Advances in Genetic Mapping and Sequencing Techniques:

A Demonstration using the Domestic Dog Model

Oliver Forman BSc

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Glasgow

School of Veterinary Medicine

September 2013

Abstract

Over the past ten years huge advances have been made in the field of genetics and genomics. Genetic mapping has evolved from laborious linkage and homozygosity based approaches to high-throughput genome-wide association studies using whole genome SNP array technology. Through massively parallel sequencing technology, gigabases of sequencing data can now be produced in a single experiment. The domestic dog has been increasingly recognised as a model for human disease and mapping of inherited disease in the domestic dog is facilitated by fixed and genetically isolated populations.

The aims of this thesis were to demonstrate advances in mapping and sequencing techniques by investigating the genetics of five inherited disorders, representing significant welfare issues in the purebred dog. An additional aim was to develop diagnostic DNA tests to identify affected individuals and asymptomatic carriers.

A parallel mapping approach was used to map two autosomal recessive conditions in the Cavalier King Charles Spaniel. The use of a single common set of controls for two independent genome-wide association studies was demonstrated as an efficient mapping strategy when studying two conditions affecting a single breed. Newly available target enrichment and massively parallel sequencing methodology was used to simultaneously sequence both disease-associated loci, with one condition acting as a control for the other.

A genome-wide homozygosity mapping approach using microsatellite markers was used to investigate spinocerebellar ataxia in the Italian Spinone. The disorder was successfully mapped to a single chromosome using six cases and six controls, and fine mapped with additional microsatellite markers. Subsequently, a progression of sequencing techniques were used to identify the disease-associated mutation, with the study highlighting the potential difficulties of using massively parallel sequencing technologies.

Spinocerebellar ataxia (or late onset ataxia) in the Parson Russell Terrier was investigated using a genome-wide association study followed by a target enriched massively parallel sequencing approach. Further sequencing was performed to reduce the large number of potential causal variants, with the entire workflow achieved in-house.

The final experimental chapter describes the use of a genome-wide mRNA sequencing (mRNA-seq) approach as a method of candidate gene sequencing of a single case of neonatal cerebellar cortical degeneration in a Beagle dog. The mRNA-seq approach demonstrates a simple, fast and cost effective method of targeted resequencing of expressed genes when a suitable tissue resource is available.

For all five disorders under investigation, disease-associated mutations were identified leading to the development of diagnostic tests. Three of the mutations were in genes not previously associated with similar conditions in humans or other model organisms.

Contents

Chapter

3. Investigation of two disorders in the Cavalier King Charles Spaniel

3.1. Background

The Cavalier King Charles Spaniel (CKCS) originated from the King Charles Spaniel in the early part of the 20th century, with a dedicated breed club forming in 1928. The popularity of the CKCS gradually increased and the breed was granted separate Kennel Club registration from the King Charles Spaniel in 1945 (Cunliffe, 2004). Today the CKCS is one of the UK's most popular breeds with 5,970 Kennel Club registrations in 2012 (The Kennel Club 2013).

In common with many other purebred dog populations, the CKCS suffers from a high incidence of inherited disease. Diseases affecting the CKCS are well documented in the scientific literature and breed health websites, with a notable focus on neurological disorders (Rusbridge, 2005) and mitral valve disease (Beardow and Buchanan, 1993) (http://www.cavalierhealth.org/, http://www.aboutcavalierhealth.com/). The health of the breed was brought to the attention of the public through two BBC documentaries entitled Pedigree Dogs Exposed, which were broadcast in August 2008 and February 2012 highlighting the painful neurological disorder syringomyelia.

This chapter describes the genetic investigation into two CKCS specific conditions, episodic falling (EF) and congenital keratoconjunctivitis sicca and ichthyosiform dermatosis (CKCSID).

3.1.1. Episodic falling

Episodic falling in the CKCS, also known as sudden collapse, muscle hypertonicity and hyperekplexia, is an exercise, excitement or stress induced syndrome caused by an increase in muscle tone and a temporary inability to relax the muscles. The condition was first reported in 1983, although had been observed in the breed since at least the early 1960's (Herrtage and Palmer, 1983). The onset age is usually between 3 and 7 months and inheritance is consistent with an autosomal recessive mode. The clinical signs are often variable between cases. Episodes can vary in severity and last from a few seconds to several minutes. Episodes often start with an increase in muscle tone, with bunny hopping movements (Herrtage and Palmer, 1983, Rusbridge, 2005) and/or presence of a deer stalker gait (Wright et al., 1986). The back may become arched and the head held close to the ground leading to collapse, either to the side or forwards. Legs may be held out in a rigid, extended fashion, although in some cases the dog may return to its feet within seconds of a collapse. In severe cases forelegs or hindlegs may become protracted until they are positioned over the top of the dog's head as shown in Figure 3.1 (Herrtage

and Palmer, 1983, Shelton, 2004). Dogs appear to remain fully conscious during an episode (Herrtage and Palmer, 1983).

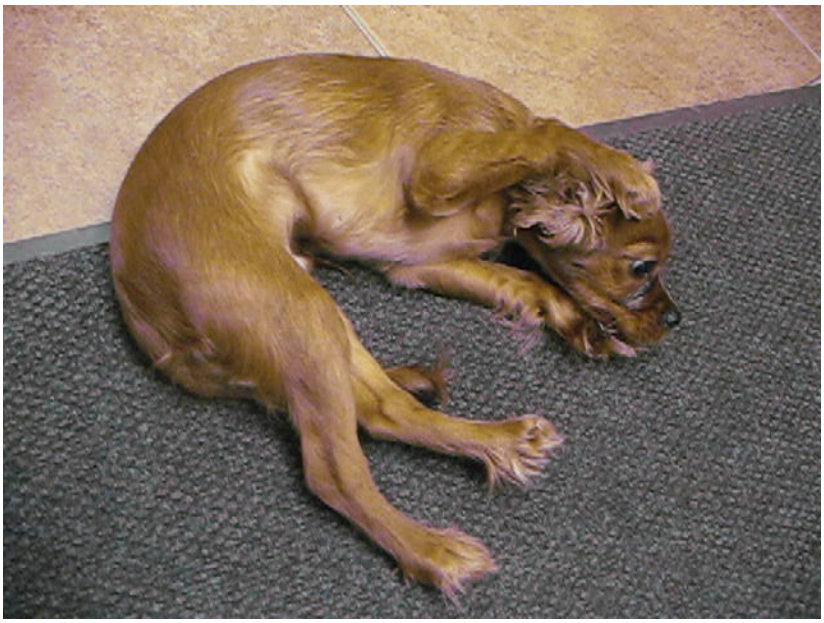


Figure 3.1 Episodic falling in a ten month old Cavalier King Charles Spaniel

Severe muscle hypertonicity in a case of EF. The dog displays an arched back and forelimbs are protracted over the head. Still image from a video provided by Dr Boaz Levitin DVM, DACVIM (Neurology).

Therapeutic agents are often used to treat EF cases, with dogs often responding well to Clonazepam (benzodiazepine) treatment. Reports in the veterinary literature describe severe cases with a high episode frequency becoming almost clinically normal after administration of the drug (Garosi et al., 2002, Shelton and Engvall, 2002).

Episodic falling shares similarities to human disorders, including hyperekplexia, Brody's myopathy and myotonia. Hyperekplexia is a disease of exaggerated startle response and increased muscle stiffness and rigidity, which shows a particularly close resemblance to EF in terms of the positive response to the drug Clonazepam. Clonazepam is thought to alleviate clinical signs by improving neurotransmission in gamma-aminobutyric acid (GABA) pathways (Tijssen et al., 1997). Mutations in several genes have been associated with hyperekplexia in humans (Table 2.2). Brody's myopathy is a disease of exercise induced muscle cramping with the inability to relax muscles (Brody, 1969). Mutations in the *ATP2A1* (ATPase, Ca++ transporting, cardiac muscle, fast twitch 1) gene have been associated with Brody's myopathy (Odermatt et al., 1996). Myotonia is described as a disease with delayed skeletal muscle relaxation after sudden and often exaggerated contraction, and exists in both autosomal recessive (Becker's disease) and dominant forms (Thomsen's disease) (Becker, 1977, Thomsen, 1876). Mutations in the *CLCN1*

(chloride channel, voltage-sensitive 1) gene have been associated with both autosomal recessive and dominant forms of the disease in humans, in the myotonic “fainting” goat, and myotonia in the Miniature Schnauzer and the Australian Cattle Dog (Beck et al., 1996, Finnigan et al., 2007, George et al., 1993, Koch et al., 1992, Rhodes et al., 1999). Characterisation of many diseases which are closely related to EF at the molecular level presents the opportunity to investigate the disease using a candidate gene approach.

3.1.2. Congenital keratoconjunctivitis sicca and ichthyosiform dermatosis

Congenital keratoconjunctivitis sicca and ichthyosiform dermatosis (CKCSID), commonly known as dry eye and curly coat syndrome, was first reported in the scientific literature in 2006 (Barnett, 2006). Clinical signs are recognisable at birth, with affected puppies having a coat that is rough or “crimped” in appearance (Figure 3.2). Affected puppies are often reported to be smaller than unaffected littermates.

Figure 3.2 Two week old puppy with CKCSID

Two week old CKCS puppy displaying clinical signs consistent with CKCSID. Note the rough appearance of the coat. (Hartley et al 2011.)

Clinical signs of keratoconjunctivitis sicca (dry eye) are apparent from eyelid opening at approximately ten days, with a reduced production of aqueous tears which can result in a discharge of tacky mucus and in severe cases ulceration of the cornea. As CKCSID affected dogs progress to adulthood, the skin can become hyperkeratinised and hyperpigmented across the ventral abdominal region. The coat is harsh and frizzy, with scaling and partial alopecia along the dorsum and flanks, which may cause the dog to scratch. Footpads also become hyperkeratinised with abnormal growth of nails and intermittent sloughing, causing lameness (Figure 3.3).

Prognosis for affected individuals is poor and the condition cannot be resolved through treatment. Many owners elect to euthanise affected dogs due to difficulties in disease management, although life expectancy is not obviously affected by the condition (Hartley et al., 2011).

Figure 3.3 Nail and footpad abnormalities in a case of CKCSID

Abnormal growth of a nail and thickening of the footpad. Images provided by Claudia Hartley.

Ichthyosis (thickened, dry and often scaled skin) has been described in other breeds of dog including the Norfolk Terrier (Credille et al., 2005), Jack Russell Terrier (Credille et al., 2009), and Golden Retriever (Grall et al., 2012), with disease-associated mutations identified for all three breeds. No syndromes reported in the human literature describe clinical signs that are entirely comparable with those seen in cases of CKCSID, although individually both keratoconjunctivitis sicca and ichthyosiform dermatosis are widely reported. Keratitis-ichthyosis-deafness (KID) syndrome, caused by mutations in the *GJB2* (gap junction protein, beta 2) gene encoding connexin-26, shows some clinical similarities to CKCSID (Richard et al., 2002, Skinner et al., 1981). A disease of woolly hair, premature tooth loss, nail dystrophy, acral hyperkeratosis and facial abnormalities has also been described in a human kindred, but unlike CKCSID no ocular clinical signs were described (van Steensel et al., 2001).

The clinical signs and progression of CKCSID in the Cavalier have been described in detail by C. Hartley and colleagues, through the diagnosis and follow up of 25 cases (Hartley et al., 2011). DNA samples were also collected and used for a subsequent candidate gene study, in which microsatellite markers were analysed for association between CKCSID and 28 canine orthologues of human disease-associated genes (Hartley et al., 2012). No association was found between CKCSID and any of the genes

investigated thus ruling out any obvious candidate genes; the study therefore progressed to a GWAS, which is described in this chapter.

3.1.3. Aims

The aims of this investigation were to identify the mutations responsible for EF and CKCSID in the CKCS. As both conditions afflict the same breed, this presented the opportunity to map both conditions in parallel using a set of cases for each disease and a single set of clinically unaffected controls. Association loci from the independent GWAS would be followed up using sequencing techniques in order to identify disease-associated mutations, allowing diagnostic assays to be developed.

3.2. Results

3.2.1. EF candidate gene study

A sample cohort of 12 EF cases and 10 controls was selected for investigation by Professor Jacques Penderis (veterinary neurologist). Canine orthologues of genes causing a similar condition to EF in humans were selected as candidate genes (see Table 2.1). For each candidate gene, two flanking microsatellite markers were identified and genotyped across the sample cohort. At least one microsatellite was required to be informative with a minor allele frequency greater than 0.1 to exclude the gene. The genotyping dataset is shown in Table 3.1 (major alleles are highlighted). For microsatellite markers segregating with EF, cases were expected to be homozygous for a single allele and controls either homozygous or heterozygous for any allele. Upon visual inspection marker GLRA1_C4_60.68 showed a pattern suggesting segregation with EF, with 9/12 cases homozygous for allele 200 and only 3/10 controls with a homozygous genotype for allele 200. The gene locus could be excluded however, by the second microsatellite marker (GLRA1_C4_60.64). No markers surrounding the other candidate genes showed a pattern of linkage disequilibrium with EF based on visual inspection. All three microsatellites genotyped around *SLC5A9* were monomorphic, so the gene could not be formally excluded. The lack of variation across the region may be suggestive of a selective sweep in the CKCS breed.

Table 3.1 Genotypes table for EF candidate genes

Gene targets and genomic coordinates are listed on the left. Major alleles are highlighted in blue or grey. Non-determined alleles are marked as n.d.

Microsatellites			Dog ID numbers and genotypes																													
			Cases												Controls																	
Gene	CFA	Mb	1001	1002	1003	1004	1015	1034	1043	1044	1053	1058	1059	1061	1025	1026	1045	1051	1052	1054	1055	1056	1057	1060
CACNL1A3	7	5.29	n.d. n.d.	284 295	295 295	284 298	285 285	285 295	n.d. n.d.	295 295	285 295	295 295	295 295	285 295	285 298	285 295	n.d. n.d.	295 295	295 295	285 295	295 295	295 295	284 295	295 295
CACNL1A3	7	5.34	284 290	284 284	290 290	284 284	272 272	272 290	n.d. n.d.	284 290	272 290	290 290	n.d. n.d.	272 290	272 286	272 290	284 290	284 290	284 290	272 290	290 290	n.d. n.d.	284 290	284 290
ATP2A1	6	21.26	223 225	223 223	227 227	227 227	223 223	223 227	223 223	223 225	223 223	225 227	223 227	223 227	227 227	223 227	225 227	223 225	223 225	223 225	223 225	223 223	223 223	223 225
ATP2A1	6	21.56	n.d. n.d.	263 265	255 265	255 255	263 263	265 265	263 265	263 265	263 265	255 265	255 265	255 263	265 265	265 265	255 271	265 271	263 271	265 271	265 271	265 265	265 265	265 265
CLCN1	16	9.26	208 208	204 212	206 212	204 212	206 212	206 206	208 208	204 204	206 206	208 208	204 208	206 206	204 212	n.d. n.d.	204 206	204 206	204 208	204 208	204 208	204 208	206 212	204 208
CLCN1	16	9.47	n.d. n.d.	99 101	101 101	99 103	101 101	101 101	n.d. n.d.	n.d. n.d.	101 101	112 112	99 112	101 101	99 103	101 101	n.d. n.d.	99 101	99 112	99 112	99 112	99 112	101 101	99 112
GLRA1	4	60.64	162 162	160 160	160 160	160 162	160 162	160 162	162 162	162 162	160 162	146 160	146 162	146 162	156 162	162 162	146 162	146 162	146 156	160 162	162 162	160 162	146 162	146 162
GLRA1	4	60.68	200 200	200 200	200 200	200 200	200 200	200 200	200 200	200 200	200 200	196 200	196 196	196 200	200 202	200 200	196 200	196 200	196 202	200 200	196 200	200 200	196 200	196 200
GLRA3	25	28.14	212 212	208 208	212 212	208 212	208 212	208 212	212 212	208 208	208 208	208 208	208 208	208 212	208 212	208 212	212 212	208 212	208 212	208 212	n.d. n.d.	212 212	208 208	208 212
GLRA3	25	28.25	147 147	144 144	147 147	144 147	n.d. n.d.	144 147	147 147	144 144	144 144	144 144	144 144	144 147	144 147	144 147	147 147	144 147	144 147	144 147	144 147	147 147	144 144	144 147
GLRB	15	57.36	126 128	126 128	128 128	126 128	128 128	128 128	126 128	126 128	128 128	119 128	119 128	119 126	128 128	126 128	128 128	128 128	128 128	115 130	128 128	126 128	128 128	128 128
GLRB	15	57.49	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 209	209 217	209 209
GPHN	8	43.92	92 92	104 104	92 92	92 104	98 100	92 98	92 92	92 92	100 100	92 98	92 92	92 98	92 98	92 98	92 92	92 98	92 92	92 104	92 104	92 92	100 104	92 92
GPHN	8	43.98	304 304	310 310	304 304	304 310	306 310	304 306	304 304	304 304	310 310	304 306	304 304	304 306	304 306	304 306	304 304	304 306	304 304	304 310	304 310	304 304	310 310	304 304
KCNE3	21	26.87	256 256	248 254	248 254	248 248	248 256	248 248	256 256	256 256	254 256	256 256	256 256	254 256	256 256	254 256	256 256	248 256	254 256	256 256	248 256	256 256	256 256	256 256
KCNE3	21	27.01	279 279	279 281	285 285	279 279	279 279	n.d. n.d.	279 279	279 279	279 285	279 279	279 279	279 285	279 279	n.d. n.d.	279 279	279 279	279 281	279 279	279 279	279 279	279 279	279 279
SCN4A	9	4.81	185 185	174 174	174 178	174 185	185 185	174 185	185 185	174 185	185 185	174 174	185 185	174 185	174 185	174 185	174 185	174 178	174 185	185 185	174 185	n.d. n.d.	174 185	174 185
SLC32A1	24	30.09	255 255	244 252	244 250	244 244	252 255	244 252	244 244	252 252	252 252	244 252	244 255	244 255	244 244	244 252	255 255	n.d. n.d.	n.d. n.d.	244 255	244 244	244 252	252 255	244 252
SLC32A1	24	30.15	n.d. n.d.	298 306	298 306	298 298	301 306	298 306	298 298	306 306	306 306	298 306	298 301	298 301	298 298	298 306	301 301	298 306	301 306	298 301	n.d. n.d.	298 306	301 306	298 306
SLC6A5	21	45.757	182 182	182 182	182 184	178 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182	182 182
SLC6A5	21	45.76	244 244	244 244	244 244	239 244	n.d. n.d.	244 244	231 244	244 244	244 244	244 244	244 244	244 244	244 244	244 244	244 244	244 244	244 244	244 244	244 244	244 244	244 244	244 244
SLC6A5	21	45.77	n.d. n.d.	176 176	176 179	176 176	175 175	175 176	n.d. n.d.	175 176	175 175	175 175	175 175	176 176	175 175	n.d. n.d.	176 176	175 176	176 176	175 176	175 176	n.d. n.d.	175 175	175 176
SLC6A5	21	45.92	n.d. n.d.	190 190	181 190	186 190	188 190	188 190	n.d. n.d.	n.d. n.d.	188 190	188 190	188 190	190 190	188 190	188 190	n.d. n.d.	188 190	190 190	188 188	188 188	188 190	188 190	190 190
SLC6A9	15	19.01	207 207	207 207	207 207	207 207	n.d. n.d.	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207	207 207
SLC6A9	15	19.012	241 241	241 241	241 241	241 241	241 241	241 241	241 241	n.d. n.d.	241 241	241 241	241 241	241 241	241 241	241 241	n.d. n.d.	241 241	241 241	241 241	241 241	241 241	241 241	241 241
SLC6A9	15	19.31	299 299	299 299	299 299	299 299	n.d. n.d.	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299	299 299

3.2.2. Genome-wide association study

3.2.2.1. Illumina CanineHD SNP array genotyping data

The 96 DNA samples sent for processing on the Illumina CanineHD SNP array genotyped successfully achieving call rates of >99%. Raw genotyping data were imported into Genome Studio for viewing and manual processing to increase the overall genotyping call rate and remove poor quality SNPs. Overall call rate was initially improved by re-clustering the SNP calls (Figure 3.4).

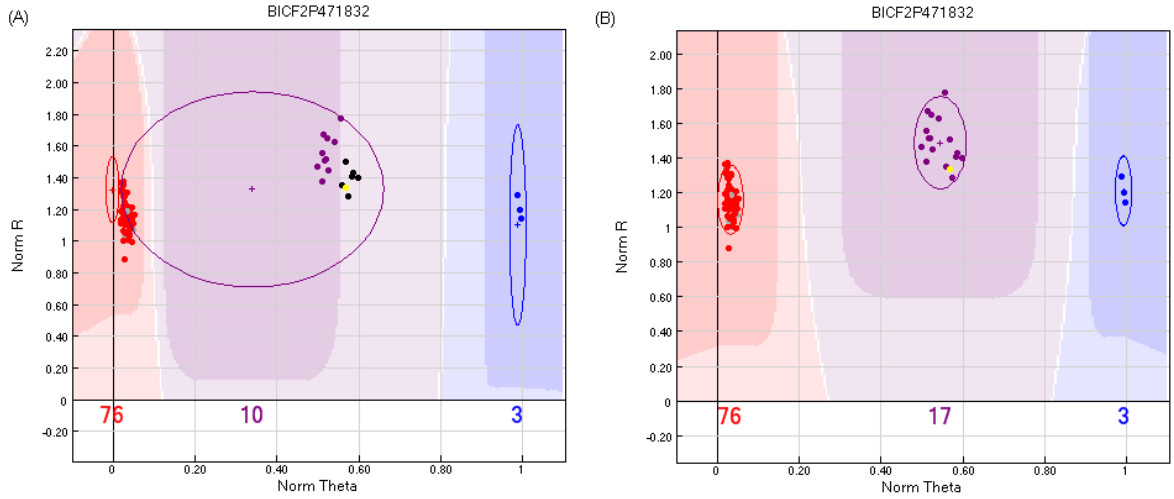


Figure 3.4 Improving call rates using the "cluster all SNPs" command

An example of data improved by the “cluster all SNPs” command in Genome Studio. Each dot on the plots represents a single SNP. The areas shaded in red, purple and blue represent the boundaries for calling alleles AA, AB and BB respectively. (A) Clustering using the predefined cluster file. (B) Clustering after the “cluster all SNPs” command has been executed. The call rate for the SNP has improved from 93% to 100%.

