

## Supplementary Materials for

### **Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs**

Bridgett M. vonHoldt, Emily Shuldiner, Ilana Janowitz Koch, Rebecca Y. Kartzinel, Andrew Hogan, Lauren Brubaker, Shelby Wanser, Daniel Stahler, Clive D. L. Wynne, Elaine A. Ostrander, Janet S. Sinsheimer, Monique A. R. Udell

Published 19 July 2017, *Sci. Adv.* **3**, e1700398 (2017)

DOI: 10.1126/sciadv.1700398

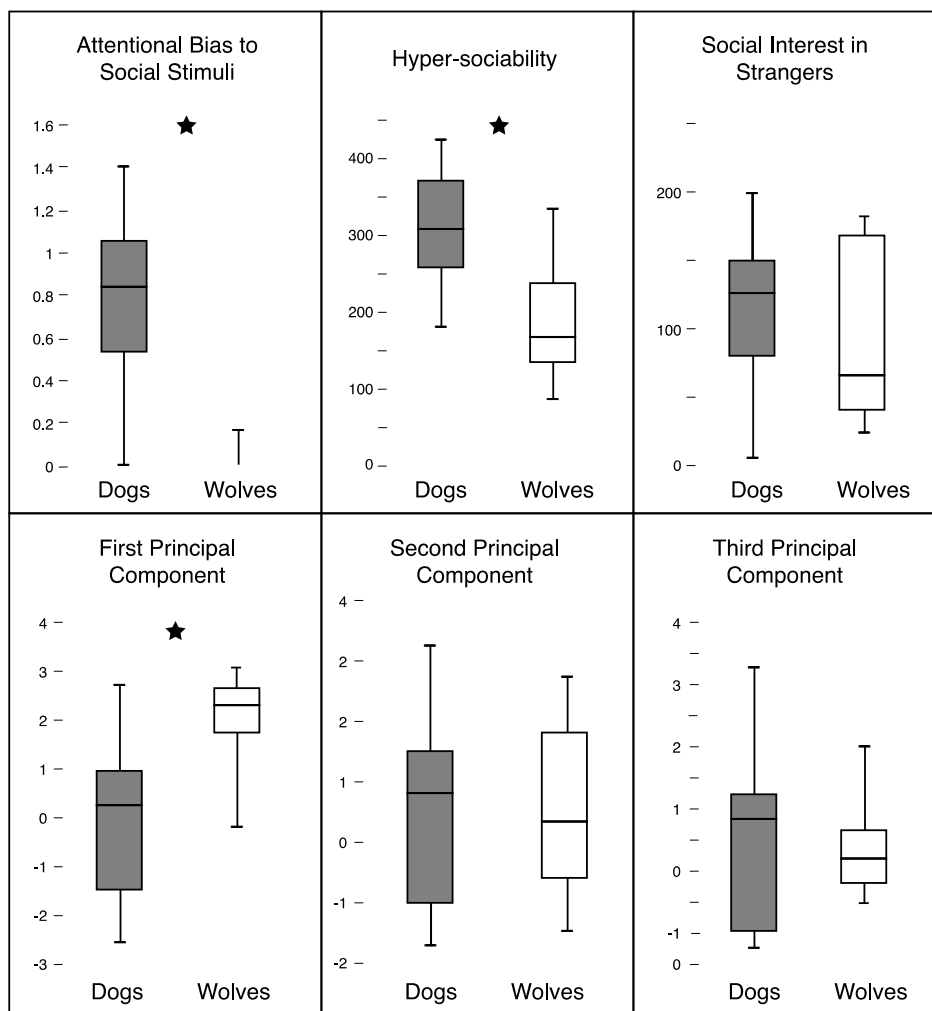
#### **The PDF file includes:**

- fig. S1. Differences between dogs and wolves for three behavioral indices used to predict the WBS phenotype.
- fig. S2. Scree plot of principal components of human-directed social behavior.
- fig. S3. Scan for positive selection using a bivariate percentile score (XP-EHH and  $F_{ST}$ ) to identify outliers (dashed line; bivariate score,  $>2$ ) indicated as sites in the 97.5th percentile.
- fig. S4. Gel electrophoresis banding patterns for three hypersociability-associated SV genotypes.
- fig. S5. A dot plot to represent the total number of insertions per population of species for each outlier locus.
- fig. S6. Plots from the ANOVA of the total number of SV insertions at four outlier loci.
- fig. S7. PCA from 25,510 unlinked genome-wide SNPs from the Affymetrix K9HDSNP array for six wolves and five dogs.
- fig. S8. SV discovery pipeline.
- fig. S9. Overlap in number of SVs identified by SVMerge, SoftSearch, and inGAP-sv.
- table S1. Raw behavioral data.
- table S2. Data for indices of human-directed social behavior.
- table S3. Loadings of first three PCs of human-directed social behavior.
- table S4. Genes in target region on canine chromosome 6.
- table S5. Sample information and the total number of raw reads compared to the number of processed reads after using cutadapt to trim/clip paired-end sequences.

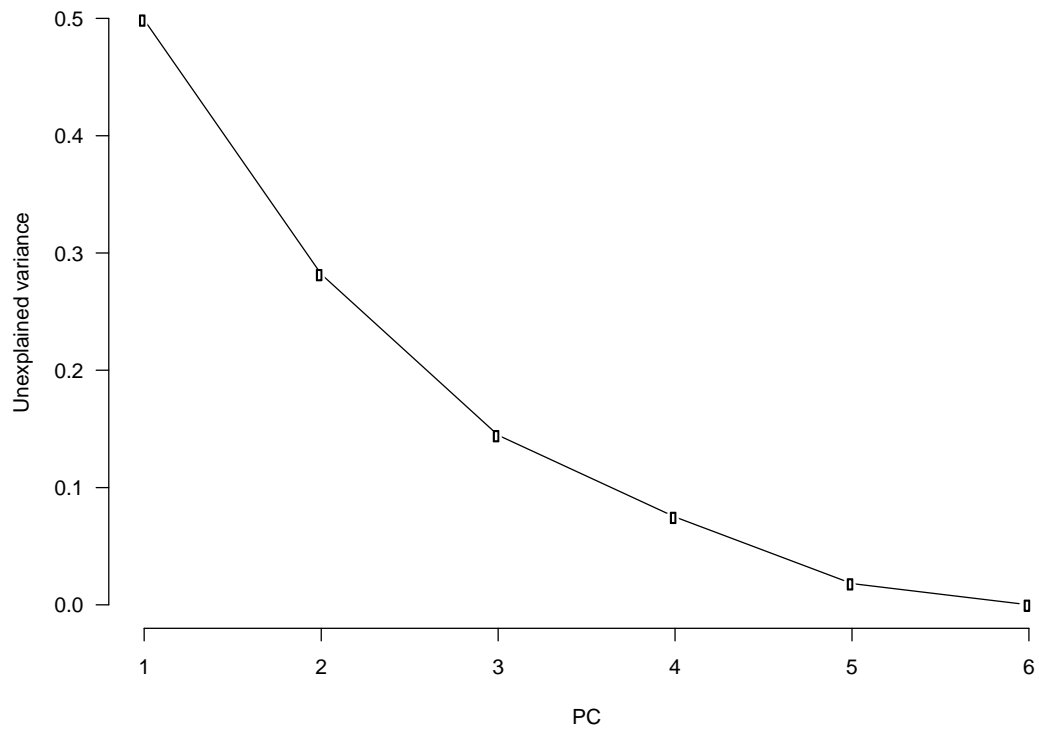
- table S6. Outlier clusters on chromosome 6 (CanFam3.1) showing signals of positive selection from XP-EHH.
- table S7. Summary of de novo annotated structural variants on canine chromosome 6.
- table S8. De novo annotated structural variants on canine chromosome 6 (coordinates based on CanFam3.1 assembly).
- table S9. Structural variant summary statistics per individual.
- table S10. Genic loci associated with indices of human-directed social behavior across dogs and wolves after inclusion of species as a covariate.
- table S11. Association to species membership.
- table S12. Predicted functional consequences of SVs.
- table S13. The significantly enriched (adjusted  $P < 0.05$ ) gene ontology term from a quantitative association test with each behavioral trait and 544,000 genome-wide SNPs.
- table S14. The significantly enriched (adjusted  $P < 0.05$ ) gene ontology term from the univariate regression analysis conducted in GEMMA with each behavioral trait and 544,000 genome-wide SNPs.
- table S15. Behavioral data and description relative to WBS.
- table S16. Parameters for in silico annotation of structural variants for the three methods SVMerge, SoftSearch, and InGAP-SV.
- table S17. Structural variant genotype per individual.
- table S18. Primer sequences used for PCR and gel-based validation of structural variants.
- Reference (99)

