS cohorts.

The fact that ALS displays heterogeneity in both phenotype and genotype complicates the development of treatments. Identifying modifier loci that influence disease severity or age of onset is important because such loci could point toward the final pathways of neurodegeneration rather than initial events, and thereby offer therapeutic opportunities that are shared across patients (8, 38).

To our knowledge, this study is the first proposing the involvement of a nuclear body protein in DM or ALS susceptibility. Establishing the role of the *SP110* gene in the pathology of DM as well as the potential role in ALS will require additional investigation, such as evaluating the effect of the risk haplotype on expression of *SOD1* and presence of SOD1 aggregates. We believe our finding suggests that variation in the immune response associated with variation in SP110 isoforms can alter the onset and progression of these diseases. Developing an understanding of the mechanism by which variation in *SP110* influences DM disease risk could help guide future clinical trials.

In conclusion, DM and ALS are fatal progressive neurodegenerative diseases with no effective treatments and where pathogenesis remains undefined. The aim of the present study was to identify genetic modifiers of disease risk in dogs that are predisposed to DM by being homozygous for the *SOD1* risk allele. We report that variants within *SP110* modify the genetic risk and age of onset of DM in PWC dogs homozygous for mutant *SOD1*, and that those variants contribute to changes in the *SP110* gene regulation and isoform ratio expressed in blood cells.
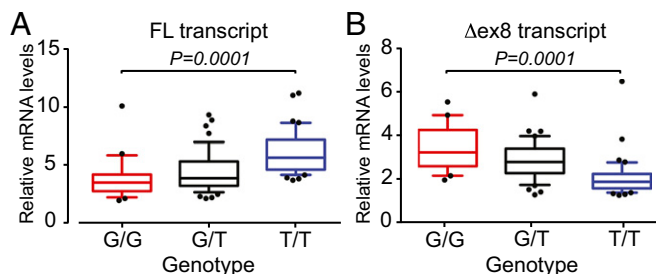
Fig. 6. *SP110* isoform expression by genotype at cfa25:45,444,053. Expression levels of the full-length (*A*; FL) and Δ exon 8 (*B*; Δex8) transcripts measured in the total RNA purified from blood of healthy NSDTRs genotyped for SNP cfa25:45,444,053 are shown. Twenty-three dogs homozygous for the G allele, 46 heterozygotes for the G and T alleles, and 43 homozygotes for the T allele were analyzed. Red boxes (risk), black boxes (heterozygotes), and blue boxes (nonrisk) represent the 25–75% interquartile range with the median, and the 10–90 percentile range with maximum and minimum values. The gene expression was normalized to the levels of the housekeeping gene TBP and analyzed using one-way ANOVA. The full-length transcript is down-regulated in the risk, whereas the Δ exon 8 transcript is up-regulated.

Methods

Samples: General. All blood samples in the study were collected from companion dogs in North America, Sweden, or Norway. Samples for genetic mapping were collected by primary care or specialist veterinarians and sent to the University of Missouri or the Broad Institute, obtained from dogs brought to the University of Missouri for euthanasia and necropsy or collected in the Canine Health Information Center DNA Bank (offa.org/chicdnabank.html). DNA was extracted from whole blood or buccal swabs as previously described (23). A presumptive diagnosis of DM was based on clinical signs, and the diagnosis was confirmed by histopathology showing a characteristic pattern of axonal degeneration, myelin loss, and gliosis in the thoracic spinal cord (16). All spinal cord tissues were examined, with phenotype blinded, by the same board-certified veterinary pathologist (G.C.J.). Sample collection protocols were approved by the University of Missouri Animal Care and Use Committee (protocols 6054 and 7349), by MIT (MIT 0910-074-13), and by the Ethical Board for Experimental Animals in Uppsala, Sweden (Dnr C138/6 and C417/12). Homozygosity for the *SOD1* risk allele was determined by pyrosequencing or TaqMan allelic discrimination as previously described (17, 23).