It was noted that SNP calls for four individuals (one Golden Retriever, two IS and one CKCS control) often clustered separately for many otherwise monomorphic SNPs. This caused reclustering errors because of the absence of a heterozygous calls cluster (Figure 3.5A). Call rates were subsequently improved by excluding the four individuals and repeating the SNP re-clustering command. The individuals were then re-included with the new cluster positions in place (Figure 3.5B). Separate clustering was expected for the non-CKCS individuals, but the CKCS control individual appeared to be an outlier and was therefore excluded from further analysis.

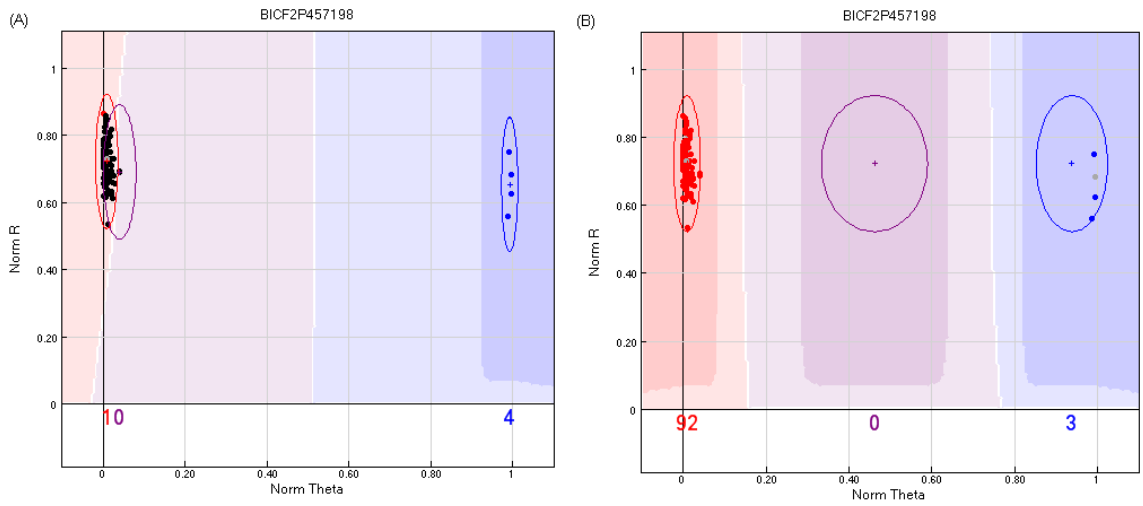


Figure 3.5 SNP clustering problems due to the absence of a heterozygous group

(A) Four individuals often formed a separate homozygous cluster in SNP genotyping data, resulting in calling errors. (B) The clustering problem was resolved by removal of the four individuals before repeating the "cluster all SNPs" command, to improve the average call rate for the entire dataset from 99.48% to 99.84%.

Re-clustered data were exported from Genome Studio into Progeny, for case-control cohort selection.

3.2.2.2. Allelic association analysis

Genotyping data for case-control cohorts were exported from Progeny for allelic association analysis. The sample set for genotyping consisted of 31 EF cases, 19 CKCSID cases, and a common set of 38 controls. SNPs with a minor allele frequency <0.05 and a genotyping call rate <0.95 were excluded from analysis. After filtering 91,427 SNPs remained for EF and 88,384 for CKCSID.

Allelic association analysis was performed using the statistical package PLINK. Strong statistical signals were seen on chromosome 7 for EF (P_{raw} = 1.9 x 10^{-14}) and chromosome 13 for CKCSID (P_{raw} = 1.2×10^{-17}). Allelic association plots for EF and CKCSID are shown in Figures 3.6 and 3.7 respectively. The top 100 SNPs in both studies were located on single chromosomes, producing single distinctive peaks on the association plots.

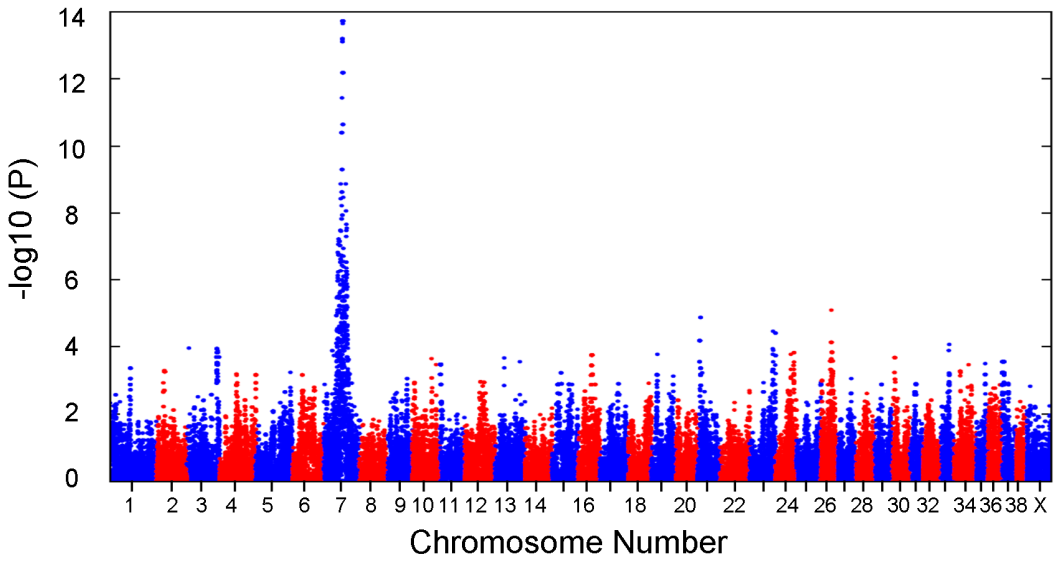


Figure 3.6 Allelic association analysis plot for EF

Allelic association analysis plot for 31 EF cases and 38 controls. Each dot represents a single SNP, with -log10 (p) values on the y-axis plotted against genome position (split into chromosomes) on the x-axis. The strongest statistical signal is on chromosome 7 (P_{raw} = 1.9 x 10^{-14}).

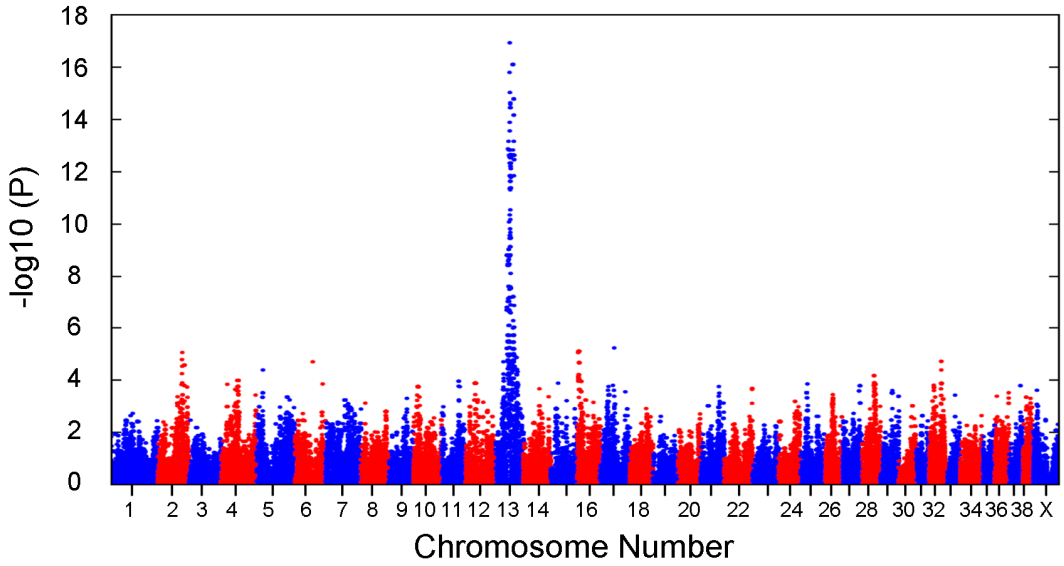


Figure 3.7 Allelic association analysis plot of CKCSID

Allelic association analysis plot for 19 CKCSID cases and 38 controls. Each dot represents a single SNP, with -log10(p) values on the y-axis plotted against genome position (split into chromosomes) on the x-axis. The strongest statistical signal is on chromosome 13 (P_{raw} = 1.2×10^{-17}).

3.2.2.3. Correction for multiple testing

Max(T) permutations analysis (100,000 permutations) was performed using PLINK to correct for multiple testing. For EF and CKCSID single statistical signals remained on chromosomes 7 and 13 respectively, which surpassed genome-wide significance at the 5% level (Figures 3.8 and 3.9). The SNPs displaying the strongest statistical signal for EF and CKCSID after permutations testing both had a P value of 1.0 x 10^{-5} (P_{genome}).

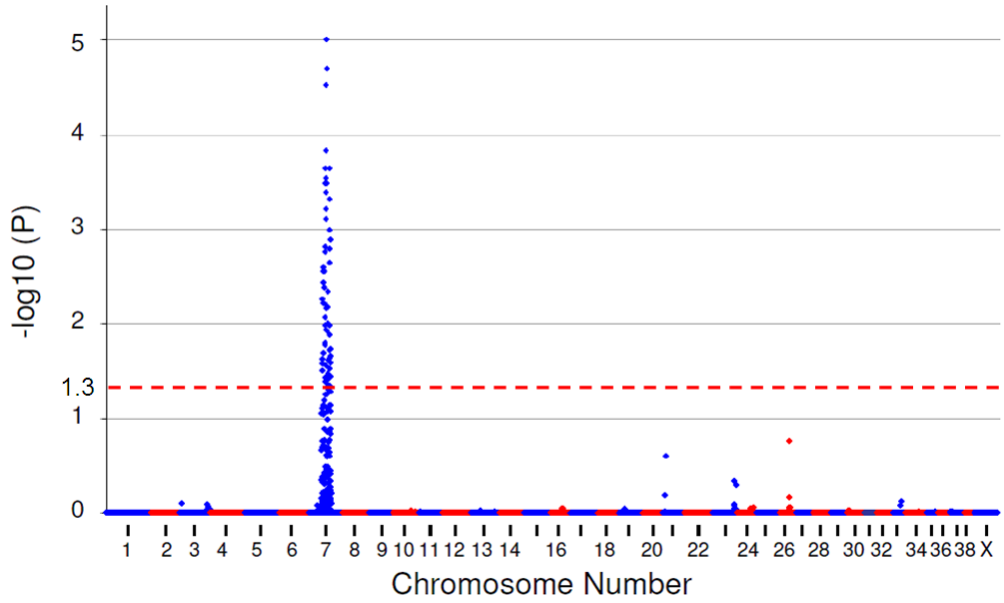


Figure 3.8 EF association analysis including max(T) permutations

Plot of P-values for the EF association study after 100,000 max(T) permutations. Pgenome = 1.0 x 10^{-5}. SNPs above the red dashed line are genome-wide significant at the 5% level.

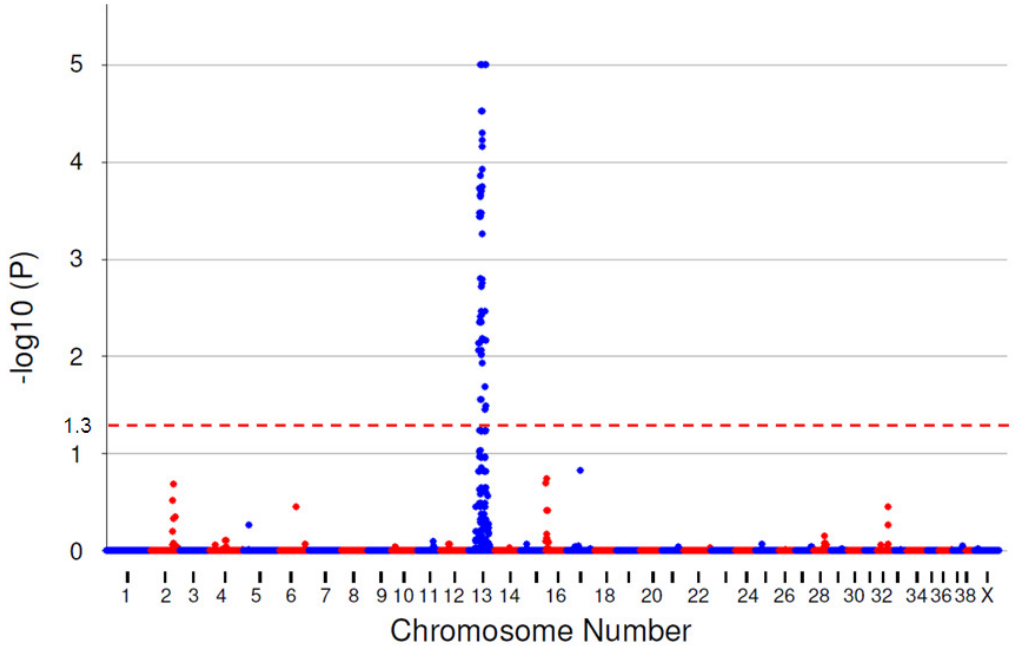


Figure 3.9 CKCSID association analysis including max(T) permutations

Plot of p-values for the CKCSID association study after 100,000 max(T) permutations. Pgenome = 1.0 x 10^{-5}. SNPs above the red dashed line are genome-wide significant at the 5% level.

3.2.3. Population stratification

Genomic inflation values based on the median chi-squared were 1.57 and 1.62 for the EF and CKCSID association analyses respectively, and were suggestive of population stratification in both datasets. Quantile-quantile (QQ) plots of observed versus expected chi-squared values were plotted as a graphical display of genomic inflation (Figure 3.10). The expected pattern for QQ plots in a dataset with no stratification is for the observed chi-squared values to match the expected chi-squared values, and closely track the line y=x. If there is a true signal in the dataset observed chi-squared values will elevate above expected values only for the highest chi-squared values. If genomic inflation is present, the observed versus expected datapoints may be consistently elevated along the line y=x.

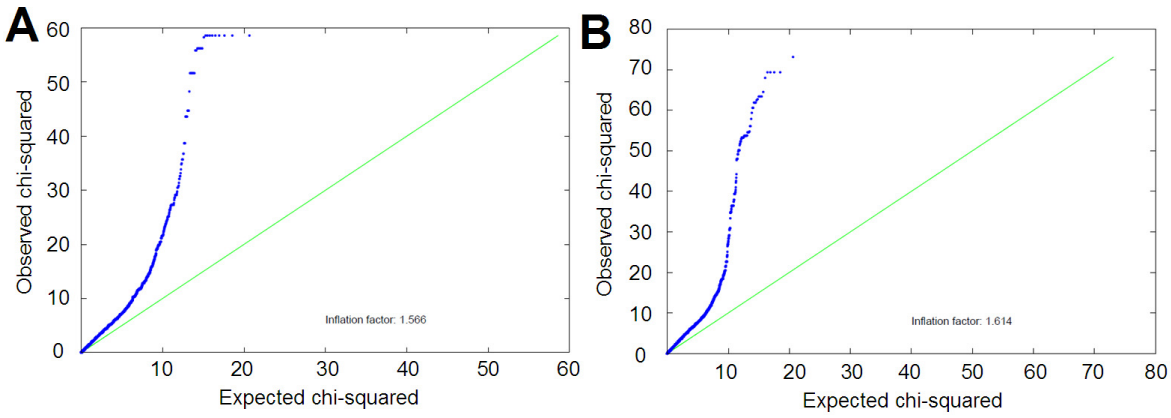


Figure 3.10 QQ plot for the EF and CKCSID SNP genotyping datasets.

QQ plots of observed (y-axis) versus expected (x-axis) chi-squared values for (A) EF and (B) CKCSID. The blue dots represent individual SNPs. The green lines have the equation y=x (ie observed chi-squared = expected chi-squared). The genomic inflation values for EF and CKCSID were 1.57 and 1.62 respectively.

Observed versus expected chi-squared coordinates for both EF and CKCSID datasets, track above the y=x line, and is further evidence of genomic inflation. Multi-dimensional scaling (MDS) plots were generated to assess the relatedness of cases and controls in the EF and CKCSID studies (Figure 3.11). Related individuals cluster closely together on MDS plots. The plots show some separately clustering cases and controls for both EF and CKCSID, which may account for the high genomic inflation values. Clustering of controls away from cases is probably due to a shared set of controls being selected for both EF and CKCSID, rather than tailoring the control set to a particular cohort of cases.

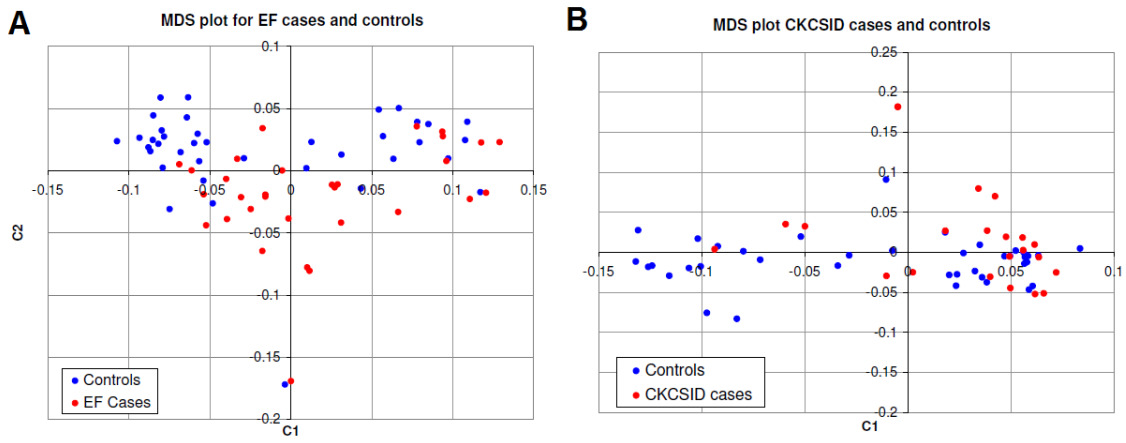


Figure 3.11 MDS plots

MDS plots for (A) EF and (B) CKCSID. Red dots represent cases and blue dots represent controls. Coordinates for the plots were calculated using pairwise identity-by-state (IBS) distance.

3.2.4. Adjusting for genomic inflation

The fast mixed model (FMM) was implemented to adjust for genomic inflation. The two top association signals remained statistically associated at 4.1×10^{-10} and 1.5×10^{-11} for the EF and CKCSID respectively (Figure 3.12).

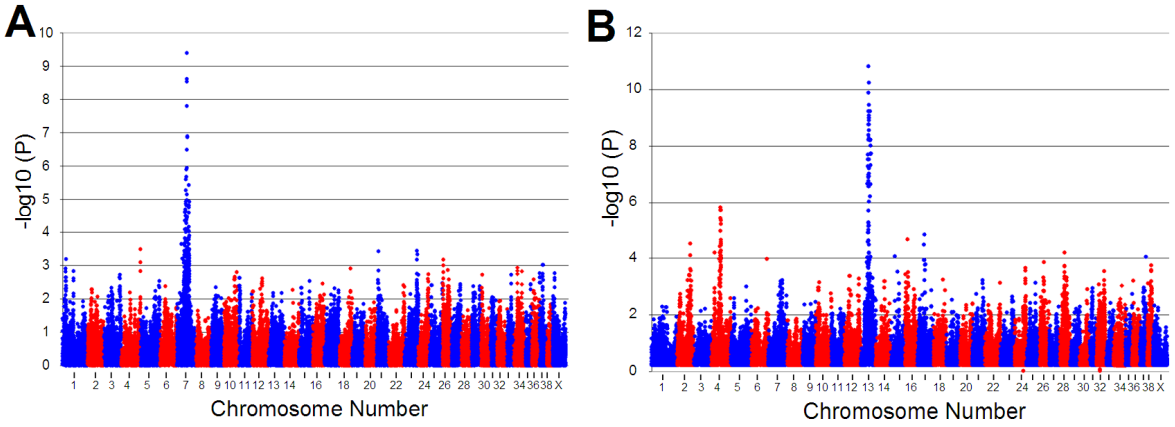


Figure 3.12 FMM corrected allelic association analysis plots

FMM adjusted allelic association plots for (A) EF and (B) CKCSID. Strong statistical signals remained on chromosomes 7 and 13 for EF and CKCSID respectively.

QQ plots are displayed for the FMM adjusted data in Figure 3.13. The genomic inflation value for the EF data after FMM correction was 0.98. Interestingly the genomic inflation value for the CKCSID data after FMM correction remained high at 1.83. The QQ plot for FMM corrected data of observed versus expected chi-squared values does however show the datapoints to more closely follow the line y=x, indicating that the values have been adjusted.