**Other Supplementary Material for this manuscript includes the following:**  
 (available at [advances.sciencemag.org/cgi/content/full/3/7/e1700398/DC1](https://advances.sciencemag.org/cgi/content/full/3/7/e1700398/DC1))

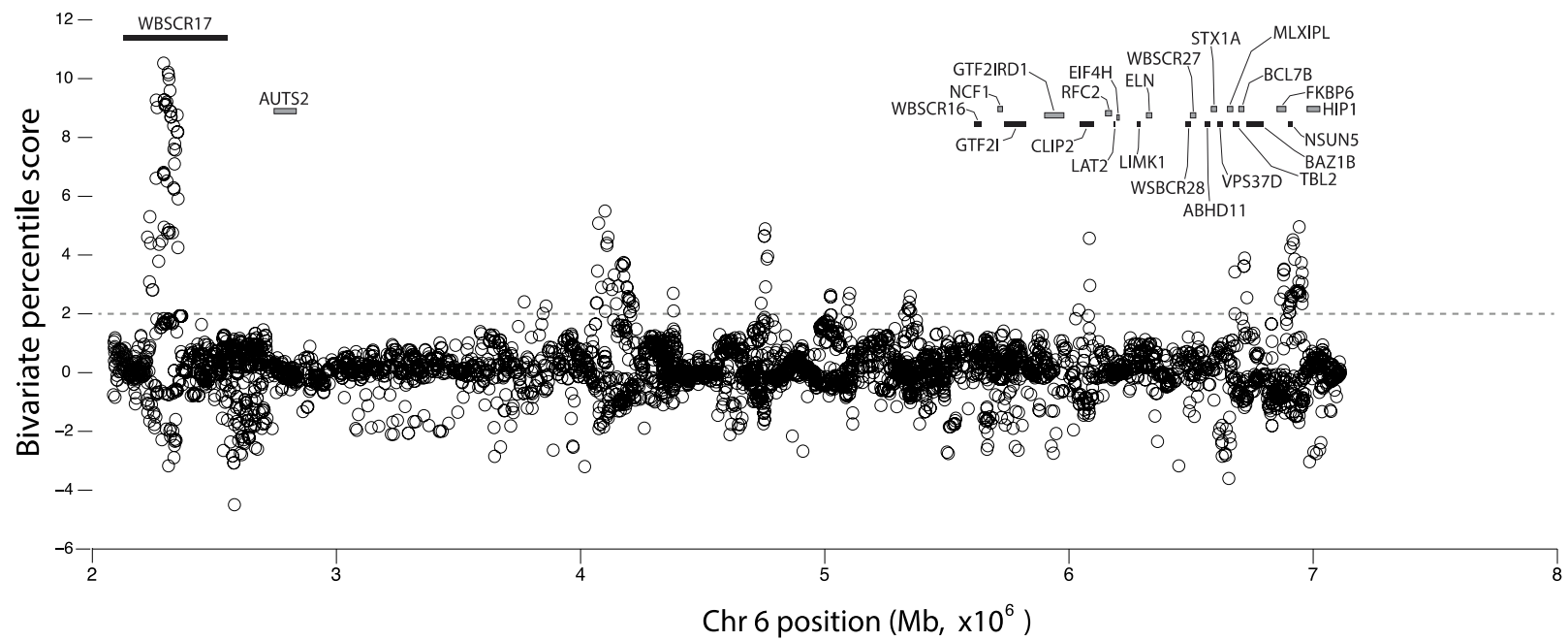
- data file S1 (Microsoft Excel format). Genotypes of the four outlier insertions in wild and domestic canids.



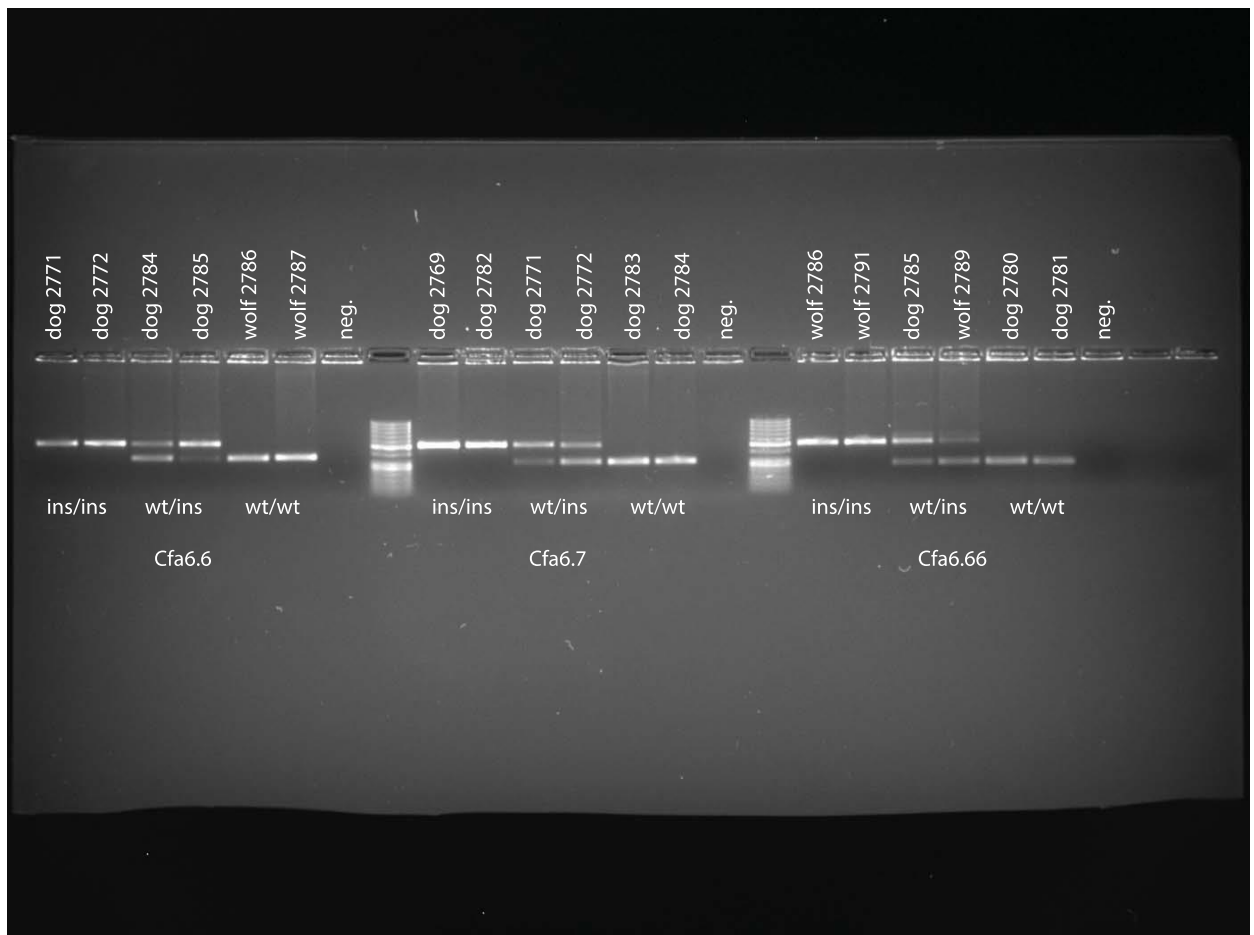
**fig. S1. Differences between dogs and wolves for three behavioral indices used to predict the WBS phenotype. Stars indicate pairwise significant differences ( $p < 0.05$ ).**