Samples: PWC. The samples used for mapping the genetic modifier in the discovery phase consisted of 15 DM-affected and 35 unaffected PWCs homozygous for the *SOD1* E40K risk allele. The affected dogs were selected to have onset of disease signs at the age of 9 y or younger. Among the affected dogs, 14 were confirmed through histopathology of the spinal cord and one dog was diagnosed through MRI ruling out other causes of the clinical signs. The unaffected dogs were free of disease signs at the ages of 11–15 y.

To replicate the findings, we studied an additional set of 32 affected and 13 unaffected PWCs homozygous for the *SOD1* risk allele. In the dogs with DM, diagnosis was confirmed by histopathology ($n = 28$), or presumptively based on clinical signs and MRI of the spinal cord ($n = 2$) or myelography ($n = 2$). The age of onset ranged from 9 to 15 y. Of the additional controls, eight were older than 11 y of age and five were reported as "older dogs without signs of DM," but exact age information was not available.

To obtain a population estimate of haplotype frequencies, 273 additional PWC dogs of varying age were genotyped. Most of these (263 of 273) dogs were homozygous for the *SOD1* risk allele, but there was no information regarding DM status.

Samples: Boxer. To examine the modifier locus in the Boxer breed, we evaluated haplotype frequencies in 15 Boxers with DM confirmed through histopathology and 10 Boxers without DM at 11 y of age, all homozygous for the *SOD1* E40K risk allele.

Samples: Other Breeds. To examine the modifier locus in additional breeds, we constructed haplotypes in a panel of 265 dogs from 85 dog breeds. Among these samples were 36 dogs that carried the *SOD1* risk allele and had a confirmed diagnosis of DM: 8 Bernese Mountain Dogs, 11 Chesapeake Bay Retrievers, 8 German Shepherd Dogs, and 9 Rhodesian Ridgebacks. No phenotypic information was available for the rest of the dogs. For gene expression studies, blood samples collected from 112 healthy NSDTRs living in Sweden or Norway were used.

GWA Analysis. Samples were genotyped using the Illumina CanineHD Genotyping BeadChip containing more than 170,000 SNP markers. The dataset was filtered for call rate in SNPs (98%) and individuals (95%), deviation from Hardy–Weinberg equilibrium in controls at $P < 1 \times 10^{-6}$, and nonvarying SNPs (minor allele frequency < 0.01). Because all samples were selected to carry two copies of the *SOD1* risk allele, this dataset was affected by some population structure. We used several strategies to control for population structure. First, to identify and remove genetic outliers in the population, multidimensional scaling plots were constructed using PLINK (54). Second, relatedness was assessed by GCTA (55), estimating the genetic relationship between all pairs of individuals in the dataset. For each pair of dogs related at >0.25 (half-sibling level) with the concordant phenotype, one dog was removed from the dataset, leaving 15 cases and 31 controls for the analysis. In the final association analysis, population structure was further controlled by using a mixed model approach in Efficient Mixed-Model Association eXpedited (EMMAX) (56) with the first two principal components calculated by GCTA as covariates. The threshold for genome-wide significance was assessed in several ways: by Bonferroni correction accounting for 170,000 markers, which defines $P < 2.9 \times 10^{-7}$ as genome-wide significant, and by plotting the 95% confidence intervals based on the distribution of beta values (effect size estimates) obtained from EMMAX. To investigate patterns of LD in the region, Haploview (57) was used to calculate pairwise estimates of r^2 values between individual SNPs.

Variant Discovery by Whole-Genome Sequencing. Three PWCs homozygous for the *SOD1* risk allele were subjected to whole-genome resequencing. These dogs were selected based on phenotype; one reached the age of 14 y without signs of DM, and the other two were phenotypically affected and histologically confirmed to have DM with onset at the age of 9 y. For each sample, a 300-bp insert size library and a 400-bp insert size library were prepared according to Illumina standard protocols. Two insert sizes were used to ensure sufficient library complexity and minimize bias from library preparation. Paired-end 100-bp reads were generated using Illumina HiSeq. Library preparation and sequencing were carried out at the University of Missouri DNA Core Facility.