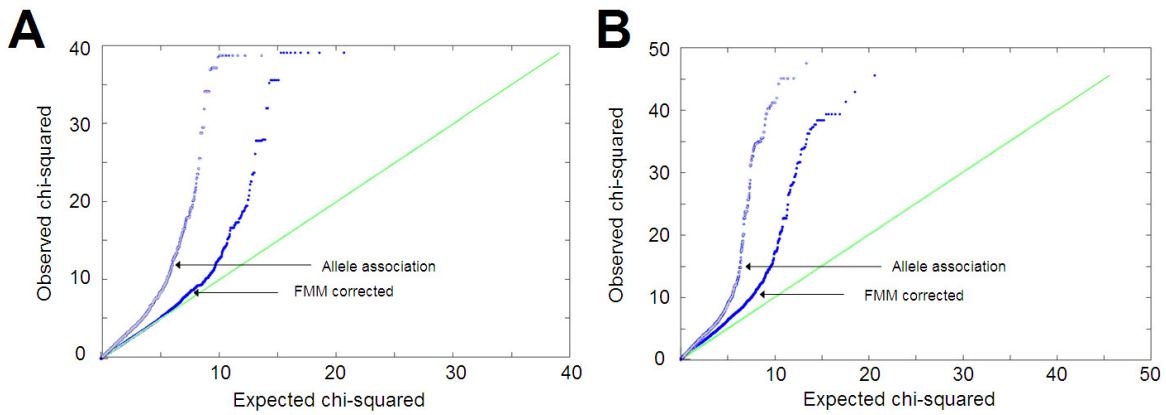


Figure 3.13 QQ plots of FMM adjusted data

QQ plots of observed (y axis) versus expected (x axis) chi-squared values for (A) EF and (B) CKCSID after adjusting for genomic inflation using the fast mixed model. The green lines track the x=y coordinates.

3.2.5. Investigation of strong statistical signals at the chromosome level

Strong statistical signals were first investigated at the chromosome level by plotting raw SNP p-values across chromosome 7 and 13 for EF and CKCSID respectively. For EF the focal point of the strong statistical region was between 40 Mb and 48 Mb on chromosome 7 (Figure 3.14). For CKCSID the statistical signal was spread over a slightly larger region of 32 Mb to 46 Mb on chromosome 13 (Figure 3.15).

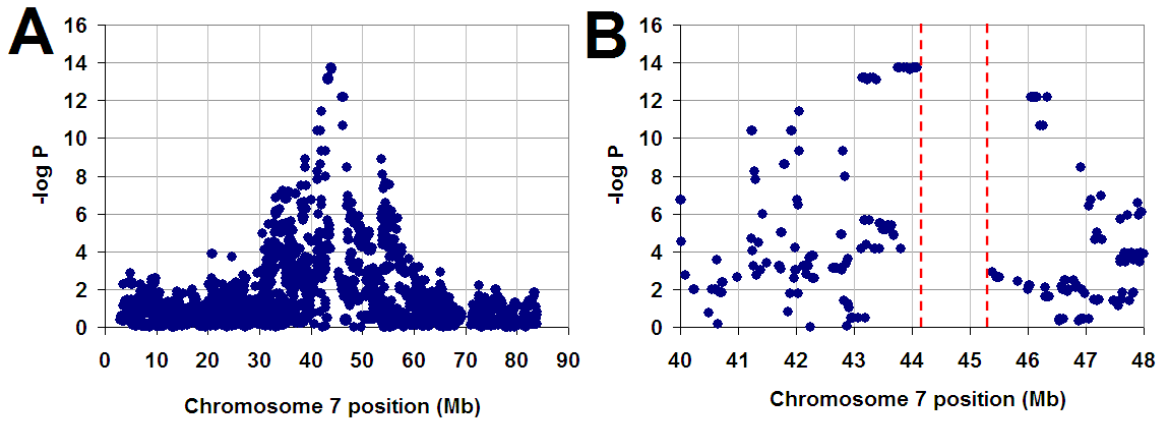


Figure 3.14 Chromosome 7 allelic association plots for the EF study

(A) Allelic association plot for EF across chromosome 7. (B) Focal point of the strong statistical signal. The signal is interrupted by a region of approximately 1 Mb containing no SNPs due to the minor allele frequency filtering parameters defined in the analysis.

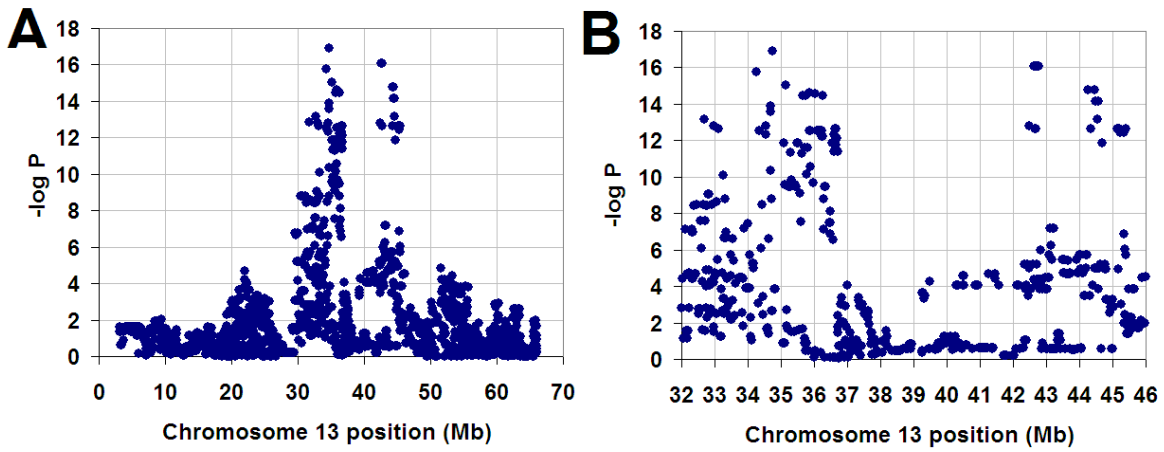
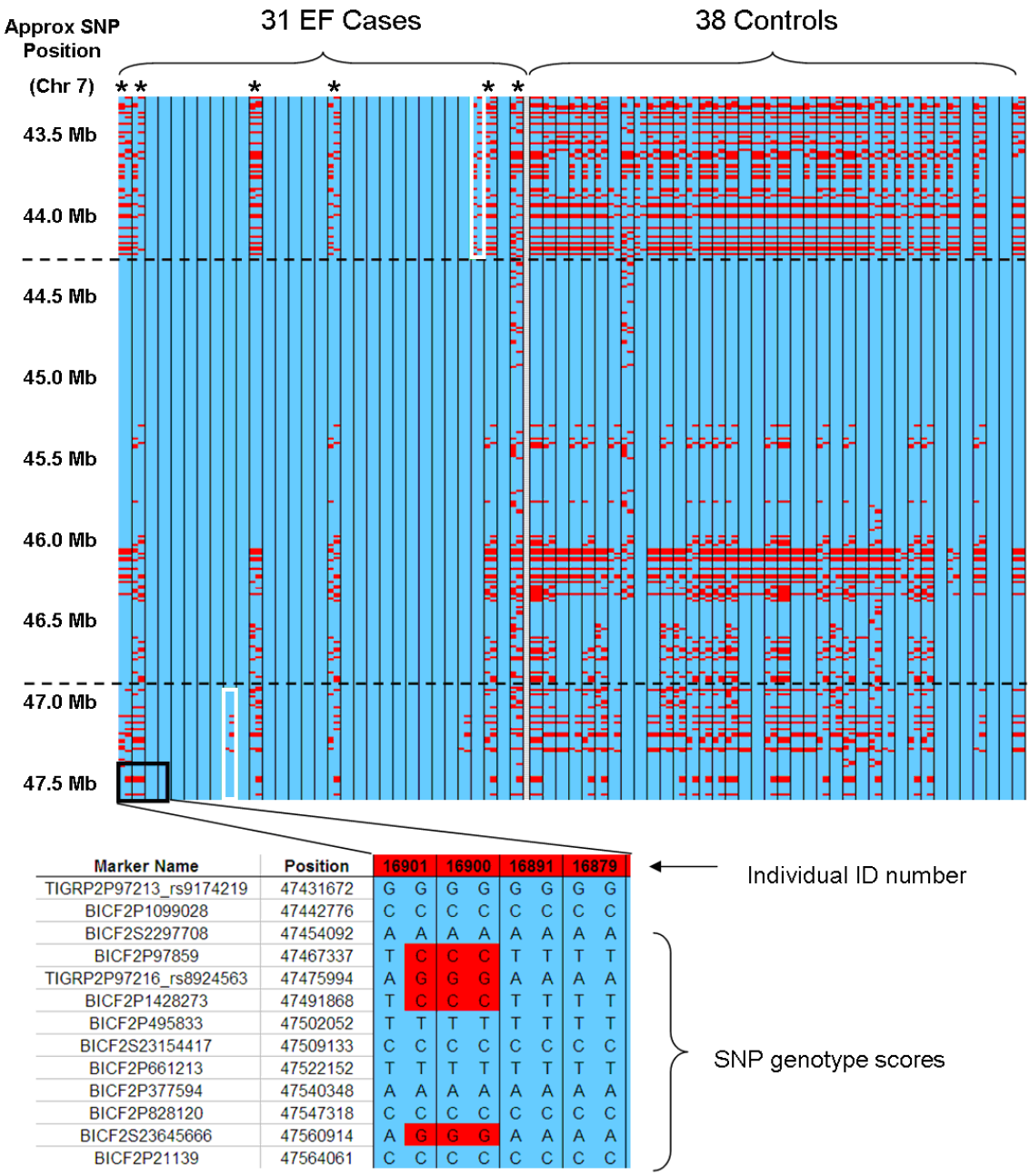


Figure 3.15 Chromosome 13 allelic association plot for the CKCSID study

(A) Allelic association plot for CKCSID across chromosome 13. (B) The focal point of the strong statistical signal.

Regions of strong association were further investigated by visualising raw SNP genotyping data across the disease-associated regions. For EF two individuals defined the shared disease-associated haplotype or "critical region" as CFA7:44,093,554-46,905,272, by a loss of shared homozygosity in cases due to recombination events (Figure 3.16). Six individuals suspected to be affected on the basis of phenotype did not share the disease-associated haplotype, and are marked with an asterisk in Figure 3.16. One of these cases was subsequently reclassified as an epilepsy case in an independent neurological work-up. The other cases could not be resolved because the dogs were deceased or owners could not be contacted. The critical region contained a 1.2 Mb haplotype that was homozygous across 67 of 69 cases and controls, which could be suggestive of a selective sweep in the CKCS breed. Two cases were critical in defining the EF disease-associated region, and it was therefore imperative that the diagnosis of these cases was correct. The case defining the upper boundary (15943) was a typical but severe case of EF, with the owner describing clinical signs during an episode of a arched back, trembling, collapse, stiff paralysed hind legs, extension of the forelimbs above the head and jaw locking. Episodes reportedly lasted from two minutes to up to an hour. A video clip was also provided with the dog presenting an episode at 4 months of age, with clinical signs consistent with EF. The sample defining the lower boundary of the region was 16867 (JP1049). Only a single episode was reported and little further information was available for this case. Owner and veterinarian details were unavailable. Sample 16155 could define the lower boundary at a slightly lower position of 47,048,914 on canine chromosome 7. This case had an initial episode at 6 months of age, presenting with collapse and hyperextension of limbs. The case was treated with and responded to Rivotril. Video footage provided with the case showed clinical signs consistent with EF.

This case therefore provided a more assured definition of the 3' boundary. The EF critical region contained 114 genes and was syntenic with human chromosome 1 (Appendix 6).



Marker Name	Position	16901		16900		16891		16879	
TIGRP2P97213_rs9174219	47431672	G	G	G	G	G	G	G	G
BICF2P1099028	47442776	C	C	C	C	C	C	C	C
BICF2S2297708	47454092	A	A	A	A	A	A	A	A
BICF2P97859	47467337	T	C	C	C	T	T	T	T
TIGRP2P97216_rs8924563	47475994	A	G	G	G	A	A	A	A
BICF2P1428273	47491868	T	C	C	C	T	T	T	T
BICF2P495833	47502052	T	T	T	T	T	T	T	T
BICF2S23154417	47509133	C	C	C	C	C	C	C	C
BICF2P661213	47522152	T	T	T	T	T	T	T	T
BICF2P377594	47540348	A	A	A	A	A	A	A	A
BICF2P828120	47547318	C	C	C	C	C	C	C	C
BICF2S23645666	47560914	A	G	G	G	A	A	A	A
BICF2P21139	47564061	C	C	C	C	C	C	C	C

Figure 3.16 EF critical region raw genotyping data

Raw genotyping data across the EF disease-associated critical region. Each column represents an individual and SNP markers are listed in rows. Cases are listed on the left, controls on the right. Major alleles are coloured in blue, minor alleles in red. Two individuals showed a loss of homozygosity due to recombination events, highlighted by a white border, defining the disease-associated haplotype as CFA7:44,093,554-46,905,272 (boundary marked by black dashed lines). Six cases appear to be apparent outliers that were not homozygous for the disease-associated haplotype (columns marked with an asterisk)

For CKCSID recombination events in two individuals defined the disease-associated haplotype as CFA13:39,648,169-42,481,707, which was shared between all cases (Figure 3.17). Diagnosis of cases was checked and confirmed in consultation with Claudia Hartley (veterinary ophthalmologist). The CKCSID critical region contained 85 genes, and was syntenic to regions of human chromosomes 4, 8 and 15 (Appendix 6).

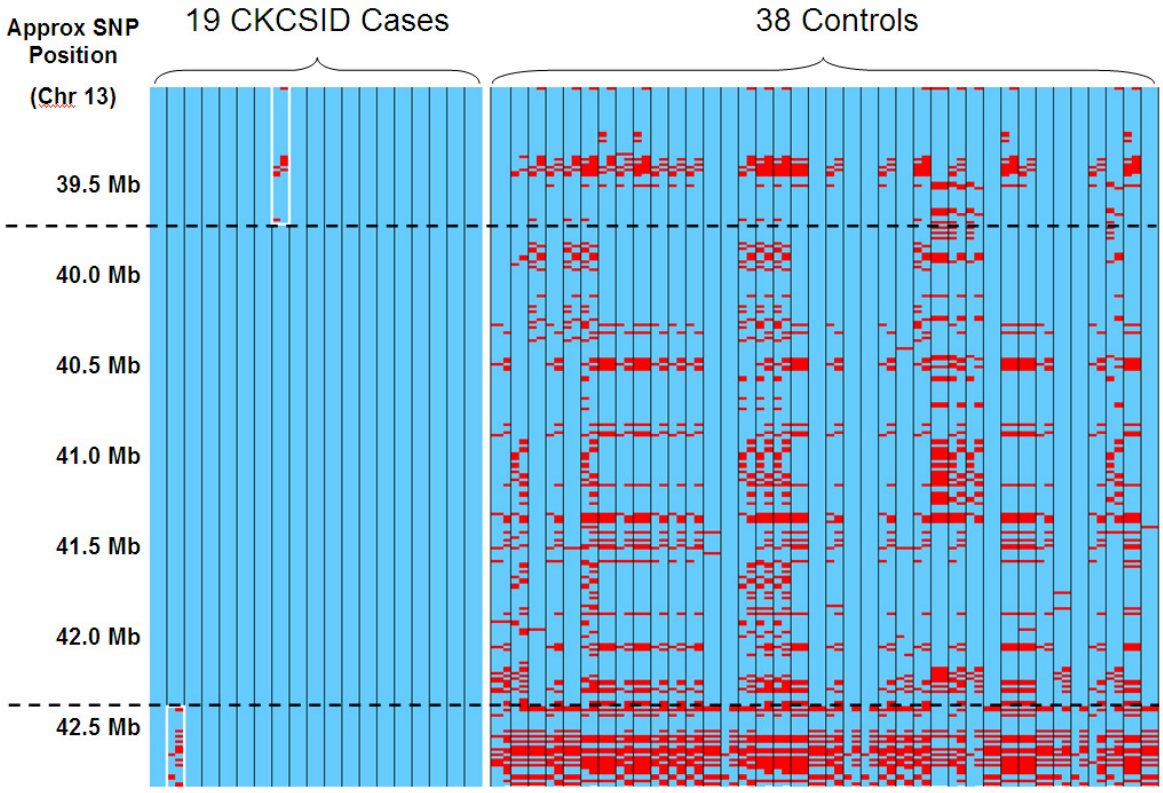


Figure 3.17 CKCSID disease-associated critical haplotype

Raw genotyping data across the CKCSID disease-associated critical region. Each column represents an individual and SNP markers are listed in rows. Cases are listed on the left, controls on the right. Major alleles are coloured in blue, minor alleles in red. Two individuals showed a loss of homozygosity due to recombination events, highlighted by a white border, defining the disease-associated haplotype as CFA13:39,648,169-42,481,707 (boundaries marked by black dashed lines).

Using the Ensembl genome browser both regions were interrogated for genes that could be potential candidates for the respective disorders. The EF critical region contained no genes that were associated with similar conditions in humans. The gene *SLURP1* was identified in the CKCSID critical region. Mutations in *SLURP1* have been associated with Mal de Meleda, an autosomal recessive skin disorder in humans, with clinical signs of transgressive palmoplantar keratoderma, keratotic skin lesions, perioral erythema, brachydactyly and nail abnormalities (Fischer et al., 2001), and was considered a good candidate gene for CKCSID that had not previously been investigated.

3.2.6. *SLURP1* sequencing

Because of the similarities between Mal de Meleda and CKCSID, the gene was exon re-sequenced in two CKCSID cases and two control individuals. No polymorphisms were identified and the gene was ruled out as being potentially causal.

3.2.7. SureSelect target enrichment and massively parallel sequencing

3.2.7.1. Probe design

Probes (RNA baits) were designed for the SureSelect solution based target enrichment system using the online tool e-array (https://earray.chem.agilent.com/earray/). Design parameters were set to capture both EF and CKCSID critical regions with 2x bait tiling (ie each base was covered by at least two RNA baits where possible) and repeat masking was applied (ie no probes were placed across known repeat regions). A total of 57,676 baits were designed across a combined region of 5,788,900 bp; 29,429 baits for the EF region and 28,247 baits for the CKCSID region, achieving a base coverage of 3,749,814 (64.8%). Bait locations were uploaded to the UCSC genome browser to assess bait coverage across exonic regions. Coverage of exons appeared to be nearly complete, although isolated examples of exons with no bait coverage were occasionally seen, often due to exons being in close proximity to repetitive elements (Figure 3.18).

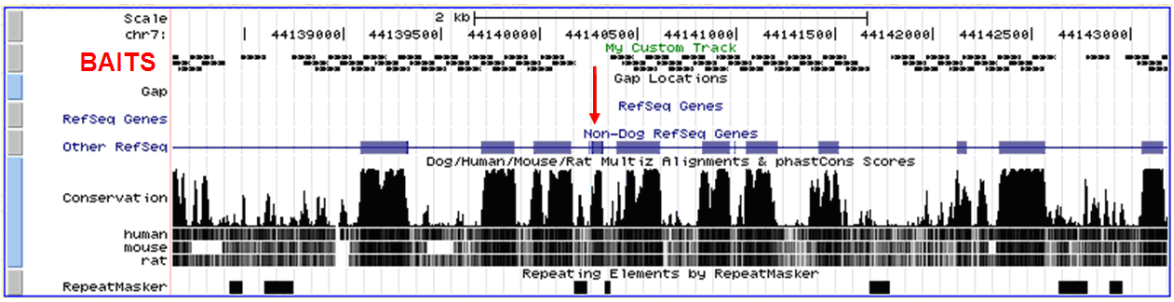


Figure 3.18 Visualisation of baits on the UCSC genome browser

Positioning of baits across a section of the EF region and displayed by the UCSC genome browser. An exon with zero bait coverage is indicated by the red arrow.

3.2.7.2. Investigation of fragmentation methods

A key stage of library preparation for next generation sequencing is fragmentation of the nucleic acid. At the time of investigation the recommended method of fragmentation for Illumina sequencing was nebulisation. To help fulfil the aim of completing bench work in-house, three methods of DNA fragmentation were tested.

3.2.7.2.1 Sonication

The results of sonication for 10, 20 and 30 cycles of 20 seconds on / 20 seconds off at full power using a cup horn sonicator are shown in Figure 3.19. Fragmentation of DNA with a size range of between 100 bp and 1,000 bp was achieved.

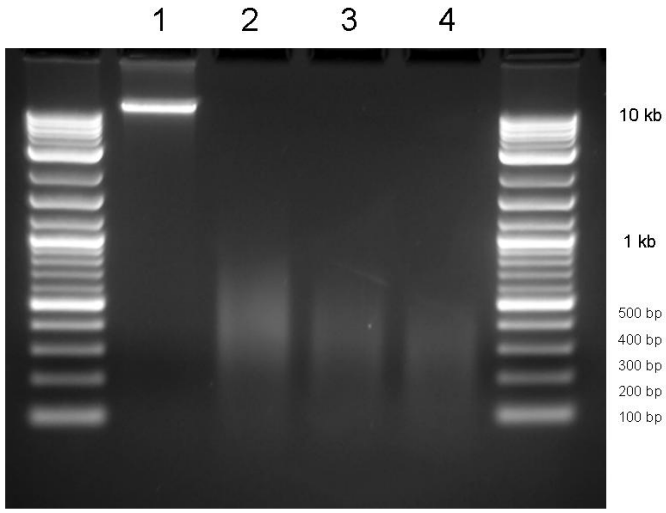


Figure 3.19 DNA sonication results

Results of sonication of high molecular weight genomic DNA (1.5% agarose gel). Lane 1 contains 200 ng non-treated high molecular weight genomic DNA. Lanes 2, 3 and 4 contain 200 ng genomic DNA sonicated for 10, 20 and 30 cycles of sonication respectively (20 seconds on / 20 seconds off at full power).