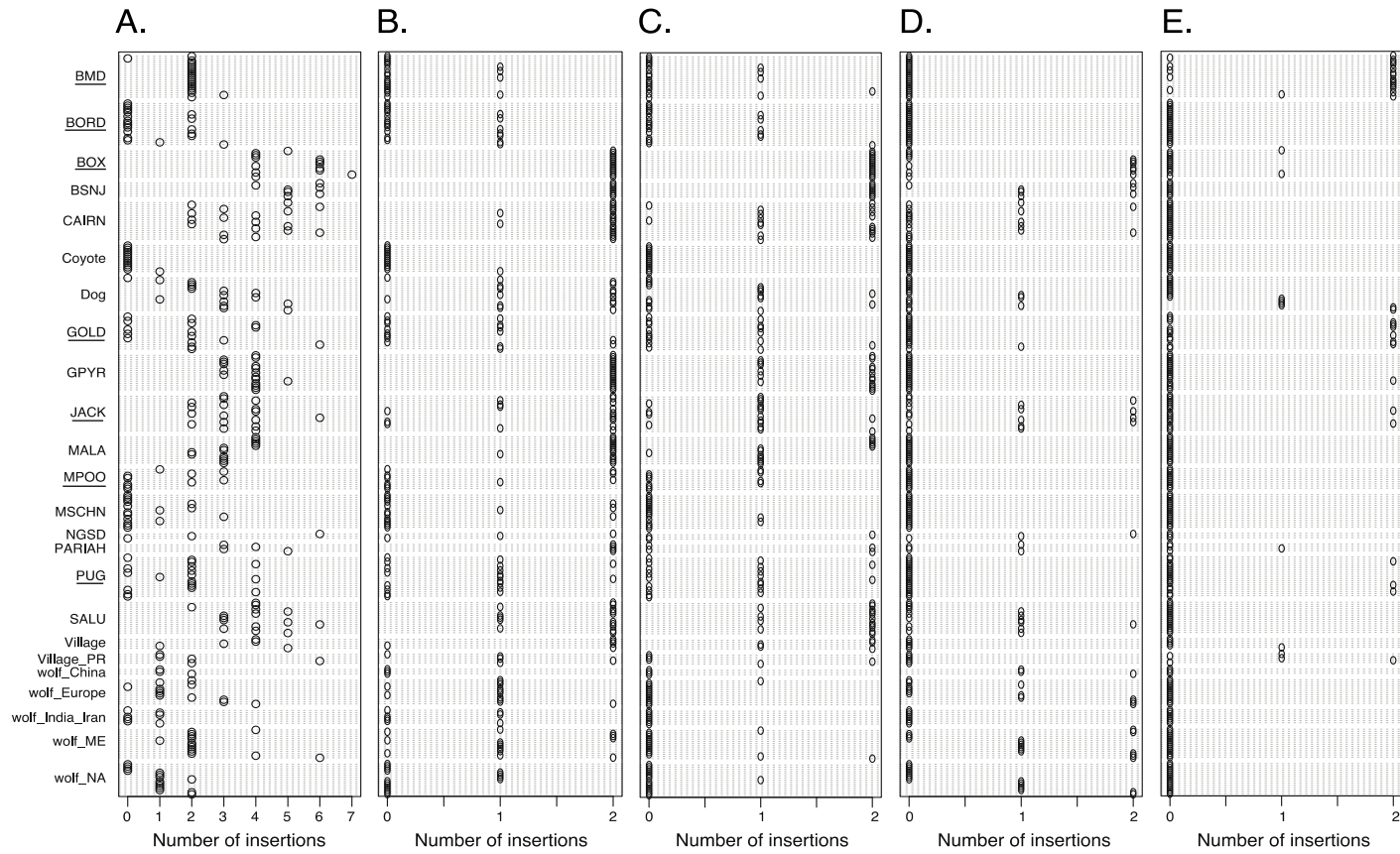
**fig. S2. Scree plot of principal components of human-directed social behavior.** Plot shows variance in original data set (table S4) explained by each PC.



**fig. S3. Scan for positive selection using a bivariate percentile score (XP-EHH and  $F_{ST}$ ) to identify outliers (dashed line; bivariate score,  $>2$ ) indicated as sites in the 97.5<sup>th</sup> percentile. Annotated genes are indicated above the plot as black and gray bars, labeled with gene names.**



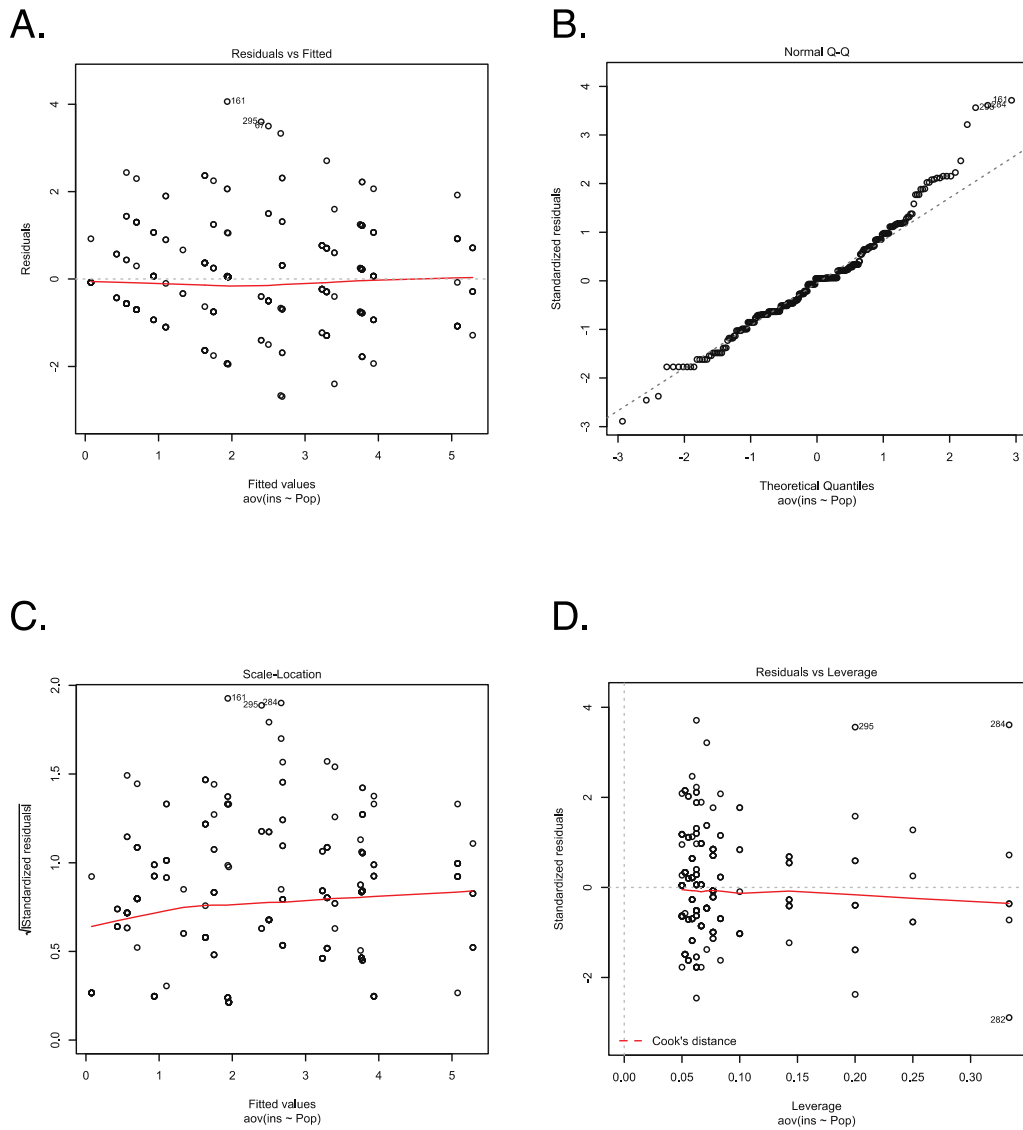
**fig. S4. Gel electrophoresis banding patterns for three hypersociability-associated SV genotypes.**