Fastq files were aligned to canFam2 using the Burrows–Wheeler Alignment (BWA) tool version 0.6.2 (58). Realignment, duplicate marking, quality recalibration, and variant calling were carried out using Genome Analysis Toolkit (GATK) version 1.4.5 (59) and Picard version 1.59 (broadinstitute.github.io/picard/). The whole-genome sequence data provided information on available variants for fine-mapping the region. The Integrative Genomics Viewer (IGV) (60) was used for visual inspection of called variants. Variants were selected based on their potential to affect function, giving priority to variants overlapping coding sequence and noncoding conserved elements, and also based on spacing to increase the resolution around the strongest SNPs in the GWAs.

Additional Genotyping. To fine-map the association signal in the PWC GWA samples, a total of 116 SNPs located in the cfa25:42,181,379–cfa25:46,659,998 region were genotyped using Sequenom iPLEX. These SNPs included 101 new SNPs selected from the sequencing data and 15 SNPs present in the GWAs. The assay was designed using MassARRAY Assay Design 3.1, and genotyping was performed at the Broad Institute's Genomics Platform according to the manufacturer's instructions. The iPLEX genotyping included all PWC GWA samples that had available DNA as well as some samples from the PWC replication cohort, the Boxer samples, and confirmed DM cases from other breeds.

Additional samples from the PWC replication cohort, the PWC population cohort, and the other breeds panel were genotyped for the five haplotype-defining SNPs by custom TaqMan SNP Genotyping Assays (Applied Biosystems). Assays were run on a LightCycler 480 (Roche Life Science) using TaqMan Universal PCR Master Mix, no AmpErase UNG, or TaqMan Genotyping Master Mix, and 45–67.5 ng of genomic DNA in a 5-μL reaction volume. Each sample was run in duplicate. End-point genotyping analysis was performed using LightCycler 480 software (version 1.5.0.39). Further details are available upon request. To ensure the consistency of the two genotyping methods, 42 samples were typed with both methods, yielding 206 of 207 concordant genotypes.

Further Association Analysis in the GWA Region Establishing the Risk Haplotype. The iPLEX data were filtered for low call rate (<90%) in SNPs or individuals and merged with the GWA data. BEAGLE version 3.3.2 (61) was used to impute missing data in the cfa25:40–50 Mb region for the total dataset using samples with both GWA data and iPLEX data as the reference database and allelic $R^2 \geq 0.8$ as the cutoff denoting successful imputation. Association was analyzed in the GWA dataset using the mixed model in EMMAX. LD (r^2) values between all variants of the associated region were obtained from Haploview. Variants within 1 Mb and in strong LD ($r^2 > 0.7$) with the top SNP were identified using PLINK LD clumping and used as input for haplotype phasing in PHASE version 2.1 (62, 63). For the final phasing of haplotypes, all PWCs ($n = 368$) with available data for the variants at positions cfa25:45,435,040, cfa25:45,437,568, cfa25:45,443,320, cfa25:45,445,891, and cfa25:45,447,628 were used, regardless of phenotype information and *SOD1* status. Differences in the frequency of affected and unaffected PWCs carrying at least one copy of the risk haplotype were assessed using the Fisher exact test, two-tailed in R.

Kaplan–Meier analysis was performed with the event defined as onset of DM signs and time to event defined as age of onset for individuals with an event. For individuals without an event, the age of ascertainment (the last time point when information regarding signs of DM was available) was used to indicate the time point for censoring. The analysis incorporated both the discovery and replication datasets (affected = 47, unaffected = 48). The analysis was performed using the *survival* package in R, and the difference between carriers and noncarriers of the risk haplotype was assessed using the log-rank test.

For samples from other breeds, genotype data for the variants at positions cfa25:45,435,040, cfa25:45,437,568, cfa25:45,443,320, cfa25:45,445,891, and

cfa25:45,447,628 were used as input for haplotype phasing in PHASE. The 25 Boxers with detailed phenotype information were analyzed separately from the other breeds that had fewer representatives.