3.2.7.2.2 Nebulisation

Nebulisation using a pressure of 24 psi and 50% glycerol buffer produced fragments in the 150–1000 bp range (Figure 3.20A). On agarose the fragments appeared most abundant in the 400–600 bp range. Only 350 µl of the 750 µl starting material could be recovered after nebulisation, indicating that half the material is lost during the shearing process, through vaporisation.

3.2.7.2.3 Double stranded DNA Fragmentase

Genomic DNA was digested for 30 minutes using NEB dsDNA Fragmentase. The process yielded fragments of 100-1,000 bp. The highest concentration of fragments was in the 100-600 bp size range (Figure 3.20B).

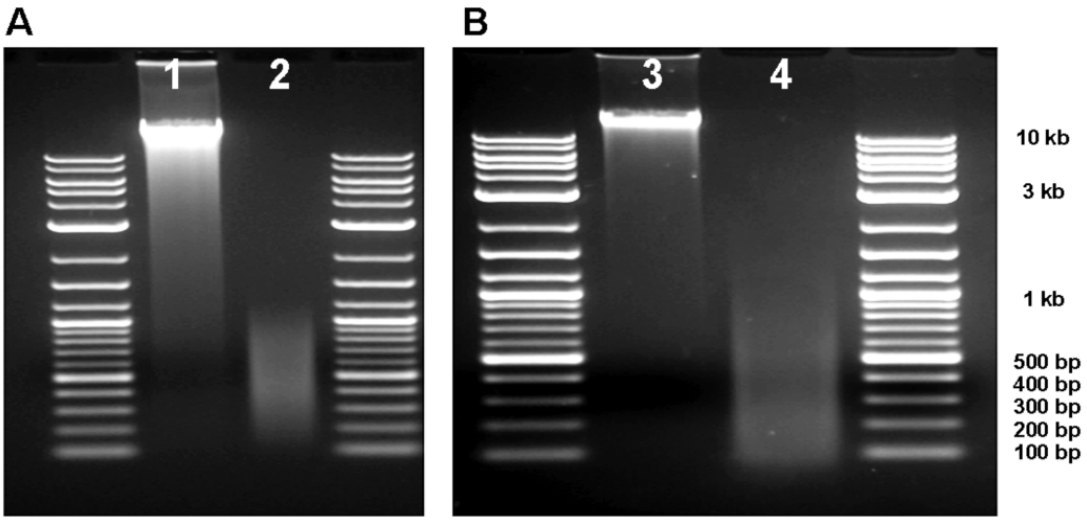


Figure 3.20 Nebulised and fragmentase treated genomic DNA

(A) Genomic DNA treated by nebulisation. Lane 1 - 200 ng high molecular weight genomic DNA. Lane 2 - 200 ng nebulised genomic DNA. (B) Genomic DNA before (lane 3) and after (lane 4) treatment with NEB dsDNA Fragmentase.

3.2.7.3. Trial library preparation and clone sequencing

Trial libraries were prepared to test in-house library preparation using the NEBnext kit. To reduce costs, primer and adapter sequences for use in the library preparation were synthesised by an oligonucleotide manufacturer, rather than being purchased from Illumina. Trial library preparation would also test the effectiveness of these custom made oligonucleotides. Three libraries were made in the trial. Libraries 1 and 2 were made using nebulised DNA. Library 3 was created from DNA treated with dsDNA Fragmentase. Libraries were agarose gel size selected after adapter ligation and amplified with Phusion polymerase using the indexing method. Library fragments were checked for correct adapter sequences by molecular cloning. Results are shown in Table 3.2.

Overall 52% of sequenced cloned fragments had perfect adapter sequences at both ends. The percentage of correctly adapted fragments was higher for Fragmentase treated DNA (77%) compared with nebulised DNA (42%). Fragmentase treatment was chosen for use in SureSelect experiments based on cloning results and the high proportion of fragments produced in the desired 100-300 bp range.

Table 3.2 Summary of the clone sequencing results.

LIBRARY No	CLONE No	PCR	SEQUENCE	ADAPTERS OK?	INSERT SIZE
1	1	3 products	NO	-	-
1	2	OK	YES	YES	319
1	3	OK	YES	NO	250
1	4	OK	YES	NO	113
1	5	OK	YES	YES	273
1	6	OK	NO	-	-
1	7	OK	NO	-	-
1	8	OK	YES	NO	201
2	1	OK	YES	NO	212
2	2	OK	YES	YES	171
2	3	OK	YES	YES	225
2	4	Faint	NO		-
2	5	OK	YES	NO	236
2	6	OK	YES	NO	251
2	7	OK	YES	YES	225
2	8	OK	YES	NO	201
3	1	No product	NO	-	-
3	2	OK	YES	YES	286
3	3	OK	YES	YES	275
3	4	OK	YES	YES	275
3	5	Two products	NO	-	
3	6	OK	YES	NO	248
3	7	OK	YES	YES	249
3	8	OK	YES	NO	258
3	9	Unexpected size	NO	-	-
3	10	OK	YES	YES	207
3	11	OK	YES	YES	274
3	12	OK	YES	NO	280
3	13	OK	YES	YES	255
3	14	OK	YES	YES	293
3	15	OK	YES	YES	322
3	16	OK	YES	YES	265

3.2.7.4. Sample selection for target enrichment

Two EF cases, two CKCSID cases and three controls were selected for potential use in the first attempt at target enrichment (Table 3.3). Cases were selected that were homozygous for the defined disease-associated haplotypes with clinical diagnosis confirmed by Jacques Penderis or Claudia Hartley. Controls were selected based on haplotype structure across the two disease-associated regions. A control with chromosome 7 and 13 haplotypes identical to disease-associated regions, but clinically unaffected with respect to EF and CKCSID was selected in an attempt to reduce the number of potential causal variants.

Table 3.3 Individuals selected for target enrichment

WTH - wild-type haplotype; DAH - disease-associated haplotype. Note control sample 16874, although clinically normal, is homozygous for the chromosome 7 and chromosome 13 disease-associated haplotypes.

Sample ID	Clinical status	CKCSID region	EF region
6823	Control	Heterozygous	Homozygous WTH
6975	**CKCSID case**	Homozygous DAH	Homozygous WTH
9916	**CKCSID case**	Homozygous DAH	Heterozygous
15943	**EF case**	Homozygous WTH	Homozygous DAH
16823	**EF case**	Homozygous WTH	Homozygous DAH
16874	Control	Homozygous DAH	Homozygous DAH
16878	Control	Homozygous WTH	Heterozygous

3.2.7.5. Library preparation including pre-capture amplification

Libraries were fragmented to the 100-600 bp range using dsDNA Fragmentase, and prepared using the NEBnext kit. Fragments in the 200-300 bp range were selected after adapter ligation. Nine cycles of pre-capture amplification were required to increase the concentration of the libraries to 147 ng/µl for the hybridisation stage, which was slightly higher than the 4-6 cycles recommended by the SureSelect protocol. The increased number of cycles was due to a change in methodologies recommended by SureSelect. In the SureSelect Illumina single-end protocol v2.0 (December 2009) the recommended fragmentation method was nebulisation and included a size selection stage. In the SureSelect Illumina paired-end protocol for multiplexed sequencing v1.0 (May 2010) the recommended fragmentation methodology had been changed to Covaris shearing with no size selection, allowing a reduction in the number of PCR cycles required. As no Covaris service was available at the time of the experiment and to keep the library preparation in-house a size selection stage was adopted. Results of the pre-capture amplification are shown in Figure 3.21.

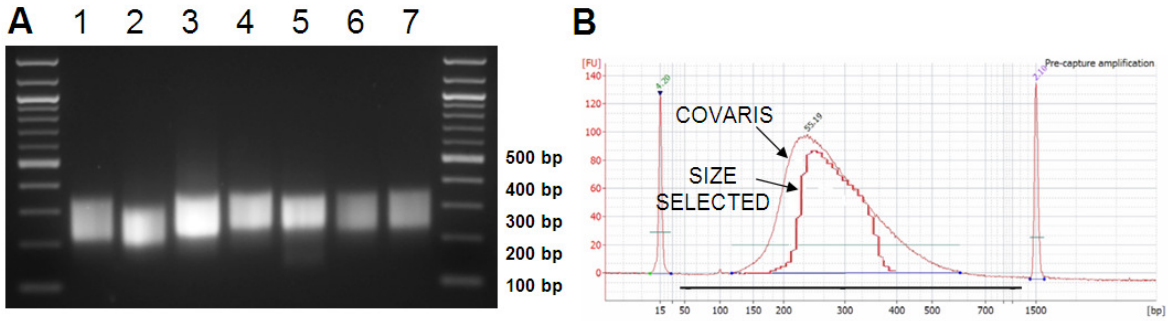


Figure 3.21 Precapture libraries

(A) Precapture libraries (147 ng) on a 2% agarose gel. Lanes 1 to 7 correspond to samples 6823, 6975, 9916, 15943, 16823, 16874 and 16878 respectively. (B) Pre-capture library 9916 was subjected to analysis on a Bioanalyser DNA 12000 chip. The size range of fragments falls inside of those seen for library preparation after Covaris shearing.

3.2.7.6. SureSelect hybridisation and post-capture amplification

Five libraries were taken forward to the enrichment stage consisting of two EF cases, two CKCSID cases and a single control which was homozygous for both the chromosome 7 and chromosome 13 disease-associated haplotypes. The hybridisation stage was followed by 13 cycles of post capture amplification to produce the final sequencing libraries, which were quantified using KAPA library quantification qPCR results in conjunction with results of accurate library sizing on the Bioanalyser. Libraries were pooled in equal amounts to 10 nM. Blunt end cloning of pooled library fragments was carried out to check end sequences and to estimate capture efficiency before libraries were sent for sequencing. A total of 23 colonies were selected for PCR and sequencing.

Results indicated a capture efficiency of 78% and an average insert size of 194 bp (Table 3.4).

Table 3.4 Summary of blunt end cloning results to estimate capture efficiency

Clone sequences not hitting target regions after BLAST searching are highlighted in yellow.

Clone No.	Index No	Chromosome	Position (Mb)	Insert size (bp)	Ends ok?
1	2	13	41.86	221	Y
2	2	13	41.85	160	Y
3	2	7	44.85	134	Y
4	4	7	45.59	231	Y
5	4	13	42.12	214	Y
6	4	7	46.08	176	Y
7	4	1	68.79	256	Y
8	4	7	45.97	148	Y
9	4	13	40.27	172	N
10	5	13	41.49	210	Y
11	5	7	45.09	166	Y
12	5	13	42.10	196	Y
13	5	7	83.59	163	N
14	5	7	44.96	221	Y
15	5	7	45.86	190	N
16	5	1	52.85	244	Y
17	6	7	46.27	189	Y
18	6	7	44.85	158	N
19	6	16	51.72	255	N
20	6	7	44.53	140	Y
21	6	13	40.27	172	Y
22	7	13	42.07	228	Y
23	7	26	16.22	213	Y

3.2.8. Illumina raw sequencing results

Illumina sequencing on the GAIIx generated a 3.47 Gb dataset consisting of 68 million reads of 51 bp in length. As expected from cloning experiments the mean insert size was ~200 bp and read share was evenly distributed amongst the chosen indices (Figure 3.22).

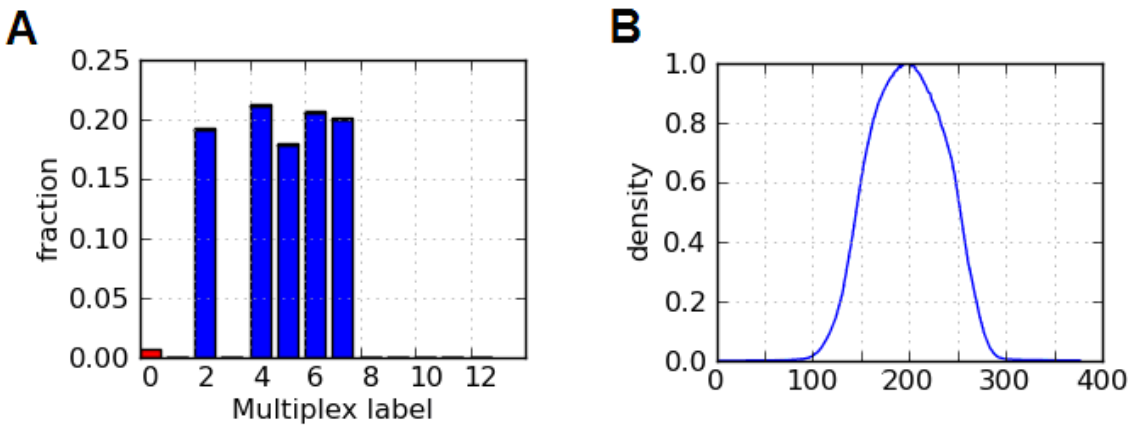


Figure 3.22 Summary histograms of CKCS Illumina sequencing

(A) Proportion of reads allocated to each index. Index 2, 4, 5, 6 and 7 were used in library preparations of samples 6975, 9916, 15943, 16823 and 16874 respectively. Label 0 represents reads that could not be demultiplexed from the raw data. (B) Insert size histogram.

3.2.9. Development of a data analysis pipeline

Illumina sequencing data for the CKCS was available as raw unaligned FASTQ files only. An example of a single read in FASTQ format is shown below.

```
@SEQUENCE_ID
TACGATCGATGCTAGCATGCATGCTTGACTGGGGACTGATCGTAGCTAGGATCGTGCAGT
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC65
```

The @ symbol which acts as a marker is followed by an identifier tag on the first line. The second line details the read sequence. The third line has a + symbol marker and the fourth line shows the per base quality score for the sequence in line two.

The two key requirements for analysis of massively parallel sequencing data are the ability to align reads to a reference genome and to call variants. A huge number of freeware programs are available online for handling and manipulating data generated by next generation sequencing platforms. Most are operated from the Linux command prompt and long strings of complex commands are often required for their operation. Additionally each program has different input requirements, and files often have to be manipulated by associated programs before they can be used to perform the desired task. This means that a long chain of complex commands have to be executed individually to achieve the desired results. To simplify the process a Perl script called “NGS analysis”, which could be run from the Linux command line, was written to sequentially process the string of required commands, after the user had been prompted to enter simplified details for the analysis. The initial user interface is shown in Figure 3.23.

3.2.9.1. Features of the NGS analysis pipeline

The NGS pipeline can handle paired-end or single-end Illumina data in FASTQ format and single end 454 data in ssq format. Reads are aligned to a reference genome using the program BWA. The best practice is to align data to a whole genome, but data can be aligned to an individual chromosome to reduce computational requirements. Datasets can also be reduced for easier handling. For instance if a target enrichment is performed across two genomic regions, alignment files can be created for each region separately to reduce file size and subsequent processing time. The pipeline will create SNP and indel calls, and there is an option to run the Ensembl Variant Effects Predictor to annotate calls with genomic information, such as consequence for variants positioned within exons.

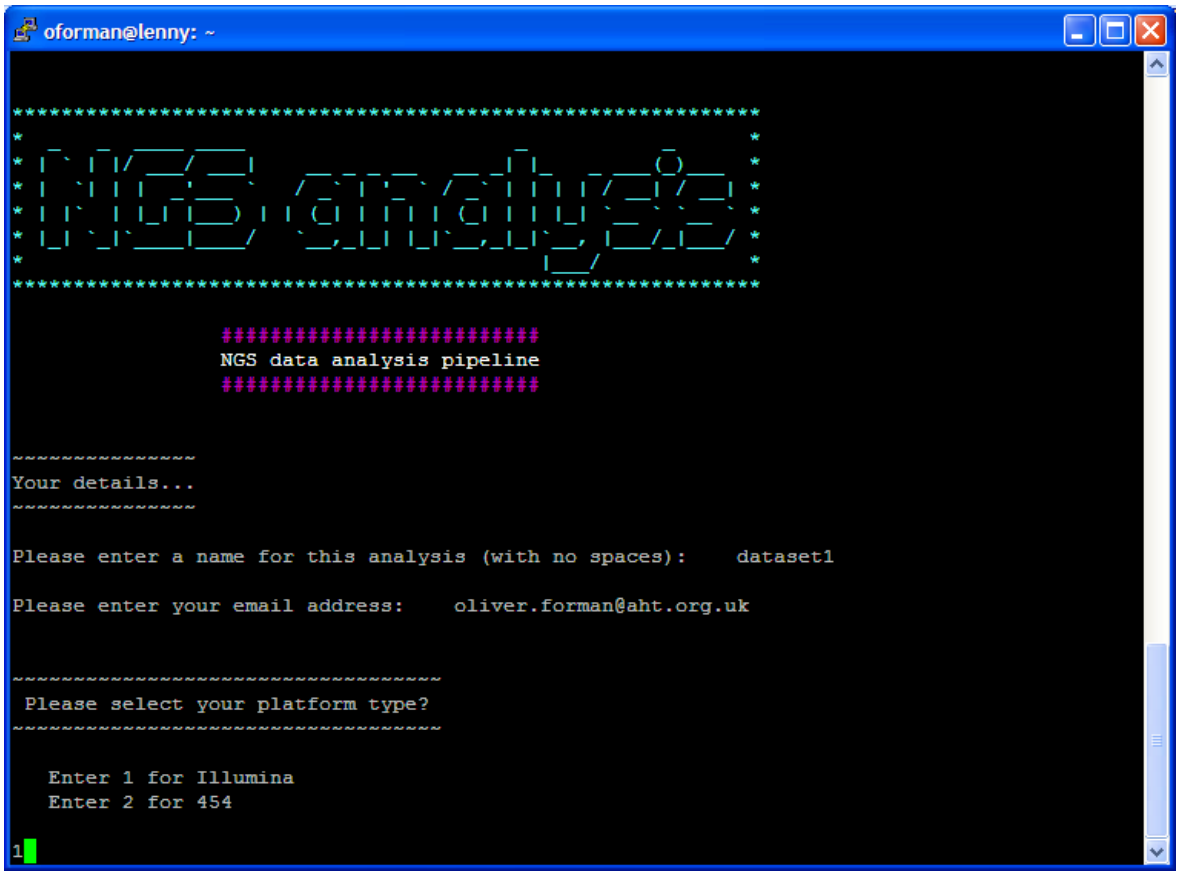


Figure 3.23 NGS analysis Perl script user interface

Screenshot of the user interface displayed at the start of the NGS analysis Perl script. The script was written to enable users to enter file names and analysis parameters, before automatic and sequential implementation of commands in the script to generate results files.

Other modules in the pipeline include removal of PCR duplicates, an option to run the structural variant analysis program Pindel, an option to run an early next generation sequencing analysis package called Maq and tools to produce alignment summary, GC bias and insert size histograms. Key files are transferred to a new results folder that is created at the end of the pipeline. A log is created during the running of the pipeline that details the analysis settings selected. If an error appears in the log file the script will automatically terminate and send the user an email notification. Email notification is also given to users on completion of the analysis, which may take several hours depending on the size of the dataset. A summary of results files is listed in Appendix 7. This list is also presented in a readme.txt file created at the end of the analysis pipeline. A workflow of options available to users of the NGS analysis pipeline is shown in Appendix 8 and a video demonstrating use of the NGS pipeline is available on the supplementary CD.

3.2.10. Sequence data analysis

The paired sequencing data files for each sample were passed through the sequence analysis pipeline. A summary table of the processed data is shown in Table 3.5.

Table 3.5 Summary statistics for sequencing data

Individual ID	No. of reads (Million)	Dataset size (Mb)	Target enrichment efficiency (%)	Percent bases achieving 10x Coverage	PCR duplicates (%)
6975	13.1	673	84.8	77.9	14.9
9916	14.6	746	85.0	79.2	9.5
15943	12.3	627	85.1	78.7	11.2
16823	14.2	723	85.7	79.0	13.8
16874	13.7	701	85.7	80.2	18.5

SNP calls across all samples were combined into a single file format using The NGS SNP Handler, an Excel macro tool written by Dr Mike Boursnell. Across the five samples a total of 13,301 SNPs were identified. The NGS SNP Handler was adapted to accept indel files after completion of the CKCS investigation. Re-analysis of the data showed that 1,149 indels were identified across the five datasets.

Compiled SNP calls from the pipeline were initially filtered by expected segregation pattern (ie homozygous in cases and either heterozygous or homozygous wild-type in controls). The positions of these candidate SNPs were used to create a Browser Extensible Data (BED) file for uploading to the Integrative Genome Viewer (IGV). The five indel files that were created by the NGS analysis pipeline were in BED format and could be loaded directly into IGV. Ensembl gene predictions were uploaded to IGV to annotate the target regions which were then scanned visually for candidate variants occurring in exonic regions. This method of analysis revealed no variants for the EF region that could be considered as potentially causal. In the CKCSID region one indel was identified in an exonic region of *FAM83H* (family with sequence similarity 83, member H) which fully segregated with the disorder. To verify the indel, the sequence alignment files (.BAM files) were uploaded to IGV and the sequence reads were viewed across the candidate locus (Figure 3.24). The two CKCSID cases were homozygous for the indel at the locus and the two EF cases and the control were wild-type homozygous.