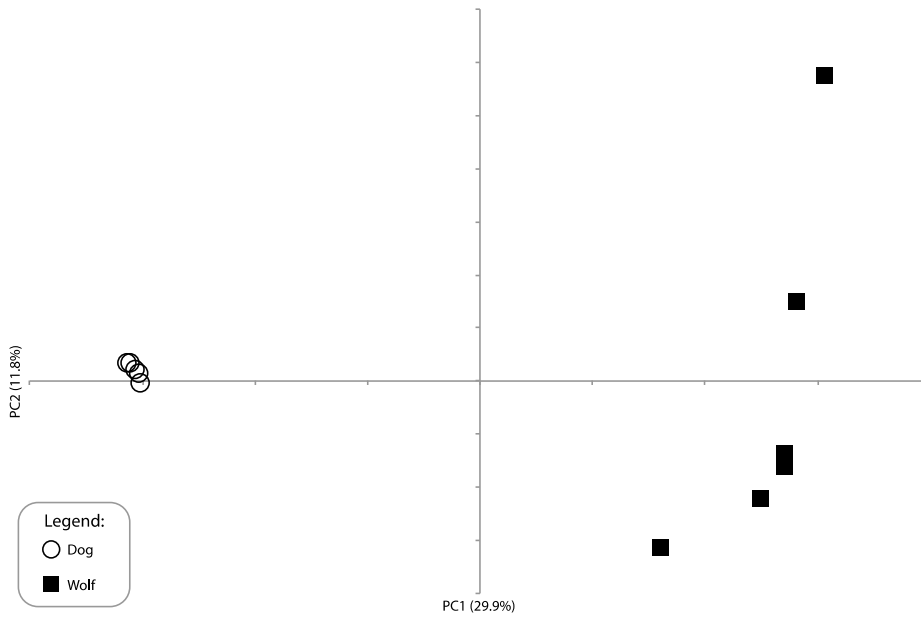
**fig. S5. A dot plot to represent the A) total number of insertions per population of species, and for each outlier locus. B) Cfa6.6, C) Cfa6.7, D) Cfa6.66, and E) Cfa6.83. Underlined breeds have the “seeks attention” behavioral stereotype. (Abbreviations: Bernese Mountain dog, BMD; Border collie, BORD; Boxer, BOX; Basenji, BSNJ; Cairn terrier, CAIRN; WBS study dogs, Dog; Golden retriever, GOLD; Great Pyrenees, GPYR; Jack Russell terrier, JACK; Alaska, malamute, MALA; Miniature poodle, MPOO;**

Miniature schnauzer, MSCHN; New Guinea singing dog, NGSD; Pariah dog, PARIAH, Saluki, SALU; Village dog, Village; Village dogs from Puerto Rico, Village\_PR; Middle East, ME; North America, NA).

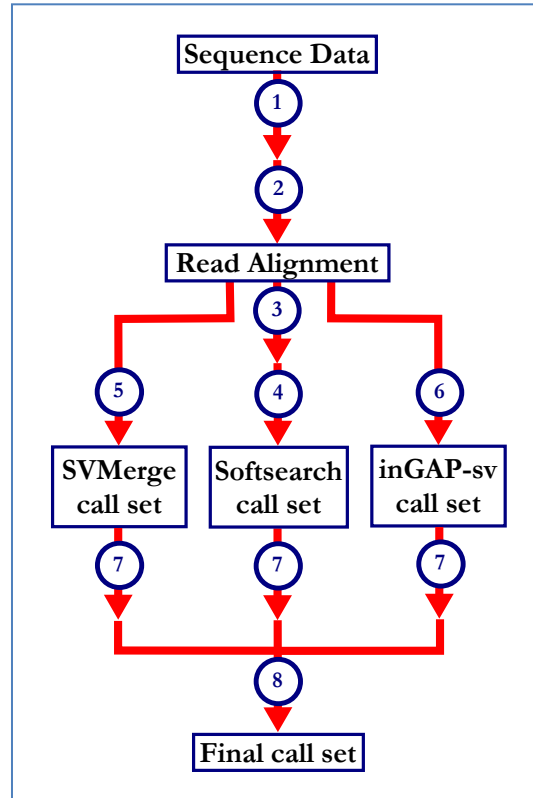




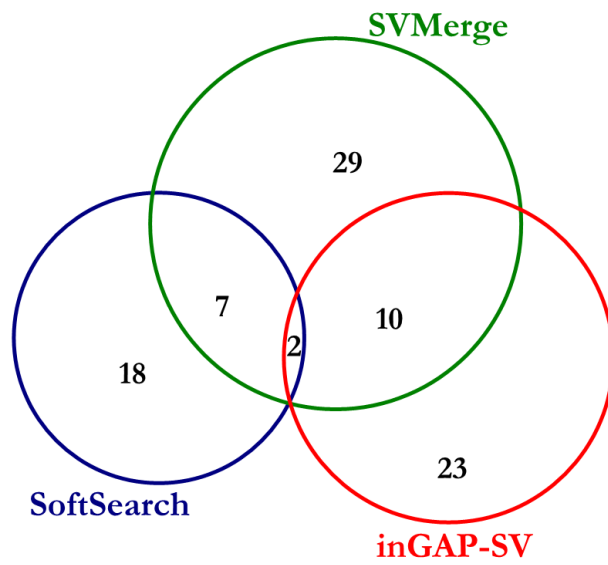
**fig. S6. Plots from the ANOVA of the total number of SV insertions at four outlier loci depend upon the population membership for A) residuals vs. fitted, B) Q-Q plot, C) scale location, and D) residuals vs. leverage.**



**fig. S7. PCA from 25,510 unlinked genome-wide SNPs from the Affymetrix K9HDSNP array for six wolves and five dogs.**



**fig. S8. SV discovery pipeline.** Numbers represent steps within the pipeline as follows:  
 1) Deplexing and quality control, 2) Alignment to reference, 3) Variant calling, 4) SoftSearch SV discovery, 5) SVMerge SV discovery, 6) inGAP-SV SV discovery, 7) Filtering of SV, and 8) Merging of filtered SVs.



**fig. S9. Overlap in number of SVs identified by SVMerge, SoftSearch, and inGAP-SV.**

**table S1. Raw behavioral data.** Dashed line separates dogs (above) from wolves (below).

Animal ID	ST-% time look box	ST-% time touch box	ST-% time look human	proximity unfamiliar passive (s)	proximity unfamiliar active (s)	proximity familiar passive (s)	proximity familiar active (s)
2768	15%	5%	4%	24.72	87.96	64.08	105.6
2769	18%	14%	25%	51.6	70.8	120	120
2770	8%	6%	17%	30	120	114	112.8
2771	9%	6%	11%	28.2	56.76	106.56	119.88
2772	4%	3%	1%	85.2	112.8	99.6	117.6
2773	69%	64%	14%	10.2	0	93.6	119.64
2774	100%	97%	0%	3.12	4.8	69.72	103.2
2775	11%	6%	4%	30	120	114	112.8
2776	4%	3%	33%	43.44	117.96	109.32	119.64
2777	5%	1%	13%	10.92	69.84	70.92	87.72
2778	5%	4%	12%	27.6	120	76.8	120
2779	9%	6%	56%	13.2	113.28	62.04	119.64
2780	10%	4%	32%	21.48	113.64	71.52	117.72
2781	20%	17%	15%	5.04	58.08	53.76	113.52
2782	25%	16%	34%	14.16	65.04	68.88	119.88
2783	19%	16%	8%	29.76	111.48	29.04	118.8
2784	18%	12%	7%	61.2	119.88	120	119.28
2785	3%	1%	85%	12	115.08	59.16	120
2786	35%	30%	1%	49.2	119.16	65.52	98.4
2787	100%	100%	0%	18.36	65.4	32.76	106.08
2788	90%	94%	0%	44.76	0	24.6	17.76
2789	97%	98%	0%	36.36	0	21.72	53.4

2790	99%	100%	0%	108.24	71.28	0	82.56
2791	83%	81%	0%	60	104.52	1.8	0
2792	100%	99%	0%	48.84	0	24.36	95.4
2793	100%	98%	0%	17.64	7.56	0.24	114.96
2794	100%	98%	0%	--	--	--	--
2795	100%	90%	0%	45.96	113.64	69.48	119.4

---

**table S2. Data for indices of human-directed social behavior.** Dashed lines separates dogs (above) from wolves (below).