Deep Sequencing of the Associated Haplotype. To detect all variants present on the associated haplotype, targeted sequencing of the haplotype region was attempted for 39 of the samples used in the GWA analysis. The region was enriched by long-range PCR; three overlapping primer pairs were designed to provide coverage of the region: fragment 1: cfa25: 45,433,855–45,440,186, amplicon size of 6,332 bp; fragment 2: cfa25: 45,439,220–45,444,832, amplicon size of 5,612 bp; and fragment 3: cfa25: 45,443,583–45,449,113, amplicon size of 5,531 bp. PCR was carried out using the Novagen KOD Hot Start DNA Polymerase kit, according to manufacturer's instructions, in a reaction volume of 25 μL using 0.4 μM of each primer and 50–100 ng of genomic DNA. Details regarding primer sequences and cycling conditions are available upon request. Post-PCR clean-up was performed using Agencourt AMPure XP Magnetic Beads (1.6×; Beckman Coulter).

Each fragment was amplified for each individual separately and analyzed on 0.8% agarose gels. Based on visual inspection, equimolar amounts of fragments 1–3 were pooled for each individual. Five hundred nanograms of the pooled DNA was fragmented by Covaris to generate 550-bp fragments and subjected to AMPure Beads clean-up (1.6×). End repair, 3′ adenylation, and adapter ligation steps were performed; each step was followed by AMPure beads clean-up. Illumina Compatible NEXTflex-96™ DNA Barcodes were added in the adapter ligation. To ensure enrichment of fragments with ligated adapters, 10 ng of barcoded DNA was amplified for seven cycles with NimbleGen TS-PCR Oligo 1 and TS-PCR Oligo 2. The amplification was followed by AMPure beads purification, and the resulting products were eluted in water. Amplification and library preparation were successful for 34 of 39 samples.

Forty nanograms of barcoded DNA from each individual was pooled and submitted for paired-end sequencing by Illumina MiSeq 2 × 250-bp reads at the SciLifeLab Uppsala SNP&SEQ Platform. Fastq files were aligned to canFam2 using BWA version 0.7.8-r455 (58). Realignment, duplicate marking, quality recalibration, and variant calling were carried out using GATK version 2.8.1 (59) and Picard version 1.92. The targeted sequence data provided genotype data for all variants detectable in the haplotype region in 34 samples from the GWA cohort. SEQscoring (64) was used to rate called variants by conservation, defined as overlap with constraint elements detected by the alignment of 29 eutherian mammals (65). SnpEff (66) was used to annotate coding variants. Prediction of the effect of amino acid substitutions on protein function was performed using SIFT (27), PolyPhen-2 (28), SNAP (29), and consensus classifier PredictSNP (30). Coordinate positions were transferred from the dog genome canFam2 to the human genome hg 19 (GRCh37) using LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver).

Association Analysis of All Variants on the Risk Haplotype. The variants from sequencing were merged with the GWA and fine-mapping data for the cfa25:40 to 50-Mb region, and BEAGLE version 3.3.2 (61) was used to impute missing data in the total dataset using samples with both GWA data and iPLEX data as the reference database and allelic $R^2 \geq 0.8$ as the cutoff denoting successful imputation. The imputed dataset was filtered for low call rate (<90%) in SNPs and merged with the full GWA dataset. Association was analyzed in the total dataset using the mixed model in EMMAX, as previously described.

Luciferase Reporter Assay. Two DNA fragments, one 251 bp long containing SNP cfa25:45,444,053 and the other 397 bp long with SNP cfa25:45,445,768, were amplified by PCR using genomic DNA obtained from dogs with known genotypes as templates and cloned in the pGL4.26 reporter vector (Promega) by EcoRV and HindIII sites. After sequence validation, the plasmids were purified using an EndoFree Plasmid Maxi Kit (Qiagen) and transfected into the Jurkat T-cell line as follows: 7 × 10⁵ cells were seeded in each well in the 24-well plates in the RPMI-1640 medium supplemented with L-glutamine and 10% (vol/vol) heat-inactivated bovine serum. Seven hundred fifty nanograms of the reporter plasmid or intact pGL4.26 vector and 50 ng of the pRL-TK (Promega) vector were mixed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol and added to each well. Thirty-six hours after transfection, cells were additionally stimulated for 12 h with 20 ng/mL PMA and 0.5 μM ionomycin, and then harvested and assayed for Firefly and Renilla luciferase activities with the Dual-Luciferase Reporter Assay System (Promega).