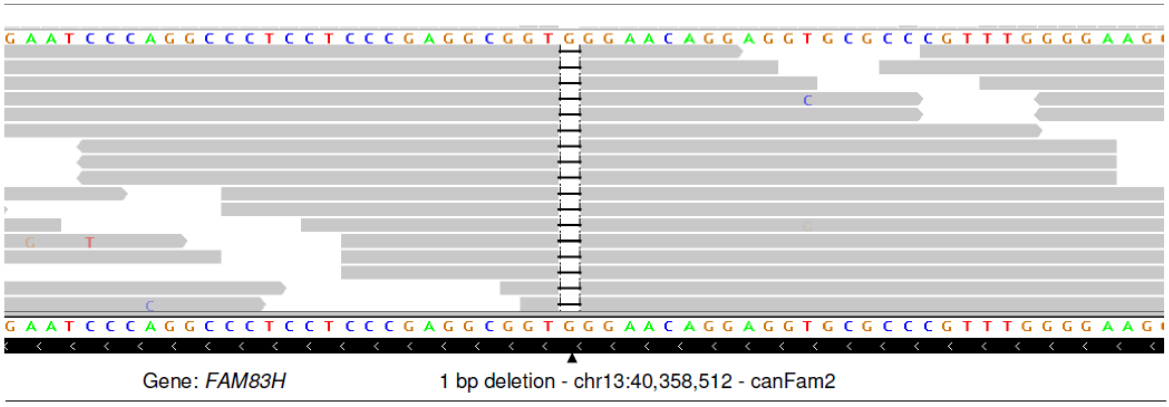


Figure 3.24 Reads from a CKCSID case aligned across the *FAM83H* candidate locus

View in IGV of reads aligned across the *FAM83H* candidate locus for a single CKCSID case, enabling the indel call to be validated. Grey bars represent individual reads. The indel is represented by a solid black horizontal line in reads.

As no potential causal SNP or indel calls had been identified for EF, the five read alignment files were loaded into IGV and the EF target region manually scanned for other potential causal variants. Three large deletions of ~6 kb, ~10 kb and ~16 kb were identified, although only the latter was situated across a coding region. The genomic positions of the three deletions within the disease-associated interval and a schematic overview of normal gene structure are shown in Appendix 6. The 16 kb deletion spanned the first three exons of the gene brevican (*BCAN*) and was potentially a full gene knock-out (Figure 3.25).

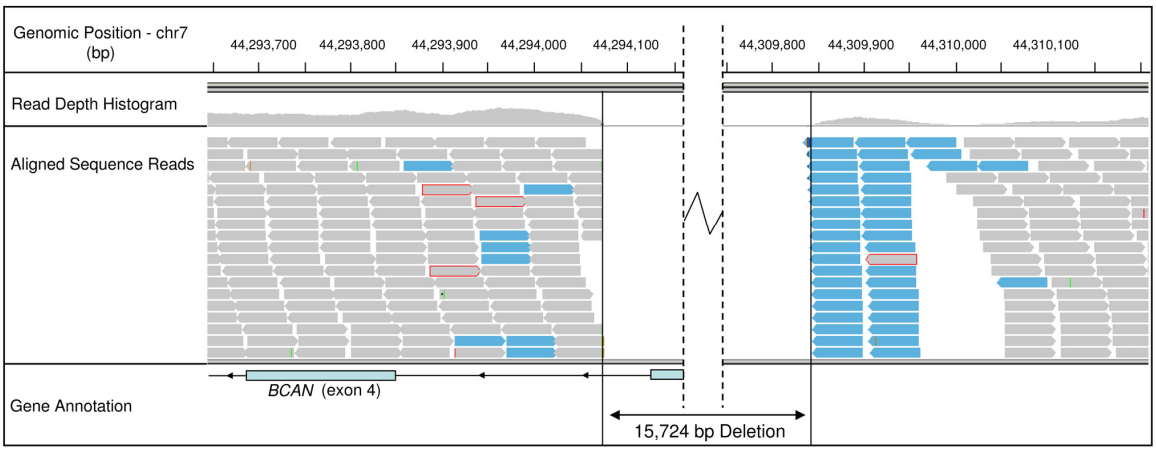


Figure 3.25 The 16 kb brevican deletion

View in IGV across the brevican deletion region. Grey bars represent individual reads. Grey bars with a red perimeter indicate reads that have not been mapped in a pair. Blue bars show read pairs with a greater than expected insert size based on selected fragment size, indicative of a large deletion.

The two EF cases were homozygous for the deletion, one CKCSID case (ID 9916) was heterozygous, and the other CKCSID case was homozygous wild-type (ID 6975). The

clinically unaffected control however was homozygous for the deletion, which was not consistent with the expected segregation pattern. In the absence of any further strong candidate mutations, the deletion was considered to be a strong potential causal variant. In order to find other potentially causal variants with a similar segregation pattern to the 16 kb deletion a second scan across the chromosome 7 target region was performed, omitting the control sample from consideration. One additional variant was identified which was predicted to cause a coding change. This was a SNP in the gene *DENND4B* (DENN/MADD domain containing 4B) which caused an arginine to a histidine amino acid substitution, which followed the same segregation pattern as the *BCAN* deletion.

3.2.11. Investigation of candidate variants

Candidate variants were assessed by genotyping the GWAS sample cohort. As the *DENND4B* mutation was identified in the Golden Retriever included as a genotyping array control and the two Italian Spinoni included for the cerebellar ataxia project (described in Chapter 4), it was considered to be a common polymorphism. The *DENND4B* SNP also showed a weaker statistical signal across a 30 EF case and 38 control cohort than the *BCAN* deletion and was ruled out as potentially causal (*DENND4B* P_{raw} = 6.81 x 10^{-11}; *BCAN* P_{raw} = 5.36 x 10^{-13}). The number of cases was reduced from 31 to 30 as a *DENND4B* genotyping result was not obtained for one EF case, and the number of successful genotypes needed to be the same for both *BCAN* and *DENND4B* to make the comparison valid. Over the sample cohort of 31 EF cases and 38 controls the P_{raw} value for the *BCAN* genotyping dataset was identical to the top statistical signal from the EF GWAS. Interestingly the *BCAN* deletion is positioned inside the region that was suggestive of a selective sweep because of the high levels of homozygosity seen in the SNP genotyping data. The *FAM83H* mutation fully segregated with CKCSID status in the cohort of 19 cases and 38 controls, producing a P_{raw} value of 2.98 x 10^{-22}, which exceeded the top statistical signal from the CKCSID GWAS of $P_{raw} = 1.2\times10^{-17}$.

To further validate the associations between the *BCAN* and *FAM83H* deletions and EF and CKCSID respectively, a panel of 308 CKCS was genotyped for both variants. This panel included the 31 EF cases, 19 CKCSID cases and 38 controls used in the GWAS analyses, and an additional 17 EF cases, 5 CKCSID cases and 198 controls. Results are shown in Table 3.6.

Table 3.6 *BCAN* and *FAM83H* genotyping results across an extended CKCS cohort

	***BCAN* Genotype**		
	(-/-)	**(-/wt)**	**(wt/wt)**
EF cases	39	3	6
EF controls	17	62	181
	***FAM83H* Genotype**		
	(-/-)	**(-/wt)**	**(wt/wt)**
CKCSID cases	24	0	0
CKCSID controls	0	38	246

In addition a panel of 341 dogs from 34 other breeds (with at least 2 dogs per breed) were assayed for both the *FAM83H* and *BCAN* mutations. All 341 dogs were homozygous wild-type for both polymorphisms. From the panel of 308 CKCS genotyped for the *BCAN* or *FAM83H* mutations, individuals that were not clinically affected, unrelated at the parent level and not related to cases at the parent level were used to estimate the mutation frequencies in the UK. From these 122 individuals, the allele frequency of both variants was estimated to be 0.08.

3.2.12. Validating mutation consequence

To validate the consequence of the two mutations, brain (cerebellum) and buccal epithelia cDNA sequencing confirmed the exon boundaries of the *BCAN* and *FAM83H* genes respectively (Genbank accession numbers JN968466–JN968467). The sequencing reads in the *BCAN* deletion region were exported and *de novo* assembled to define the exact deletion breakpoints (Figure 3.26). Assembly revealed the deletion to be 15,724 bp with a small insertion of 5 bp spanning the deletion breakpoints. The *FAM83H* single base deletion is in exon 5, and is predicted to truncate the peptide from 1,151 to 582 amino acids, with 257 aberrant amino acids at the C terminal.

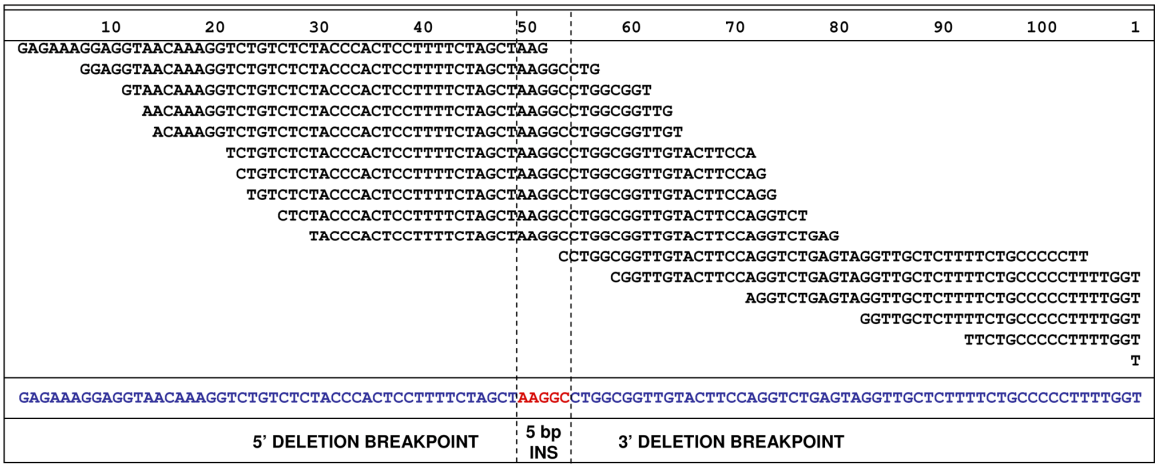


Figure 3.26 *De novo* assembly of reads across the *BCAN* deletion

Sequence reads mapping to the deletion breakpoints were exported and *de novo* assembled to define the deletion breakpoints.

3.2.13. Expression analysis

Quantitative reverse transcription PCR (qRT-PCR) was used to assess *FAM83H* and *BCAN* expression levels in canine skin and brain (cerebellum) tissues, using *ACTB* (beta actin) as a control gene. *BCAN*, *FAM83H* and *ACTB* reaction efficiencies were estimated at 97.5%, 95.7% and 94.3% respectively, with standard curve r^2 values all > 0.99. *BCAN* expression was confirmed in the brain, but was not detected in the skin. A similar level of *FAM83H* expression was detected in both skin and brain (Figure 3.27). *FAM83H* expression was also detected in footpad and buccal epithelia by RT-PCR.

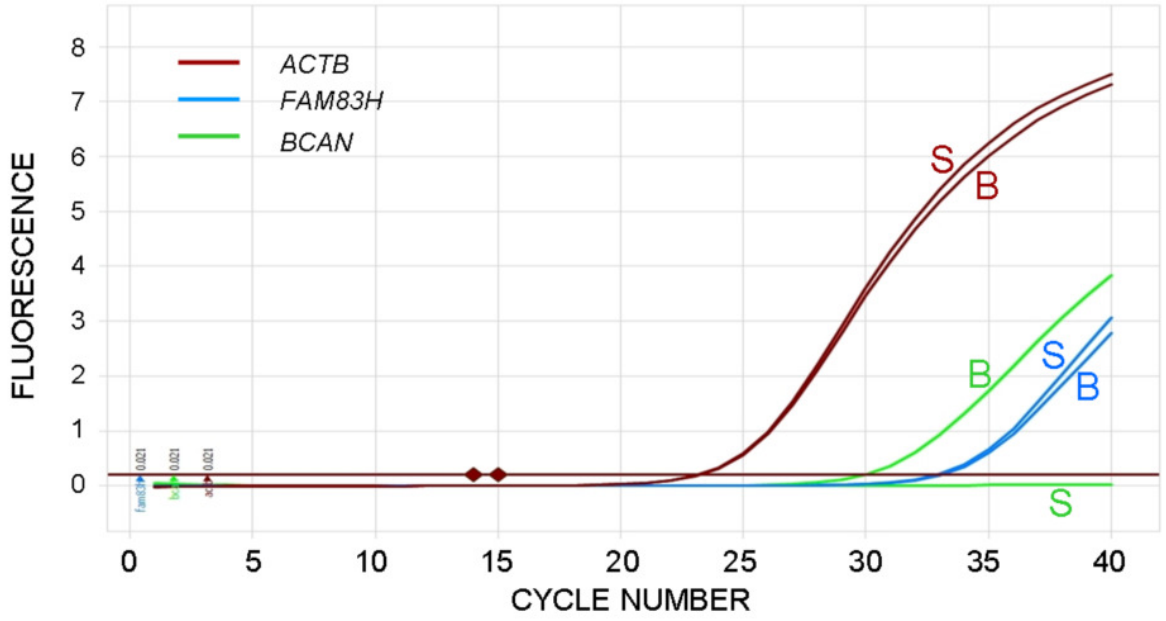
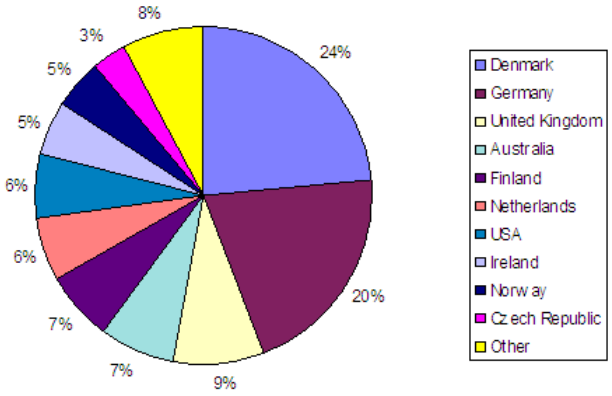


Figure 3.27 qRT-PCR assessment of *FAM83H* and *BCAN* levels

Limited expression analysis of *FAM83H* and *BCAN* using qRT-PCR, with *ACTB* as a control gene. (S = Skin; B = Brain).

3.2.14. Diagnostic DNA testing

A diagnostic DNA test to assay for both the *FAM83H* and *BCAN* mutations simultaneously was launched in April 2011, offered by the AHT Genetics Services department (www.ahtdnatesting.co.uk/). In the first year of testing, 2,457 samples were tested for CKCSID and 2,469 for EF (some owners requested individual disease testing). Dogs with two copies of the disease-associated allele were defined as affected, with one copy of the disease-associated allele as carriers, and with no copies of the disease-associated allele as normal. Four affected and 196 carrier individuals were identified for CKCSID and 96 affected and 539 carriers were identified for EF. Samples originated from 26 countries worldwide. A summary of testing is shown in Figure 3.28.



	CKCSID	EF
AFFECTED	4	96
CARRIER	196	539
NORMAL	2257	1834
Total	2457	2469

Figure 3.28 First year results of EF and CKCSID DNA testing

Ten countries submitting the most samples for EF and CKCSID testing.

3.3. Comments and conclusions

3.3.1. EF candidate gene study

Although no EF associated genes were identified, the investigation successfully ruled out many genes previously associated with hypertonicity disorders. When many strong candidate genes are available, a candidate gene study is often a logical first step to take before undertaking an expensive GWAS approach. The absence of an association with a candidate gene, may suggest that the disorder is caused by a previously unassociated gene.

3.3.2. Parallel mapping of EF and CKCSID by GWAS

The fact that EF and CKCSID were CKCS specific presented the opportunity to map both conditions in parallel using a single set of controls. This approach helped to reduce cost and decreased the overall timeframe for the studies as all DNA samples could be sent for processing on the CanineHD SNP array in a single batch. The disadvantage of the approach was that controls could not be tailored to the individual case sets. For this reason a largely unrelated set of controls was chosen, which may have contributed to the high genomic inflation values seen for both studies, indicating population stratification. Association analysis results for both studies however revealed strong statistical signals on single chromosomes, which remained after corrective analysis, indicating that the results were not largely affected by the genomic inflation, and confirmed the involvement of single, large effect, high penetrance genes, suggesting autosomal recessive inheritance. The same level of genomic inflation in the study of complex disease may have had a more significant impact on the results, where association signals on several chromosomes can be seen, and confidence in the results is crucial.

On analysis of the raw genotyping data there were clear homozygous disease-associated haplotypes for EF and for CKCSID, defined by single recombination events in two individuals, resulting in a loss of shared homozygosity. Amongst the dogs originally defined as EF cases, based on their clinical signs, there were six outliers that were not homozygous for the *BCAN* deletion, highlighting that diagnosis of the condition is not simple and that other conditions, such as epilepsy, exist in the breed which may show a similar clinical presentation.

3.3.3. Target enrichment and massively parallel sequencing

In the absence of any strong candidate genes in the EF critical region and after exclusion of *SLURP1,* a candidate gene in the CKCSID critical region, it was decided that the fastest

and most cost effective method of exploring the two gene rich critical regions would be to use a newly available solution based target enrichment system (SureSelect) and a massively parallel sequencing approach (Illumina/Solexa). The solution based target enrichment system could be performed in-house, negating the need to outsource the work.

The first stage of library preparation is the DNA fragmentation. Out of the three methods that were trialled an enzymatic method was chosen. This method appeared to give the most fragments in the desired range after processing, which is critically important especially as loss of fragments at the size selection stage may lead to additional PCR cycles being required. Any duplicates in sequencing data as a result of PCR are omitted during analysis, reducing the dataset size and eventual read depth. Enzymatic fragmentation also allowed several DNA samples to be processed in parallel, which was more time effective and ensured all samples were subjected to identical treatment.

The entire library preparation was performed in-house. Accurate sizing of libraries on an Agilent Bioanalyser was outsourced due to unavailability of resources. Average library fragment size could have been assessed by agarose gel electrophoresis, but more precise sizing can be calculated using the Bioanalyser, which is important for determining final library concentration, especially when pooling libraries for multiplexed sequencing.

Data were analysed in-house using a Perl script to run a pipeline to manipulate sequence files, align reads to the genome and make variant calls. The sequencing experiment was successful and a high level of target enrichment was achieved, with low levels of PCR duplicates. Bait coverage of target regions was limited to 65% because of the high levels of repetitive elements, particularly SINEs (short interspersed nuclear elements), present in the dog genome. The number of bases achieving at least 10x coverage was 79% however, and near complete exonic coverage of the critical region was achieved.

By browsing through read alignments to target regions in IGV, which was annotated with filtered SNP and indel positions, deletions in *BCAN* and *FAM83H* were identified as strong candidate mutations for EF and CKCSID respectively. The dataset was the first large-scale target enriched sequencing project to be processed by the NGS analysis pipeline, and highlighted areas of the analysis tool that could be improved. As both candidate mutations were identified by manual browsing, measures were subsequently made to make data processing more automated. Firstly the variant effects predictor module was added to the NGS analysis pipeline for annotation of SNP and indel calls with gene

information. Running indel calls through the predictor would have flagged the 1 bp *FAM83H* deletion as a frameshift mutation, warranting further investigation. Additionally a copy number variation (CNV) analysis Perl script was written to compare the number of aligned reads between two samples across overlapping genomic windows, and would have detected the 16 kb *BCAN* deletion without manual browsing of the sequence alignments.

3.3.4. Candidate mutations and phenotype concordance

EF has a variable phenotype in the CKCS and 17 out of 56 dogs that were homozygous for the *BCAN* deletion were reportedly not affected by EF, which suggests that the disorder may be influenced by variation in environmental stimuli and potential variants in modifier genes. As EF is an exercise-induced condition, differences in levels of activity among affected dogs may account for some of the phenotypic variation. One dog in the study that was homozygous for the *BCAN* mutation but did not display clinical signs consistent with EF was reported by its owner to be “docile and unexcitable”, suggesting, for this dog at least, insufficient environmental stimuli were provided to trigger the condition. In addition, nine out of 48 EF cases were not homozygous for the *BCAN* mutation. An extensive neurological assessment would be required to mitigate against misclassification of these cases, but this is typically not possible for the majority of canine patients due to expense or lack of owner consent. The CKCS breed is also affected by idiopathic epilepsy, and the EF episodes may often be difficult to distinguish from epileptic seizures, as a definitive diagnostic test is not available for either condition (Rusbridge, 2005). It is therefore possible that for some cases epilepsy or other neurological conditions have wrongly been diagnosed as EF, although it is formally possible that there may be a second, genetically distinct form of EF in the CKCS. The EF disease-associated region was not completely resequenced due to the repetitive nature of around 35% of the sequence, and although unsequenced regions were largely non-coding, potential causal mutations could be situated within these regions and therefore would not have been identified in the current study.

Results for *FAM83H* genotyping were fully concordant with CKCSID disease status across the extended sample cohort, showing that the CKCSID cases can be precisely diagnosed and that the condition has a simple autosomal recessive mode of inheritance, with no modifiers.

No mutant *BCAN* or *FAM83H* alleles were detected among 341 dogs from 34 other breeds. This suggests that the two mutations are limited to the CKCS breed, although only a small selection of dogs was tested from just a subset of all dog breed populations.

Additional dogs would need to be screened to formally conclude that the mutation is not present in any other breeds.

3.3.5. *BCAN*

The EF-associated gene *BCAN* encodes brevican, which is one of the central nervous system specific members of the hyaluronan-binding chondroitin sulphate proteoglycan family (Yamada et al., 1994). Brevican is important in the organisation of the nodes of Ranvier in myelinated large diameter axons (Bekku et al., 2009) and disruption of this region results in a delay in axonal conduction (Bekku et al., 2010). Interestingly the gene *HAPLN2* (hyaluronan and proteoglycan link protein 2) is tandemly arranged upstream of *BCAN*. *HAPLN2* encodes Bral1, a brain specific hyaluronan and protoglycan link protein and is co-localised with brevican and versican V2 to form complexes at the nodes of Ranvier (Bekku et al., 2010). The *BCAN* deletion moves the 3′ UTR of *HAPLN2* to within 2 kb of exon 4 of *BCAN*. Expression analysis would be required to fully establish whether the ~16 kb deletion causes a complete knock-out of the *BCAN* gene and to investigate any potential effects on *HAPLN2* expression.