<i>Animal ID</i>	<i>ABS</i>	<i>HYP</i>	<i>SIS</i>	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>
2769	0.864	362.40	122.4	-1.321	0.453	-0.764
2771	0.823	311.40	84.96	-1.041	-0.211	-1.030
2772	0.326	415.20	198	-0.944	2.244	-0.936
2773	0.424	223.44	10.2	0.214	-1.634	-1.214
2774	0.000	180.84	7.92	1.289	-1.681	-1.083
2775	0.548	376.80	150	-1.412	0.645	-0.802
2776	1.246	390.36	161.4	-1.973	0.638	-0.001
2777	1.031	239.40	80.76	-0.520	-0.470	0.089
2778	0.995	344.40	147.6	-1.320	0.367	-0.106
2779	1.188	308.16	126.48	-1.920	-0.739	1.395
2780	1.067	324.36	135.12	-1.518	-0.156	0.563
2781	0.704	230.40	63.12	-0.423	-1.059	-0.032
2782	0.863	267.96	79.2	-0.983	-0.987	0.278
2783	0.585	289.08	141.24	-0.416	0.239	0.397
2784	0.538	420.36	181.08	-1.330	1.498	-0.984
2785	1.393	306.24	127.08	-2.545	-1.124	2.356
2786	0.167	332.28	168.36	-0.174	1.155	-0.108
2787	0.000	222.60	83.76	1.250	-0.689	-0.187
2788	0.000	87.12	44.76	3.086	0.020	0.532
2789	0.000	111.48	36.36	2.687	-0.496	0.131
2790	0.000	262.08	179.52	2.404	2.168	0.415
2791	0.000	166.32	164.52	2.690	1.662	1.815
2792	0.000	168.60	48.84	2.224	-0.400	-0.474
2793	0.000	140.40	25.2	1.995	-1.442	-0.249

**table S3. Loadings of first three PCs of human-directed social behavior.**

Behavior	PC1	PC2	PC3
Proportion of time look human	-0.386	-0.269	0.631
Proportion of time look object	0.536	-0.153	-0.066
Proximity unfamiliar passive	0.153	0.782	-0.061
Proximity unfamiliar active	-0.400	0.490	0.339
Proximity familiar passive	-0.444	0.084	-0.554
Proximity familiar active	-0.429	-0.213	-0.414

**table S4. Genes in target region on canine chromosome 6.** Positions are from canfam3.1 genome build.

Gene Start (bp)	Gene End (bp)	Gene Name	Reference
2132919	2563654	WBSCR17	19
2749188	2831960	AUTS2	19
5606042	5632471	WBSCR16	99
5700832	5719439	NCF1	99
5722967	5811965	GTF2I	99
5885985	5963867	GTF2IRD1	99
6028219	6090774	CLIP2	99
6136604	6159026	RFC2	99
6164647	6171316	LAT2	99
6180064	6192330	EIF4H	99
6264548	6285910	LIMK1	99
6304623	6321742	ELN	99
6472520	6474969	WBSCR28	99
6488348	6492871	WBSCR27	99
6493513	6494145	CLDN4	99
6534229	6535237	CLDN3	99
6550966	6556271	ABHD11	99
6574691	6581692	STX1A	99
6595294	6595974	DNAJC30	99
6602419	6606916	VPS37D	99
6633050	6652557	MLXIPL	99
6656046	6674950	TBL2	99
6680478	6701631	BCL7B	99
6709697	6778186	BAZ1B	99
6782774	6784558	FZD9	99
6836043	6868433	FKBP6	99
6887596	6899406	NSUN5	99



**table S5. Sample information and the total number of raw reads compared to the number of processed reads after using *cutadapt* to trim/clip paired-end sequences.** Average sequence coverage is for target region chromosome 6 (2,031,491-7,215,670 bp). (Abbreviations: female, F; North America, NA; male, M)

Sample ID	Species Membership	Breed	Age	Sex	No. of raw reads	No. of reads post-processing	Prop. of reads dropped	Mean library insert size (bp)	Mean coverage*
2769	Domestic dog	Mix	4	F	20127088	20066328	0.003	304	15.4
2771	Domestic dog	Mix	2	F	18501656	18448868	0.003	271	14.5
2772	Domestic dog	Russian Terrier	5	F	17305060	17251544	0.003	304	12.6
2773	Domestic dog	Dachshund	6	F	21494600	21434496	0.003	281	17.7
2774	Domestic dog	Weimaraner	6	M	20697840	20634212	0.003	256	15.6
2775	Domestic dog	Mix	6	M	19885276	19821756	0.003	275	16.2
2776	Domestic dog	Golden Retriever	10	F	19910172	19847944	0.003	279	14.3
2777	Domestic dog	Labrador Retriever	3	M	30074496	29994680	0.003	259	17.2
2778	Domestic dog	Mix	2	M	24926916	24854488	0.003	274	18.5
2779	Domestic dog	Mix	1	M	19283016	19227644	0.003	282	13.8
2780	Domestic dog	Mix	11	F	18985824	18928344	0.003	272	12.2
2781	Domestic dog	Mix	6	M	22208112	22140900	0.003	269	18.4
2782	Domestic dog	Saluki	2	M	20127088	20066328	0.003	304	15.4
2783	Domestic dog	Mix	3	M	18501656	18448868	0.003	271	14.5
2784	Domestic dog	Mix	2	M	18294764	18238376	0.003	261	11.7
2785	Domestic dog	Mix	4	F	19086720	19034004	0.003	261	15.1
2786	Gray wolf	NA	8	F	19333428	19275892	0.003	264	13.9
2787	Gray wolf	NA	3	M	21307104	21243032	0.003	262	16.0
2788	Gray wolf	NA	7	M	20928148	20866492	0.003	263	13.8
2789	Gray wolf	NA	3	F	22880760	22818112	0.003	270	17.5
2790	Gray wolf	NA	14	F	19837444	19779788	0.003	264	16.6

2791	Gray wolf	NA	8	F	20472512	20415648	0.003	276	14.5
2792	Gray wolf	NA	2	F	23722756	23652336	0.003	267	18.2
2793	Gray wolf	NA	3	M	21855032	21776192	0.004	258	17.3

---

\*After PCR duplications were removed.

**table S6. Outlier clusters on chromosome 6 (CanFam3.1) showing signals of positive selection from XP-EHH.** Abbreviations: bivariate percentile score, BPS; observed heterozygosity,  $H_O$ ).  $P$ -values from a 1-tailed t-test of unequal variance are provided in parentheses.

Cluster ID	Start	Stop	Median BVS	No. outlier SNPs	Average $H_O$ in dog/wolf (t test $p$ -value)	Signal of selection in genome of:	Genes
6.1	2,226,371	2,352,136	7.34	54	0.01/0.37 ( $7.4 \times 10^{-38}$ )	Dog	<i>WBSCR17</i>
6.2	3,769,529	3,858,530	2.27	3	0.40/0.00 ( $1.4 \times 10^{-3}$ )	Wolf	
6.3	4,064,791	4,215,649	2.89	36	0.24/0.00 ( $5.8 \times 10^{-6}$ )	Wolf	
6.4	4,739,462	4,766,313	3.96	7	0.11/0.25 ( $1.3 \times 10^{-2}$ )	Dog	
6.5	5,023,335	5,102,019	2.58	8	0.04/0.47 ( $3.4 \times 10^{-4}$ )	Dog	
6.6	5,341,689	5,351,682	2.17	8	0.03/0.59 ( $7.9 \times 10^{-6}$ )	Dog	
6.7	5,351,682	6,085,558	2.96	3	0.25/0.0 (0.012)	Wolf	<i>CLIP2</i>
6.8	6,679,302	6,728,731	3.62	5	0.00/0.38 ( $7.4 \times 10^{-8}$ )	Dog	<i>BAZ1B</i>
6.9	6,866,332	6,955,315	2.74	35	0.43/0.10 ( $7.4 \times 10^{-8}$ )	Wolf	<i>FKBP6, NSUN5</i>

**table S7. Summary of de novo annotated structural variants on canine chromosome 6.**

Substituent SV		SVMerge	SoftSearch	InGAP-sv	Total
Detection Programs:					
Number of Structural Variants	Raw	120	126	112	358
	Post-Filtering	96	111	70	277
	Merged				89

**table S8. De novo annotated structural variants on canine chromosome 6**

(coordinates based on CanFam3.1 assembly). Abbreviations: D\_I, deletion-insertion; DEL, deletion; DUP, duplication; INS, insertion; INV, inversion; TRA, translocation; SV, structural variant; f, frequency.