The experiment was repeated three times with four technical replicates for each plasmid. The unpaired Student's t test was used for statistical analysis.

RNA Extraction from NSDTR Blood and cDNA Synthesis. For gene expression studies, 112 healthy NSDTRs genotyped for the SNP cfa25:45,444,053 were used. Briefly, total RNA from blood drawn directly in Tempus Blood RNA tubes (Applied Biosystems) was purified using the Tempus Spin RNA Isolation Reagent Kit (Applied Biosystems) according to the manufacturer's instructions. cDNA synthesis was performed at 42 °C for 80 min using 2 μg of RNA, 5 μM oligo-dT primer, Moloney Murine Leukemia Virus Reverse Transcriptase (MuLV-RT), and RNase inhibitor in the buffer supplemented with 5 mM MgCl₂ and 1 mM dNTPs. All reagents were purchased from Applied Biosystems. The reaction was stopped by heat inactivation at 95 °C for 5 min, and cDNA was diluted to 15 ng/μL.

Quantitative RT-PCR. Before gene expression studies, the transcripts of the canine SP110 gene were annotated by PCR with primers matching different exons. Two major isoforms, the full-length and Δ exon 8, with in-frame deletion of exon 8, were detected in the canine blood cells. Further, the mRNA levels for the two major SP110 transcripts, the total SP110 transcripts, and the total SP140 gene transcripts were measured by quantitative real-time PCR using SYBR Green for signal detection. Gene-specific primers are shown in *SI Appendix*, Table S2. After initial denaturation at 95 °C for 5 min, 45 cycles (denaturation at 95 °C for 15 s, annealing at 62 °C for 15 s, and polymerization at 72 °C for 25 s) were carried out. PCR buffer was supplemented with 1.5 mM MgCl₂, 200 μM dNTPs (each), primers, SYBRGreen (Molecular Probes), 15 ng of cDNA, and 0.5 units of Platinum Taq polymerase (Invitrogen). Expression levels were normalized to the levels of the TBP gene. All experiments were run in triplicate. Correlation of gene expression with genotypes was performed using one-way ANOVA tests in PRISM 6 (GraphPad Software).

EMSA. EMSAs were performed to test for differential DNA binding between the nonrisk and risk alleles at three SNPs: cfa25:45,443,320, cfa25:45,444,053, and cfa25:45,445,768. Probes were designed as 31-bp duplex oligonucleotides with the SNP at the midpoint. The probes were labeled with 5′ biotin (Integrated DNA Technologies). Nuclear extract from the Jurkat cell line was purchased from Active Motif.

EMSAs were performed using the Thermo Scientific LightShift Chemiluminescent EMSA Kit, following the manufacturer's instructions with minor adjustments. Briefly, for both the nonrisk and risk probes for each SNP, three reactions were performed: (i) control with labeled probe and no nuclear extract; (ii) 5 μg of nuclear extract and 20 fmol of labeled probe; and (iii) 5 μg of nuclear extract, labeled probe, and a 200-fold molar excess of unlabeled probe as a competitor. The reaction mixes were incubated for 40 min at 4 °C and then run at room temperature on a prerun native 6% polyacrylamide gel (Invitrogen) at 100 V for 60 min. The DNA was then transferred to a nylon membrane using the XCell II Blot Module (Invitrogen), and UV cross-linked. Detection was performed using stabilized streptavidin-horseradish peroxidase conjugate, followed by a stable peroxide/Luminol/enhancer solution.

1. Kiernan MC, et al. (2011) Amyotrophic lateral sclerosis. *Lancet* 377(9769):942–955.
2. Lindblad-Toh K, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069):803–819.
3. Patterson DF (2000) Companion animal medicine in the age of medical genetics. *J Vet Intern Med* 14(1):1–9.
4. Rosen DR, et al. (1993) Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362(6415):59–62.
5. Abel O, et al. (2013) Development of a Smartphone App for a Genetics Website: The Amyotrophic Lateral Sclerosis Online Genetics Database (ALSoD). *JMIR Mhealth Uhealth* 1(2):e18.