Mutations in *BCAN* have not previously been associated with a disease phenotype and brevican-deficient mice are viable, fertile, physiologically normal, display normal behaviour and have a normal life expectancy (Brakebusch et al., 2002). However, the absence of any apparent abnormalities in brevican-deficient mice may relate to an absence of episode triggers within the environment the mice were maintained in. In dogs, episodes are induced by exercise or excitement and it is highly likely that mice will not exercise at a sufficiently high intensity within their routine laboratory environment.

EF is a condition that becomes self-limiting and can self-rectify in some cases, with some dogs becoming clinically normal after a period of months to years of being clinically affected. It is interesting to speculate that this might be due to compensatory effects of other chondroitin sulphate proteoglycans in the brain, in particular versican V2 (Bekku et al., 2009) taking over the role of brevican, although the effect could also be due to modified owner and/or dog behaviour in response to the episodes, such as a change in exercise levels or the avoidance of trigger events once these have been identified.

An identical mutation in *BCAN* has recently been associated with EF in the CKCS, by an independent research group (Gill et al., 2011). This independent investigation helps to further validate the association of *BCAN* with EF.

3.3.6. *FAM83H*

Several mutations in the CKCSID-associated gene *FAM83H* have been associated with autosomal-dominant hypocalcification amelogenesis imperfecta (ADHAI) in humans, which is a disease of faulty tooth enamel formation (Kim et al., 2008). To date the mutations associated with ADHAI have all been found in exon 5 of *FAM83H* and are either nonsense or frameshift mutations leading to a premature stop codon after a sequence of aberrant amino acids. Further to this, mutations in the 5′ region of exon 5 appear to result in a generalised phenotype, affecting all teeth, compared to mutations occurring in the 3′ region, which appear to give a localised phenotype, with just a subset of teeth being affected (Urzua et al., 2011). The canine mutation is at a position within the gene that would predict a more generalised phenotype. Anecdotal evidence suggests that CKCSID cases do show clinical signs of tooth disease, although this is a post-hoc observation and would require further investigation to determine the exact nature of the dental problems.

The CKCSID phenotype suggests that *FAM83H* has an important role in skin development and regulation, in addition to enamel formation, at least in the dog. Limited expression analysis has revealed that *FAM83H* is expressed in canine skin, and also in the brain (cerebellum), footpad and buccal epithelia, in concordance with previous reports that *FAM83H* may be ubiquitously expressed (Kim et al., 2008). Species-specific differences in gene expression and function have not currently been investigated and no significant skin or nail phenotypes have been associated with ADHAI in human patients. In humans all *FAM83H* mutations reported to date have been dominant, and no human patients with homozygous *FAM83H* mutations have been reported. In contrast the canine mutation is recessive and heterozygous dogs do not have a discernable phenotype, so it is interesting to speculate that the gene is playing a different role in enamel formation between the two species and that human patients may present additional phenotypes, similar to CKCSID, if a deleterious homozygous *FAM83H* mutation was identified.

3.3.7. Summary

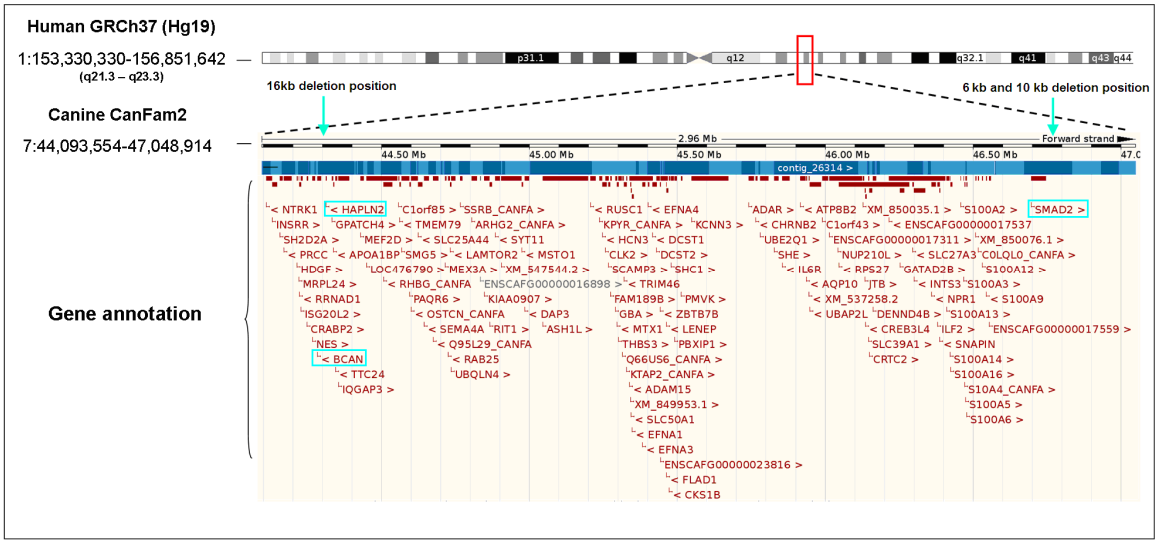
Mutations in *BCAN* and *FAM83H* have been identified which strongly associate with EF and CKCSID respectively, using an efficient parallel mapping and simultaneous sequencing approach. Neither of the two genes had been previously associated with similar disease phenotypes in other species. The discovery of these mutations may suggest potential novel biological functions for *FAM83H* and *BCAN*, although formal proof of this would require further functional data to confirm the causality of the two mutations with respect to their associated disease phenotypes. The study illustrates how two disease phenotypes in a single dog breed can be investigated using a very modest

sample set to successfully identify disease-associated mutations, using resources available in-house where possible to maximise cost effectiveness and efficiency.

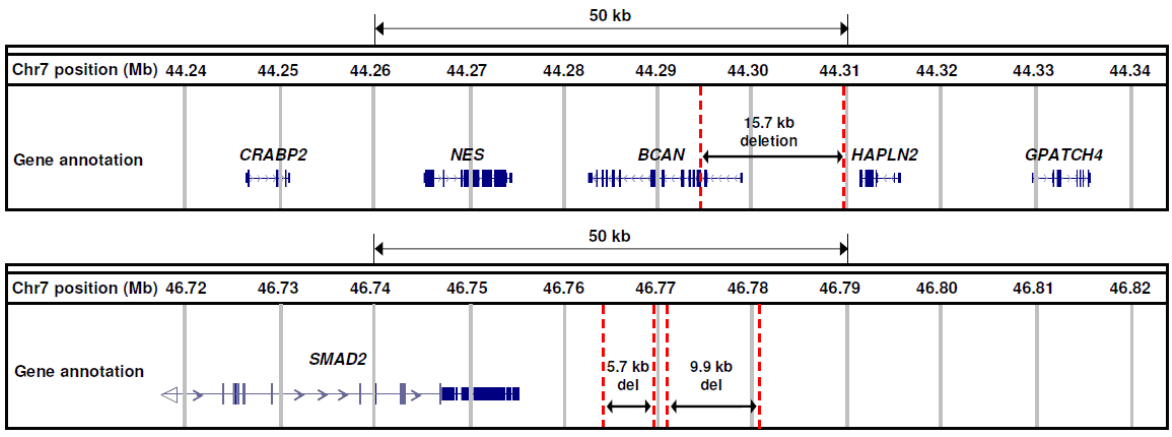
Appendix 6 CKCS critical regions, features and human synteny

EF disease-associated interval

Deletion positions are indicated by the blue arrows, with genes in close proximity highlighted with a blue box.

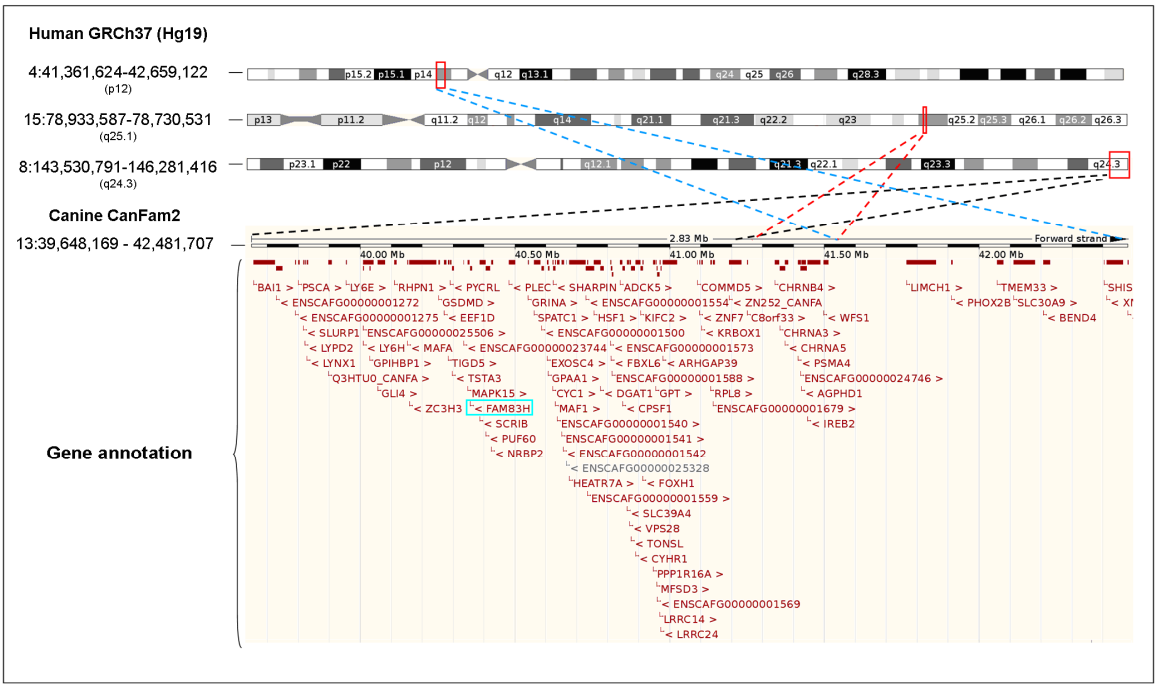


Schematic overview of the genes in close proximity to the three identified deletions.



CKCSID disease-associated interval

FAM83H is highlighted with a blue box.



List of Tables

List of Figures

List of Appendices

Acknowledgments

I would like to thank the Animal Health Trust for giving me the opportunity to study towards my PhD and all the amazing people that work there. In particular Louise Pettitt for being a great friend and for making work such good fun. I would like to thank Dr Mike Boursnell for introducing me to Linux and Perl, and being enormously patient and helpful, and Graham Newland for IT support.

I would like to thank my University of Glasgow supervisor Professor Jacques Penderis for helping to initiate my PhD and for help and support throughout my studies. I am particularly thankful to my supervisor at the Animal Health Trust, Dr Cathryn Mellersh, for her leadership and being incredibly patient and supportive throughout.

I would like to thank my parents and family for their love and support and I am especially thankful to my wife Roxanna for being loving and supportive, believing in me and always being there for me. I'm proud of everything we have achieved together.

Publications

Forman, O. P., Penderis, J., Hartley, C., Hayward, L. J., Ricketts, S. L. & Mellersh, C. S. (2012) Parallel mapping and simultaneous sequencing reveals deletions in *BCAN* and *FAM83H* associated with discrete inherited disorders in a domestic dog breed. *PLoS Genetics*, 8, e1002462.

Forman, O. P., De Risio, L., Stewart, J., Mellersh, C. S. & Beltran, E. (2012) Genome-wide mRNA sequencing of a single canine cerebellar cortical degeneration case leads to the identification of a disease associated *SPTBN2* mutation. *BMC Genetics*, 13, 55.

Forman, O. P., De Risio, L., & Mellersh, C. S. (2013) Missense mutation in *CAPN1* is associated with spinocerebellar ataxia in the Parson Russell Terrier dog breed. *PLoS One,* 8, e64627.

Forman, O. P., De Risio, L., Matiasek, K., Platt, S. R. & Mellersh, C. S. (2013) Spinocerebellar ataxia in the Italian Spinone dog breed is associated with an intronic GAA repeat expansion in *ITPR1*. *Acta Neuropathologica* (*In preparation*).

Declaration

I declare that this thesis is my own original work, and has not been submitted for an award at any other university. Contributions from service providers and individuals are acknowledged in the text.

Oliver Forman
9th September 2013

Abbreviations

ADHAI	Autosomal-dominant hypocalcification *amelogenesis imperfecta*
AHT	Animal Health Trust
BAC	Bacterial artificial chromosome
BAEPs	Brainstem auditory-evoked potentials
BAER	Brainstem auditory-evoked responses
BAM	Binary sequence alignment/map
BCAN	Brevican
BED	Browser extensible data
BLAST	Basic local alignment search tool
BLOSUM	Blocks substitution matrix
BWA	Burrows-Wheeler aligner
C2L	C2 like domain
CAPN1	Calcium dependent cysteine protease, calpain1
ChIP-seq	Chromatin immunoprecipitation sequencing
CKCS	Cavalier King Charles Spaniel
CKCSID	Congenital keratoconjunctivitis sicca and ichthyosiform dermatosis
CNV	Copy number variation
CSKDD	Committee for the standardisation of the karyotype in the domestic dog
Ct	Threshold cycle
DENND4B	DENN/MADD domain containing 4B
DMD	Dystrophin gene
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EB	Empty basket
EDTA	Ethylenediaminetetraacetic acid
EF	Episodic falling
ENCODE	Encyclopedia of DNA elements
ESTs	Expressed sequence tags
FFPE	Formalin fixed paraffin embedded
FISH	Fluorescent *in situ* hybridisation
FMM	Fast mixed model
FXN	Frataxin
GA	Genome Analyser
GABA	Gamma-aminobutyric acid
GATK	Genome Analysis ToolKit
GBS	Genotyping-by-sequencing
GENO	SNP genotyping frequency
GSP	Gene specific primer
GWAS	Genome-wide association study

HCRTR2	Hypocretin (orexin) receptor 2
IBS	Identity-by-state
IGV	Integrative Genomics Viewer
Indel	Insertion deletion polymorphism
IP3	Inositol triphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITPR1	Inositol 1,4,5-trisphosphate receptor, type 1
JRT	Jack Russell Terrier
KID	Keratitis-ichthyosis-deafness syndrome
LB	Lysogeny broth
LD	Linkage disequilibrium
LOA	Late onset ataxia
LOD	Logarithm of odds
MAF	Minor allele frequency
MDS	Multidimensional scaling
MIND	Percentage SNP missingness per individual
MPSS	Massively parallel signature sequencing
mRNA-seq	Genome-wide mRNA sequencing
NCCD	Neonatal cerebellar cortical degeneration
NGS	Next generation sequencing
NHGRI	National Human Genome Research Institute
OMIA	Online Mendelian Inheritance in Animals
OMIM	Online Mendelian Inheritance in Man
PBS	Phosphate-buffered saline
PBS-T	Phosphate-buffered saline/0.1% Tween 20
PCD	Protease core domain
PC	Purkinje cells
PCR	Polymerase chain reaction
PEF	Penta EF-hand calcium binding domain
PGM	Personal Genome Machine
Polyphen	Polymorphism Phenotyping
PPi	Pyrophosphate group
PRCD	Progressive rod-cone degeneration
PRT	Parson Russell Terrier
PTP	PicoTitre plate
QPCR	Quantitative polymerase chain reaction
QRT-PCR	Quantitative reverse transcription polymerase chain reaction
QQ	Quantile-quantile
RACE	Rapid amplification of cDNA ends
RH-mapping	Radiation hybrid mapping
RIN	RNA integrity number
RNA	Ribonucleic acid

RPE65	Retinal pigment epithelium-specific 65 kDa protein
RT-PCR	Reverse transcription PCR
SAM	Sequence Alignment/Map
SBS	Sequencing by synthesis
SCA	Spinocerebellar ataxia
SCAR	Spinocerebellar ataxia recessive
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SHFT	Smooth-Haired Fox Terrier
SIFT	Sorting Intolerant from Tolerant
SINEs	Short interspersed nuclear elements
SNP	Single nucleotide polymorphism
SOC medium	Super optimal broth with catabolite repression medium
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
SPTBN2	Beta-III spectrin
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
UTR	Untranslated region
w/t	Wild-type

References

Acland, G. M., Aguirre, G. D., Ray, J., Zhang, Q., Aleman, T. S., Cideciyan, A. V., Pearce-Kelling, S. E., Anand, V., Zeng, Y., Maguire, A. M., Jacobson, S. G., Hauswirth, W. W. & Bennett, J. (2001) Gene therapy restores vision in a canine model of childhood blindness. *Nat Genet,* 28**,** 92-5.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. & Sunyaev, S. R. (2010) A method and server for predicting damaging missense mutations. *Nat Methods,* 7**,** 248-9.

Asan, N. F., Xu, Y., Jiang, H., Tyler-Smith, C., Xue, Y., Jiang, T., Wang, J., Wu, M., Liu, X., Tian, G., Yang, H. & Zhang, X. (2011) Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol,* 12**,** R95.

Astle, W. & Balding, D. J. (2009) Population structure and cryptic relatedness in genetic association studies. *Statistical Science,* 24**,** 451–471.

Azam, M., Andrabi, S. S., Sahr, K. E., Kamath, L., Kuliopulos, A. & Chishti, A. H. (2001) Disruption of the mouse mu-calpain gene reveals an essential role in platelet function. *Mol Cell Biol,* 21**,** 2213-20.

Banerjee, S. & Hasan, G. (2005) The InsP3 receptor: its role in neuronal physiology and neurodegeneration. *Bioessays,* 27**,** 1035-47.

Bannai, H., Fukatsu, K., Mizutani, A., Natsume, T., Iemura, S., Ikegami, T., Inoue, T. & Mikoshiba, K. (2004) An RNA-interacting protein, SYNCRIP (heterogeneous nuclear ribonuclear protein Q1/NSAP1) is a component of mRNA granule transported with inositol 1,4,5-trisphosphate receptor type 1 mRNA in neuronal dendrites. *J Biol Chem,* 279**,** 53427-34.

Barnett, K. C. (2006) Congenital keratoconjunctivitis sicca and ichthyosiform dermatosis in the cavalier King Charles spaniel. *J Small Anim Pract,* 47**,** 524-8.

Barrett, A. J. & Rawlings, N. D. (2001) Evolutionary lines of cysteine peptidases. *Biol Chem,* 382**,** 727-33.

Bauer, P., Schols, L. & Riess, O. (2006) Spectrin mutations in spinocerebellar ataxia (SCA). *Bioessays,* 28**,** 785-7.

Beardow, A. W. & Buchanan, J. W. (1993) Chronic mitral valve disease in cavalier King Charles spaniels: 95 cases (1987-1991). *J Am Vet Med Assoc,* 203**,** 1023-9.

Beck, C. L., Fahlke, C. & George, A. L., Jr. (1996) Molecular basis for decreased muscle chloride conductance in the myotonic goat. *Proc Natl Acad Sci U S A,* 93**,** 11248-52.

Becker, R. E. (1977) Myotonia congenita and syndromes associated with myotonia. *Top Hum Gen,* Vol. III. .

Bekku, Y., Rauch, U., Ninomiya, Y. & Oohashi, T. (2009) Brevican distinctively assembles extracellular components at the large diameter nodes of Ranvier in the CNS. *J Neurochem,* 108**,** 1266-76.

Bekku, Y., Vargova, L., Goto, Y., Vorisek, I., Dmytrenko, L., Narasaki, M., Ohtsuka, A., Fassler, R., Ninomiya, Y., Sykova, E. & Oohashi, T. (2010) Bral1: its role in diffusion barrier formation and conduction velocity in the CNS. *J Neurosci,* 30**,** 3113-23.

Bennett, J., Ashtari, M., Wellman, J., Marshall, K. A., Cyckowski, L. L., Chung, D. C., McCague, S., Pierce, E. A., Chen, Y., Bennicelli, J. L., Zhu, X., Ying, G. S., Sun, J., Wright, J. F., Auricchio, A., Simonelli, F., Shindler, K. S., Mingozzi, F., High, K. A. & Maguire, A. M. (2012) AAV2 gene therapy readministration in three adults with congenital blindness. *Sci Transl Med,* 4**,** 120ra15.

Bennett, V. & Baines, A. J. (2001) Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues. *Physiol Rev,* 81**,** 1353-92.

Berti, P. J. & Storer, A. C. (1995) Alignment/phylogeny of the papain superfamily of cysteine proteases. *J Mol Biol,* 246**,** 273-83.

Bjorck, G., Dyrendahl, S., Olsson, S. E. (1957) Hereditary ataxia in Smooth-haired Fox Terriers. *Vet Rec,* 69**,** 87--92.

Bjorck, G., Mair, W., Olsson, S. E., Sourander, P. (1962) Hereditary ataxia in Fox Terriers. *Acta Neuropath,* 1**,** 45-48.

Boersma, A., Zonnevylle, H., Sanchez, M. A. & Diaz Espineira, M. (1995) Progressive ataxia in a rottweiler dog. *Vet Q,* 17**,** 108-9.

Bonfield, J. K., Smith, K. & Staden, R. (1995) A new DNA sequence assembly program. *Nucleic Acids Res,* 23**,** 4992-9.

Bornman, D. M., Hester, M. E., Schuetter, J. M., Kasoji, M. D., Minard-Smith, A., Barden, C. A., Nelson, S. C., Godbold, G. D., Baker, C. H., Yang, B., Walther, J. E., Tornes, I. E., Yan, P. S., Rodriguez, B., Bundschuh, R., Dickens, M. L., Young, B. A. & Faith, S. A. (2012) Short-read, high-throughput sequencing technology for STR genotyping. *Biotechniques*.

Brakebusch, C., Seidenbecher, C. I., Asztely, F., Rauch, U., Matthies, H., Meyer, H., Krug, M., Bockers, T. M., Zhou, X., Kreutz, M. R., Montag, D., Gundelfinger, E. D. & Fassler, R. (2002) Brevican-deficient mice display impaired hippocampal CA1 long-term potentiation but show no obvious deficits in learning and memory. *Mol Cell Biol,* 22**,** 7417-27.