Locus ID	Type	Start	Size (bp)	f(Dogs)	f(Wolves)	Gene
Cfa6.1	DEL	2,095,386	638,909	0.06	0.00	<i>WBSCR17</i> , <i>GLNT9</i> , <i>AUTS2</i>
Cfa6.2	INS	2,140,817	341	0.00	0.13	<i>WBSCR17</i>
Cfa6.3	INS	2,141,493	76	0.25	0.38	<i>WBSCR17</i>
Cfa6.4	INS	2,205,140	11	0.06	0.00	<i>WBSCR17</i>
Cfa6.5	DEL	2,432,140	1,800,34 2	0.06	0.00	<i>WBSCR17</i> , <i>AUTS2</i>
Cfa6.6	DEL	2,521,650	198	0.00	0.88	<i>WBSCR17</i>
Cfa6.7	DEL	2,546,359	235	0.06	0.38	<i>WBSCR17</i>
Cfa6.8	INS	2,583,455	58	0.00	0.88	
Cfa6.9	DEL	2,625,969	218	0.06	0.63	
Cfa6.10	INS	2,689,902	7	0.00	0.13	
Cfa6.11	DUP	2,734,984	2,347,33 4	0.00	0.13	<i>AUTS2</i>
Cfa6.12	DEL	3,009,985	627,053	0.00	0.13	<i>AUTS2</i>
Cfa6.13	DUP	3,010,060	626,977	0.00	0.13	<i>AUTS2</i>
Cfa6.14	DUP	3,010,100	2,121,75 1	0.00	0.13	<i>AUTS2</i>
Cfa6.15	DUP	3,012,589	2,239,16 7	0.00	0.13	<i>AUTS2</i>
Cfa6.16	DEL	3,018,553	687	0.06	0.00	<i>AUTS2</i>
Cfa6.17	DEL	3,209,048	1,279	0.25	0.00	<i>AUTS2</i>
Cfa6.18	DEL	3,241,300	694	0.06	0.00	<i>AUTS2</i>
Cfa6.19	DEL	3,452,567	715	0.06	0.00	<i>AUTS2</i>
Cfa6.20	DEL	3,452,997	294	0.31	0.25	<i>AUTS2</i>
Cfa6.21	INS	3,589,775	0	0.06	0.00	<i>AUTS2</i>
Cfa6.22	DUP	3,633,739	1,338,74 9	0.25	0.13	<i>AUTS2</i>
Cfa6.23	INS	3,875,594	351	0.06	0.00	<i>AUTS2</i>
Cfa6.24	DEL	3,976,412	223	0.31	0.00	
Cfa6.25	TRA	3,986,929	513	0.56	0.88	
Cfa6.26	DEL	3,986,940	536	0.88	0.63	
Cfa6.27	INS	4,194,480	22	0.00	0.38	
Cfa6.28	INS	4,231,965	3	0.31	0.13	
Cfa6.29	DEL	4,232,120	351	0.50	0.38	

Cfa6.30	INS	4,272,655	13	0.13	0.00	
Cfa6.31	DUP	4,312,149	638	0.13	0.00	
Cfa6.32	INS	4,358,450	368	0.81	0.75	
Cfa6.33	DUP	4,369,908	888	0.13	0.00	
Cfa6.34	DEL	4,370,055	448	0.00	0.13	
Cfa6.35	DEL	4,370,264	246	0.25	0.63	
Cfa6.36	DUP	4,377,190	859	0.13	0.00	
Cfa6.37	DEL	4,486,820	470	0.00	0.13	
Cfa6.38	DUP	4,514,366	1,067	0.13	0.00	
Cfa6.39	DEL	4,514,621	551	0.06	0.13	
Cfa6.40	TRA	4,514,643	573	0.31	0.50	
Cfa6.41	DEL	4,691,721	216	0.00	0.25	
Cfa6.42	INS	4,766,960	73	0.13	0.50	
Cfa6.43	INS	4,767,241	65	0.06	0.00	
Cfa6.44	INS	4,767,367	363	0.88	1.00	
Cfa6.45	DUP	4,792,086	646	0.13	0.00	
Cfa6.46	INV	4,839,392	3,328	0.00	0.13	
Cfa6.47	DEL	4,842,442	225	0.25	0.88	
Cfa6.48	DUP	4,858,013	622	0.13	0.00	
Cfa6.49	DUP	4,910,429	578	0.13	0.00	
Cfa6.50	INS	5,042,932	366	0.38	0.88	
Cfa6.51	DEL	5,065,830	14,796	0.06	0.00	
Cfa6.52	DEL	5,089,612	213	0.00	0.13	
Cfa6.53	DUP	5,213,911	810	0.13	0.00	
Cfa6.54	DEL	5,214,176	381	0.44	0.63	
Cfa6.55	DEL	5,277,159	2,053	0.00	0.25	
Cfa6.56	INS	5,318,593	30	0.31	0.50	
Cfa6.57	DEL	5,337,423	226	0.00	0.75	
Cfa6.58	INS	5,346,453	17	0.00	0.13	
Cfa6.59	INS	5,634,780	456	0.94	0.88	<i>CBX3</i>
Cfa6.60	D_I	5,646,194	118	0.44	0.38	
Cfa6.61	INS	5,646,321	294	0.13	0.00	
Cfa6.62	INS	5,646,326	21	0.06	0.00	
Cfa6.63	INS	5,646,624	3	0.00	0.13	
Cfa6.64	INS	5,652,383	4	0.06	0.13	
Cfa6.65	TRA	5,682,203	226	0.00	0.13	<i>GTF2IRD2</i>
Cfa6.66	DEL	5,753,703	290	0.69	0.13	<i>GTF2I</i>
Cfa6.67	TRA	5,753,734	265	0.13	0.00	<i>GTF2I</i>
Cfa6.68	DEL	5,820,166	231	0.38	0.63	<i>GTF2I</i>
Cfa6.69	INS	5,844,759	49	0.31	0.50	
Cfa6.70	INS	5,845,117	3	0.00	0.13	
Cfa6.71	INS	5,859,196	3	0.06	0.00	
Cfa6.72	DEL	5,902,332	715	0.19	0.13	<i>GTF2IRD1</i>

Cfa6.73	DEL	6,016,951	254	0.06	0.00	
Cfa6.74	TRA	6,016,969	207	0.06	0.00	
Cfa6.75	INS	6,289,609	345	0.13	0.25	
Cfa6.76	DEL	6,399,238	266	0.00	0.13	
Cfa6.77	DEL	6,400,822	216	0.31	0.75	
Cfa6.78	DEL	6,522,289	234	0.25	0.38	
Cfa6.79	DEL	6,718,167	263	0.25	0.38	<i>BAZIB</i>
Cfa6.80	DEL	6,718,220	10,206	0.00	0.13	<i>BAZIB</i>
Cfa6.81	INS	6,795,640	0	0.06	0.00	
Cfa6.82	INS	6,889,833	10	0.00	0.38	<i>NSUN5</i>
Cfa6.83	DEL	6,914,106	222	0.19	0.50	<i>POM121</i>
Cfa6.84	TRA	6,947,879	260	0.06	0.00	
Cfa6.85	DEL	6,947,889	247	0.00	0.13	
Cfa6.86	INS	7,156,514	336	0.25	0.13	
Cfa6.87	DEL	7,194,197	174	0.00	0.13	
Cfa6.88	DEL	7,378,268	479	0.13	0.00	<i>STYXLI</i>
Cfa6.89	INS	7,378,904	2	0.06	0.00	<i>STYXLI</i>

---

**table S9. Structural variant summary statistics per individual.**

Sample ID	Species Membership	# SVs	# Nucleotides Affected by SVs	% Target Region Affected by SVs
2769	Domestic Dog	20	9,853	0.19%
2771	Domestic Dog	10	2,949	0.06%
2772	Domestic Dog	9	1,339,332	25.83%
2773	Domestic Dog	16	3,287	0.06%
2774	Domestic Dog	18	641,634	12.38%
2775	Domestic Dog	15	3,330	0.06%
2776	Domestic Dog	13	2,939	0.06%
2777	Domestic Dog	14	4,264	0.08%
2778	Domestic Dog	16	1,339,843	25.84%
2779	Domestic Dog	12	3128	0.06%
2780	Domestic Dog	7	1,801,588	34.75%
2781	Domestic Dog	22	1341938	25.89%
2782	Domestic Dog	20	9,853	0.19%
2783	Domestic Dog	9	2,709	0.05%
2784	Domestic Dog	13	4,178	0.08%
2785	Domestic Dog	18	1,356,209	26.16%
Average Across Dogs		15	491,690	9.48%
2786	Gray Wolf	17	3,089	0.06%
2787	Gray Wolf	19	633,662	12.22%
2788	Gray Wolf	9	1,349,796	26.04%
2789	Gray Wolf	31	6,642	0.13%
2790	Gray Wolf	27	2,351,731	45.36%
2791	Gray Wolf	20	2,240,258	43.21%
2792	Gray Wolf	23	2,124,153	40.97%
2793	Gray Wolf	25	6,491	0.13%
Average Across Wolves		21	1,089,478	21.02%
Average Across all Animals		17	690,952	13.33%