6. Turner MR, et al. (2013) Controversies and priorities in amyotrophic lateral sclerosis. *Lancet Neurol* 12(3):310–322.

7. Andersen PM, Al-Chalabi A (2011) Clinical genetics of amyotrophic lateral sclerosis: What do we really know? *Nat Rev Neurol* 7(11):603–615.

8. Renton AE, Chiò A, Traynor BJ (2014) State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* 17(1):17–23.

9. DeJesus-Hernandez M, et al. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72(2):245–256.

10. Renton AE, et al.; ITALSGEN Consortium (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72(2):257–268.

11. Majounie E, et al.; Chromosome 9-ALS/FTD Consortium; French Research Network on FTLD/FTLD/ALS; ITALSGEN Consortium (2012) Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *Lancet Neurol* 11(4):323–330.

12. van Blitterswijk M, et al. (2012) Genetic overlap between apparently sporadic motor neuron diseases. *PLoS One* 7(11):e48983.

13. Cirulli ET, et al.; FALS Sequencing Consortium (2015) Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* 347(6229):1436–1441.

14. Averill DR, Jr (1973) Degenerative myelopathy in the aging German Shepherd dog: Clinical and pathologic findings. *J Am Vet Med Assoc* 162(12):1045–1051.

15. Coates JR, et al. (2007) Clinical characterization of a familial degenerative myelopathy in Pembroke Welsh Corgi dogs. *J Vet Intern Med* 21(6):1323–1331.

16. Coates JR, Wininger FA (2010) Canine degenerative myelopathy. *Vet Clin North Am Small Anim Pract* 40(5):929–950.

17. Awano T, et al. (2009) Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. *Proc Natl Acad Sci USA* 106(8):2794–2799.

18. March PA, et al. (2009) Degenerative myelopathy in 18 Pembroke Welsh Corgi dogs. *Vet Pathol* 46(2):241–250.

19. Ogawa M, et al. (2014) Neuronal loss and decreased GLT-1 expression observed in the spinal cord of Pembroke Welsh Corgi dogs with canine degenerative myelopathy. *Vet Pathol* 51(3):591–602.

20. Shelton GD, et al. (2012) Degenerative myelopathy associated with a missense mutation in the superoxide dismutase 1 (SOD1) gene progresses to peripheral neuropathy in Pembroke Welsh corgis and boxers. *J Neurol Sci* 318(1-2):55–64.

21. Morgan BR, Coates JR, Johnson GC, Bujnak AC, Katz ML (2013) Characterization of intercostal muscle pathology in canine degenerative myelopathy: A disease model for amyotrophic lateral sclerosis. *J Neurosci Res* 91(12):1639–1650.

22. Morgan BR, Coates JR, Johnson GC, Shelton GD, Katz ML (2014) Characterization of thoracic motor and sensory neurons and spinal nerve roots in canine degenerative myelopathy, a potential disease model of amyotrophic lateral sclerosis. *J Neurosci Res* 92(4):531–541.

23. Zeng R, et al. (2014) Breed distribution of SOD1 alleles previously associated with canine degenerative myelopathy. *J Vet Intern Med* 28(2):515–521.

24. Wininger FA, et al. (2011) Degenerative myelopathy in a Bernese Mountain Dog with a novel SOD1 missense mutation. *J Vet Intern Med* 25(5):1166–1170.

25. Crisp MJ, Beckett J, Coates JR, Miller TM (2013) Canine degenerative myelopathy: Biochemical characterization of superoxide dismutase 1 in the first naturally occurring non-human amyotrophic lateral sclerosis model. *Exp Neurol* 248:1–9.

26. Axelsson E, et al. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.

27. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4(7):1073–1081.

28. Adzhubei IA, et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7(4):248–249.

29. Bromberg Y, Rost B (2007) SNAP: Predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res* 35(11):3823–3835.

30. Bendl J, et al. (2014) PredictSNP: Robust and accurate consensus classifier for prediction of disease-related mutations. *PLOS Comput Biol* 10(1):e1003440.