Bralic, M., Stemberga, V. & Stifter, S. (2012) Introduction of calpain inhibitors in traumatic brain injury: a novel approach? *Med Hypotheses,* 79**,** 358-60.

Breen, M., Jouquand, S., Renier, C., Mellersh, C. S., Hitte, C., Holmes, N. G., Cheron, A., Suter, N., Vignaux, F., Bristow, A. E., Priat, C., McCann, E., Andre, C., Boundy, S., Gitsham, P., Thomas, R., Bridge, W. L., Spriggs, H. F., Ryder, E. J., Curson, A., Sampson, J., Ostrander, E. A., Binns, M. M. & Galibert, F. (2001) Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Res,* 11**,** 1784-95.

Breen, M., Reimann, N., Bosma, A. A., Landon, D., Zijlstra, C., Bartnitzke, S., Switonski, M., Long, S. E., de Haan, N. A. & Binns, M. M. (1998) *Standardisation of the chromosome nos. 22-38 of the dog (Canis familiaris) with the use of chromosome painting probes. Proceedings of the 13th European Colloquium on Cytogenetics of Domestic Animals. June 1-6. Hungarian Academy of Sciences, Budapest, Hungary.*

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S. R., Moon, K., Burcham, T., Pallas, M., DuBridge, R. B., Kirchner, J., Fearon, K., Mao, J. & Corcoran, K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol,* 18**,** 630-4.

Brody, I. A. (1969) Muscle contracture induced by exercise. A syndrome attributable to decreased relaxing factor. *N Engl J Med,* 281**,** 187-92.

Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Canizares, J., Koutnikova, H., Bidichandani, S. I., Gellera, C., Brice, A., Trouillas, P., De Michele, G., Filla, A., De Frutos, R., Palau, F., Patel, P. I., Di Donato, S., Mandel, J. L., Cocozza, S., Koenig, M. & Pandolfo, M. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science,* 271**,** 1423-7.

Carmichael, S., Griffiths, I. R. & Harvey, M. J. (1983) Familial cerebellar ataxia with hydrocephalus in bull mastiffs. *Vet Rec,* 112**,** 354-8.

Chieffo, C., Stalis, I. H., Van Winkle, T. J., Haskins, M. E. & Patterson, D. F. (1994) Cerebellar Purkinje's cell degeneration and coat color dilution in a family of Rhodesian Ridgeback dogs. *J Vet Intern Med,* 8**,** 112-6.

Clark, R. G., Hartley, W. J., Burgess, G. S., Cameron, J. S. & Mitchell, G. (1982) Suspected inherited cerebellar neuroaxonal dystrophy in collie sheep dogs. *N Z Vet J,* 30**,** 102-3.

Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S. & Bayley, H. (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol,* 4**,** 265-70.

Clarkson, Y. L., Gillespie, T., Perkins, E. M., Lyndon, A. R. & Jackson, M. (2010) Beta-III spectrin mutation L253P associated with spinocerebellar ataxia type 5 interferes with binding to Arp1 and protein trafficking from the Golgi. *Hum Mol Genet,* 19**,** 3634-41.

Coates, J. R., O'Brien, D. P., Kline, K. L., Storts, R. W., Johnson, G. C., Shelton, G. D., Patterson, E. E. & Abbott, L. C. (2002) Neonatal cerebellar ataxia in Coton de Tulear dogs. *J Vet Intern Med,* 16**,** 680-9.

Copeland, H., Dukes-McEwan, J., Sargan, D., Kennedy, L., Starkey, M., Hendricks, A. & Callanan, S. (2008) LUPA - studying human diseases using dog genetics. *Vet Rec,* 163**,** 550.

Cossee, M., Schmitt, M., Campuzano, V., Reutenauer, L., Moutou, C., Mandel, J. L. & Koenig, M. (1997) Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. *Proc Natl Acad Sci U S A,* 94**,** 7452-7.

Credille, K. M., Barnhart, K. F., Minor, J. S. & Dunstan, R. W. (2005) Mild recessive epidermolytic hyperkeratosis associated with a novel keratin 10 donor splice-site mutation in a family of Norfolk terrier dogs. *Br J Dermatol,* 153**,** 51-8.

Credille, K. M., Minor, J. S., Barnhart, K. F., Lee, E., Cox, M. L., Tucker, K. A., Diegel, K. L., Venta, P. J., Hohl, D., Huber, M. & Dunstan, R. W. (2009) Transglutaminase 1-deficient recessive lamellar ichthyosis associated with a LINE-1 insertion in Jack Russell terrier dogs. *Br J Dermatol,* 161**,** 265-72.

Cremers, F. P., van den Hurk, J. A. & den Hollander, A. I. (2002) Molecular genetics of Leber congenital amaurosis. *Hum Mol Genet,* 11**,** 1169-76.

Cunliffe, J. (2004) *Cavalier King Charles Spaniel*, Kennel Club Books.

Darke, P. G. & Kelly, D. F. (1976) Correspondence: Cerebellar ataxia in the kerry blue terrier. *Vet Rec,* 98**,** 307.

Deelman, L. E., Jonk, L. J. & Henning, R. H. (1998) The isolation and characterization of the promoter of the human type 1 inositol 1,4,5-trisphosphate receptor. *Gene,* 207**,** 219-25.

Demarchi, F. & Schneider, C. (2007) The calpain system as a modulator of stress/damage response. *Cell Cycle,* 6**,** 136-8.

Derrien, T., Vaysse, A., Andre, C. & Hitte, C. (2012) Annotation of the domestic dog genome sequence: finding the missing genes. *Mamm Genome,* 23**,** 124-31.

Ding, Z. L., Oskarsson, M., Ardalan, A., Angleby, H., Dahlgren, L. G., Tepeli, C., Kirkness, E., Savolainen, P. & Zhang, Y. P. (2011) Origins of domestic dog in Southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity,* 108, 507-14.

Durr, A., Cossee, M., Agid, Y., Campuzano, V., Mignard, C., Penet, C., Mandel, J. L., Brice, A. & Koenig, M. (1996) Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N Engl J Med,* 335**,** 1169-75.

Dutt, P., Croall, D. E., Arthur, J. S., Veyra, T. D., Williams, K., Elce, J. S. & Greer, P. A. (2006) m-Calpain is required for preimplantation embryonic development in mice. *BMC Dev Biol,* 6**,** 3.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J. & Turner, S. (2009) Real-time DNA sequencing from single polymerase molecules. *Science,* 323**,** 133-8.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. & Mitchell, S. E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One,* 6**,** e19379.

Epplen, C., Epplen, J. T., Frank, G., Miterski, B., Santos, E. J. & Schols, L. (1997) Differential stability of the (GAA)n tract in the Friedreich ataxia (STM7) gene. *Hum Genet,* 99**,** 834-6.

Faddis, B. T., Hasbani, M. J. & Goldberg, M. P. (1997) Calpain activation contributes to dendritic remodeling after brief excitotoxic injury *in vitro*. *J Neurosci,* 17**,** 951-9.

Finnigan, D. F., Hanna, W. J., Poma, R. & Bendall, A. J. (2007) A novel mutation of the *CLCN1* gene associated with myotonia hereditaria in an Australian cattle dog. *J Vet Intern Med,* 21**,** 458-63.

Fischer, J., Bouadjar, B., Heilig, R., Huber, M., Lefevre, C., Jobard, F., Macari, F., Bakija-Konsuo, A., Ait-Belkacem, F., Weissenbach, J., Lathrop, M., Hohl, D. & Prud'homme, J. F. (2001) Mutations in the gene encoding SLURP-1 in Mal de Meleda. *Hum Mol Genet,* 10**,** 875-80.

Furuichi, T., Simon-Chazottes, D., Fujino, I., Yamada, N., Hasegawa, M., Miyawaki, A., Yoshikawa, S., Guenet, J. L. & Mikoshiba, K. (1993) Widespread expression of inositol 1,4,5-trisphosphate receptor type 1 gene (*Insp3r1*) in the mouse central nervous system. *Recept Chan,* 1**,** 11-24.

Gandini, G., Botteron, C., Brini, E., Fatzer, R., Diana, A. & Jaggy, A. (2005) Cerebellar cortical degeneration in three English bulldogs: clinical and neuropathological findings. *J Small Anim Pract,* 46**,** 291-4.

Gao, Y., Perkins, E. M., Clarkson, Y. L., Tobia, S., Lyndon, A. R., Jackson, M. & Rothstein, J. D. (2011) beta-III spectrin is critical for development of purkinje cell dendritic tree and spine morphogenesis. *J Neurosci,* 31**,** 16581-90.

Garosi, L. S., Platt, S. & Shelton, G. D. (2002) Hypertonicity in the Cavalier King Charles Spaniel. *J Vet Intern Med,* 16**,** 330.

George, A. L., Jr., Crackower, M. A., Abdalla, J. A., Hudson, A. J. & Ebers, G. C. (1993) Molecular basis of Thomsen's disease (autosomal dominant myotonia congenita). *Nat Genet,* 3**,** 305-10.

Gill, J. L., Tsai, K. L., Krey, C., Noorai, R. E., Vanbellinghen, J. F., Garosi, L. S., Shelton, G. D., Clark, L. A. & Harvey, R. J. (2011) A canine *BCAN* microdeletion associated with episodic falling syndrome. *Neurobiol Dis,* 45**,** 130-6.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S. & Nusbaum, C. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol,* 27**,** 182-9.

Grall, A., Guaguere, E., Planchais, S., Grond, S., Bourrat, E., Hausser, I., Hitte, C., Le Gallo, M., Derbois, C., Kim, G. J., Lagoutte, L., Degorce-Rubiales, F., Radner, F. P., Thomas, A., Kury, S., Bensignor, E., Fontaine, J., Pin, D., Zimmermann, R., Zechner, R., Lathrop, M., Galibert, F., Andre, C. & Fischer, J. (2012) *PNPLA1* mutations cause autosomal recessive congenital ichthyosis in golden retriever dogs and humans. *Nat Genet,* 44**,** 140-7.

Grantham, R. (1974) Amino acid difference formula to help explain protein evolution. *Science,* 185**,** 862-4.

Hanzlicek, D., Kathmann, I., Bley, T., Srenk, P., Botteron, C., Gaillard, C. & Jaggy, A. (2003) [Cerebellar cortical abiotrophy in American Staffordshire terriers: clinical and pathological description of 3 cases]. *Schweiz Arch Tierheilkd,* 145**,** 369-75.

Hara, K., Shiga, A., Nozaki, H., Mitsui, J., Takahashi, Y., Ishiguro, H., Yomono, H., Kurisaki, H., Goto, J., Ikeuchi, T., Tsuji, S., Nishizawa, M. & Onodera, O. (2008) Total deletion and a missense mutation of *ITPR1* in Japanese SCA15 families. *Neurology,* 71**,** 547-51.

Hartley, C., Barnett, K. C., Pettitt, L., Forman, O. P., Blott, S. & Mellersh, C. S. (2012) Congenital keratoconjunctivitis sicca and ichthyosiform dermatosis in Cavalier King Charles spaniel dogs-part II: candidate gene study. *Vet Ophthalmol*, 15, 327-32.

Hartley, C., Donaldson, D., Smith, K. C., Henley, W., Lewis, T. W., Blott, S., Mellersh, C. & Barnett, K. C. (2011) Congenital keratoconjunctivitis sicca and ichthyosiform dermatosis in 25 Cavalier King Charles spaniel dogs - part I: clinical signs, histopathology, and inheritance. *Vet Ophthalmol*. 5, 315-26

Hartley, W. J. & Palmer, A. C. (1973) Ataxia in Jack Russell terriers. *Acta Neuropathol,* 26**,** 71-4.

Hedan, B., Thomas, R., Motsinger-Reif, A., Abadie, J., Andre, C., Cullen, J. & Breen, M. (2011) Molecular cytogenetic characterization of canine histiocytic sarcoma: A spontaneous model for human histiocytic cancer identifies deletion of tumor suppressor genes and highlights influence of genetic background on tumor behavior. *BMC Cancer,* 11**,** 201.

Henikoff, S. & Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A,* 89**,** 10915-9.

Hernandez, D., Francois, P., Farinelli, L., Osteras, M. & Schrenzel, J. (2008) De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res,* 18**,** 802-9.

Herrtage, M. E. & Palmer, A. C. (1983) Episodic falling in the cavalier King Charles spaniel. *Vet Rec,* 112**,** 458-9.

Higgins, J. J., Nee, L. E., Vasconcelos, O., Ide, S. E., Lavedan, C., Goldfarb, L. G. & Polymeropoulos, M. H. (1996) Mutations in American families with spinocerebellar ataxia (SCA) type 3: SCA3 is allelic to Machado-Joseph disease. *Neurology,* 46**,** 208-13.

Higgins, R. J., LeCouteur, R. A., Kornegay, J. N. & Coates, J. R. (1998) Late-onset progressive spinocerebellar degeneration in Brittany Spaniel dogs. *Acta Neuropathol,* 96**,** 97-101.

Hillbertz, N. H. & Andersson, G. (2006) Autosomal dominant mutation causing the dorsal ridge predisposes for dermoid sinus in Rhodesian ridgeback dogs. *J Small Anim Pract,* 47**,** 184-8.

Hitte, C., Derrien, T., Andre, C., Ostrander, E. A. & Galibert, F. (2004) CRH_Server: an online comparative and radiation hybrid mapping server for the canine genome. *Bioinformatics,* 20**,** 3665-7.

Holleran, E. A., Ligon, L. A., Tokito, M., Stankewich, M. C., Morrow, J. S. & Holzbaur, E. L. (2001) beta III spectrin binds to the Arp1 subunit of dynactin. *J Biol Chem,* 276**,** 36598-605.

Holmes, S. E., O'Hearn, E. E., McInnis, M. G., Gorelick-Feldman, D. A., Kleiderlein, J. J., Callahan, C., Kwak, N. G., Ingersoll-Ashworth, R. G., Sherr, M., Sumner, A. J., Sharp, A. H., Ananth, U., Seltzer, W. K., Boss, M. A., Vieria-Saecker, A. M., Epplen, J. T., Riess, O., Ross, C. A. & Margolis, R. L. (1999) Expansion of a novel CAG trinucleotide repeat in the 5' region of *PPP2R2B* is associated with SCA12. *Nat Genet,* 23**,** 391-2.

Hosfield, C. M., Elce, J. S., Davies, P. L. & Jia, Z. (1999) Crystal structure of calpain reveals the structural basis for Ca(2+)-dependent protease activity and a novel mode of enzyme activation. *EMBO J,* 18**,** 6880-9.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I. & Clamp, M. (2002) The Ensembl genome database project. *Nucleic Acids Res,* 30**,** 38-41.

Ikeda, Y., Dick, K. A., Weatherspoon, M. R., Gincel, D., Armbrust, K. R., Dalton, J. C., Stevanin, G., Durr, A., Zuhlke, C., Burk, K., Clark, H. B., Brice, A., Rothstein, J. D., Schut, L. J., Day, J. W. & Ranum, L. P. (2006) Spectrin mutations cause spinocerebellar ataxia type 5. *Nat Genet,* 38**,** 184-90.

Iwaki, A., Kawano, Y., Miura, S., Shibata, H., Matsuse, D., Li, W., Furuya, H., Ohyagi, Y., Taniwaki, T., Kira, J. & Fukumaki, Y. (2008) Heterozygous deletion of *ITPR1*, but not *SUMF1*, in spinocerebellar ataxia type 16. *J Med Genet,* 45**,** 32-5.

Jackson, M., Song, W., Liu, M. Y., Jin, L., Dykes-Hoberg, M., Lin, C. I., Bowers, W. J., Federoff, H. J., Sternweis, P. C. & Rothstein, J. D. (2001) Modulation of the

neuronal glutamate transporter EAAT4 by two interacting proteins. *Nature,* 410**,** 89-93.

Kampfl, A., Posmantur, R., Nixon, R., Grynspan, F., Zhao, X., Liu, S. J., Newcomb, J. K., Clifton, G. L. & Hayes, R. L. (1996) mu-calpain activation and calpain-mediated cytoskeletal proteolysis following traumatic brain injury. *J Neurochem,* 67**,** 1575-83.

The Kennel Club http://www.thekennelclub.org.uk/ (Accesssed on 18th May 2012)

The Kennel Club http://www.thekennelclub.org.uk/download/7748/top20breedreg.pdf (Accesssed on 16th April 2013)

Kent, M., Glass, E. & deLahunta, A. (2000) Cerebellar cortical abiotrophy in a beagle. *J Small Anim Pract,* 41**,** 321-3.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. (2002) The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

Kim, J. W., Lee, S. K., Lee, Z. H., Park, J. C., Lee, K. E., Lee, M. H., Park, J. T., Seo, B. M., Hu, J. C. & Simmer, J. P. (2008) *FAM83H* mutations in families with autosomal-dominant hypocalcified amelogenesis imperfecta. *Am J Hum Genet,* 82**,** 489-94.

Kirkness, E. F., Bafna, V., Halpern, A. L., Levy, S., Remington, K., Rusch, D. B., Delcher, A. L., Pop, M., Wang, W., Fraser, C. M. & Venter, J. C. (2003) The dog genome: survey sequencing and comparative analysis. *Science,* 301**,** 1898-903.

Kirkwood, K. L., Homick, K., Dragon, M. B. & Bradford, P. G. (1997) Cloning and characterization of the type I inositol 1,4,5-trisphosphate receptor gene promoter. Regulation by 17beta-estradiol in osteoblasts. *J Biol Chem,* 272**,** 22425-31.

Koch, M. C., Steinmeyer, K., Lorenz, C., Ricker, K., Wolf, F., Otto, M., Zoll, B., Lehmann-Horn, F., Grzeschik, K. H. & Jentsch, T. J. (1992) The skeletal muscle chloride channel in dominant and recessive human myotonia. *Science,* 257**,** 797-800.

Koide, R., Kobayashi, S., Shimohata, T., Ikeuchi, T., Maruyama, M., Saito, M., Yamada, M., Takahashi, H. & Tsuji, S. (1999) A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? *Hum Mol Genet,* 8**,** 2047-53.

Koob, M. D., Moseley, M. L., Schut, L. J., Benzow, K. A., Bird, T. D., Day, J. W. & Ranum, L. P. (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat Genet,* 21**,** 379-84.

Kumar, S., Tao, C., Chien, M., Hellner, B., Balijepalli, A., Robertson, J. W., Li, Z., Russo, J. J., Reiner, J. E., Kasianowicz, J. J. & Ju, J. (2012) PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. *Sci Rep,* 2**,** 684.

Lander, E. & Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet,* 11**,** 241-7.

Lathrop, G. M. & Lalouel, J. M. (1984) Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet,* 36**,** 460-5.

Le Meur, G., Stieger, K., Smith, A. J., Weber, M., Deschamps, J. Y., Nivard, D., Mendes-Madeira, A., Provost, N., Pereon, Y., Cherel, Y., Ali, R. R., Hamel, C., Moullier, P. & Rolling, F. (2007) Restoration of vision in RPE65-deficient Briard dogs using an AAV serotype 4 vector that specifically targets the retinal pigmented epithelium. *Gene Ther,* 14**,** 292-303.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

Li, H., Ruan, J. & Durbin, R. (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*, 11, 1851-8.

Lieberman, K. R., Cherf, G. M., Doody, M. J., Olasagasti, F., Kolodji, Y. & Akeson, M. (2010) Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. *J Am Chem Soc,* 132**,** 17961-72.

Lin, L., Faraco, J., Li, R., Kadotani, H., Rogers, W., Lin, X., Qiu, X., de Jong, P. J., Nishino, S. & Mignot, E. (1999) The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell,* 98**,** 365-76.

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., 3rd, Zody, M. C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., DeJong, P. J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C. W., Cook, A., Cuff, J., Daly, M. J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K. P., Parker, H. G., Pollinger, J. P., Searle, S. M., Sutter, N. B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature,* 438**,** 803-19.

Lindblad, K., Savontaus, M. L., Stevanin, G., Holmberg, M., Digre, K., Zander, C., Ehrsson, H., David, G., Benomar, A., Nikoskelainen, E., Trottier, Y., Holmgren, G., Ptacek, L. J., Anttinen, A., Brice, A. & Schalling, M. (1996) An expanded CAG repeat sequence in spinocerebellar ataxia type 7. *Genome Res,* 6**,** 965-71.

Lise, S., Clarkson, Y., Perkins, E., Kwasniewska, A., Sadighi Akha, E., Schnekenberg, R. P., Suminaite, D., Hope, J., Baker, I., Gregory, L., Green, A., Allan, C., Lamble, S., Jayawant, S., Quaghebeur, G., Cader, M. Z., Hughes, S., Armstrong, R. J., Kanapin, A., Rimmer, A., Lunter, G., Mathieson, I., Cazier, J. B., Buck, D., Taylor, J. C., Bentley, D., McVean, G., Donnelly, P., Knight, S. J., Jackson, M., Ragoussis, J. & Nemeth, A. H. (2012) Recessive mutations in SPTBN2 implicate beta-III spectrin in both cognitive and motor development. *PLoS Genet,* 8**,** e1003074.

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J. & Turner, D. J. (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods,* 7**,** 111-8.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. & Rothberg, J. M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

Matsumoto, M., Nakagawa, T., Inoue, T., Nagata, E., Tanaka, K., Takano, H., Minowa, O., Kuno, J., Sakakibara, S., Yamada, M., Yoneshima, H., Miyawaki, A., Fukuuchi, Y., Furuichi, T., Okano, H., Mikoshiba, K. & Noda, T. (1996) Ataxia and epileptic seizures in mice lacking type 1 inositol 1,4,5-trisphosphate receptor. *Nature,* 379**,** 168-71.