**table S10. Genic loci associated with indices of human-directed social behavior across dogs and wolves after inclusion of species as a covariate.**

Phenotype	Locus ID	SV Type	Position (Mb)	$\beta$ (se)	% variation explained	<i>p</i> -value	Candidate Gene
ABS	Cfa6.66	Deletion	5.75	0.23 (0.091)	5.76	$2.33 \times 10^{-4}$	<i>GTF2I</i>
	Cfa6.7	Deletion	2.54	0.10 (0.081)	0.58	$9.56 \times 10^{-4}$	<i>WBSCR17</i>
	Cfa6.3	Insertion	2.14	0.081 (0.076)	0.50	$1.06 \times 10^{-3}$	<i>WBSCR17</i>
SIS	Cfa6.66	Deletion	5.75	-9.7 (32)	1.80	$1.67 \times 10^{-3}$	<i>GTF2I</i>

**table S11. Association to species membership.**

Locus ID	$\chi^2$	<i>p</i> -value	Odds Ratio
Cfa6.3	0.3345	0.563	1.615
Cfa6.6	16.39	0.00005155	NA
Cfa6.7	3.409	0.06484	7.154
Cfa6.8	16.39	0.00005155	NA
Cfa6.9	7.714	0.005479	14.09
Cfa6.17	2.182	0.1396	0
Cfa6.20	0.08362	0.7724	0.7714
Cfa6.22	0.4465	0.504	0.4667
Cfa6.24	2.791	0.09481	0
Cfa6.25	1.172	0.279	1.988
Cfa6.26	0.6969	0.4038	0.5844
Cfa6.27	6.4	0.01141	NA
Cfa6.28	0.8571	0.3545	0.36
Cfa6.29	0.2359	0.6272	0.6923
Cfa6.32	0.04356	0.8347	0.8769
Cfa6.35	2.462	0.1167	3.182
Cfa6.40	0.6154	0.4328	1.8
Cfa6.42	3.429	0.06408	5
Cfa6.44	0.1678	0.682	1.286
Cfa6.47	5.897	0.01517	5.444
Cfa6.50	3.376	0.06616	3.37
Cfa6.54	0.5	0.4795	1.623
Cfa6.56	0.6154	0.4328	1.8
Cfa6.57	13.71	0.0002128	NA
Cfa6.59	0.04196	0.8377	0.8815
Cfa6.60	0.06316	0.8016	0.8242
Cfa6.66	4.5	0.03389	0.1273
Cfa6.68	0.9435	0.3314	1.97
Cfa6.69	0.6154	0.4328	1.8
Cfa6.72	0.1364	0.7119	0.6444
Cfa6.75	0.5455	0.4602	2.143
Cfa6.77	2.889	0.08916	3.24
Cfa6.78	0.3345	0.563	1.615
Cfa6.79	0.3345	0.563	1.615
Cfa6.82	6.4	0.01141	NA
Cfa6.83	2.091	0.1482	3.222
Cfa6.86	0.4465	0.504	0.4667

**table S12. Predicted functional consequences of SVs.** Consequences predicted using Ensembl Variant Effect Predictor.

Consequence	Description	IMPACT rating	# of SVs
Transcript ablation	Deleted region includes a transcript feature	High	3
Start lost	Changes at least one base of start codon	High	7
Transcript amplification	Amplification of region containing a transcript	High	5
Coding sequence variant	Changes the coding sequence of a gene	Modifier	10
Feature truncation	Reduces genomic feature relative to reference	Modifier	16
Feature elongation	Located within a regulatory region	Modifier	12
Non-coding transcript variant	Transcript variant of a non-coding RNA gene	Modifier	3
Intronic variant	Located in an intron	Modifier	32
5' UTR variant	Located in 5' UTR	Modifier	8
3' UTR variant	Located in 3' UTR	Modifier	1
Upstream variant	Located 5' of a gene	Modifier	9
Downstream variant	Located 3' of a gene	Modifier	4
Intergenic variant	Located >5 kb from a gene	Modifier	40

**table S13. The significantly enriched (adjusted  $P < 0.05$ ) gene ontology term from a quantitative association test with each behavioral trait and 544,000 genome-wide SNPs.** Abbreviations: biological process, BP; number of reference genes in the category, C; expected number in category, E; molecular function, MF; number of genes in the gene set and category, O; ratio of enrichment, R.

Behavioral trait	Database	Name	C	O	E	R	$p$ -value	Adjusted $p$ -value
ABS	MF	Phosphoric ester hydrolase activity	203	10	2.58	3.87	0.0003	0.0483
HYP	BP	Cell development	899	208	136	1.53	$1.47^{-11}$	$5.61^{-8}$
	BP	Generation of neurons	575	144	87	1.66	$9.84^{-11}$	$1.25^{-7}$
	BP	Cellular developmental process	1679	342	254	1.36	$8.42^{-11}$	$1.27^{-7}$
	BP	System development	2098	410	317.4	1.29	$2.27^{-10}$	$1.44^{-7}$
	BP	Cell differentiation	1556	319	235.4	1.36	$2.16^{-10}$	$1.44^{-7}$
	BP	Anatomical structure development	2445	467	369.9	1.26	$2.22^{-10}$	$1.44^{-7}$
	BP	Neuron differentiation	506	129	77	1.68	$4.46^{-10}$	$2.13^{-7}$
	BP	Nervous system development	892	201	134.9	1.49	$4.14^{-10}$	$2.13^{-7}$
	BP	Neurogenesis	626	151	94.7	1.59	$6.25^{-10}$	$2.65^{-7}$
	BP	Multicellular organismal development	2298	439	347.6	1.26	$1.11^{-9}$	$4.24^{-7}$
	MF	GTPase regulatory activity	167	51	25.14	2.03	$2.49^{-7}$	$6.71^{-5}$
	MF	Small GTPase regulatory activity	123	41	18.5	2.21	$2.83^{-7}$	$6.71^{-5}$
	MF	Protein binding	5688	948	856.4	1.11	$1.10^{-7}$	$6.71^{-5}$
	MF	Phosphoric ester hydrolase activity	203	58	30.6	1.9	$4.78^{-7}$	$8.5^{-5}$
	MF	Nucleoside-triphosphatase regulatory activity	172	51	25.9	1.97	$6.85^{-7}$	$9.74^{-5}$
	MF	Guanyl-nucleotide exchange factor activity	77	27	11.6	2.33	$1.04^{-5}$	0.0009
MF	Ion channel activity	242	62	36.43	1.7	$1.05^{-5}$	0.0009	