31. Wu C, et al. (2009) BioGPS: An extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 10(11):R130.

32. Hoeppner MP, et al. (2014) An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. *PLoS One* 9(3):e91172.

33. Desmet F-O, et al. (2009) Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37(9):e67.

34. Reese MG, Eeckman FH, Kulp D, Haussler D (1997) Improved splice site detection in Genie. *J Comput Biol* 4(3):311–323.

35. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31(13):3568–3571.

36. Leslie R, O'Donnell CJ, Johnson AD (2014) GRASP: Analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* 30(12):i185–i194.

37. Chiò A, et al. (2009) A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. *Hum Mol Genet* 18(8):1524–1532.

38. Ahmeti KB, et al. (2013) Age of onset of amyotrophic lateral sclerosis is modulated by a locus on 1p34.1. *Neurobiol Aging* 34(1):357.e7–357.e19.

39. Ben-David E, Shifman S (2012) Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. *PLoS Genet* 8(3):e1002556.

40. Clement AM, et al. (2003) Wild-type nonneuronal cells extend survival of SOD1 mutant motor neurons in ALS mice. *Science* 302(5642):113–117.

41. Uhlén M, et al. (2015) Proteomics. Tissue-based map of the human proteome. *Science* 347(6220):1260419.

42. Bloch DB, et al. (2000) Sp110 localizes to the PML-Sp100 nuclear body and may function as a nuclear hormone receptor transcriptional coactivator. *Mol Cell Biol* 20(16):6138–6146.

43. Lallemand-Breitenbach V, de Thé H (2010) PML nuclear bodies. *Cold Spring Harb Perspect Biol* 2(5):a000661.

44. Roscioli T, et al. (2006) Mutations in the gene encoding the PML nuclear body protein Sp110 are associated with immunodeficiency and hepatic veno-occlusive disease. *Nat Genet* 38(6):620–622.

45. Wang T, Ong P, Roscioli T, Cliffe ST, Church JA (2012) Hepatic veno-occlusive disease with immunodeficiency (VODI): First reported case in the U.S. and identification of a unique mutation in Sp110. *Clin Immunol* 145(2):102–107.

46. Bloch DB, et al. (2012) Decreased IL-10 production by EBV-transformed B cells from patients with VODI: Implications for the pathogenesis of Crohn disease. *J Allergy Clin Immunol* 129(6):1678–1680.

47. Cliffe ST, et al. (2012) Clinical, molecular, and cellular immunologic findings in patients with SP110-associated veno-occlusive disease with immunodeficiency syndrome. *J Allergy Clin Immunol* 130(3):735.e6–742.e6.

48. Lee MN, et al. (2013) Identification of regulators of the innate immune response to cytosolic DNA and retroviral infection by an integrative approach. *Nat Immunol* 14(2):179–185.

49. Amit I, et al. (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* 326(5950):257–263.

50. Mankouri J, et al. (2010) Optineurin negatively regulates the induction of IFNbeta in response to RNA virus infection. *PLoS Pathog* 6(2):e1000778.

51. Appel SH, Beers DR, Henkel JS (2010) T cell-microglial dialogue in Parkinson's disease and amyotrophic lateral sclerosis: Are we listening? *Trends Immunol* 31(1):7–17.

52. de Castro E, et al. (2006) ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* 34(Web Server Issue):W362–W365.

53. Vinayagam A, et al. (2011) A directed protein interaction network for investigating intracellular signal transduction. *Sci Signal* 4(189):rs8.

54. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.

55. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* 88(1):76–82.

56. Kang HM, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42(4):348–354.

57. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263–265.

58. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.

59. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.

60. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.

61. Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 84(2):210–223.

62. Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73(5):1162–1169.

63. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68(4):978–989.

64. Truvé K, et al. (2011) SEQscoring: A tool to facilitate the interpretation of data generated with next generation sequencing technologies. *EMBnet Journal* 17(1):38–45.

65. Lindblad-Toh K, et al.; Broad Institute Sequencing Platform and Whole Genome Assembly Team; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team; Genome Institute at Washington University (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478(7370):476–482.

66. Cingolani P, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6(2):80–92.

67. Kent WJ, et al. (2002) The human genome browser at UCSC. *Genome Res* 12(6):996–1006.

68. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.