Matsuura, T., Yamagata, T., Burgess, D. L., Rasmussen, A., Grewal, R. P., Watase, K., Khajavi, M., McCall, A. E., Davis, C. F., Zu, L., Achari, M., Pulst, S. M., Alonso, E.,

Noebels, J. L., Nelson, D. L., Zoghbi, H. Y. & Ashizawa, T. (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat Genet,* 26**,** 191-4.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res,* 20**,** 1297-303.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. & Cunningham, F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics,* 26**,** 2069-70.

Mellersh, C. (2008) Microsatellite-based candidate gene linkage analysis studies. *Methods Mol Biol,* 439**,** 75-86.

Mellersh, C. S., Hitte, C., Richman, M., Vignaux, F., Priat, C., Jouquand, S., Werner, P., Andre, C., DeRose, S., Patterson, D. F., Ostrander, E. A. & Galibert, F. (2000) An integrated linkage-radiation hybrid map of the canine genome. *Mamm Genome,* 11**,** 120-30.

Mellersh, C. S., Langston, A. A., Acland, G. M., Fleming, M. A., Ray, K., Wiegand, N. A., Francisco, L. V., Gibbs, M., Aguirre, G. D. & Ostrander, E. A. (1997) A linkage map of the canine genome. *Genomics,* 46**,** 326-36.

Mellersh, C. S., Pettitt, L., Forman, O. P., Vaudin, M. & Barnett, K. C. (2006) Identification of mutations in HSF4 in dogs of three different breeds with hereditary cataracts. *Vet Ophthalmol,* 9**,** 369-78.

Minouchi, O. (1928) The spermatogenesis of the dog, with special reference to meiosis. *Jap J Zoology,* 1**,** 255-268.

Miyadera, K., Kato, K., Boursnell, M., Mellersh, C. S. & Sargan, D. R. (2012) Genome-wide association study in RPGRIP1(-/-) dogs identifies a modifier locus that determines the onset of retinal degeneration. *Mamm Genome,* 23**,** 212-23.

Miyoshi, Y., Yamada, T., Tanimura, M., Taniwaki, T., Arakawa, K., Ohyagi, Y., Furuya, H., Yamamoto, K., Sakai, K., Sasazuki, T. & Kira, J. (2001) A novel autosomal dominant spinocerebellar ataxia (SCA16) linked to chromosome 8q22.1-24.1. *Neurology,* 57**,** 96-100.

National Health Service http://www.nhs.uk/conditions/ataxia (Accesssed on 16th April 2013).

Neff, M. W., Broman, K. W., Mellersh, C. S., Ray, K., Acland, G. M., Aguirre, G. D., Ziegle, J. S., Ostrander, E. A. & Rine, J. (1999) A second-generation genetic linkage map of the domestic dog, Canis familiaris. *Genetics,* 151**,** 803-20.

Ng, P. C. & Henikoff, S. (2001) Predicting deleterious amino acid substitutions. *Genome Res,* 11**,** 863-74.

Nibe, K., Kita, C., Morozumi, M., Awamura, Y., Tamura, S., Okuno, S., Kobayashi, T. & Uchida, K. (2007) Clinicopathological features of canine neuroaxonal dystrophy and cerebellar cortical abiotrophy in Papillon and Papillon-related dogs. *J Vet Med Sci,* 69**,** 1047-52.

Nishino, S., Ripley, B., Overeem, S., Lammers, G. J. & Mignot, E. (2000) Hypocretin (orexin) deficiency in human narcolepsy. *Lancet,* 355**,** 39-40.

Odermatt, A., Taschner, P. E., Khanna, V. K., Busch, H. F., Karpati, G., Jablecki, C. K., Breuning, M. H. & MacLennan, D. H. (1996) Mutations in the gene-encoding SERCA1, the fast-twitch skeletal muscle sarcoplasmic reticulum Ca^{2+} ATPase, are associated with Brody disease. *Nat Genet,* 14**,** 191-4.

Oetting, W. S., Lee, H. K., Flanders, D. J., Wiesner, G. L., Sellers, T. A. & King, R. A. (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics,* 30**,** 450-8.

OMIA http://omia.angis.org.au/ (Accesssed on 18 May 2012)

Orr, H. T., Chung, M. Y., Banfi, S., Kwiatkowski, T. J., Jr., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P. & Zoghbi, H. Y. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet,* 4**,** 221-6.

Ovodov, N. D., Crockford, S. J., Kuzmin, Y. V., Higham, T. F., Hodgins, G. W. & van der Plicht, J. (2011) A 33,000-year-old incipient dog from the Altai Mountains of Siberia: evidence of the earliest domestication disrupted by the Last Glacial Maximum. *PLoS One,* 6**,** e22821.

Pang, J. F., Kluetsch, C., Zou, X. J., Zhang, A. B., Luo, L. Y., Angleby, H., Ardalan, A., Ekstrom, C., Skollermo, A., Lundeberg, J., Matsumura, S., Leitner, T., Zhang, Y. P. & Savolainen, P. (2009) mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol Biol Evol,* 26**,** 2849-64.

Parker, H. G., Kukekova, A. V., Akey, D. T., Goldstein, O., Kirkness, E. F., Baysac, K. C., Mosher, D. S., Aguirre, G. D., Acland, G. M. & Ostrander, E. A. (2007) Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res,* 17**,** 1562-71.

Perkins, E. M., Clarkson, Y. L., Sabatier, N., Longhurst, D. M., Millward, C. P., Jack, J., Toraiwa, J., Watanabe, M., Rothstein, J. D., Lyndon, A. R., Wyllie, D. J., Dutia, M. B. & Jackson, M. (2010) Loss of beta-III spectrin leads to Purkinje cell dysfunction recapitulating the behavior and neuropathology of spinocerebellar ataxia type 5 in humans. *J Neurosci,* 30**,** 4857-67.

Peyron, C., Faraco, J., Rogers, W., Ripley, B., Overeem, S., Charnay, Y., Nevsimalova, S., Aldrich, M., Reynolds, D., Albin, R., Li, R., Hungs, M., Pedrazzoli, M., Padigaru, M., Kucherlapati, M., Fan, J., Maki, R., Lammers, G. J., Bouras, C., Kucherlapati, R., Nishino, S. & Mignot, E. (2000) A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. *Nat Med,* 6**,** 991-7.

Pierson, T. M., Adams, D., Bonn, F., Martinelli, P., Cherukuri, P. F., Teer, J. K., Hansen, N. F., Cruz, P., Mullikin For The Nisc Comparative Sequencing Program, J. C., Blakesley, R. W., Golas, G., Kwan, J., Sandler, A., Fuentes Fajardo, K., Markello, T., Tifft, C., Blackstone, C., Rugarli, E. I., Langer, T., Gahl, W. A. & Toro, C. (2011) Whole-exome sequencing identifies homozygous *AFG3L2* mutations in a spastic ataxia-neuropathy syndrome linked to mitochondrial m-AAA proteases. *PLoS Genet,* 7**,** e1002325.

Pignol, B., Auvin, S., Carre, D., Marin, J. G. & Chabrier, P. E. (2006) Calpain inhibitors and antioxidants act synergistically to prevent cell necrosis: effects of the novel dual inhibitors (cysteine protease inhibitor and antioxidant) BN 82204 and its pro-drug BN 82270. *J Neurochem,* 98**,** 1217-28.

Porreca, G. J., Zhang, K., Li, J. B., Xie, B., Austin, D., Vassallo, S. L., LeProust, E. M., Peck, B. J., Emig, C. J., Dahl, F., Gao, Y., Church, G. M. & Shendure, J. (2007) Multiplex amplification of large sets of human exons. *Nat Methods,* 4**,** 931-6.

Priat, C., Hitte, C., Vignaux, F., Renier, C., Jiang, Z., Jouquand, S., Cheron, A., Andre, C. & Galibert, F. (1998) A whole-genome radiation hybrid map of the dog genome. *Genomics,* 54**,** 361-78.

PubMeb National Center for Biotechnology Information, U.S. http://www.ncbi.nlm.nih.gov/pubmed/ (Accesssed on 18th May 2010 2010)

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. & Sham, P. C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81**,** 559-75.

Rhodes, T. H., Vite, C. H., Giger, U., Patterson, D. F., Fahlke, C. & George, A. L., Jr. (1999) A missense mutation in canine *C1C-1* causes recessive myotonia congenita in the dog. *FEBS Lett,* 456**,** 54-8.

Richard, G., Rouan, F., Willoughby, C. E., Brown, N., Chung, P., Ryynanen, M., Jabs, E. W., Bale, S. J., DiGiovanna, J. J., Uitto, J. & Russell, L. (2002) Missense mutations in *GJB2* encoding connexin-26 cause the ectodermal dysplasia keratitis-ichthyosis-deafness syndrome. *Am J Hum Genet,* 70**,** 1341-8.

Richard, I., Broux, O., Allamand, V., Fougerousse, F., Chiannilkulchai, N., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C. & et al. (1995) Mutations in

the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell,* 81**,** 27-40.

Riess, O., Schols, L., Bottger, H., Nolte, D., Vieira-Saecker, A. M., Schimming, C., Kreuz, F., Macek, M., Jr., Krebsova, A., Macek, M. S., Klockgether, T., Zuhlke, C. & Laccone, F. A. (1997) SCA6 is caused by moderate CAG expansion in the alpha1A-voltage-dependent calcium channel gene. *Hum Mol Genet,* 6**,** 1289-93.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. & Mesirov, J. P. (2011) Integrative genomics viewer. *Nat Biotechnol,* 29**,** 24-6.

Rohdin, C., Ludtke, L., Wohlsein, P. & Jaderlund, K. H. (2010) New aspects of hereditary ataxia in smooth-haired fox terriers. *Vet Rec,* 166**,** 557-60.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem,* 242**,** 84-9.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T. & Bustillo, J. (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature,* 475**,** 348-52.

Rozen, S. & Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol,* 132**,** 365-86.

Rusbridge, C. (2005) Neurological diseases of the Cavalier King Charles spaniel. *J Small Anim Pract,* 46**,** 265-72.

Saido, T. C., Yokota, M., Nagao, S., Yamaura, I., Tani, E., Tsuchiya, T., Suzuki, K. & Kawashima, S. (1993) Spatial resolution of fodrin proteolysis in postischemic brain. *J Biol Chem,* 268**,** 25239-43.

Sakaguchi, G., Orita, S., Naito, A., Maeda, M., Igarashi, H., Sasaki, T. & Takai, Y. (1998) A novel brain-specific isoform of beta spectrin: isolation and its interaction with Munc13. *Biochem Biophys Res Commun,* 248**,** 846-51.

Sakamoto, N., Chastain, P. D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J. D. & Wells, R. D. (1999) Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. *Mol Cell,* 3**,** 465-75.

Sambrook, J. (2001) Molecular Cloning: A Laboratory Manual
New York, Cold Spring Harbor Laboratory Press.

Sanchez-Carpintero Abad, R., Sanmarti Vilaplana, F. X. & Serratosa Fernandez, J. M. (2007) Genetic causes of epilepsy. *Neurologist,* 13**,** S47-51.

Sanpei, K., Takano, H., Igarashi, S., Sato, T., Oyake, M., Sasaki, H., Wakisaka, A., Tashiro, K., Ishida, Y., Ikeuchi, T., Koide, R., Saito, M., Sato, A., Tanaka, T., Hanyu, S., Takiyama, Y., Nishizawa, M., Shimizu, N., Nomura, Y., Segawa, M., Iwabuchi, K., Eguchi, I., Tanaka, H., Takahashi, H. & Tsuji, S. (1996) Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet,* 14**,** 277-84.

Sargan, D. R. (2004) IDID: inherited diseases in dogs: web-based information for canine inherited disease genetics. *Mamm Genome,* 15**,** 503-6.

Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods,* 7**,** 575-6.

Seppala, E. H., Koskinen, L. L., Gullov, C. H., Jokinen, P., Karlskov-Mortensen, P., Bergamasco, L., Baranowska Korberg, I., Cizinauskas, S., Oberbauer, A. M., Berendt, M., Fredholm, M. & Lohi, H. (2012) Identification of a novel idiopathic epilepsy locus in Belgian shepherd dogs. *PLoS One,* 7**,** e33549.

Shelton, G. D. (2004) Muscle pain, cramps and hypertonicity. *Vet Clin North Am Small Anim Pract,* 34**,** 1483-96.

Shelton, G. D. & Engvall, E. (2002) Muscular dystrophies and other inherited myopathies. *Vet Clin North Am Small Anim Pract,* 32**,** 103-24.

Shendure, J. & Ji, H. (2008) Next-generation DNA sequencing. *Nat Biotechnol,* 26**,** 1135-45.

Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., Wang, M. D., Zhang, K., Mitra, R. D. & Church, G. M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science,* 309**,** 1728-32.

Skinner, B. A., Greist, M. C. & Norins, A. L. (1981) The keratitis, ichthyosis, and deafness (KID) syndrome. *Arch Dermatol,* 117**,** 285-9.

Sorimachi, H., Ishiura, S. & Suzuki, K. (1997) Structure and physiological function of calpains. *Biochem J,* 328**,** 721-32.

Stankewich, M. C., Gwynn, B., Ardito, T., Ji, L., Kim, J., Robledo, R. F., Lux, S. E., Peters, L. L. & Morrow, J. S. (2010) Targeted deletion of betaIII spectrin impairs synaptogenesis and generates ataxic and seizure phenotypes. *Proc Natl Acad Sci U S A,* 107**,** 6022-7.

Steinberg, H. S., Troncoso, J. C., Cork, L. C. & Price, D. L. (1981) Clinical features of inherited cerebellar degeneration in Gordon setters. *J Am Vet Med Assoc,* 179**,** 886-90.

Steinlein, O. K. (2008) Genetics and epilepsy. *Dialogues Clin Neurosci,* 10**,** 29-38.

Storey, E., Gardner, R. J., Knight, M. A., Kennerson, M. L., Tuck, R. R., Forrest, S. M. & Nicholson, G. A. (2001) A new autosomal dominant pure cerebellar ataxia. *Neurology,* 57**,** 1913-5.

Street, V. A., Bosma, M. M., Demas, V. P., Regan, M. R., Lin, D. D., Robinson, L. C., Agnew, W. S. & Tempel, B. L. (1997) The type 1 inositol 1,4,5-trisphosphate receptor gene is altered in the opisthotonos mouse. *J Neurosci,* 17**,** 635-45.

Sutter, N. B., Eberle, M. A., Parker, H. G., Pullar, B. J., Kirkness, E. F., Kruglyak, L. & Ostrander, E. A. (2004) Extensive and breed-specific linkage disequilibrium in *Canis familiaris. Genome Res,* 14**,** 2388-96.

Switonski, M., Reimann, N., Bosma, A. A., Long, S., Bartnitzke, S., Pienkowska, A., Moreno-Milan, M. M. & Fischer, P. (1996) Report on the progress of standardization of the G-banded canine (*Canis familiaris*) karyotype. Committee for the Standardized Karyotype of the Dog (*Canis familiaris*). *Chromosome Res,* 4**,** 306-9.

Tan, Y., Dourdin, N., Wu, C., De Veyra, T., Elce, J. S. & Greer, P. A. (2006) Conditional disruption of ubiquitous calpains in the mouse. *Genesis,* 44**,** 297-303.

Thannickal, T. C., Moore, R. Y., Nienhuis, R., Ramanathan, L., Gulyani, S., Aldrich, M., Cornford, M. & Siegel, J. M. (2000) Reduced number of hypocretin neurons in human narcolepsy. *Neuron,* 27**,** 469-74.

Thomsen, J. (1876) Tonische kraempfe in willkuerlich beweglichen muskeln in folge von ererbter psychischer disposition: Ataxia muscularis? *Arch. Psychiat. Nervenkr,* 6**,** 702-718.

Tijssen, M. A., Schoemaker, H. C., Edelbroek, P. J., Roos, R. A., Cohen, A. F. & van Dijk, J. G. (1997) The effects of clonazepam and vigabatrin in hyperekplexia. *J Neurol Sci,* 149**,** 63-7.

Tonelli, A., D'Angelo, M. G., Salati, R., Villa, L., Germinasi, C., Frattini, T., Meola, G., Turconi, A. C., Bresolin, N. & Bassi, M. T. (2006) Early onset, non fluctuating spinocerebellar ataxia and a novel missense mutation in *CACNA1A* gene. *J Neurol Sci,* 241**,** 13-7.

Touma, E., Kato, S., Fukui, K. & Koike, T. (2007) Calpain-mediated cleavage of collapsin response mediator protein(CRMP)-2 during neurite degeneration in mice. *Eur J Neurosci,* 26**,** 3368-81.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. & Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc,* 7**,** 562-78.

Urzua, B., Ortega-Pinto, A., Morales-Bozo, I., Rojas-Alcayaga, G. & Cifuentes, V. (2011) Defining a new candidate gene for amelogenesis imperfecta: from molecular genetics to biochemistry. *Biochem Genet,* 49**,** 104-21.

van de Leemput, J., Chandran, J., Knight, M. A., Holtzclaw, L. A., Scholz, S., Cookson, M. R., Houlden, H., Gwinn-Hardy, K., Fung, H. C., Lin, X., Hernandez, D., Simon-Sanchez, J., Wood, N. W., Giunti, P., Rafferty, I., Hardy, J., Storey, E., Gardner, R. J., Forrest, S. M., Fisher, E. M., Russell, J. T., Cai, H. & Singleton, A. B. (2007) Deletion at ITPR1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans. *PLoS Genet,* 3**,** e108.

van der Merwe, L. L. & Lane, E. (2001) Diagnosis of cerebellar cortical degeneration in a Scottish terrier using magnetic resonance imaging. *J Small Anim Pract,* 42**,** 409-12.

van Steensel, M. A., Koedam, M. I., Swinkels, O. Q., Rietveld, F. & Steijlen, P. M. (2001) Woolly hair, premature loss of teeth, nail dystrophy, acral hyperkeratosis and facial abnormalities: possible new syndrome in a Dutch kindred. *Br J Dermatol,* 145**,** 157-61.

van Tongern, S. E., van Vonderen, I. K., van Nes, J. J. & van den Ingh, T. S. (2000) Cerebellar cortical abiotrophy in two Portuguese Podenco littermates. *Vet Q,* 22**,** 172-4.

Vanhaesebrouck, A., Franklin, R., Van Ham, L. & Bhatti, S. (2012) Hereditary ataxia, myokymia and neuromyotonia in Jack Russell terriers. *Vet Rec,* 171**,** 131-2.

Vignaux, F., Hitte, C., Priat, C., Chuat, J. C., Andre, C. & Galibert, F. (1999) Construction and optimization of a dog whole-genome radiation hybrid panel. *Mamm Genome,* 10**,** 888-94.

Vonholdt, B. M., Pollinger, J. P., Lohmueller, K. E., Han, E., Parker, H. G., Quignon, P., Degenhardt, J. D., Boyko, A. R., Earl, D. A., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J. C., Mosher, D. S., Spady, T. C., Elkahloun, A., Geffen, E., Pilot, M., Jedrzejewski, W., Greco, C., Randi, E., Bannasch, D., Wilton, A., Shearman, J., Musiani, M., Cargill, M., Jones, P. G., Qian, Z., Huang, W., Ding, Z. L., Zhang, Y. P., Bustamante, C. D., Ostrander, E. A., Novembre, J. & Wayne, R. K. (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature,* 464**,** 898-902.

Wang, W. & Kirkness, E. F. (2005) Short interspersed elements (SINEs) are a major source of canine genomic diversity. *Genome Res,* 15**,** 1798-808.

Wessmann, A., Goedde, T., Fischer, A., Wohlsein, P., Hamann, H., Distl, O. & Tipold, A. (2004) Hereditary ataxia in the Jack Russell Terrier--clinical and genetic investigations. *J Vet Intern Med,* 18**,** 515-21.

Wheeler, S. & Rusbridge, C. (1996) Neurological syndrome in Italian spinones. *Vet Rec,* 138**,** 216.

Wright, J. A., Brownlie, S. E., Smyth, J. B., Jones, D. G. & Wotton, P. (1986) Muscle hypertonicity in the cavalier King Charles spaniel--myopathic features. *Vet Rec,* 118**,** 511-2.

Yamada, H., Watanabe, K., Shimonaka, M. & Yamaguchi, Y. (1994) Molecular cloning of brevican, a novel brain proteoglycan of the aggrecan/versican family. *J Biol Chem,* 269**,** 10119-26.

Yasuba, M., Okimoto, K., Iida, M. & Itakura, C. (1988) Cerebellar cortical degeneration in beagle dogs. *Vet Pathol,* 25**,** 315-7.

Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics,* 25**,** 2865-71.

Yuzbasiyan-Gurkan, V., Blanton, S. H., Cao, Y., Ferguson, P., Li, J., Venta, P. J. & Brewer, G. J. (1997) Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington terriers. *Am J Vet Res,* 58**,** 23-7.

Zangerl, B., Goldstein, O., Philp, A. R., Lindauer, S. J., Pearce-Kelling, S. E., Mullins, R. F., Graphodatsky, A. S., Ripoll, D., Felix, J. S., Stone, E. M., Acland, G. M. & Aguirre, G. D. (2006) Identical mutation in a novel retinal gene causes progressive

rod-cone degeneration in dogs and retinitis pigmentosa in humans. *Genomics,* 88**,** 551-63.

Zhou, Q., Ben-Efraim, I., Bigcas, J. L., Junqueira, D., Wiedmer, T. & Sims, P. J. (2005) Phospholipid scramblase 1 binds to the promoter region of the inositol 1,4,5-triphosphate receptor type 1 gene to enhance its expression. *J Biol Chem,* 280**,** 35062-8.