	MF	Phospholipid binding	225	59	33.9	1.74	$7.92^{-6}$	0.0009
	MF	Substrate-specific channel activity	246	62	37	1.67	$1.81^{-5}$	0.0013
	MF	Binding	7867	1244	1184.4	1.05	$1.86^{-5}$	0.0013
	CC	Synapse	180	60	27.1	2.21	$5.04^{-10}$	$2.22^{-7}$
	CC	Cell periphery	1588	314	239.1	1.31	$1.32^{-8}$	$1.94^{-6}$
	CC	Cell projection	569	135	85.7	1.58	$1.27^{-8}$	$1.94^{-6}$
	CC	Plasma membrane	1512	298	227.7	1.31	$5.01^{-8}$	$5.51^{-6}$
	CC	Neuron projection	246	68	37	1.84	$1.96^{-7}$	$1.72^{-5}$
	CC	Proteinaceous extracellular matrix	169	51	25.5	2	$3.72^{-7}$	$2.73^{-5}$
	CC	Axon	123	38	18.5	2.05	$6.09^{-6}$	0.0003
	CC	Extracellular matrix	243	63	36.7	1.72	$5.80^{-6}$	0.0003
	CC	Basement membrane	65	24	9.8	2.45	$1.17^{-5}$	0.0006
	CC	Dendrite	99	31	14.9	2.08	$3.19^{-5}$	0.0014
SIS	BP	Cell development	899	215	150.8	1.43	$4.72^{-9}$	$1.80^{-5}$
	BP	Cell adhesion	429	113	72	1.57	$1.99^{-7}$	0.0003
	BP	Biological adhesion	430	113	72.1	1.57	$2.27^{-7}$	0.0003
	BP	Transmission of nerve impulse	266	76	44.6	1.7	$7.74^{-7}$	0.0004
	BP	Multicellular organismal signaling	269	77	45.1	1.71	$6.04^{-7}$	0.0004
	BP	Single-organism behavior	229	68	38.4	1.77	$6.44^{-7}$	0.0004
	BP	Nervous system development	892	203	149.6	1.36	$7.46^{-7}$	0.0004
	BP	Neurogenesis	626	147	105	1.4	$5.02^{-6}$	0.0021
	BP	Cellular developmental process	1679	345	281.7	1.22	$4.31^{-6}$	0.0021
	BP	Neuron differentiation	509	123	85.4	1.44	$7.32^{-6}$	0.0026
	MF	Protein binding	5688	1086	977.7	1.11	$3.16^{-9}$	$2.42^{-6}$
	MF	Binding	7867	1432	1352.3	1.06	$7.2^{-8}$	$2.76^{-5}$
	MF	Kinase activity	396	96	68	1.41	0.0002	0.0383
	MF	Peptide hormone binding	7	6	1.2	4.99	0.0002	0.0383
	MF	Protein tyrosine kinase activity	81	27	13.9	1.94	0.0003	0.0383

MF	Calcium-release channel activity	10	7	1.7	4.07	0.0003	0.0383
CC	Neuron projection	246	77	41.2	1.87	$8.93^{-9}$	$4.07^{-9}$
CC	Cell projection	569	141	95.3	1.48	$2.97^{-7}$	$6.77^{-5}$
CC	Synapse	180	57	30.1	1.89	$4.99^{-7}$	$7.58^{-5}$
CC	Basement membrane	65	27	10.9	2.48	$1.88^{-6}$	0.0002
CC	Proteinaceous extracellular matrix	169	51	28.3	1.8	$9.27^{-6}$	0.0008
CC	Axon	123	40	20.6	1.94	$1.22^{-5}$	0.0009
CC	Dendrite	99	33	16.6	1.99	$3.94^{-5}$	0.0026
CC	Cell periphery	1588	320	265.9	1.2	$5.25^{-5}$	0.003
CC	Extracellular matrix	243	62	40.7	1.52	0.0003	0.0152
CC	Plasma membrane	1512	299	253.2	1.18	0.0004	0.0182

---

**table S14. The significantly enriched (adjusted  $P < 0.05$ ) gene ontology term from the univariate regression analysis conducted in *GEMMA* with each behavioral trait and 544,000 genome-wide SNPs. HYP had no significant GO categories enriched.**

Abbreviations: biological process, BP; number of reference genes in the category, C; expected number in category, E; molecular function, MF; number of genes in the gene set and category, O; ratio of enrichment, R.

Behavioral trait	Database	Name	C	O	E	R	$p$ -value	Adjusted $p$ -value
ABS	BP	Regulation of neuron maturation	2	2	0.02	97.14	0.0001	0.0377
	BP	Negative regulation of neuron maturation	2	2	0.02	97.14	0.0001	0.0377
	CC	Synapse	180	8	1.79	4.47	0.0004	0.0400
SIS	CC	Cell periphery	1588	30	15.10	1.99	$8.88^{-5}$	0.0090
	CC	Plasma membrane	1512	28	14.38	1.95	0.0002	0.0101
	CC	Cell junction	309	10	2.94	3.40	0.0007	0.0236

**table S15. Behavioral data and description relative to WBS.**

Behavior	Quantified task and information on WBS	Calculation	Reference
Attentional bias towards social stimuli ( <i>ABS</i> )	<ul style="list-style-type: none"><li>• Higher proportion of time referencing familiar human</li><li>• Lower proportion of time looking at object</li><li>• Lower proportion of time physically contacting object</li></ul>	The ratio of the proportion of time spent looking at the experimenter to the sum of the proportion of time spent looking at the experimenter plus the proportion of time spent looking at the puzzle box in the solvable task test.	22
Hyper-sociability ( <i>HYP</i> )	<ul style="list-style-type: none"><li>• Higher proportion of time spent in proximity of familiar or unfamiliar human</li></ul>	Sum of the time spent in proximity to the experimenter in each phase of the sociability test.	22, 36
Social interest in strangers ( <i>SIS</i> )	<ul style="list-style-type: none"><li>• Higher proportion of time in proximity with unfamiliar human</li></ul>	Sum of the time spent in proximity to the experimenter in the two unfamiliar phases of the sociability test.	22, 36



**table S16. Parameters for in silico annotation of structural variants for the three methods SVMerge, SoftSearch, and InGAP-SV. A) SVMerge, B) SoftSearch, and C) InGAP-SV.**

**A.**

Parameter	Flag	Parameter Definition	Default Value	Value Used
BDconfParams	-c	Number of standard deviations away from mean insert size for read pair mapping to be considered discordant; Passed to BreakDancer	4	4
	-n	Number of observations required to estimate mean and standard deviation of insert size; Passed to BreakDancer	10000	10000
Bdparams	-c	Number of standard deviations away from mean insert size for read pair mapping to be considered discordant; Passed to BreakDancer	3	3
	-m	Maximum SV size callable; Passed to BreakDancer	5000000	5000000
BD copynum	-q	Minimum mapping quality used in SV determination; Passed to BreakDancer	25	25
		Ploidy of Organism; Passed to BreakDancer	2	2
PDoptParams	-x	Maximum SV size callable; Passed to Pindel. Note: 5 corresponds to 32,368bp	5	5
	-v	Minimum Inversion size callable; Passed to Pindel	1000	1000
SECqual		Minimum mapping quality used in SV determination; Passed to SECluster	20	25
SECmin		Minimum number of reads in either the forward or reverse cluster, when clusters are paired.	5	5
SECminCluster		Minimum number of reads to form a single-end forward or reverse cluster.	3	5
SECmax		Maximum number of reads allowed in a cluster.	500	500
Filtering Step	BDscore	Score cut-off for data from BreakDancer	25	25

	BDRs	Minimum number of supporting read pairs for data from BreakDancer	2	2
	PDscore	Score cut-off for data from Pindel	30	30
	PDsupports	Minimum number of supporting read pairs for data from Pindel	10	10
	Hashlen	Hash-length for assembly; Passed to Velvet	29	29
Library Insert Size		Average insert size	NA	See Table S2

**B.**

Parameter	Parameter Definition	Default Value	Value Used
q	Minimum mapping quality used in SV determination	20	25
l	Minimum length of soft-clipped segment used in SV determination	10	10
r	Minimum soft-clipped read depth used in SV determination	10	10
m	Minimum number of discordant read pairs to support soft clipped event	10	10
s	Number of standard deviations away from mean insert size for read pair mapping to be considered discordant	4	4
d	Minimum distance between soft-clipped segments and discordant read pairs	300	300

**C.**

Parameter	Parameter Definition	Default Value	Value Used
Min quality	Minimum mapping quality used in SV determination	10	25
Min PE support	Minimum number of discordantly mapped read pairs to support SV	4	4
Min SE support	Minimum number of singly-mapped read pairs to support SV	4	4
Max SV size	Maximum SV size callable	100000	1000000
X of std dev	Number of standard deviations away from mean insert size for read pair mapping to be considered discordant	3	3





**table S18. Primer sequences used for PCR and gel-based validation of structural variants.**

Locus	Primer sequence	Amplicon size (bp):	
		Insertion	Wildtype
Cfa6.6	Forward: CCCCTTCAGCCAGCATATAA Reverse: TTCTCTGGGCTGTCTGGACT	555	357
Cfa6.6 (internal)	Forward: AAGTTTCTCTGATGGAAAACACA Reverse: GGTGGCTGGAAATTTTCAGTAG	278	90
Cfa6.7	Forward: TGGAGCCATGATTAGGAAGG Reverse: TAAGGAAGGACCCCATTTCC	504	269
Cfa6.66	Forward: TGCTGCTTCATGTTCTGTGA Reverse: TGGTGCATTAGCTTTGGTTG	505	215
Cfa6.83	Forward: AACCACAGGAACAAAACCTCA Reverse: CTCCTGTTGGACATTTGGA	400	184